UNIVERSIDADE FEDERAL DE UBERLÂNDIA

INSTITUTO DE BIOTECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOQUÍMICA

EVALUATING ADVANCED STRATEGIES IN PLANT BREEDING USING SIMULATION: GENOMIC SELECTION WITH RAPID-CYCLING, SPEED BREEDING AND GENE EDITING

Aluno: Rafael Augusto Vieira

Orientador: Prof. Dra. Ana Paula Oliveira Nogueira

Co-Orientador: Prof. Dr. Roberto Fritsche-Neto

UBERLÂNDIA-MG

2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

INSTITUTO DE BIOTECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOQUÍMICA

EVALUATING ADVANCED STRATEGIES IN PLANT BREEDING USING SIMULATION: GENOMIC SELECTION WITH RAPID-CYCLING, SPEED BREEDING AND GENE EDITING

Aluno: Rafael Augusto Vieira

Orientador: Prof. Dra. Ana Paula Oliveira Nogueira

Co-Orientador: Prof. Dr. Roberto Fritsche-Neto

Tese apresentada à Universidade Federal de Uberlândia como parte dos requisitos para obtenção do Título de Doutor em Genética e Bioquímica (Área Genética)

UBERLÂNDIA-MG

2025

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Genética e
Bioquímica
Av. Pará 1720, Bloco 2E, Sala 244 - Bairro Umuarama, Uberlândia-MG, CEP 38400-902
Telefone: +55 (34) 3225-8438 - www.ppggb.ibtec.ufu.br - ppggb@ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Genética e Bioquímica
Defesa de:	Doutorado Acadêmico - PPGGB.

Data:	Quinze de abril de dois mil e vinte e cinco.	Hora de início:	13:30h	Hora de encerramento:	17:43h

Matrícula do Discente:	12313GBI014
Nome do Discente:	Rafael Augusto Vieira
Título do Trabalho:	Evaluating Advanced Strategies in Plant Breeding using Simulation: Genomic Selection with Rapid-cycling, Speed Breeding and Gene Editing.
Área de concentração:	Genética
Linha de pesquisa:	Genética, Epigenética, Biologia e Melhoramento de Plantas e Animais.
Projeto de Pesquisa de vinculação:	Estratégias de seleção em programas de melhoramento genético de soja por meio de simulações estocásticas.

Aos quinze dias do mês de abril de dois mil e vinte e cinco, às 13:30horas, reuniu-se via web conferência pela Plataforma Microsoft Teams, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Genética e Bioquímica do Instituto de Biotecnologia da Universidade Federal de Uberlândia, assim composta: Dr. Júlio César do Vale Silva, Dr. José Felipe Gonzaga Sabadin, Dr. Renan Santos Uhdre, Dr. Fernando Cezar Juliatti e Drª. Ana Paula Oliveira Nogueira, como presidente da mesa e demais convidados presentes conforme lista de presença. Iniciando os trabalhos o (a) presidente da mesa, Drª. Ana Paula Oliveira Nogueira apresentou a Comissão Examinadora e o (a) candidato (a), agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa de Pós-graduação em Genética e Bioquímica. A seguir o (a) senhor (a) presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o (a) candidato (a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu os conceitos finais. Em face do resultado obtido, a Banca Examinadora considerou o candidato (a):

APROVADO.

Esta defesa de Tese de Doutorado é parte dos requisitos necessários à obtenção do título de Doutor. O competente diploma será expedido após cumprimento dos

demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU. Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Ana Paula Oliveira Nogueira, Professor(a) do Magistério Superior**, em 15/04/2025, às 17:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Júlio César do Vale Silva, Usuário Externo**, em 15/04/2025, às 17:50, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **José Felipe Gonzaga Sabadin, Usuário Externo**, em 15/04/2025, às 17:51, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Renan Santos Uhdre, Usuário Externo**, em 15/04/2025, às 17:51, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Fernando Cezar Juliatti, Usuário Externo**, em 22/04/2025, às 20:59, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6175538** e o código CRC **5596A43A**.

"Tenha coragem de seguir o seu coração e sua intuição. Todo resto é secundário."

— Steve Jobs

Dedico à minha família, cujo amor e apoio inabaláveis foram fundamentais.

AGRADECIMENTOS

À minha família, pelo amor, paciência e por sempre ser meu porto seguro.

À Francielli, minha esposa, por sua compreensão, parceria ao longo dos anos e por construirmos juntos um caminho repleto de aprendizado e crescimento.

À Isabelli e Manuela, cuja alegria e amor tornam cada desafio mais leve e cada conquista mais especial.

Aos meus pais, por todo amor e pelos valores que, mais do que ensinados, foram vividos e compartilhados.

Aos meus orientadores, Prof. Dra. Ana Paula e Prof. Dr. Roberto pela confiança e contribuições.

À Pós-Graduação em Genética e Bioquímica da Universidade Federal de Uberlândia, pelo aprendizado.

ÍNDICE

APRESENTAÇÃO

Alcançar ganhos genéticos significativos em *pipelines* de melhoramento, garantindo ao mesmo tempo retornos econômicos, exige a geração de ideias e a experimentação para avaliar a utilidade e relação custo-benefício na entrega de retorno sobre o investimento.

Nesse contexto, as simulações permitem explorar cenários e lidar com as limitações dos testes empíricos, que podem ser excessivamente caros e demorados. Essa capacidade de replicar cenários permite a identificação de tendências centrais e fornece *insights* valiosos.

No capítulo um, revisamos e resumimos as principais descobertas dos principais estudos de simulação sobre características quantitativas no melhoramento de plantas. Além disso, organizamos essas descobertas em esquemas de melhoramento, que podem servir como orientação para melhoristas atuais e futuros.

No capítulo dois, simulamos e analisamos estratégias de melhoramento dentro de uma estrutura de programa de melhoramento de soja ao longo de 10 ciclos de seleção, a fim de avaliar seu impacto no ganho genético e na diversidade genética. Essas estratégias incluíram abordagens, tanto isoladas quanto integradas, de seleção genômica, *rapid-cyling*, *speed breeding* e edição gênica.

CHAPTER I - OPTIMIZING THE SELECTION OF QUANTITATIVE TRAITS IN PLANT BREEDING USING SIMULATION

ABSTRACT

This review summarizes findings from simulation studies on quantitative traits in plant breeding and translates these insights into practical schemes. As agricultural productivity faces growing challenges, plant breeding is central to addressing these issues. Simulations use mathematical models to replicate biological conditions, bridging theory and practice by validating hypotheses early and optimizing genetic gain and resource use. While strategies can improve trait value, they reduce genetic diversity, making a combination of approaches essential. Studies emphasize the importance of aligning strategy with trait heritability and selection timing and maintaining genetic diversity while considering genotype-environment interactions to avoid biases in early selection. Using markers accelerates breeding cycles when marker placement is precise, foreground and background selection are balanced, and QTL are effectively managed. Genomic selection increases genetic gains by shortening breeding cycles and improving parent selection, especially for low heritability traits and complex genetic architectures. Regular updates of training sets are critical, regardless of genetic architecture. Bayesian methods perform well with fewer genes and in early breeding cycles, while BLUP is more robust for traits with many QTL, and RR-BLUP proves flexible across different conditions. Larger populations lead to greater gains when clear objectives and adequate germplasm are available. Accuracy declines over generations, influenced by genetic architecture and population size. For low heritability traits, multi-trait analysis improves accuracy, especially when correlated with high heritability traits. Updates including top-performing candidates, but conserving variability enhances gains and accuracy. Low-density genotyping and imputation offer cost-effective alternatives to high-density genotyping, achieving comparable results. Targeting populations optimizes genetic relationships, further improving accuracy and breeding outcomes. Evaluating genomic selection reveals a balance between short-term gains and long-term potential and rapid-cycling genomic programs excel. Diverse approaches

preserve rare alleles, achieve significant gains, and maintain diversity, highlighting the trade-offs in optimizing breeding success.

RESUMO

Esta revisão resume os achados de estudos de simulação sobre características quantitativas no melhoramento de plantas e traduz essas percepções em esquemas práticos. À medida que a produtividade agrícola enfrenta desafios crescentes, o melhoramento de plantas torna-se central para abordar essas questões. As simulações utilizam modelos matemáticos para replicar condições biológicas, conectando teoria e prática ao validar hipóteses precocemente e otimizar o ganho genético e o uso de recursos. Embora estratégias possam melhorar o valor das características, elas reduzem a diversidade genética, tornando essencial a combinação de diferentes abordagens. Estudos enfatizam a importância de alinhar a estratégia com a herdabilidade da característica e o momento da seleção, além de manter a diversidade genética, considerando interações genótipo-ambiente para evitar vieses na seleção precoce. O uso de marcadores acelera os ciclos de melhoramento quando o posicionamento dos marcadores é preciso, a seleção de primeiro plano e de fundo estão equilibradas e os QTL são gerenciados de forma eficaz. A seleção genômica aumenta os ganhos genéticos ao encurtar os ciclos de melhoramento e aprimorar a seleção de parentais, especialmente para características de baixa herdabilidade e arquiteturas genéticas complexas. Atualizações regulares dos conjuntos de treinamento são fundamentais, independentemente da arquitetura genética. Métodos Bayesianos apresentam bom desempenho quando há poucos genes e em estágios iniciais do melhoramento, enquanto o BLUP é mais robusto para características com muitos QTL, e o RR-BLUP se mostra flexível em diferentes condições. Populações maiores resultam em maiores ganhos quando há objetivos claros e germoplasma adequado disponível. A precisão diminui ao longo das gerações, sendo influenciada pela arquitetura genética e pelo tamanho populacional. Para características de baixa herdabilidade, a análise multicaracterística melhora a precisão, especialmente quando correlacionada com características de alta herdabilidade. Atualizações que incluem os melhores candidatos, mas conservam a variabilidade, aumentam os ganhos e a precisão. A genotipagem de baixa densidade combinada com imputação

oferece alternativas custo-efetivas à genotipagem de alta densidade, alcançando resultados comparáveis. O direcionamento de populações otimiza as relações genéticas, melhorando ainda mais a precisão e os resultados do melhoramento. A avaliação da seleção genômica revela um equilíbrio entre ganhos de curto prazo e potencial de longo prazo, com programas genômicos de ciclos rápidos se destacando. Abordagens diversas preservam alelos raros, alcançam ganhos significativos e mantêm a diversidade, destacando os *trade-offs* na otimização do sucesso do melhoramento.

INTRODUCTION

First, this review will summarize the primary findings of main simulation studies on quantitative traits in plant breeding. Second, we further organize these collective discoveries into appropriate breeding schemes that can serve as orientation for both present and potential breeders.

Plant breeding and its significance in agriculture

Agricultural impact

Global agriculture faces a convergence of critical challenges, including food and fuel crises, climate fluctuations, and a range of biotic and abiotic stresses. These factors collectively threaten crop yields, creating significant risks to food security and agricultural sustainability. Addressing these challenges requires innovative strategies that consider the complex interactions between genetic, environmental, and economic factors. By deepening our understanding and management of these stresses, we can develop resilient agricultural systems capable of withstanding the pressures of modern crop production.

The growing population and urgent demand for food security necessitate significant increases in both yield potential and crop quality, nutritional composition of agricultural products. This challenge is compounded by shifting consumption patterns and climate variability that further strain agricultural productivity. The need for advancements in crop productivity and resilience is indispensable and can be met, in part, by developing crop varieties and hybrids that thrive in different climatic conditions without sacrificing yield. By harnessing genetic diversity and employing innovative techniques, plant breeding plays an important role in this effort.

The evolution of plant breeding, from early domestication to modern precision breeding, showcases the synergy of scientific inquiry, technological progress, and agricultural practice. By integrating cutting-edge techniques—such as marker-assisted selection, genomic selection, selection indexes, gene editing, artificial intelligence, process automation, data analytics, epigenomics, phenomics, enviromics, and biotechnology—plant breeding has transformed into a precise science. It is now a powerful engine for innovation, driving the development of high-performing crop varieties and addressing evolving agricultural challenges.

Historical context

The evolution from traditional practices to scientifically informed breeding strategies unfolds through distinct eras characterized by technological breakthroughs and scientific discoveries, each shaping the refinement and application of breeding methodologies in agriculture. At present, plant breeding plays a determining role in global food security, sustainability, and adaptation to climate fluctuations, demonstrating the enduring legacy of Mendel's discoveries in shaping agricultural practices worldwide.

In the beginning, however, plant breeding originated from the early domestication of numerous plant species by ancient civilizations, where rudimentary selection practices focused on enhancing plant compatibility with agricultural systems. These initial steps can be named Breeding 1.0 and laid the foundation for human societies to transition from nomadic lifestyles to settled agricultural communities, representing a significant moment in human history.

Afterward, Breeding 2.0 marked a new era with the introduction of controlled crossbreeding techniques. This period, starting in the 19th century, saw the discovery of fundamental laws of inheritance, Gregor Mendel's groundbreaking work in 1860 elucidated the principles of genetic inheritance through pea plant experiments and provided a systematic framework for understanding how traits are passed down from generation to generation. Concurrently, Francis Galton's investigations into the relationship between offspring and parents established the groundwork for quantitative genetics, highlighting statistical approaches to studying heredity.

Following the rediscovery of Mendel's laws and Galton's investigations, plant breeding experienced a transformative shift in the early twentieth century. Breeders began selecting and crossing superior individuals, advancing beyond the unrefined methods prevalent until then. In some species, this progression moved away from the reliance on selfed seed harvesting or vegetative propagation for cultivar development, practices that were previously employed without the knowledge of genetic principles (Hickey et al., 2017). The transition was driven by a newfound understanding of genetics, enabling breeders to manipulate traits intentionally and their genetic diversity through controlled crosses to capitalize on genetic recombination and variation systematically, then resulting in improved cultivars.

By moving beyond reliance on random natural variation, breeders could strategically target desired characteristics like yield, disease resistance, and nutritional quality. By integrating Mendelian genetics with practical breeding techniques, researchers from that time laid the foundation for modern plant breeding. These methodologies continue to evolve, incorporating advancements in molecular biology, genomics, statistics, and biotechnology to accelerate breeding progress further. Examining the critical milestones and innovations across these eras brings to light an ongoing advancement driven by scientific inquiry and the commitment to sustainable food production.

Starting from the early 20th century, in 1903, Wilhelm Johannsen introduced the theory of pure line breeding, emphasizing the importance of maintaining genetically uniform populations to preserve and enhance desirable traits. Around the same time, Henry Shull's pioneering work on heterosis, or hybrid vigor, revealed the benefits of crossing genetically diverse individuals to enhance offspring productivity. These developments set the stage for more sophisticated breeding strategies aimed at improving crop yields.

By 1908, the Hardy-Weinberg equilibrium, formulated independently by G.H. Hardy and Wilhelm Weinberg, established a foundational law of population genetics, providing insights into genetic variation within populations. Concurrently, geneticist Albert Nilsson-Ehle's development of the genealogical or pedigree method in 1910 enabled breeders to trace the inheritance patterns of traits over multiple generations, refining selection methods that remain in use today.

At the end of the 1910s, advancements in breeding methodologies occurred when Ronald Fisher, Sewall Wright, and J.B.S. Haldane contributed to population genetics in 1918, laying the theoretical foundation for understanding genetic drift, gene flow, and natural selection within breeding populations.

The 1930s saw the rise of quantitative genetics, spearheaded by researchers like Jay Lush. His work in 1935 introduced statistical methods to assess genetic contributions to phenotypic variation, revolutionizing breeding by quantifying the heritability of traits crucial for selection.

Technical innovations in breeding methodologies continued to evolve throughout the 20th century. In 1939, the development of the Single Seed Descent (SSD) method by L.C. Goulden enabled rapid generation turnover in plant breeding programs, accelerating the selection process. Concurrently, the advancement by Comstock, Robinson, and Harvey in 1949 refined the recurrent selection method, providing a systematic framework for enhancing complex traits across successive breeding cycles, whether influenced by additive or non-additive genetic mechanisms.

Advancements in statistical modeling further enhanced breeding strategies. In 1950, C.R. Henderson's introduction of genetic values as random effects laid the foundation for Best Linear Unbiased Prediction (BLUP) methods years later. These statistical tools, further developed by Henderson, enabled breeders to estimate breeding values with greater accuracy, optimizing selection decisions.

In the 1950s, innovations in plant breeding continued to diversify. Ralph Chase's methods for developing double haploid lines in 1952 provided a means to rapidly achieve homozygous lines, facilitating the study and manipulation of specific genetic traits. In essence, the inaugural implementation of the first speed breeding technique was realized. Simultaneously, the elucidation of DNA's double-helix structure by James Watson and Francis Crick in 1953 set the stage for molecular genetics, laying the groundwork for understanding how genetic information is encoded and transmitted.

The early 1970s marked the onset of genetic engineering and recombinant DNA technology in plants. Paul Berg's groundbreaking experiments in 1972 marked a pivotal moment in manipulating genetic material, laying the groundwork for genetically modified organisms (GMOs). This technological leap opened new

possibilities for adding innovative traits to crops, improving their resilience and productivity.

Statistical advancements continued to refine breeding methodologies. Robert Henderson's development of Best Linear Unbiased Prediction (BLUP) methods in 1975 provided breeders with robust statistical tools to predict breeding values based on phenotype and, later, genetic markers, improving the accuracy and efficiency of selection decisions.

The 1980s marked the commercialization of biotechnology applications in agriculture. In the 1990s, the pioneering work of Russell Lande and Stephen Thompson in developing molecular markers for selecting desirable traits significantly accelerated breeding progress. This advancement enabled the identification and selection of plants based on specific genetic fingerprints, thereby enhancing the efficiency of breeding programs.

As a continuity of the previous advancement, Breeding 3.0 expanded upon these foundations by integrating advanced phenotyping and genotyping technologies. This era facilitated the selection of both native and transgenic genes, enhancing breeders' ability to predict and optimize breeding outcomes at both population and individual plant levels and incorporating statistical learnings at practical and commercial levels.

The period between 1989 and 2001 marked a paradigm shift with the introduction of genomic selection. The seminal studies by Fernando and Grossman (1989) and Bernardo (1998), associating molecular markers with Best Linear Unbiased Predictors (BLUPs), and by Meuwissen et al. in 2001, leveraging genomic information to predict an individual's genetic merit for breeding purposes, established new standards in plant breeding.

These advancements expanded the marked-assisted selection approach and brought genomic selection to the forefront, revolutionizing breeding strategies by enabling the early identification of superior genotypes, accelerating genetic gains, and reducing breeding cycles. After this pivotal moment in history, the integration of genomic selection, enhanced breeding methods, and their derivatives, and the utilization of double-haploid techniques became standard practices for numerous crops globally.

In turn, Breeding 4.0 (or Precision Breeding) represents the current frontier, characterized by high-capacity phenotyping and genotyping capabilities, integration of artificial intelligence and automation to streamline breeding processes and new techniques such as gene editing, and accelerated plant cycles. Genomics, epigenomics, phenomics, and enviromics are integral to this approach, enriching breeding strategies with unprecedented precision and adaptability. In general, some of these capabilities have become routine in breeding programs, whilst others promise unparalleled precision, efficiency, and scalability.

In recent years, the landscape of genetic manipulation has been significantly altered by the advent of advanced genome editing technologies. While the theoretical framework for manipulating genomes has been under exploration since the 1970s, recent years have witnessed pivotal advancements in the practical application of these techniques, with the introduction of CRISPR-Cas9 genome editing in 2013 (Hickey et al., 2017).

The concept behind genome editing is elegantly straightforward: targeted alterations to specific base pairs within the genomic sequence, are achieved through methods that entail the deletion, substitution, or addition of genetic material (Hickey et al., 2016). These methods operate through the precise induction of double-stranded breaks at targeted loci, followed by the subsequent repair process. Variations among these techniques are evident in their mechanisms for DNA recognition and cleavage, as well as in considerations of efficiency, precision, cost-effectiveness, and the complex landscape of intellectual property.

Contemporary innovations such as envirotyping and enviromics, introduced by researchers like Cooper et al. (2014) and Tassinari et al. (2020), have significantly expanded the scope of breeding strategies. These methodologies integrate environmental data and omics technologies to tailor breeding programs to specific environmental conditions. They contribute significantly to addressing one of the primary challenges in plant breeding: the genotype-by-environment interaction, thereby enhancing the ability to improve crop resilience and adaptability more effectively.

In the field of breeding for genetic gain, the journey for accelerated breeding cycles, as demonstrated by the emergence of speed breeding techniques by Watson et al. in 2018, has markedly streamlined the breeding process. These

methodologies leverage controlled environments and optimized growth conditions to expedite generation turnover times, facilitating prompt evaluation and selection of promising genotypes. Moreover, the growth and developmental processes of plants are linked to environmental cues that stimulate hormone synthesis, thereby modulating cell cycle dynamics and gene expression to enhance mitosis and cellular differentiation rates. This integration augments genetic gain efficiency over time, ultimately diminishing the duration necessary to complete a breeding cycle.

The evolution of plant breeding, from its origins in early domestication to its contemporary applications in precision agriculture, exemplifies a dynamic synergy of scientific inquiry, technological advancement, and agricultural practice. In essence, plant breeding focuses on harnessing available genetic resources to cultivate superior genotypes, utilizing advanced methods to maximize genetic gain efficiently and economically (Li et al., 2012).

By integrating knowledge from several scientific fields and cutting-edge techniques—such as marker-assisted selection, genomic selection, gene editing, artificial intelligence, algorithms, process automation, epigenomics, phenomics, enviromics, and biotechnological approaches—plant breeding has evolved from an art into a highly efficient and precise science. This transformation is making it a well-oiled engine for developing high-performing crop varieties through innovation and scalable processes, effectively addressing evolving agricultural challenges.

Strategies and objectives

Plant breeding is a continuous process focused on improving genetic traits through systematic selection and crossbreeding. A modern selfing plant breeding program typically involves: (i) selecting individuals with desirable traits for crossbreeding; (ii) employing methods to achieve homozygosity, either fully or partially; and (iii) utilizing recurrent selection to recycle high-merit individuals while evaluating them in field trials. This dual strategy ensures the development of new cultivars with improved yield, performance, disease resistance, and quality while continuously improving the genetic foundation of the breeding population.

The simplest selection method is based on phenotypic values, but relying solely on field trials to estimate breeding values is risky, as they may not fully represent the target environment. In this context, since the breeding value determines an

individual's genetic contribution to future generations, effective selection depends on how well phenotype can represent breeding values. To select superior genotypes with high agronomic performance across diverse environments, breeders must test numerous candidates either through multi-year, multi-location phenotypic trials or by genotyping to estimate breeding values. However, many agronomic traits are complex, influenced by multiple quantitative trait loci (QTL), and their expression is heavily impacted by environmental factors.

Genomic estimated breeding values (GEBVs) and genomic selection offer new opportunities to increase breeding efficiency. GEBVs combine genomic and phenotypic data to predict an individual's genetic merit for specific traits. In genomic selection, candidates are genotyped to obtain GEBVs, bypassing the need for phenotyping. GEBVs are derived from a prediction equation based on the genetic proximity of candidates to a training population, which has been both genotyped and phenotyped.

Genomic predictions rely on accurate estimates of genomic relationships. Implementing genomic selection begins with creating a training population that provides phenotypic data for target traits and genome-wide DNA marker genotypes (Meuwissen et al., 2001). The genotype and phenotype data develop a prediction equation that fits each marker's effect on the trait. When markers are in sufficient linkage disequilibrium with causal variations of the trait, they capture a significant proportion of the associated genetic variance, allowing the prediction of GEBVs without phenotype collection. Many breeding programs today incorporate genomic data into various phases, improving selection accuracy, shortening breeding cycles, and accelerating early-generation selection (Hickey et al., 2014).

Several strategies integrate genomic selection into breeding programs. For example, rapid-cycle recurrent genomic selection promises significant increases in genetic gains (Santantonio et al., 2020). Combining genomic selection with speed breeding can further reduce breeding cycle time. Genome editing introduces new variations, while marker-assisted selection and transgenic methods enhance breeding efficiency. Nonetheless, effective plant improvement requires balancing genetic gain with time and cost (Hickey et al., 2014). Whether employing phenotypic, genomic, or combined selection approaches, optimizing strategies is crucial to maximizing genetic gain while managing costs and operational efficiency.

Improving multiple traits simultaneously is a key challenge in plant breeding, as the product's value often depends on interrelated traits. Trait selection methods include tandem selection, which sequentially improves traits across generations, and independent culling, which rejects individuals failing predefined standards for any trait. A widely used approach is multi-trait selection through a selection index. With genomic selection, multi-trait analysis enhances breeding accuracy, especially for low heritability traits linked to high heritability ones (Hayashi and Iwata, 2013).

Assuming clear breeding objectives, rigorous selection criteria, a diverse pool of germplasm, and well-defined target environments, the efficacy of a breeding program predominantly depends on the efficient utilization of available resources to respond to selection, also known as genetic gain. Optimizing either genomic or phenotypic selection approaches within breeding pipelines is crucial. Substantial genetic progress requires considering several factors, including the number of breeding cycles, population size, parent selection, genomic prediction accuracy, genetic diversity, and cost management.

Simulations and their role in plant breeding

Definition and applications

In plant breeding and genetics, simulations use mathematical models to emulate biological conditions and investigate specific problems (Figure 1). These models are either deterministic or stochastic (Covarrubias-Pazaran et al. 2022). Deterministic models rely on equations from quantitative genetics to predict selection responses using parameters like selection intensity, heritability, and accuracy. However, due to its complexity, they have limitations in accounting for breeding processes such as crossing, generation advancement, and genetic introgression. Stochastic simulations, on the other hand, generate genotypic and phenotypic data for each genetic entity, making them more applicable to breeding stages like recombination, evaluation, and selection.

Simulations bridge theory and practice by computationally modeling breeding strategies to optimize genetic gain, minimize genetic variance loss, and ensure resource efficiency. They compare strategies like phenotypic, marker-assisted, and genomic selection over various timeframes, incorporating early- and late-stage processes. Simulations also identify optimal selection factors, such as intensity,

parental lines, and family sizes, while considering genotype-by-environment interactions, inheritance, and gene effects. By validating hypotheses prior to real-world testing, simulations streamline transitions from phenotypic to marker-assisted and genomic selection.

Limitations and advantages

One limitation of breeding scheme simulations is the potential for random error, reflecting real-life processes like meiosis and genotype evaluation, where simulations often introduce random deviations from the genotypic value (Jannink et al., 2024). These simulations can also be computationally intensive. Genomic predictions require large-scale linear mixed models, which are time-consuming for both real and simulated data. Incorporating training data from previous selection cycles is possible, but computing limitations may arise when dealing with large datasets (Ramasubramanian and Beavis, 2021).

Simulation accuracies often exceed real-world conditions due to factors like error-free molecular markers, absence of epistasis, and limited germplasm exchange (Gaynor et al., 2017). Therefore, careful interpretation is necessary when applying simulation results to practical breeding. Real-world breeding programs also account for multiple traits—such as agronomic performance, disease resistance, and end-use quality—adding complexity to simulations, which must represent multi-trait selection accurately.

Despite these challenges, simulations offer significant advantages. They allow the exploration of various scenarios, genetic models, and methods, helping identify efficient paths to target cultivars (Li et al., 2012). This is particularly important in genomic selection, where balancing resources, program size, and genetic gain is critical. Simulations provide insights into factors affecting genetic gain, prediction accuracy, and cost-effectiveness under different conditions (Hickey et al., 2014; Gorjanc et al., 2018).

While field trials are essential, they provide only a snapshot of outcomes, and random effects may distort results (Li et al., 2012). Simulations, with multiple replications, offer more reliable probabilistic assessments and can test a range of input parameters to evaluate their impact on selection efficiency. They also account for multistage selection and inbreeding rates (Hassanpour et al., 2023).

Modern simulation is characterized by the capacity to model meiotic processes, including crossing over and crossover interference, through coalescent and gene-drop methods, which are employed in backward-in-time and forward-in-time simulations (Hickey and Gorjanc, 2012). The coalescent method generates whole-chromosome founder haplotypes with linkage disequilibrium and allele frequency that align with a specified population genetic model (Hickey and Gorjanc, 2012). Conversely, the gene-drop method simulates the creation of new haplotypes from original founder haplotypes by modeling genetic recombination during meiosis. This process is guided by a genetic map and incorporates the gamma model, which accounts for crossover interference (Hickey and Gorjanc, 2012).

Overall, simulations have become crucial in plant breeding, enabling the evaluation of genetic gain and comparison of breeding strategies by incorporating gene information, crossing schemes, population size, and selection intensity. This helps optimize breeding parameters to maximize gains, while efficiently allocating resources (Li et al., 2012; Hickey et al., 2014; Gorjanc et al., 2018). Simulations also aid in gene mapping, validating new methods, and assessing factors like marker density and QTL heritability (Li et al., 2012).

Translating breeding strategies into digital schemes requires abstraction to represent biological and operational processes while addressing programming limitations. Breeding involves iterative cycles of selection, crossing, and evaluation, which must be encoded into simulations. Challenges include balancing computational efficiency with biological realism, handling large datasets, and modeling diverse genetic architectures. Simulations often assume idealized conditions—error-free markers, no genotype-environment interactions, or perfect linkage maps—which may not align with real-world complexities. Interpreting results requires critical evaluation of these assumptions. Practical integration involves validating insights with field data, incorporating feedback loops, and refining models to align predictions with real-world outcomes. Simulated genetic gain represents a theoretical maximum, often reduced by attrition during practical implementation.

Diversity, background recovery, pre-breeding and bridging to elite germplasm

A simulated study by Fu (2015) assessed how selection impacts genetic diversity, confirming that plant breeding can reduce diversity as expected. Although artificial

selection enhanced the genetic value of targeted traits, it also decreased genetic diversity and heterozygosity (DoVale et., 2021). The breeding schemes studied showed similar effects, either preserving or accelerating diversity loss, with half-sib mating proving the most effective for achieving higher genetic gains while minimizing reductions in diversity. The trends in trait improvement and diversity loss were consistent for progeny sizes of 20 and 50, offering important insights for managing these factors.

Gorjanc et al. (2016) investigated different strategies and genetic parameters for initiating pre-breeding programs with selected landrace populations for integration into elite maize breeding programs. Their results indicated that starting pre-breeding with landraces could significantly influence genetic merit. Higher genetic merit was found when the founding population had low diversity, high within-landrace diversity, or substantial heritability (h^2 = 0.50). Although testcross initiation provided the highest genetic merit, it mainly reconstructed the elite genome and did not incorporate the desired traits from landraces. The authors recommended random mating between initial landrace × elite individuals across several generations to recombine genetic segments and break their linkage. Krenzer et al. (2024) emphasized the importance of simulation-designed pre-breeding crossing schemes for maintaining genetic variation and ensuring long-term success before implementing general combining ability-based selection in hybrid breeding programs.

Building on this, Allier et al. (2020) advanced the understanding of pre-breeding strategies by investigating the integration of new germplasm into elite lines using genomic selection and optimal cross-selection in simulations. Their study introduced donors with various performance levels, showing that recurrent introductions of improved donors could maintain genetic diversity and enhance both mid- and long-term performance. Considering a bridging step resulted significantly higher mid- and long-term genetic gain even when introducing low performing donors. Further, direct introductions of donors with a large performance gap compared to elite germplasm did not generate gains. For donors with an intermediate performance gap, both direct introductions and bridging steps provided increased long-term gains. Additionally, a genomic selection model trained on both bridging and breeding progeny improved prediction accuracy within introduction families. This suggests

that excessive selection for performance during the bridging phase might favor the elite genome, limiting the incorporation of new favorable alleles, as noted by Gorjanc et al. (2016).

Cowling et al. (2017), Villiers et al. (2024), and Platten and Fritsche-Neto (2023) highlight the complex balance required between genetic diversity, selection intensity, and long-term genetic gains. Cowling et al. demonstrated the advantages of optimal contribution selection (OCS) over truncation selection, with OCS yielding higher index values and maintaining lower coancestry across various population sizes and selection pressures. Villiers et al., building on the importance of genetic diversity underscored by Cowling et al., introduced optimal haplotype stacking (OHS) as a superior strategy for preserving diversity while achieving genetic gains. Their results, particularly in small populations, revealed that even a single OHS cycle could outperform OCS and truncation selection in long-term outcomes, emphasizing its potential as a complement to other methods.

Platten and Fritsche-Neto (2023) further nuanced this discussion by focusing on strategies to integrate new genetic material into breeding programs. Their findings showed that while introducing fewer parents (e.g., 10 rather than 20) minimized penalties and enhanced fixation efficiency, genetic diversity and background recovery remained critical factors. This aligns with Villiers et al.'s emphasis on diversity preservation and Cowling et al.'s observation of plateauing gains under traditional truncation selection. Notably, Platten and Fritsche-Neto's recommendation to prioritize GEBV-based selection parallels the targeted approaches of OCS and OHS, highlighting how tailored strategies optimize genetic outcomes under different constraints. Together, these studies illustrate the trade-offs inherent in balancing genetic diversity, selection efficiency, and long-term genetic progress. While Cowling et al. advocate for OCS in avoiding diversity erosion over extended cycles, Villiers et al. demonstrate how OHS can bolster diversity and gains in the short term, even when combined with truncation selection. Meanwhile, Platten and Fritsche-Neto emphasize the necessity of strategic parent selection to achieve rapid fixation of desirable traits. In conclusion, balancing genetic gain and diversity remains a critical challenge in plant breeding. Strategies like optimal contribution selection and optimal haplotype stacking show potential to address this issue. Integrating landraces or other germplasm donors with good performance can

improve long-term gains, but careful planning is necessary. Ultimately, combining strategic selection approaches is crucial for optimizing breeding outcomes while maintaining genetic variability, as simplistic methods are inadequate for navigating this complex process.

The findings from this section are illustrated in Figure 2.

Phenotypic selection and classical methods, including gene effects

Casali and Tigchelaar (1975) conducted pioneering simulations to investigate breeding methods, focusing on single seed descent (SSD) combined with pedigree and mass selection. Their research found that SSD, whether used alone or with mass selection, better preserved genetic variability for line selection in the F6 generation. Pedigree and mass selection were effective for high heritability traits in early generations (F2 and F3), while SSD and F6 selection maximized genetic gain for low heritability traits. Their work demonstrated that early generations are effective for traits with medium or high heritability, while traits with low heritability benefit from later selection. These findings laid the groundwork for modern breeding practices.

Complementary, Van Oeveren and Stam (1992) examined breeding strategies for self-pollinating crops using simulations. They compared early selection, which estimates cross means and variances in the F3 generation and selects lines based on these estimates, with SSD, which delays selection until the F6 generation. SSD performed comparably to early selection with more crosses and outperformed early selection at low heritability due to early prediction inaccuracies. Differences between early selection and SSD decreased when traits were influenced by many loci. SSD was less effective with fewer than 100 F2 plants, and the number of segregating loci had a significant impact on selection outcomes.

These studies underline a critical principle in plant breeding: the choice of breeding strategy should be matched to trait heritability and selection timing. Early selection methods are advantageous for traits with high heritability, while SSD is more effective for traits with low heritability when selection is delayed. They highlight the importance of maintaining genetic diversity and considering genotype-environment interactions to avoid biases in early phenotypic selection. These findings were based on simulated data and predate the adoption of marker-assisted and genomic selection techniques.

Building on this, Wang et al. (2003) compared two prevalent breeding strategies—modified pedigree selection and selected bulk selection—using genetic models that accounted for epistasis, pleiotropy, and genotype-environment interactions. Their simulations indicated that bulk selection achieved a 3.3% higher genetic gain than modified pedigree selection, required one-third less land, and produced 60% fewer families, highlighting its superior efficiency. The study emphasized that selected bulk selection not only resulted in marginally higher genetic gains but also offered significant cost-effectiveness advantages, which is crucial for modern commercial breeding programs.

In a subsequent study, Wang et al. (2004) examined the impact of partial dominance in wheat and found that it had minimal effect on genetic advancement compared to pure additive models. The partial dominance effects were small and difficult to detect through covariance across various mating schemes, suggesting limited applicability in inbred breeding. These observations may be specific to wheat or reflect limitations in the simulations' ability to capture such gene effects. Additionally, Wang et al. (2007) compared three marker selection schemes for integrating nine genes into a single genotype and recommended a top cross scheme. This approach, involving equal selection intensity at three stages (top cross F1, top cross F2, and doubled haploid), minimized the total number of lines needed to achieve the breeding objective.

Continuing the focus on maximizing selection response and genetic gain, Bernardo (2003) analyzed the impact of parental selection, the number of breeding populations, and population size on optimal breeding strategies. Using computer simulations with 2,000 recombinant inbreds for a quantitative trait controlled by 100 additive loci with varying heritabilities, Bernardo found that the highest selection responses were achieved by maximizing the number of breeding populations. The ability to identify high-performing breeding populations before making crosses was more critical than balancing the number of populations and their size.

As emphasized by Bernardo, successful breeding depends on strategic parental selection before establishing breeding populations. A widely accepted recommendation is to use genetically divergent yet superior parents to broaden the genetic base and enhance target traits.

Covarrubias-Pazaran et al. (2022) simulated various parent combinations to provide practical recommendations for breeding programs. Their model, designed for cassava but applicable to other crops, found that overlapping cohorts for recycling—using a mixed crossing block with parents from both preliminary and advanced yield trials—consistently led to higher genetic gains. However, the number of parents used was critical; using more than 30 parents reduced genetic gain compared to fewer. For 60 years, involving 16 to 32 parents proved most effective, with more crosses generally enhancing genetic gains. Within a 20-year timeframe, fewer crosses with more progeny per cross yielded higher gains, with an optimal setup being 8–16 parents, 24 crosses, and approximately 68 progenies per cross annually. Over 60 years, the optimal configuration was 16–32 parents, 60 crosses, and around 30 progeny per cross. The study recommended using 15–30 parents recycled from both preliminary and advanced yield trials, with 40 crosses and 40 progenies per year, a strategy that could be adapted to other crops.

Supporting these findings, Lanzl et al. (2023) demonstrated through simulations that mating designs producing large biparental families from few disjoint crosses risk generating progenies with strong covariances between QTL pairs on different chromosomes. Their research revealed that a single recombination round effectively disrupts both positive and negative covariances between QTL pairs, providing practical insight for optimizing breeding programs.

In summary, optimizing genetic gain in breeding programs involves strategic parental selection and careful management of breeding populations. Using overlapping cohorts for recycling, along with an optimal number of parents and crosses, can significantly enhance genetic gain. Identifying high-performing breeding populations prior to initiating crosses is essential. Additionally, the optimal number of parental lines involved in breeding crosses varies depending on the breeding timeframe. Specifically, a larger number of parents is beneficial for achieving long-term genetic gains, whereas fewer parents are recommendable for attaining higher short-term, immediate gains. These strategies, validated through simulations, apply to both phenotypic and genomic approaches, improving breeding efficiency across various crops.

Backcross is a common method in plant breeding for transferring specific traits from one genetic entity to another while retaining desirable characteristics of the

original line. Another common approach in developing breeding crosses involves emphasizing the importance of one parent over the other. In this method, the F1 generation is backcrossed with the superior parent to enhance its desirable traits, either to recover the parental genetic background or directly derive recombinant lines for the development of new cultivars.

Wang et al. (2009) used computer simulations to show that single backcrossing with a selected bulk is more effective than other crossing and selection methods. This approach is particularly useful for preserving or enhancing the adaptation of recurrent parents while transferring most of the desired donor genes. However, repeated backcrossing may not be necessary if molecular markers can identify the transferred genes, especially when donor parents exhibit good adaptation. The study identified three conditions under which single backcrossing with bulk selection until F6 is highly effective: (1) when traits being transferred are controlled by multiple genes, (2) when donor parents have genes that can improve adaptation in recipient parents, and (3) when conventional phenotypic selection is used, or individual genotypes cannot be easily identified. These insights are valuable for introgressing elite germplasm in rapid-cycle breeding programs, enhancing both efficiency and outcomes.

Continuing on this topic, backcrossing is used to introduce traits or genes into important crops, such as corn. This process includes four stages: single-event introgression, event pyramiding, trait fixation, and version testing. The goal is to ensure that at least one version of the final product performs as well as the original. Sun et al. (2015) investigated two critical aspects of this process through simulations: (i) the impact of residual donor germplasm (non-recurrent parent) in the converted germplasm and (ii) the effect of creating multiple versions for each parental line conversion. Their research found that a residual donor germplasm range of about 5-8% of the genome was associated with a high success rate in conversion. Additionally, the success rate increased with the number of versions produced, with 3-5 versions providing the best outcomes. These findings offer practical insights for market-assisted backcrossing in other crops, emphasizing the importance of managing residual donor germplasm and generating multiple versions during germplasm conversion for specific traits.

The findings from this section are illustrated in Figure 3.

Marker-assisted selection methods and QTL studies

In the early development of marker-assisted selection (MAS), simulations were primary utilized to assess its feasibility as either a complement or replacement for conventional breeding methods. The secondary goal was to optimize MAS integration into breeding strategies for both allogamous and autogamous crops.

Simulation studies by Edwards and Page (1994) and Van Berloo and Stam (2001) assessed the effectiveness of MAS compared to phenotypic selection. They investigated various genetic structures, parental populations, and crop species with different reproductive systems. At this point in history, when researchers assessed the potential of applying MAS to quantitative traits, studies found that while MAS led to rapid initial gains, these improvements diminished significantly after three to five cycles. MAS was particularly effective in utilizing genetic diversity in more heterozygous populations, outperforming phenotypic selection. However, its effectiveness was limited by linkage disequilibrium between markers and QTL, being effective for traits controlled by fewer QTL, such as disease resistance and growth habits in autogamous plants, but not yield.

Despite these limitations, MAS at that time offered notable advantages, including the ability to perform two selection cycles per year for faster initial improvements, with conventional methods potentially taking over in later stages (Edwards and Page, 1994). MAS also provided more accurate predictions of superior parental combinations for multiple traits compared to phenotypic selection (Van Berloo and Stam, 2001). Thus, while MAS generally outperformed phenotypic selection, its initial gains highlighted the value of integrating MAS with traditional breeding methods to maximize genetic potential.

Hospital et al. (1997) further explored MAS through simulations and found that its response at low heritability levels was more variable compared to phenotypic selection. Nonetheless, MAS's efficiency increased with larger population sizes, even for low heritability traits. MAS was effective in fixing QTL with large effects in early generations. However, while fixing QTL alleles with small effects, the advantage was offset by a higher fixation rate of unfavorable background alleles or an increased likelihood of linkage drag, making MAS less efficient than phenotypic selection over time. MAS operates by targeting favorable alleles linked to specific

markers, independent of a trait's heritability (h²). Nevertheless, heritability indirectly affects the efficacy of MAS, as it reflects the proportion of phenotypic variation explained by genetic factors. Traits with high h² facilitate more precise marker-trait associations, enhancing the reliability of selection, whereas traits with low h² may experience reduced accuracy due to a stronger influence of environment. Consequently, while MAS is not directly contingent on h², the strength of the genetic signal remains a determinant of its success.

Overall, these studies emphasize MAS's substantial initial benefits in plant breeding, particularly in capturing genetic diversity and facilitating rapid gains in early selection cycles. MAS is especially effective in more heterozygous populations and for specific traits such as disease resistance. However, the reliance on linkage disequilibrium between markers and QTL presents challenges, limiting MAS's effectiveness for complex traits like yield.

In managing and optimizing marker-assisted selection (MAS), Lande and Thompson (1990) highlighted that the effectiveness of MAS and the advantages of molecular markers depend on genetic parameters and the selection strategy used. They emphasized that large population sizes can enhance the detection of additive genetic variance through marker associations, accelerating quantitative trait improvement.

Defining practical population sizes is essential for cost-effective breeding. Ooijen (1992) investigated the use of adjacent marker pairs to improve QTL linkage information within specific chromosomal segments. The study demonstrated that with 200 backcross or F2 individuals, there is a high probability of identifying QTL accounting for at least 5% of the total variance. Although QTL with larger genetic effects are mapped more precisely, challenges remain for traits controlled by multiple minor effect genes. Ooijen concluded that QTL mapping accuracy is influenced by trait heritability, gene number, interactions, and marker distribution. To enhance QTL effect estimation, he recommended increasing sample sizes and testing environments.

Bernardo and Charcosset (2006) argued that targeting QTL with large effects is more beneficial than including those with smaller effects. Their findings indicated that marker-assisted recurrent selection programs are most effective when focusing on traits controlled by a moderately large number of QTL, approximately 40. For

traits governed by fewer genes, having more genetic information correlates with a higher selection response. In contrast, for traits influenced by 40 or 100 genes, including all genes in the selection model results in a lower selection response (Bernardo and Charcosset, 2006). Although these conclusions are now considered intuitive, they were significant in shaping subsequent research and practical MAS applications in plant breeding.

Gimelfarb and Lande (1994) confirmed that using genetic markers can effectively utilize linkage disequilibrium between markers and QTL generated by crossing inbred lines. They cautioned that merely increasing the number of markers does not necessarily improve selection efficiency. Instead, the population size under selection was identified as a more critical factor for MAS efficiency, aligning with Lande and Thompson's findings (1990). Darvasi and Soller (1997) further refined molecular marker spacing in backcross and F2 linkage experiments through simulations, providing formulas for calculating the optimal number of markers based on factors like population size, and allele substitution effects, and explained variance. Their recommendations offered practical guidance for MAS strategy design.

In autogamous crops, MAS has been shown to improve selection outcomes for obtaining superior genotypes compared to traditional methods (Van Berloo and Stam, 1998). This aligns with earlier findings suggesting MAS can outperform phenotypic selection techniques. Van Berloo and Stam highlighted the advantages of MAS when dominant alleles at QTL are linked in a coupling phase—where loci on the same chromosome are co-inherited more frequently. However, the benefits were moderated by uncertainties in QTL map positions, underscoring the need for precise QTL mapping. Charmet et al. (1997) further supported this by finding that precise QTL location significantly impacts selection when using inbred line parents to maximize favorable allele accumulation. Fine-tuning QTL following initial identification remains crucial in modern crop breeding.

Marker-assisted selection (MAS) offers improved accuracy in parental selection and potentially quicker breeding cycles but tends to be less efficient over time, particularly as the risk of fixing unfavorable alleles increases in later generations. These observations suggest that MAS should complement traditional breeding methods to maximize genetic potential and ensure long-term improvement. Kuchel et al. (2005) investigated integrating restricted backcrossing and doubled haploid

(DH) techniques in MAS breeding strategies. They found that applying MAS at all three stages of breeding was effective for achieving high frequencies of desired outcomes and combining favorable traits. However, they identified that applying MAS at two specific stages (BC1F1 and haploid) was optimal. This approach increased genetic gain compared to phenotypic selection and reduced overall costs by 40%.

A prevalent application of MAS is optimizing backcross phases in breeding programs. Early simulations were crucial in refining this method. Viicher et al. (1996) demonstrated that genetic markers were effective for introgressing specific alleles and selecting the desired background in backcross populations derived from inbred lines. Their simulations showed that markers spaced 10–20 cM apart could shorten the breeding process by one to two generations compared to random or phenotypic selection methods. This time reduction is essential for achieving genetic gain and delivering products efficiently.

Hospital and Charcosset (1997) found that for generations 1 to 3, background selection on both carrier and non-carrier chromosomes with equal weights is optimal. Selection for recipient parent marker alleles should include both carrier and non-carrier chromosomes. They indicated that manipulating no more than four QTL simultaneously is advisable, requiring approximately 1600 individuals to achieve a target genotype with 99% confidence, though using a pyramidal design can reduce this number to 580.

Ribaut et al. (2002) similarly demonstrated that the selection response in BC1 is significantly enhanced when the selectable population size is fewer than 50, with diminishing returns above 100. Efficient selection is achieved by targeting loci in BC1 and BC2 combined with background selection in BC3. Specifically, a population size of 10 in BC1 and BC2, and 100 in BC3, can reduce the donor genome to below 5%, which is often desirable. Peng et al. (2014) recommended a selection scheme involving five generations of MAS for integrating up to 15 transgenic events. This involves selecting for the event of interest in BC1 through BC3 with a population size of 600, and for both the event and recurrent parent germplasm in BC4 and BC5 with a population size of 400, using a selection intensity of 1% for all generations.

These findings are summarized in Figure 4. The integration of MAS has significantly enhanced the efficiency of backcross breeding programs, particularly

for allele introgression and background optimization. Strategic marker use can expedite breeding cycles, improving genetic gains and product delivery. Key aspects include precise marker positioning, balancing foreground and background selection, and managing the number of QTL. Effective background selection in early generations and controlling QTL numbers are essential for maximizing efficiency. Additionally, optimizing selectable population sizes underscores the strategic role of MAS in achieving targeted genetic outcomes.

As noted, the scarcity of marker-assisted selection (MAS) methods and QTL studies in the literature over the past decade reflects the substantial evolution in breeding technologies. This shift is largely attributed to the emergence of genomic selection (GS).

Genomic selection

Emerging from the limitations of marker-assisted selection (MAS), genomic selection has become the foundational methodology in modern plant breeding. Since its introduction, computer simulations have been essential in demonstrating and optimizing genomic selection to achieve greater and more sustainable genetic gains. These insights are summarized in Figure 5.

Two key simulation-based studies established the framework for genomic selection before the term was widely recognized (Bernardo, 1999; Meuwissen et al., 2001). Meuwissen et al. (2001) established the foundational principles of genomic selection. Their simulations used approximately 50,000 marker haplotypes derived from a limited number of phenotypic records within a simulated genome of 1000 cM, with markers spaced at 1 cM intervals. They found that predicting genetic values from markers could significantly accelerate genetic gain, especially when combined with reproductive techniques to shorten generation intervals. Their research also demonstrated that genomic selection could overcome the limitations of traditional MAS, such as its restricted variance capture, overestimated QTL effects, and discrepancies between mapping and breeding populations.

Evaluating genomic selection versus phenotypic and marker-assisted selection

The synthesis of findings across various studies highlights the superior performance of genomic selection compared to traditional phenotypic selection and

MAS. Although genomic selection accelerates the depletion of genetic diversity, it consistently yields higher genetic gains per unit of time due to reduced breeding cycle durations and improved accuracy in parent selection. This advantage is particularly notable under conditions of low trait heritability and complex genetic architectures. Regular updates to training sets are essential to maintain the effectiveness of genomic selection, regardless of genetic architecture (Muleta et al., 2019; Lin et al., 2023). The integration of doubled haploid populations and high-throughput phenotyping further enhances the efficiency and stability of genomic selection, establishing it as a crucial tool in modern breeding programs. Despite its benefits, genomic selection accelerates the depletion of genetic variance compared to phenotypic selection.

Bernardo (1999) bridges the gap between MAS and genomic selection. He assessed the predictive accuracy of breeding values derived from phenotype data alone versus a combination of phenotype and markers. His findings showed that both methods were similarly effective in predicting untested single crosses, across a range of 10 to 100 QTL and trait heritabilities of 0.4 or 0.6. Bernardo and Yu (2007) further demonstrated that genomic selection provided a 6 to 43% greater response compared to marker-assisted recurrent selection, with this effect being more pronounced with a higher number of QTL and lower heritability.

Simulation studies consistently show that genomic selection outperforms phenotypic selection for traits with low heritability and polygenic architecture, such as yield, regardless of population sizes or selection cycles (Gaynor et al., 2017; Muleta et al., 2019; Ali et al., 2020; Silva et al., 2021; Lubanga et al., 2022; Chiaravallotti et al., 2023; Fritsche-Neto et al., 2024). This advantage is particularly evident in early selection cycles, with genomic selection reaching a performance plateau after fewer cycles (Gaynor et al., 2017; Muleta et al., 2019; Ali et al., 2020; Tessema et al., 2020; Lubanga et al., 2022; Chiaravallotti et al., 2023; Fritsche-Neto et al., 2024).

As expected, genomic selection leads to a more significant reduction in genetic diversity compared to phenotypic selection (Gaynor et al., 2017; Muleta et al., 2019; Tessema et al., 2020; Silva et al., 2021; Sabadin et al., 2022). This reduction is influenced by factors such as population size and genetic architecture but can be mitigated by retaining a larger number of individuals for future generations and

incorporating new breeding materials from outside the program. Lin et al. (2023) found that parental populations of 15 exhibited the highest allele fixation, particularly when combined with single-seed descent and speed breeding. Larger population sizes can enhance selection intensity while preserving genetic variance, leading to significant genetic gains and reduced depletion of genetic variance (Muleta et al., 2019; Silva et al., 2021; Fritsche-Neto et al., 2024).

In practical breeding applications, the size of breeding populations is often constrained by limitations in phenotyping and genotyping capabilities. Silva et al. (2021) observed that the number of families and their sizes significantly impact genetic gains over 50 and 100 breeding cycles. More families contribute to increased long-term genetic gains, while larger population sizes are associated with enhanced short-term gains. Despite these constraints, genomic selection generally offers more advantages. Even when genomic and phenotypic selection yields similar genetic gains, the ability of genomic selection to bypass phenotyping costs often makes it preferable (Lin et al., 2023).

However, some simulation studies suggest that phenotypic selection can outperform marker-based selection strategies, particularly under epistatic genetic models and reduced genotype-by-environment interaction (Ali et al., 2020; Peixoto et al., 2023). Muleta et al. (2019) found that in oligogenic architectures—where a few genes with large effects control the trait—the number of genomic selection cycles affected short-term genetic gains. Their results indicated that while more cycles of genomic selection were beneficial for maximizing genetic gain, phenotypic selection achieved higher long-term gains before the fifth breeding cycle. Phenotypic selection also maintained greater genetic variance during early selection years, while prediction accuracy for traits with oligogenic architecture declined more rapidly than for traits with polygenic architecture (Muleta et al., 2019).

Addressing the enhancement of multiple traits simultaneously in plant breeding requires a comprehensive approach. The genomic selection index—a linear combination of genomic estimated breeding values leveraging genomic markers to predict net genetic merit and select parents from a non-phenotyped population— has proven effective in delivering greater genetic gains per unit of time compared to approaches based solely on phenotypic data (Ceron-Rojas et al., 2015).

Gaynor et al. (2017) utilized simulations to propose a two-part genomic selection program comprising: (i) a recycling component for rapid population improvement, and (ii) a product development component for generating commercial products and updating genomic models. Their study showed that this scheme more than doubled genetic gain compared to conventional breeding programs due to shorter breeding cycles. Similarly, Sabadin et al. (2022) and Fritsche-Neto et al. (2024) reported that genomic selection achieved genetic gains 0.5 to 6 times higher than phenotypic selection, attributed to reduced breeding cycle duration and enhanced selection accuracy. Tessema et al. (2020) and Li et al. (2022) reached similar conclusions, emphasizing that shorter breeding cycles were key to the improved effectiveness of genomic selection.

Breeding strategies utilizing genomic selection also demonstrated more stable genetic gains despite genotype-by-environment interactions, whereas traditional programs showed greater variability in genetic gains (Gaynor et al., 2017; Peixoto et al., 2023). This stability arises from using genomic predictions that incorporate multi-year historical data and markers designed to minimize genotype-by-environment interaction effects, ensuring optimal cohort selection (Bernal-Vasquez et al., 2022). Conventional programs, in contrast, rely on phenotypic data from a limited number of years. Bernal-Vasquez et al. (2022) showed that employing a kinship matrix to model breeding values over multiple years, even with single-year data, improved the outcomes in breeding programs.

Gaynor et al. (2017) also highlighted that applying genomic selection at the head-row or progeny stage enhances parent selection accuracy, crucial for successful breeding. Sabadin et al. (2022) demonstrated that applying genomic selection to F2 and evaluating F4 families phenotypically resulted in higher genetic gains compared to other methods, including exclusive phenotypic selection and rapid cycling from parental selection in F2. They also found that preserving a larger number of individuals (48-parent scenario) helped reduce genetic variance depletion and allowed for long-term genetic gains. Lin et al. (2023) supported these findings, noting that postponing line derivation to later generations is advantageous for traits with low heritability, offering a product development perspective. Silva et al. (2021) suggested that parental selection at the F4 stage is often treated as an early recycling generation due to limited and less reliable phenotypic data.

In summary, these authors advocate for the strategic use of genomic selection, emphasizing later-stage parental selection, for example, F4 generation (Gaynor et al., 2017; Silva et al., 2021; Sabadin et al., 2022; Lin et al., 2023). Their research highlights the benefits of this approach in terms of precision and long-term genetic gains while preserving a larger genetic pool. They argue that later-stage selection provides more stable gains and maintains genetic diversity better than early-stage selection.

In contrast, Li et al. (2022) found that selecting parents early in the breeding cycle, at stages such as F1 or F2, yielded substantially higher genetic gains compared to selection at later stages. They emphasize that earlier-stage genomic selection can yield higher immediate genetic gains, despite the moderate prediction accuracy. This result aligns with Peixoto et al. (2023) corroborated this by showing that early parental recycling, whether based on genomic or phenotypic selection, produced greater and faster genetic gains compared to scenarios where parental selection occurred after test cross-evaluation.

In summary, the synthesis of simulation studies emphasizes the advantage of selecting parents for recycling as early as possible while considering selective accuracy and trait-related breeding value estimates. Breeding programs equipped with high-throughput phenotyping, extensive genotyping, or other advanced methods can benefit from rapid-cycling strategies to maximize genetic gain. However, it is crucial to set specific accuracy or heritability targets for selecting parents to effectively address the complexities of breeding, such as specific objectives, available genetic diversity, and the balance between short-term and long-term gains.

For crops utilizing doubled haploids, rapid cycling within the population improvement component of two-stage breeding programs is particularly advantageous for managing heterozygosity and advancing generations efficiently. In crops where doubled haploid techniques are unavailable, it may be necessary to incorporate an additional stage with phenotypic assessments before yield testing to eliminate poorly segregating individuals and expedite the advancement of promising ones.

Evaluating different models in breeding value estimation

Best Linear Unbiased Prediction (BLUP) is fundamental in genomic selection, each offering unique advantages in optimizing prediction accuracy across complex genetic architectures. BLUP, utilized within mixed linear models, seeks to minimize bias by accounting for both fixed and random effects. On the other hand, Bayesian methods integrate prior knowledge with observed data, allowing for continual refinement of probability estimates to address uncertainty.

Meuwissen et al. (2001) provided early insights into these methods, demonstrating that while BLUP achieved a prediction accuracy of 0.73, Bayesian approaches, particularly those incorporating marker haplotypes, improved accuracy to 0.85. This early evidence underscored Bayesian methods' potential to enhance genomic selection performance.

Subsequent research has shown comparable performance between BLUP-based methods (e.g., gBLUP, RR-BLUP) and Bayesian methods (e.g., BayesB, LASSO), with minor differences in accuracy and genetic gains across various models (Technow et al., 2012; Wang et al., 2015; Yao et al., 2018; Silva et al., 2021; Fernández-González et al., 2023). Bayesian approaches, particularly BayesB and Bayesian LASSO, have demonstrated superior accuracy and genetic gains in models accounting for dominance effects, especially when traits are influenced by fewer genes (Technow et al., 2012; Wang et al., 2018). BayesB, in particular, is noted for its effectiveness in capturing genetic gains and accelerating outcomes in early breeding cycles (Ramasubramanian and Beavis, 2021).

Certain BLUP variants, such as sBLUP (Super BLUP) and cBLUP (Compressed BLUP), demonstrate distinct advantages depending on the heritability of the traits. When compared to Bayesian LASSO and gBLUP in prediction accuracy, sBLUP outperformed these methods for Mendelian (simple) traits, whereas cBLUP proved more effective for traits with low heritability. In the cBLUP method, particularly for traits with low heritability, grouping individuals based on kinship and maximum likelihood (clustering) proved advantageous for estimating breeding values. This approach outperformed other well-known methods, likely due to replacing individual kinship with group kinship, which enhances estimation accuracy and mitigates the challenges associated with low heritability. However, it does not offer similar benefits under conditions of high heritability. In another study, RR-BLUP (Ridge Regression

BLUP) demonstrated a slight advantage over other methods for traits influenced by numerous minor genes, making it a viable alternative (Wang et al., 2015).

Best Linear Unbiased Prediction (BLUP) is extensively applied in the estimation of breeding values, renowned for its simplicity and robustness in managing large populations with polygenic traits—characterized by the cumulative influence of numerous genes with small effects. A notable strength of BLUP is its capacity to minimize bias by simultaneously accounting for fixed and random effects, thereby delivering consistent predictions across varying scenarios. Specific adaptations, such as cBLUP, exhibit heightened efficiency when applied to traits with high heritability. However, BLUP is not without limitations. The method encounters challenges in accommodating complex genetic architectures, particularly those involving significant dominance effects, and exhibits reduced accuracy for traits with low heritability. In such cases, alternative approaches, such as Bayesian methods, often outperform BLUP. Furthermore, its reliance on additive genetic models restricts its adaptability to scenarios involving non-additive genetic interactions or traits controlled by a small number of major genes.

Bayesian methods offer an effective alternative for breeding value estimation, particularly for traits influenced by a few major genes or dominance effects, where they often outperform traditional approaches. Their flexibility allows the integration of prior knowledge and adaptation to diverse genetic models. Methods like BayesB and Bayesian LASSO are especially useful in early breeding cycles, capturing genetic gains efficiently. However, Bayesian methods are computationally intensive, which can be a limitation in large-scale breeding programs, and reliance on prior specifications may introduce bias if not properly defined. While Bayesian methods excel in specific genetic contexts, BLUP remains a reliable tool for polygenic traits with numerous small-effect genes, ensuring consistent performance in various breeding scenarios.

In conclusion, the choice between BLUP and Bayesian methods should be driven by the genetic architecture, population size, and specific breeding objectives. Bayesian methods, such as BayesB, are particularly beneficial for traits with fewer genes and in early breeding cycles. In contrast, BLUP methods are robust for traits with extensive QTL. RR-BLUP stands out for its flexibility and consistent performance across diverse conditions. Thus, no single model excels universally,

emphasizing the need for tailored approaches based on the breeding program's specific characteristics and goals.

Analyzing the impact of genetic parameters and architectures on gains

Increasing the size of breeding populations typically leads to greater genetic gains with genomic selection, provided that breeding objectives are well-defined and there is adequate. Extensive research has consistently indicated this trend (Lorenz, 2013; Muleta, et al., 2019; Silva et al., 2021). However, it is important to note that the accuracy of genomic recurrent selection predictions tends to decline over generations. This decline is influenced by the trait's genetic architecture and population size (Muleta et al., 2019). The population size can fluctuate due to various breeding operations such as truncation selection, which reduces genetic diversity and narrows the gene pool over generations. Additionally, other breeding operations, like crossing and inbreeding, may further impact the effective population size and the genetic variability available for selection. As these factors alter the genetic structure of the population, they can lead to reduced accuracy in genomic predictions, especially when selecting for complex traits with low heritability.

For traits with genetic architectures involving 40 or 400 QTL, the selection response often plateaus after approximately 10 to 15 cycles (Ramasubramanian, and Beavis, 2021). Conversely, traits with more complex genetic architectures, such as those involving 4,289 QTL, can sustain selection responses for up to 30 to 40 cycles before encountering limitations (Ramasubramanian and Beavis, 2021). In this context, the numerous QTLs frequently exhibit significant interactions with the environment, ultimately reducing the heritability of the trait.

Heritability plays a critical role in determining genetic gain, with higher heritability leading to greater gains across various models and parental selection methods (Yao et al., 2018). Traits with high heritability demonstrate greater prediction accuracy, as a larger proportion of the phenotypic variation is attributed to genetic factors, and the corresponding QTL exhibit reduced interaction with environmental influences. In such cases, multi-trait analysis can improve the accuracy of breeding value predictions for traits with low heritability, particularly when these traits are correlated with high-heritability traits, as compared to single-trait analysis (Hayashi and Iwata, 2013). However, simulation studies investigating the advantages of multi-trait

analysis remain limited, partly constrained by programming capabilities. To improve the accuracy of genomic selection, increasing the number of initial parents in the breeding cycle can be advantageous (Chiaravallotti et al., 2023). Modeling marker effects as population-specific is beneficial under low linkage disequilibrium and enhances prediction accuracy at lower marker densities (Technow et al., 2012). Additionally, incorporating dominance effects into models can substantially improve prediction accuracy, particularly for populations with convergent parentage (Technow et al., 2012). The effectiveness of genomic selection in introgressing large QTL is influenced by the trait's genetic architecture, breeding strategy, and the number of initial parents, highlighting the complexity involved in optimizing genomic selection methods (Chiaravallotti et al., 2023).

Analyzing the impact of different training set configurations

The studies highlight the importance of optimizing training populations and marker densities in genomic selection to enhance prediction accuracy and genetic gain. In general, using lines in the test set that are closely related to the training population improves genomic breeding value (GEBV) accuracy by having a close relationship (Hickey et al., 2014; Muleta et al., 2019; Ramasubramanian and Beavis, 2021; Sabadin et al., 2022; Chiaravallotti et al., 2023). When closely related to the test set, smaller, well-structured training sets with low to mid-density markers effectively capture genetic information, especially with high linkage disequilibrium (Gorjanc et al., 2017). Regular updates to the training population with top-performing lines further boost genetic gain and prediction accuracy.

Expanding training and prediction sets with low-density genotyping and imputation is a cost-effective strategy to improve selection responses. Although low-density genotyping may initially reduce prediction accuracy, strategic imputation and training set updates can counter this, achieving results comparable to high-density genotyping. Targeting relevant populations enhances genetic relationships between training and prediction sets, leading to better prediction accuracy and breeding outcomes.

Simulation studies indicate that for biparental populations closely related to selection candidates, using 200–500 markers and about 1,000 phenotypes can achieve effective breeding value accuracy (Hickey et al., 2014; Gorjanc et al., 2017).

Increasing marker density beyond 10,000 markers provides no additional benefit (Hickey et al., 2014). Phenotyping only a subset of doubled haploid (DH) lines is usually sufficient to predict the performance of the remaining lines. However, while studying single biparental population of doubled haploid lines, Lorenz (2013) found that including all DH lines in the training set significantly improves prediction accuracy compared to excluding part of them, suggesting that phenotyping a subset might be less effective. Budget constraints should guide these decisions.

Training sets with full-sibling and half-sibling families generally achieve high prediction accuracy (Melchinger et al., 2023a; Melchinger et al., 2023b). Including unrelated families is discouraged due to potential negative impacts on prediction outcomes (Melchinger et al., 2023a; Melchinger et al., 2023b). To avoid these issues, unrelated families should be excluded, and a moderate number of genotypes per family (about 50) should be maintained. Training with full-sibling families, compared to half-sibling families, leads to greater selection gains in hybrid breeding programs, particularly where specific combining ability (SCA) effects are crucial.

There are two main strategies for genotyping: high-density genotyping, which is accurate but costly, and lower-density genotyping combined with imputation, which is more cost-effective yet retains acceptable precision. Gorjanc et al. (2017) found that accurate imputation is a cost-effective approach by (i) reducing genotyping costs per individual for both training and prediction sets, (ii) increasing selection intensity through larger prediction sets, and (iii) improving prediction accuracy by enlarging training sets. Using nearly 100 DHs, they reported imputation accuracies of 0.71 across families and 0.42 within families with high-density genotypes. Imputation led to a modest decline across families (from 0.71 to 0.66) compared to within families, from 0.42 to 0.34 (Gorjanc et al., 2017), likely due to the substantial genetic diversity represented and its connection to the training set. Accuracy improved most when imputed genotypes were used for both training and prediction sets (Gorjanc et al., 2017). For cross-family predictions, accuracy plateaued with marker densities between 100 and 50 low-density markers, while for within-family predictions, the plateau occurred between 200 and 100 low-density markers (Gorjanc et al., 2017).

Both modelling for cross-pollinated genomes, Gorjanc et al. (2017) and DoVale et al. (2022) offer complementary perspectives on the trade-offs between low-

density genotyping, imputation, and marker density in breeding programs. Gorjanc et al. demonstrated the effectiveness of expanding training and prediction sets through low-density genotyping and imputation, showing that the use of low-density genotypes, followed by imputation with as few as five segregating markers per chromosome, could achieve prediction accuracy and selection response comparable to high-density genotyping. They found that the response plateaued when 100 low-density markers were used, indicating diminishing returns with increased marker numbers. Imputation with about 20 segregating markers per chromosome still yielded satisfactory results in selection response and prediction accuracy favorable results, especially in non-phenotyped families, where genotyping with 50 low-density markers achieved a minimum accuracy of 0.3. In contrast, Do Vale et al. (2022) highlighted that increasing marker numbers can mitigate the decay of linkage disequilibrium (LD) over cycles, potentially improving long-term selection responses. They argued that while low-density marker panels offer benefits such as reduced multicollinearity, faster computation times, and cost-efficiency, high-density panels ultimately provide superior long-term gains, especially when the training set is not frequently updated. Their analysis suggests that the advantage of marker reduction through LD-based strategies is limited to a single cycle, as the accuracy of selection declines significantly in subsequent cycles, emphasizing the long-term value of high-density genotyping. Ultimately, the choice of marker strategy depends on the specific breeding objectives, resource constraints, and the need for sustainable improvements across multiple cycles.

Prediction accuracy is influenced by population structure, trait heritability, training set size, and the precision of genomic relationships at QTL regions. Regular updates are crucial for maintaining effective breeding programs (Hickey et al., 2014; Cuyabano et al., 2024). Research indicates that prediction accuracy typically declines within the first two cycles (Fritsche-Neto et al., 2024) and updating training sets generally enhances accuracy by preserving linkage disequilibrium between markers and QTL (Muleta et al., 2019; Ramasubramanian and Beavis, 2021; Sabadin et al., 2022; Chiaravallotti et al., 2023).

Trait heritability is a significant factor in the loss of prediction accuracy (Muleta et al., 2019; Ramasubramanian and Beavis, 2021). Ramasubramanian and Beavis (2021) found that genomic prediction models performed best when updated with

data from up to 14 selection cycles, with optimal cycles varying by QTL count: 10 cycles for 40 QTLs and 30 cycles for 4289 QTLs. Results also depended on QTL effect sizes and whether effects were additive or dominant. Larger training sets, incorporating diverse conditions, offered more effective guidance for future selection decisions. Their study also revealed that static training sets when using genomic selection combined with a chain-rule mating design, can potentially capture up to 62% of the greatest genetic potential , even with advanced methods like RR-REML and Bayesian approaches (Ramasubramanian and Beavis, 2021). It worth mentioning that they considered 20 family designed with a common parent in all crosses with 100 F5 lines of family size. Expanding both the training population size and the number of replications improved prediction accuracy, with population size being particularly beneficial when genotyping costs were low, or heritability was high. Consistent with Lorenz (2013), phenotyping all doubled haploid (DH) lines resulted in greater genetic gains, making this approach preferable when feasible. Additionally, dispersing populations across multiple environments improved selection outcomes (Lorenz, 2013).

Sabadin et al. (2022) found that integrating top-performing parental lines from each cycle into long-term breeding schemes—while maintaining genetic diversity to balance allele frequencies and better fit the prediction model—yielded the highest genetic gains. They recommended retaining data from the most recent three generations to optimize genetic relationships between the training set and the target population. This method, referred as TSGPO, was noted for its ease of implementation and consistent performance.

Fernández-González et al. (2023) demonstrated that integrating test set information into the training set—referred to as targeted optimization—achieved the highest accuracies, particularly for low heritability traits like yield. For untargeted optimization approach , they advised that minimizing the average genetic relationship within the training set as a more reliable strategy than maximizing the relationship between the training and test sets. This finding is particularly noteworthy, as it challenges the widely accepted assumption that maximizing the relationship between training and test sets is essential for enhancing prediction accuracy (Wallace et al., 2018). Instead, it highlights that a universal approach may not be suitable in all cases.

Evaluation of different genomic selection strategies

Evaluating genomic selection strategies reveals a complex balance between short-term gains and long-term genetic potential. Rapid-cycling genomic selection programs, which initiate selection and parental recycling earlier, often surpass traditional phenotypic selection methods. This advantage arises from the reduced time required for parental recycling and the development of new cultivars, with both the rapid-cycling process and the genomic selection components contributing significantly to this efficiency. Advanced strategies such as Optimal Population Value Selection (OPV), Genomic Optimal Contribution Selection (GOCS), Optimal Haploid Value (OHV), and Weighted Genomic Selection (WGS) effectively preserve rare favorable alleles, achieve significant genetic gains, and ensure long-term genetic diversity, highlighting the trade-offs in optimizing breeding success.

Gaynor et al. (2017) proposed a two-part genomic selection program consisting of (i) a rapid population improvement component through genetic gain recycling, and (ii) a product development component that creates commercial products and updates genomic selection models with new data. Their research demonstrated that genetic gains were 1.31 to 1.46 times greater than those achieved with traditional genomic selection methods. The approach involving recycling parental lines from head-rows for population improvement yielded superior results, although these gains were inversely related to the depletion of genetic variability in subsequent seasons. Addressing this challenge will be critical for future strategies.

Gorjanc et al. (2017) tackled the challenge of balancing short-term and long-term genetic gains by proposing optimal cross-selection within rapid recurrent genomic selection programs. Their approach enhanced the efficiency of converting genetic diversity into genetic gain, reducing the loss of genetic diversity and the decline in genomic prediction accuracy that often accompanies rapid cycling.

Goiffon et al. (2017) found that the Optimal Population Value (OPV) method achieved superior genetic gains over the first ten generations compared to other methods. OPV and Genotype Building (GB) showed better long-term responses than basic genomic selection, WGS, and OHV. However, genomic selection and WGS provided rapid and significant improvements in GEBVs before reaching a plateau by generation 4, proving valuable in specific contexts.

OPV, GB, and OHV were more effective at preserving genetic variance within breeding populations over time compared to genomic selection or WGS. OPV, which evaluates the genetic merit of selection candidates rather than using truncation selection, is particularly useful for large, commercially significant cohorts. OPV and OHV also excel at maintaining rare favorable alleles, which might otherwise be underrepresented in training sets due to their infrequent occurrence or unique characteristics.

Beukelaer et al. (2017) compared Genomic Optimal Contribution Selection (GOCS), which limits genomic relationships among selection candidates, and Weighted Genomic Selection (WGS) with basic genomic selection. They found that GOCS and WGS achieved similar long-term genetic gains and inbreeding rates, but WGS slightly reduced short-term genetic gain compared to basic genomic selection. Alternative strategies such as IND-HE (focusing on heterozygosity) and IND-RA (focusing on rare alleles) demonstrated superior long-term gains and a better balance between genetic merit and diversity than GOCS or WGS, remaining effective across varying trait heritabilities and initial training population sizes. As outlined by Beukelaer et al. (2017), IND-HE aims to balance genetic gain and expected heterozygosity, seeking to control the inbreeding rate, defined as the relative decline in expected heterozygosity calculated from SNP markers. Additionally, IND-RA focuses on preserving rare alleles by incorporating them into the selection process.

Ramasubramanian and Beavis (2021) found that WGS produced greater responses in longer selection cycles compared to unweighted basic genomic selection, although it initially yielded lower response rates than phenotypic selection and basic genomic selection. They concluded that the effectiveness of genomic selection depends on the number of simulated QTL and trait heritability, with RR-BLUP offering better responses early on and WGS excelling in later cycles. Combining strategies can leverage different approaches' strengths, mitigating costs, errors, and time requirements associated with phenotypic selection.

Akdemir et al. (2019), Moeinizade et al. (2019), Silva et al. (2021), Ramasubramanian and Beavis (2021), and Sabadin et al. (2022) contribute to the growing understanding of optimized selection strategies in breeding programs, offering insights into the balance between short-term genetic gains, long-term

sustainability, and genomic advancements. Akdemir et al. demonstrated that multi-objective optimized parental proportion approaches, which balanced genetic variance and genomic estimated breeding values , yielded 20-30% higher outcomes over extended breeding cycles, outperforming index-based selection methods.

Silva et al. (2021) and Ramasubramanian and Beavis (2021) focused on selection methods within breeding cycles, highlighting the nuances of family-based selection. Silva et al. found that while across-family selection benefited short-term gains, within-family selection proved more effective for long-term improvements, with optimal results emerging at selection intensities between 7.5% and 10%. Their work emphasized that using 100 families with 150 individuals per family maximized genetic gains over ten cycles. Ramasubramanian and Beavis, in a similar vein, noted that genomic selection through RR-BLUP combined with recombination of high-merit parents within a hub network design was particularly effective in the first 5 to 10 cycles, further supporting the utility of family-based and genomic strategies for early-cycle improvements.

Building on these findings, Sabadin et al. (2022) refined genomic selection methods, demonstrating that a two-stage approach combining genomic selection in the F2 generation with phenotypic selection in the F4 produced 20% higher genetic gains than fast-recycling methods that skipped phenotyping. Their research highlighted the trade-offs between increasing selection cycles and the necessity of updating training sets to maintain selection effectiveness. This two-stage approach proved more advantageous than rapid cycling methods that relied solely on genomic selection in the early stages, reinforcing the importance of a balanced approach to maximize genetic gain over time.

In hybrid crop breeding, programs using high-density markers (10,000 or more) and QTL-level genotyping achieved greater genetic gains and heterosis compared to those using low-density and SNP markers (De Jong et al., 2023). Genomic approaches for predicting single-crosses were found to be more effective and simplified compared to rapid-cycling recurrent genomic selection and cyclical tester updates (Fritsche-Neto et al., 2024), though combined applications of these methods hold potential.

Multi-trait selection

Hayashi and Iwata (2013) and Akdemir et al. (2019) explore multi-trait genomic selection, offering complementary insights into improving prediction accuracy and genetic gains while managing program constraints. Hayashi and Iwata demonstrated that multi-trait Bayesian analysis, especially for low-heritability traits correlated with high-heritability traits, enhances genomic breeding value predictions. Although multi-trait analysis did not consistently outperform single-trait analysis for uncorrelated traits, its effectiveness in leveraging trait correlations makes it valuable for breeding programs targeting interrelated traits.

Akdemir et al. (2019) advanced this by incorporating multi-objective optimization frameworks, such as the MOOB approach, to achieve sustainable gains across multiple traits. Their method showed 20–30% higher long-term gains compared to traditional approaches like tandem and index selection. While Hayashi and Iwata focused on trait correlations, Akdemir et al. highlighted the importance of maintaining genetic diversity and optimizing parental contributions to ensure breeding population viability. The MOOB framework balances genetic variance, accuracy, and gains, addressing practical challenges such as the difficulty of assigning economic weights in index selection.

Both studies emphasize strategies that accommodate the complexities of multi-trait breeding. Hayashi and Iwata's findings align with Akdemir et al.'s conclusion that methods exploiting trait interdependence or optimizing multiple objectives outperform traditional single-trait approaches. Moreover, the MOOB framework's focus on genetic diversity preservation complements Hayashi and Iwata's emphasis on accurate prediction, offering a comprehensive approach for breeding programs targeting long-term sustainability and trait improvement. Together, these approaches demonstrate the potential of advanced genomic tools and optimization techniques to address the multi-dimensional challenges of modern breeding.

Inbreeding, heterosis and epistasis

Studies highlight the challenge of balancing genetic gain with inbreeding in genomic selection, emphasizing the need for refined approaches. Beukelaer et al. (2017) found that genomic optimal contribution selection (GOCS) and genomic selection (GS) offer comparable long-term gains but struggle with inbreeding control. GOCS fails to achieve an optimal trade-off, while Goiffon et al. (2017)

showed that weighted genomic selection (WGS) better controls inbreeding by targeting rare favorable alleles, offering a better balance. However, methods like GOCS and IND-OC lead to higher long-term gains but incur short-term penalties on progress.

Sabadin et al. (2022) noted that GS accelerates inbreeding in self-pollinated crops, depleting genetic variance more quickly than phenotypic selection. These findings underscore the need for updated training sets and strategies to preserve diversity while maximizing gains. GOCS and WGS each offer benefits but need continuous refinement to optimize long-term genetic improvement.

Advancements in genomic prediction models for hybrid breeding highlight the impact of marker density. De Jong et al. (2023) reported that high-density SNP markers generated 1.72 times more heterosis than low-density markers and 2.06 times more than QTL genotypes. Moreover, initial germplasm grouping is unnecessary to achieve high heterosis (Cowling et al., 2020). Models incorporating dominance effects improved heterosis by 1.44 times, emphasizing the value of capturing dominance interactions. Combining high-density markers with dominance-inclusive models enhanced prediction accuracy and genetic progress. Additionally, optimizing tester selection from previous cycles advanced genetic composition, complementing heterotic patterns (Melchinger and Frisch, 2023b).

Epistatic genetic architectures also affect selection strategies. Alia et al. (2020) showed that, under additive models, GS outperformed phenotypic selection (PS), but PS was more effective under epistatic models. Wang et al. (2004) found that dominance and epistasis slow genetic progress in wheat breeding, as epistasis stabilizes variance but hinders allele fixation. This suggests breeders should balance genomic and phenotypic approaches, integrating genetic architecture insights to optimize long-term selection outcomes.

Cost analysis

Hospital and Charcosset (1997) emphasized the benefits of combining foreground and background selection in marker-assisted backcross programs for introgressing quantitative trait loci (QTL). Pyramidal designs, where QTL are monitored sequentially, optimize genomic similarity with the recipient parent while minimizing costs when paired with appropriate population sizes. Kuchel et al. (2005)

demonstrated the economic advantage of integrating marker-assisted selection (MAS) at the BC1F1 and haploid stages in wheat breeding, showing a 40% cost reduction and greater genetic gain compared to phenotypic methods. These findings highlight the importance of aligning genotyping and phenotyping for maximum efficiency.

In genomic selection, Lorenz (2013) and Hickey et al. (2014) focused on maximizing genetic gain through effective resource allocation. Lorenz's simulations showed that increasing population size and phenotyping all DH lines maximized genetic gain under constrained heritability and genotyping costs. Hickey et al. suggested that combining phenotypic data from related populations improves prediction accuracy while reducing time and cost penalties. Recent advancements in low-density genotyping and imputation, as demonstrated by Gorjanc et al. (2017), can reduce genotyping costs by up to 87% with minimal loss in prediction accuracy. Using as few as 50 segregating markers per genome can yield a high return on investment (5.67 times the baseline), making genomic selection more accessible for early-generation selection.

Genotype-by-environment interaction (GxE)

Studies by Buntaran et al. (2022) and Li et al. (2022) emphasize the need for refined selection strategies to address genotype-by-environment (GxE) interactions in genomic selection. Buntaran et al. showed that adjusting selection to environmental factors like drought or excess rainfall is critical for maintaining genetic progress, especially when environmental fluctuations bias breeding value estimates. Li et al.'s simulations focused on complex traits, such as grain yield and disease resistance, where GxE plays a significant role.

The impact of model parameters, selection intensities, and breeding strategies on long-term outcomes is further explored by Silva et al. (2021) and Covarrubias-Pazaran et al. (2022). Silva et al. highlighted GxE variance as a source of non-heritable variation, stressing the need for precise GS model parameterization to prevent genetic erosion. Covarrubias-Pazaran et al. simulated genotype-by-year and genotype-by-location interactions, offering valuable insights for improving program efficiency. However, better representation of GxE, particularly interactions across years, is still needed in simulation studies.

GxE models significantly impact genetic progress by enabling the use of environment-specific genetic performance. In the short term, they enhance trait prediction and optimize selection strategies for specific environments. In the long term, they support developing resilient germplasm with stability across diverse conditions. Strategies to leverage GxE include using selection indices tailored to environments, prioritizing traits with moderate to high heritability, incorporating genomic predictions, and conducting multi-environment trials to align GxE insights with breeding goals.

Software tools

Simulations are crucial in breeding programs, with several R packages offering comprehensive solutions for various needs. AlphaSimR is highly cited for its flexibility in modeling complex genetic architectures, selection schemes, and long-term strategies. Adam is essential for studies involving structured populations, tracking genetic inheritance in crossing programs. Synbreed excels in genomic prediction and analysis using SNP data. MoBPS is a modular and versatile tool for accurately simulating hybrid and genomic breeding programs. QU-LINE, a pioneer in simulating quantitative traits, remains valuable for modeling additive and epistatic effects in classical genetics. PedigreeSim specializes in simulating crosses and genetic recombination in structured populations, often used for inheritance and lineage management studies.

CONCLUSION

In plant breeding and genetics, simulations serve as mathematical models that replicate real-world biological conditions to address specific challenges and phenomena. Their growing importance stems from the necessity to optimize resource use in relation to program scale and genetic gains. The complexities of multi-trait selection make simulations crucial and challenging, as breeding pipelines are dynamic, influenced by business decisions, advancements in genotyping technologies, and fluctuations in resource allocation from year to year.

Simulations offer significant advantages by allowing researchers to explore diverse scenarios, genetic models, and methodologies, helping them identify the most effective strategies for developing target cultivars. These models facilitate the

evaluation of genetic gain under various conditions, simulating entire or partial breeding programs. This capability enables comparisons of different breeding strategies by integrating genetic data, crossing schemes, propagation methods, population sizes, selection intensities, and the number of generations involved.

As breeding programs evolve and adopt new methodologies, the demand for simulations to assess their benefits and explore alternatives remains critical. Future research in this area is likely to focus increasingly on genetic diversity, pre-breeding efforts, and advances in genomic selection strategies. Simulations will continue to play a pivotal role in shaping the future of plant breeding by providing insights that help refine breeding pipelines and maximize genetic gains.

Studies highlight the need to align breeding strategies with trait heritability, selection timing, and genetic diversity, while accounting for genotype-environment interactions to minimize bias in early selection. Genomic selection enhances genetic gains by accelerating breeding cycles and optimizing parent selection, particularly for traits with low heritability and complex genetic structures. Regular updates to training sets are essential, irrespective of the genetic architecture. Bayesian methods are effective in early breeding stages with fewer genes, while BLUP is more suitable for traits with multiple QTL, and RR-BLUP offers flexibility across varying conditions.

Larger populations yield greater gains when clear objectives and appropriate germplasm are present. Updating with top-performing parentals improves accuracy and gains. Low-density genotyping, combined with imputation, provides a cost-effective alternative to high-density genotyping with comparable results. Evaluating genomic selection requires balancing short-term gains with long-term potential. Rapid-cycling programs are particularly effective.

Future research in plant breeding and genetics using simulations is poised to explore several promising avenues to enhance models and their applications. One key focus area may involve refining simulations to better capture genotype-environment interactions, particularly in the context of climate challenges.

Another potential direction for research is the expansion of simulations to support multi-genomic selection frameworks. As omics technologies continue to advance, future studies may aim to integrate multiple layers of data, such as epigenetics and transcriptomics, into simulations. This approach would establish a connection with

the previously mentioned genotype-environment enhancement opportunity, while also contributing to a more comprehensive understanding of the mechanisms through which genetic networks influence trait expression.

There is significant potential in combining simulation models with artificial intelligence and machine learning algorithms. These technologies could be harnessed to optimize breeding pipelines by automatically identifying patterns, predicting outcomes, and dynamically adjusting selection criteria in real time. This approach would allow breeding programs to become increasingly adaptive and data-informed, facilitating faster genetic gains and greater breeding efficiency, with the primary focus on delivering value to farmers and consumers.

Finally, another critical focus is expanding simulations to support multi-trait selection frameworks, while refining key breeding parameters such as parental number, population size, and selection timing in both phenotypic and genomic programs. It is worth noting that the exploration of various genomic selection strategies remains a relevant and continually evolving area of research, as this approach becomes increasingly routine in breeding programs.

REFERENCES

Akdemir, D., Beavis, W., Fritsche-Neto, R., Singh, A. K., and Isidro-Sánchez, J. (2019). Multi-objective optimized genomic breeding strategies for sustainable food improvement. Hered. 122, 672–683. doi: 10.1038/s41437-018-0147-1

Ali, M., Zhang, L., DeLacy, I., Arief, V., Dieters, M., Pfeiffer, W. H. et al. (2020). Modeling and simulation of recurrent phenotypic and genomic selections in plant breeding under the presence of epistasis. Crop J. 8, 866–877. doi: 10.1016/j.cj.2020.04.002

Allier, A., Teyssèdre, S., Lehermeier, C., Moreau, L., and Charcosset, A. (2020). Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genom. 21, 349. doi: 10.1186/s12864-020-6756-0

Bernal-Vasquez, A. M., Buntaran, H., Gordillo, A., Sahr, M., Wimmer, V., and Piepho, H. (2022). Assessing the response to genomic selection by simulation Valentin and

Hans-Peter. Theor. Appl. Genet. 135, 2891–2905. doi: 10.1007/s00122-022-04157-1

Bernardo, R. (1999). Marker-assisted best linear unbiased prediction of single-cross performance. Crop Sci. 39, 1277–1282.

Bernardo, R. (2003). Parental selection, number of breeding populations, and size of each population in inbred development. Theor. Appl. Genet. 107, 1252–1256. doi: 10.1007/s00122-003-1375-0

Bernando, R., and Charcosset, A. (2006). Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. Rex Bernardo, Alain. Crop Sci. 46, 614–621. doi: 10.2135/cropsci2005.05-0088

Bernardo, R., and Yu, J. (2007). Prospects for genome-wide selection for quantitative traits in maize. Crop Sci. 47, 1082–1090. doi: 10.2135/cropsci2006.11.0690

Beukelaer, H., Badke, Y., Fack, V., and Meyer, G. (2017). Moving beyond managing realized genomic relationship in long-term genomic selection. Genetics 206, 1127–1138. doi: 10.1534/genetics.116.194449

Casali, V. W. D., and Tigchelaar, E. C. (1975). Computer simulation studies comparing pedigree, bulk, and single seed descent selection in self pollinated populations. J. Amer. Soc. Hort. Sci. 100, 364–367. doi: 10.21273/JASHS.100.4.364

Ceron-Rojas, J. J., Crossa, J., Arief, V. N., Basford, K., Rutkoski, J., Jarquín, D. et al. (2015). A genomic selection index applied to simulated and real data. G3 5, 2155–2164. doi: 10.1534/g3.115.019869

Charmet, G., Robert, N., Perretant, M. R., Gay, G., Sourdille, P., Groos, C. et al. (1999). Marker-assisted recurrent selection for cumulating additive and interactive

QTL in recombinant inbred lines. Theor. Appl. Genet. 99, 1143–1148. doi: 10.1080/13102818.2014.995920

Chiaravallotti, I., Lin, J., Arief, V., Jahufer, Z., Osorno, J. M., McClean, P. et al. (2023). Simulations of multiple breeding strategy scenarios in common bean for assessing genomic selection accuracy and model updating. The Plant Genome 17, doi: 10.1002/tpg2.20388

Covarrubias-Pazaran, G., Gebeyehu, Z., Gemenet, D., Werner, C., Labroo, M., Sirak, S. et al. (2022). Breeding Schemes: What Are They, How to Formalize Them, and How to Improve Them? Front. Plant Sci. 12, doi: 10.3389/fpls.2021.791859

Cowling, W. A., Li, L., Siddique, K. H. M., Henryon, M., Berg, P., Banks, R. G. et al. (2017). Improving response in genomic selection with a population-based selection strategy: optimal population value selection. J. Exp. Bot. 68, 1927–1939. doi: 10.1534/genetics.116.197103

Cowling, W. A., Gaynor, R. C., Antolin, R., Gorjanc, G., Edwards, S. M., Powell, O. et al. (2020). In silico simulation of future hybrid performance to evaluate heterotic pool formation in a self-pollinating crop. Sci. Rep. 10, 4037. doi: 10.1038/s41598-020-61031-0

Cuyabano, B. C. D., Boichard, D., and Gondro, C. (2024). Expected values for the accuracy of predicted breeding values accounting for genetic differences between reference and target populations. Genet. Sel. Evol. 56. doi: 10.1186/s12711-024-00876-9

Darvasi, A., and Soller, M. (1997). A simple method to calculate resolving power and confidence interval of QTL map location. Behav. Genet. 27, 125–132. doi: 10.1023/A:1025685324830

De Jong, G., Powell, O., Gorjanc, G., Hickey, J. M., and Gaynor, C. (2023). Comparison of genomic prediction models for general combining ability in early

stages of hybrid breeding programs. Crop Sci. 63, 3338–13355. doi: 10.1002/csc2.21105

DoVale, J. C., Carvalho, H. F., Sabadin, F., and Fritsche-Neto, R. (2022). Genotyping marker density and prediction models effects in long-term breeding schemes of cross-pollinated crops. Theor. Appl. Genet. 135, 4523–4539. doi: 10.1007/s00122-022-04236-3

Edwards, M. D., and Page, N. J. Evaluation of marker-assisted selection through computer simulation. Theor. Appl. Genet. 88, 376–382. doi: 10.1007/BF00223648

Fernández-González, J., Akdemir, D., and Sánchez, J. I. (2023). A comparison of methods for training population optimization in genomic selection. Theor. Appl. Genet. 136. doi: 10.1007/s00122-023-04265-6

Fritsche-Neto, R., Ali, J., De Asis, E. J., Allahgholipour, M., and Labroo, M. R. (2024). Improving hybrid rice breeding programs via stochastic simulations: number of parents, number of hybrids, tester update, and genomic prediction of hybrid performance. Theor. Appl. Genet. 137. doi: 10.1007/s00122-023-04508-6

Fu, Y. Understanding crop genetic diversity under modern plant breeding. (2015). Theor. Appl. Genet. 128, 2131–2142. doi: 10.1007/s00122-023-04508-6

Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R. et al. (2017). A two-part strategy for using genomic selection to develop inbred lines. Crop Sci. 56, 2372–2386. doi: 10.2135/cropsci2016.09.0742

Gimelfarb, A., and Lande, R. (1994). Simulation of marker assisted selection in hybrid populations. Genetics Res. 63, 39–47. doi: 10.1017/s0016672300032067

Goiffon, M., Kusmec, A., Wang, L., Hu, G., and Schnable, P. S. (2017). Improving response in genomic selection with a population-based selection strategy: optimal

population value selection. Genetics 206, 1675–1682. doi: 10.1534/genetics.116.197103

Gorjanc, G., Jenko, J., Hearne, S. J., and Hickey, J. M. (2016). Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations. BMC Genom. 17. doi: 10.1186/s12864-015-2345-z

Gorjanc, G., Battagin, M., Dumasy, J., Antolin, R., Gaynor, C. R., and Hickey, J. M. (2017). Prospects for cost-effective genomic selection via accurate within-family imputation. Crop Sci. 57, 216-228. doi: 10.2135/cropsci2016.06.0526

Gorjanc, G., Gaynor, C. R., and Hickey, J. M. (2018). Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theor. Appl. Genet. 131. doi: 10.1007/s00122-018-3125-3

Hassanpour, A., Geibel, J., Simianer, H., and Pook, T. (2023). Optimization of breeding program design through stochastic simulation with kernel regression. Henner Torsten. G3 6. doi: 10.1093/g3journal/jkad217

Hayashi, T., and Iwata, H. (2013). A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinform. 31. doi: 10.1186/1471-2105-14-34

Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B. M et al. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci. 54, 1476–1488. doi: 10.2135/cropsci2013.03.0195

Hickey, J. M. and Gorjanc, G. (2012). Simulated data for genomic selection and genome-wide association studies using a combination of coalescent and gene drop methods. G3 2, 425–427. doi: 10.1534/g3.111.001297

Hickey, J. M., Chiurugwi, T., Mackay, I., and Powell, W. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nat. Genet. 49, 1297–1303. doi: 10.1038/ng.3920

Hospital. F., and Charcosset, A. (1997). Marker-assisted introgression of quantitative trait loci. Genetics 147, 1469–1485. doi: 10.1093/genetics/147.3.1469

Hospital, F., Moreau, L., Lacoudre, F., Charcosset, A., and Gallais, A. (1997). More on the efficiency of marker-assisted selection. Theor. Appl. Genet. 95, 1181–1189. doi: 10.1007/s001220050679

Jannink, J., Astudillo, R, and Frazier, P. (2024). Insight into a two-part plant breeding scheme through Bayesian optimization of budget allocations. Crop Sci. doi: 10.1002/csc2.21124

Krenzer, D., Frisch, M, Beckmann, K., Kox, T., Flachenecker, C., Abbadi, A. et al. (2024). Simulation-based establishment of base pools for a hybrid breeding program in winter rapeseed. Theor. Appl. Genet. 137. doi: 10.1007/s00122-023-04519-3

Kuchel, H., Ye, G., Fox, R., and Jefferies, S. (2005). Genetic and economic analysis of a targeted marker-assisted wheat breeding strategy. Mol. Breeding 16, 67–78. doi: 10.1007/s11032-005-4785-7

Lanzl, T., Melchinger, A. E., and Schön, C. (2023). Influence of the mating design on the additive genetic variance. in plant breeding populations. Theor. Appl. Genet. 136. doi: 10.1007/s00122-023-04447-2

Li, Y., Kaur, S., Pembleton, L. W., Valipour-Kahrood, H., Rosewarne, G. M., and Daetwyler, H. D. (2022). Strategies of preserving genetic diversity while maximizing genetic response from implementing genomic selection in pulse breeding programs. Theor. Appl. Genet. 135, 1813–1828. doi: 10.1007/s00122-022-04071-6

Li, X., Zhu, C., Wang, J., and Yu, J. (2012). Computer simulation in plant breeding. Adv. Agron., 116, 219–264. doi: 10.1016/B978-0-12-394277-7.00006-3

Lin, J. Arief, V., Jahufer, Z., Osorno, J., McClean, P., Jarquin, D. et al. (2023). Simulations of rate of genetic gain in dry bean breeding programs. Theor. Appl. Genet. 136. doi: 10.1007/s00122-023-04244-x

Lorenz, A. J. (2013). Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: a simulation experiment. G3 3, 481–491. doi: 10.1534/g3.112.004911

Lubanga, N., Massawe, F., Mayes, S., Gorjanc, G., and Bančič., J. (2022). Genomic selection strategies to increase genetic gain in tea breeding programs. The Plant Genome 16. doi: 10.1002/tpg2.20282

Melchinger, A. E., Fernando, R., Stricker, C., Schön, C., and Auinger, H. (2023a). Genomic prediction in hybrid breeding: I. Optimizing the training set design. Theor. Appl. Genet. 136. doi: 10.1007/s00122-023-04413-y

Melchinger, A. E., and Frisch, M. Genomic prediction in hybrid breeding: II. Reciprocal recurrent genomic selection with full-sib and half-sib families. (2023b). Theor. Appl. Genet. 136. doi: 10.1007/s00122-023-04446-3

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. (2001). Genetics 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Moeinizade, S., Hu, G., Wang, L., and Schnable, P. S. (2019). Optimizing selection and mating in genomic selection with a look-ahead approach: an operations research framework. G3 9, 2123-2133. doi: 10.1534/g3.118.200842

Muleta, K. T., Pressoir, G., and Morris, G. P. (2019). Optimizing genomic selection for a sorghum breeding program in Haiti: a simulation study. G3 7, 391–401. doi: 10.1534/g3.118.200932

Ooijen, J. W. V. (1992). Accuracy of mapping quantitative trait loci in autogamous species. Theor. Appl. Genet. 84, 803-811. doi: 10.1007/BF00227388

Peixoto, M. A., Coelho, I. F., Leach, K. A., Bhering, L. L., and Resende, M. F. R. (2023). Simulation-based decision-making and implementation of tools in hybrid crop breeding pipelines. Crop Sci. 64, 110–125. doi: 10.1002/csc2.21139

Platten, J. D, and Fritsche-Neto, R. (2023). Optimizing quantitative trait loci introgression in elite rice germplasms: Comparing methods and population sizes to develop new recipients via stochastic simulations. Plant Breed. 142, 439–448. doi: 10.1111/pbr.13118448

Peng, T., Sun, X., and Mumm, R. H. (2014). Optimized breeding strategies for multiple trait integration: I. Minimizing linkage drag in single event introgression. Mol Breeding 33, 89–104. doi: 10.1007/s11032-013-9936-7

Ramasubramanian, V., and Beavis, W. D. (2021). Strategies to assure optimal trade-offs among competing objectives for the genetic improvement of soybean. Front. Plant Sci. 12. doi: 10.3389/fgene.2021.675500

Ribaut, J. M., Jiang, C., and Hoisington, D. (2002). Simulation experiments on efficiencies of gene introgression by backcrossing. Crop Sci. 42, 557–565. doi: 10.2135/cropsci2002.5570

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

Santantonio, N., Atanda, S. A., Beyene, Y., Varshney, R. K., Olsen, M., Jones, E. et al. (2020). Strategies for effective use of genomic information in crop breeding programs serving Africa and South Asia. Front. Plant Sci. 11. doi: 10.3389/fpls.2020.00353

Silva, E. D. B., Xavier, A., and Faria, M. V. (2021). Impact of genomic prediction model, selection intensity, and breeding strategy on the long-term genetic gain and genetic erosion in soybean breeding. Front. Plant Sci. 12. doi: 10.3389/fgene.2021.637133

Sun, X., and Mumm, R. H. (2015). Optimized breeding strategies for multiple trait integration: III. Parameters for success in version testing. Mol. Breeding 35. doi: 10.1007/s11032-015-0397-z

Technow, T., Riedelsheimer, C., Schrag, T. A., and Melchinger, A. E. (2012). Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theor. Appl. Genet. 125, 1181–1194. doi: 10.1007/s00122-012-1905-8

Tessema, B. B., Liu, H., Sørensen, A. C., Andersen, J. R., and Jensen, J. (2020). Strategies using genomic selection to increase genetic gain in breeding programs for wheat. Front. Plant Sci. 11. doi: 10.3389/fgene.2020.578123

Van Berloo, R., and Stam, P. (1992). Comparative simulation studies on the effects of selection for quantitative traits in autogamous crops: early selection versus single seed descent. Hered. 69, 342–351.

Van Berloo, R., and Stam, P. (1998). Marker-assisted selection in autogamous RIL populations: a simulation study. Theor. Appl. Genet. 196, 147–154. doi: 10.1007/s001220050721

Van Berloo, R., and Stam, P. (2001). Simultaneous marker-assisted selection for multiple traits in autogamous crops. Theor. Appl. Genet. 102, 1107–1112. doi: 10.1007/s001220000518

Viicher, P. M., Haley, C. S., and Thompson, R. (1996). Marker-assisted introgression in backcross breeding programs. Genetics 144, 1923–1932.

Villiers, K., Voss-Fels, K. P., Dinglasan, E., Jacobs, B., Hickey, L., and Hayes, B. J. (2024). Evolutionary computing to assemble standing genetic diversity and achieve long-term genetic gain. The Plant Genome 17. doi: 10.1002/tpg2.20467

Wallace, J. G., Rodgers-Melnick, E., and Buckle, E. S. On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. Annu Rev Genet. 23, 421-444. doi: 10.1146/annurev-genet-120116-024846

Wang, J., Ginkel, M., Podlich, D., Ye, G., Trethowan, R., Pfeiffer, W. et al. (2003). Comparison of two breeding strategies by computer simulation. Crop Sci. 43, 1764-1773. doi: 10.2135/cropsci2003.1764

Wang, J., Ginkel, M., Trethowan, R., Ye, G., DeLacy, I., Podlich, D. et al. (2004). Simulating the effects of dominance and epistasis on selection response in the CIMMYT wheat breeding program using QuCim. Crop Sci. 44, 2006–2018. doi: 10.2135/cropsci2004.2006

Wang, J., Chapman, S. C., Bonnett, D. G., Rebetzke, G. J., and Crouch, J. (2007). Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Sci 47, 582–588. doi: 10.2135/cropsci2006.05.0341

Wang, J., Singh, R. P., Braun, H., and Pfeiffer, W. H. (2009). Investigating the efficiency of the single backcrossing breeding strategy through computer simulation. Theor. Appl. Genet. 118, 683–694. doi: 10.1007/s00122-008-0929-6

Wang, J., Zhou, Z., Zhang, Z., Li, H., Liu, H., Liu, D. (2018). Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. Hered. 121, 648–662. doi: 10.1038/s41437-018-0075-0

Wang, X., Yang, Z., and Xu, C. (2015). A comparison of genomic selection methods for breeding value prediction. Sci. Bull. 60, 925–935. doi: 10.1007/s11434-015-0791-2

Yao, J., Zhao, D., Chen, X., Zhang, Y., and Wang, Z. (2018). Use of genomic selection and breeding simulation in cross prediction for improvement of yield and quality in wheat (Triticum aestivum L.). Crop J. 6, 353–365. doi: 10.1016/j.cj.2018.05.003
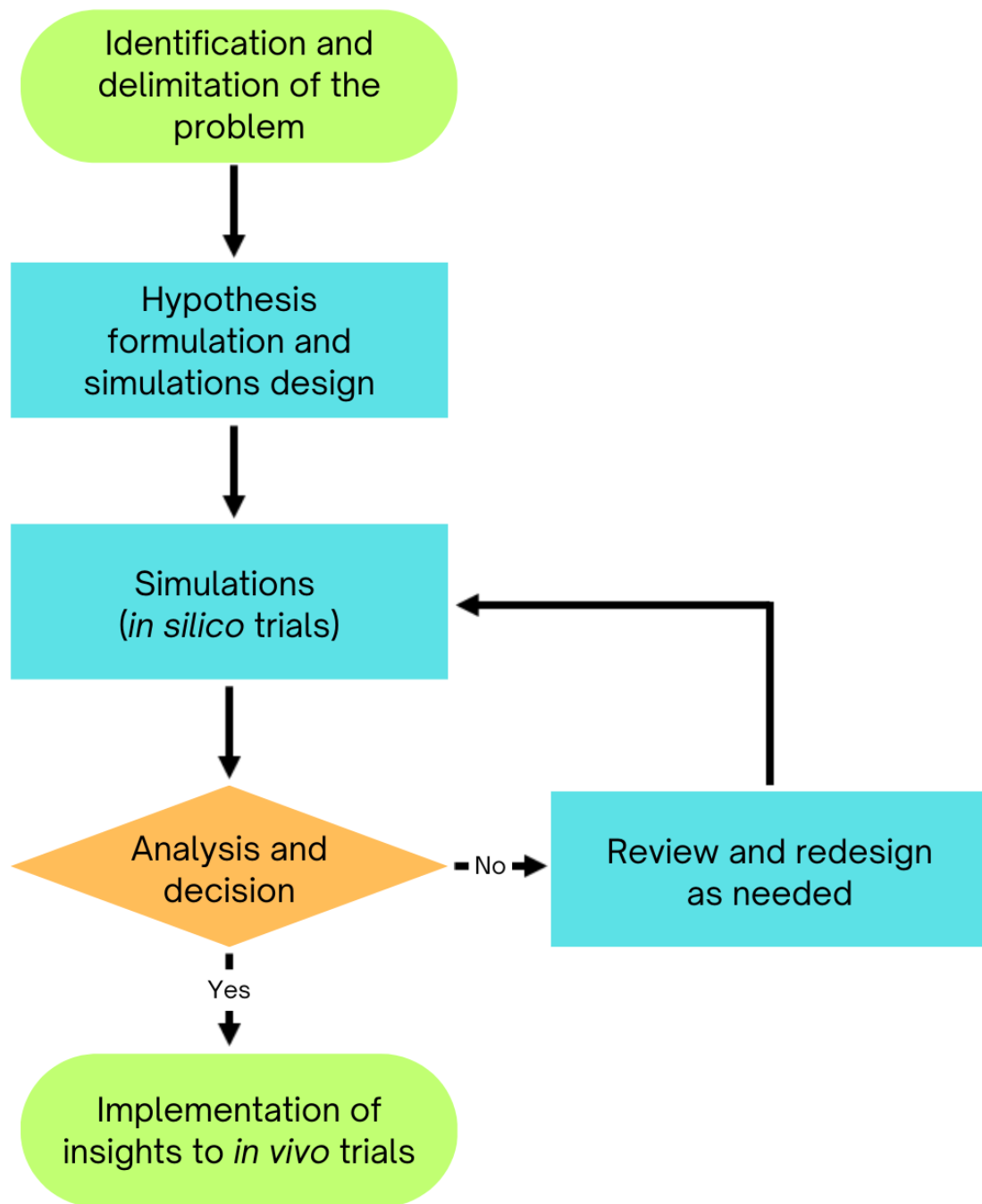
Figure 1. Diagram of simulation models in plant breeding, illustrating how mathematical models replicate biological conditions to bridge theoretical concepts with practical breeding applications.
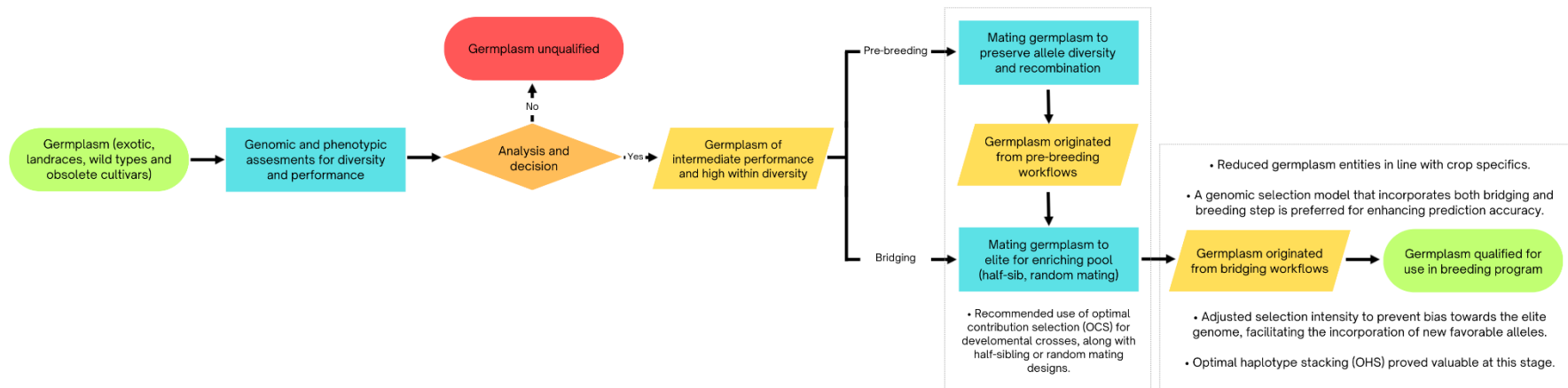
Figure 2. Flowchart depicting the application of simulation-derived strategies for managing genetic diversity, pre-breeding processes, and transitioning from theoretical models to elite germplasm.
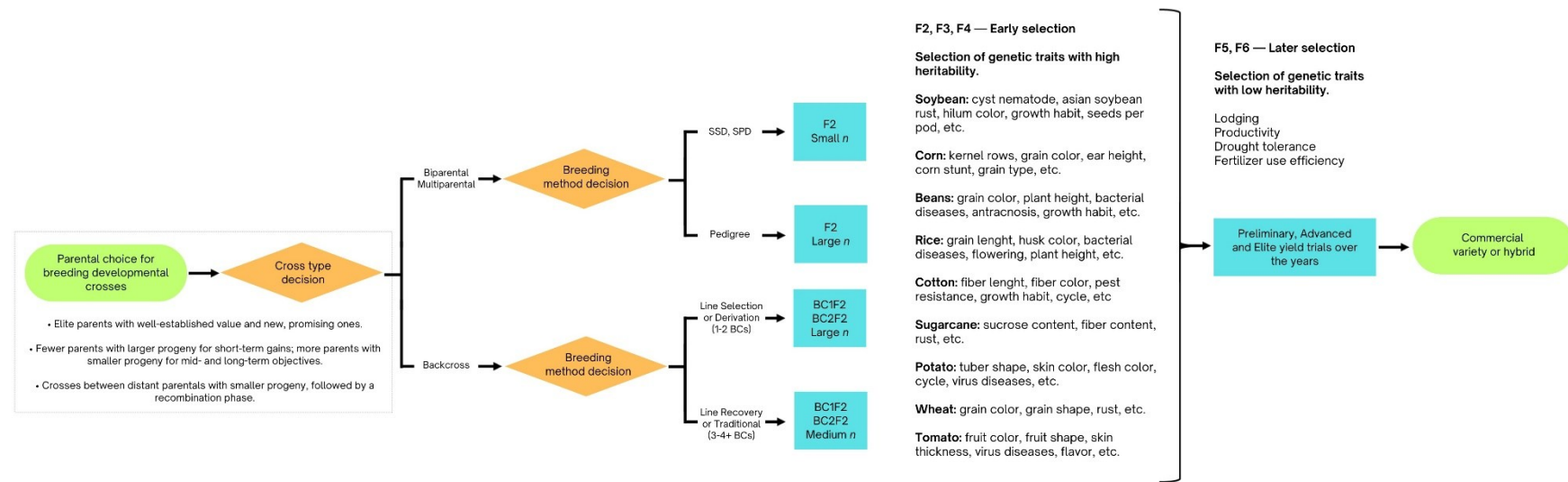
Figure 3. Flowchart demonstrating practical insights for phenotypic selection and traditional breeding methods, as informed by simulation studies. It outlines how these methods can be optimized to balance trait selection and genetic diversity.
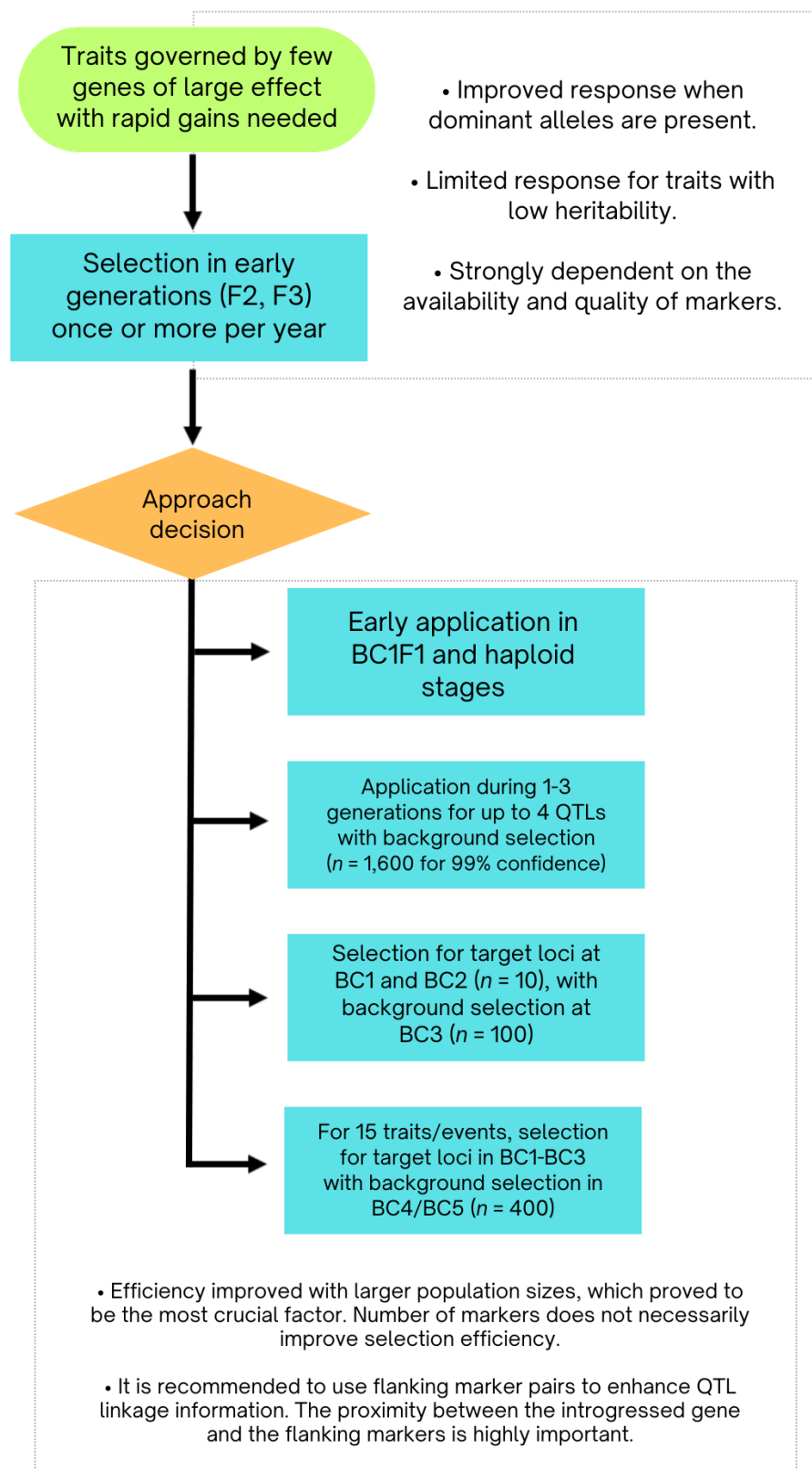
Figure 4. Flowchart showcasing marker-assisted selection and QTL management strategies derived from simulation research.
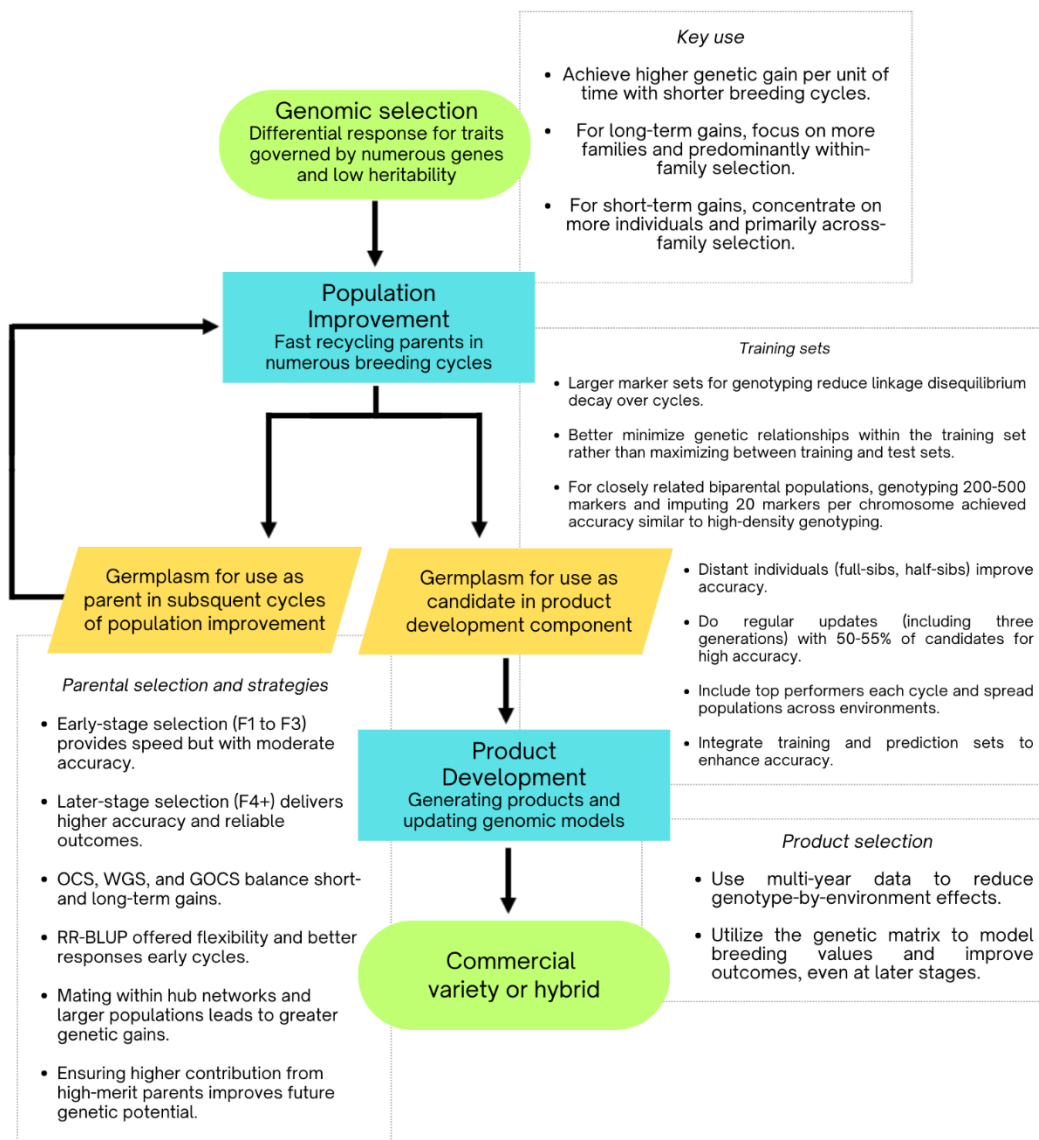
Figure 5. Flowchart illustrating practical schemes for implementing genomic selection, based on insights from simulation studies.

CHAPTER II - GENETIC GAIN AND DIVERSITY IN SOYBEAN THROUGH GENOMIC SELECTION WITH RAPID-CYCLING, SPEED BREEDING AND GENE EDITING

ABSTRACT

This study investigates various breeding strategies on genetic gain and diversity in soybean breeding programs. We simulated six distinct strategies—from phenotypic selection to advanced genomic selection integrating rapid-cycling, gene editing, and speed breeding—within a general soybean breeding framework. Grain yield was simulated under an infinitesimal model. The results demonstrate that genomic strategies consistently outperform the phenotypic approach, with average annual genetic gains of 99 kg/ha for basic genomic strategies and 110 kg/ha for the most advanced strategies, combining rapid-cycling, 50 QTL editing, and speed breeding. These strategies led to substantial increases in yield, surpassing phenotypic gains by up to 153% in the commercial deployment stage. Moreover, genomic strategies integrating rapid-cycling and gene editing showed improved stability and consistency in genetic gains across breeding cycles, with reduced fluctuations compared to basic genomic strategies. While genomic strategies resulted in higher genetic gains, they also contributed to a gradual reduction in genetic diversity, particularly in smaller cohorts at the commercial deployment stage. The most advanced strategies, such as rapid-cycling with extensive QTL editing, exhibited a higher rate of allele fixation, potentially limiting long-term adaptability. In contrast, the preliminary stage, which maintained larger population sizes, showed greater preservation of genetic diversity, emphasizing the importance of early-stage breeding practices in safeguarding future adaptability. These findings underscore the potential of genomic and gene editing technologies to accelerate genetic gain in soybean. They also highlight the need for balanced strategies that integrate genetic improvement and the conservation of diversity, ensuring long-term sustainability in crop breeding.

RESUMO

Este estudo investiga o impacto de várias estratégias de melhoramento no ganho genético e na diversidade do genoma em programas de melhoramento de soja ao

longo de 10 ciclos de melhoramento. Simulamos seis estratégias distintas — que variam desde a seleção fenotípica até técnicas genômicas avançadas integrando *rapid-cycling*, edição de genes e *speed breeding*. A população fundadora consistiu de 300 indivíduos, modelados usando um mapa genético com 6.000 sítios de segregação. O rendimento de grãos foi simulado sob um modelo infinitesimal. Os resultados demonstram que as estratégias genômicas superam consistentemente a abordagem fenotípica, com ganhos genéticos anuais médios de 99 kg/ha para estratégia genômica básica e 110 kg/ha para a estratégia mais avançada, combinando *rapid-cycling*, edição de 50 QTLs e *speed breeding*. Essa estratégia resultou em substanciais ganhos genéticos, superando os ganhos fenotípicos em até 153% na fase de lançamento comercial. Além disso, as estratégias genômicas que integraram *rapid-cycling* e edição de genes mostraram maior estabilidade e consistência no progresso genético ao longo dos ciclos de melhoramento, com flutuações reduzidas em comparação com as estratégias genômicas básicas. Embora as estratégias genômicas tenham resultado em maiores ganhos genéticos, também contribuíram para uma redução gradual na diversidade genética, particularmente em nos produtos que alcançaram a fase de lançamento comercial. As estratégias mais avançadas, como *rapid-cycling* com edição de QTLs, exibiram uma taxa maior de fixação de alelos. Em contraste, a fase preliminar, que manteve tamanhos de população maiores, mostrou maior preservação da diversidade genética. Esses resultados enfatizam o potencial das tecnologias genômicas e de edição de genes para acelerar o ganho genético em programas de melhoramento de soja. No entanto, também ressaltam a necessidade de estratégias equilibradas que integrem tanto o aprimoramento genético quanto a conservação da diversidade, garantindo a sustentabilidade a longo prazo.

INTRODUCTION

Soybean [*Glycine max* (L.)] is the most important source of protein for animal feed and a significant source of oil for biofuels and industrial applications, with Brazil, the United States, Argentina, China, India, Paraguay, and Canada as the main producing countries (USDA, 2025). In these countries and globally, soybean is the primary agricultural crop, valued for its versatility and economic returns, supporting thousands of jobs directly and indirectly.

The success of soybean cultivation is linked to the development of technologies and research. The maintenance and growth of soybean's importance depend on the efficiency and improvement of public and private breeding programs, which aim to obtain improved cultivars for grain yield, cycle, and adaptation to growing conditions—allowing scalability in different regions and profitability—and defensive traits, such as resistance to diseases and pests—ensuring the stability of grain production.

When foundational elements such as clear breeding objectives, stringent selection criteria, a rich germplasm pool, and defined target environments are established, the success of a breeding program lies in strategically utilizing resources to maximize genetic gain and minimize diversity erosion. However, achieving significant genetic gains in breeding pipelines while ensuring economic returns requires ideation and experimentation with selection approaches to determine their usefulness and cost-effectiveness in delivering a return on investment.

In this context, simulations demonstrate their utility by efficiently exploring numerous scenarios and addressing the constraints of empirical testing, which can be excessively costly and time-consuming. This capacity to replicate thousands of scenarios allows for the identification of central tendencies and provides valuable insights.

The future of plant breeding is integrated and technology-driven, with simulations playing a role in guiding and demonstrating the value of these approaches. As breeding programs evolve with the adoption of new technologies and methodologies, the need for simulations to assess their benefits and explore alternatives remains. Today, the ability of simulations to mimic diverse genetic strategies makes them essential for maximizing genetic gain, preserving genetic diversity, and optimizing resource efficiency.

We simulated and analyzed various breeding strategies within a general soybean breeding program framework over 10 breeding cycles to evaluate their impact on genetic gain and genome diversity. These strategies included both single and integrated approaches utilizing genomic selection, rapid cycling, speed breeding, and gene editing.

MATERIAL AND METHODS

Simulation parameters

The founder population comprised 300 inbred individuals, simulated using a genetic map with 20 chromosomes totaling 2,283 cM (Grant et al., 2010). Each chromosome included 300 segregating sites, resulting in a total of 6,000 sites, representing the BARCSoySNP6K array from Illumina Inc.

The simulated trait, grain yield, was selected to represent the primary objective of soybean breeding programs worldwide. We employed an infinitesimal model, assuming that 70% of the segregating sites (i.e., 210 per chromosome) had a non-zero effect on the trait. The simulations initiated with an average yield of 3,000 kg/ha and a genetic variance of eight times the average.

In this study, we considered additive, epistatic, and genotype-by-environment variance, utilizing the *addTraitAEG* function from AlphaSimR to simulate the phenotypic values.

Burn-in phase

The burn-in phase was implemented to establish a common foundation for all breeding strategies, comprising five complete cycles of a phenotypic breeding program, as described below.

In the year 1, (i) a total of 1,225 breeding crosses (F1s) were created from 300 founder parental lines; (ii) the F1 plants were self-pollinated to generate F2 populations; and, (iii) The F2 were self-pollinated to produce 25 progeny per populations, resulting in a total of 30,625 F3 progenies. During cycles 1, 2, and 3 of the burn-in phase, the parents for breeding crosses were exclusively limited to the founder population. Beginning with cycle 4, selected individuals from the preliminary stage were the parental lines, establishing a consistent selection of 300 newly selected individuals for F1s each cycle.

In the year 2, (iv) the F3 were self-pollinated to generate F4 progenies; (v) the F4 were self-pollinated to produce F5 progenies; and, (vi) each F5 was tested in progeny rows, with a heritability of 0.10 (Silva et al., 2011), resulting in the selection of 6,000 candidates to enter the first stage of multi-environment field testing, referred to as the preliminary stage.

In the years 3, 4, and 5, (vii) the preliminary stage, from headrows, involved 10 locations, selecting 300 candidates from the initial 6,000, achieving a selection intensity of 5% with a heritability of 0.30; (viii) the advanced stage was conducted with 40 locations, selecting 50 candidates from the 300 chosen in the previous stage, a 17% selection intensity with a heritability of 0.40; and, (ix) the final stage, named as the elite stage, focused on selecting 10 candidates from the 50 for commercial deployment as new cultivars, with a selection intensity of 20%, and 80 locations with a heritability of 0.40.

Rapid-cycling

After the first preliminary stage was completed and its selection became available, a rapid-cycling framework was established during cycles 4 and 5 of the burn-in phase to initiate the rapid-cycling breeding strategies in cycle 6, the beginning of the breeding phase of this study.

This framework was divided into two rounds: round 1, in which 300 candidates selected from the most recent preliminary stage were combined in 1,225 F1 crosses to generate 9,800 rapid-cycling F3 progenies (8 each cross), genomically selecting the top 150 F3; and round 2, in which the 150 rapid-cycling F3 progenies selected in round 1 were used to create 1,225 F1 crosses, resulting in an additional 9,800 rapid-cycling F3 progenies. Genomic selection was then applied to identify the top 150 progenies, which formed the round 2 of rapid-cycling F3 progenies.

Considering the appropriate years needed to obtain these parents from rapid-cycling and make them available for participation in F1 developmental crosses within the product development segment of the program, the rapid-cycling F3 from round 1 were available two cycles after starting, and the rapid-cycling F3 from round 2 were available three cycles after starting. For example, rapid-cycling round 1, derived from preliminary individuals in cycle 4 (still in the burn-in phase), became available to generate new F1s in cycle 6 (which initiates the rapid-cycling breeding strategies in this study), while rapid-cycling round 2 required one additional year to complete and became available for cycle 7. The same framework was applied throughout the years of the rapid-cycling breeding strategies in this study.

Breeding phase

After 5 cycles of burn-in, six distinct breeding strategies were implemented to represent the diverse methodologies currently and potentially employed in soybean breeding programs worldwide. The strategies were designated as follows: (i) phenotypic, (ii) genomic, (iii) genomic with rapid-cycling, (iv) genomic with double rapid-cycling, (v) genomic with rapid-cycling, 10 best QTL edited and speed breeding, (vi) genomic with rapid-cycling, 50 best QTL edited and speed breeding. These strategies were implemented from cycle 6 to 15, totaling 10 complete breeding cycles.

To enable comparison among breeding strategies, the parentals for the 1,225 balanced F1 crosses consisted of the 300 individuals from the most recent preliminary trial available in each breeding cycle, with each cross producing 25 F3 progenies. The same pipeline sizes were maintained across the stages for the methods examined: 6,000 candidates at the preliminary stage, 300 candidates at the advanced stage, 50 candidates at the elite stage, and finally, 10 commercial deployments selected from the elite. Further details are provided in the following paragraphs, as well as in Supplemental Tables 1 to 5 and Figure 1.

The phenotypic strategy extended the breeding program from the burn-in phase, maintaining consistent parameters. This approach applied phenotypic selection at each stage, from the progeny test (30,625 F5 progenies) to the commercial release of 10 new cultivars per year within a 5-year framework.

In turn, the genomic strategy relied entirely on genomic estimated breeding values (GEBVs) derived from the most informative training set, which included only the lines used as parents and incorporated up-to-date phenotypic data. Selection at the F3 stage (n=30,625) for entry into the preliminary stage (n=6,000), as well as candidate advancement to advanced trials (n=300), elite trials (n=50), and ultimately commercial deployment (n=10), was guided by GEBVs, with genomic selection fully replacing phenotypic selection at each stage.

As an extension, the genomic strategy with rapid-cycling retained all elements of the genomic strategy while integrating the key advantage of rapid-cycling frameworks—the continuous availability of newly recycled parents for breeding crosses. In this approach, the number of parents used in breeding crosses remained the same (n=300); however, the composition changed, with half of the parents

(n=150) selected from the most recent preliminary trial and the other half (n=150 F3) derived from round 1 of rapid-cycling.

Following that, the genomic strategy with double rapid-cycling preserved the core elements of the previous strategy while incorporating an additional feature— parental selection from round 2 of rapid-cycling. In this approach, the parental composition for breeding crosses was adjusted to include one-third (n=100) from the most recent preliminary trial, one-third (n=100) from round 1 of rapid-cycling, and the remaining one-third (n=100) from round 2 of rapid-cycling.

Additionally, two other strategies were developed based on the genomic strategy with rapid-cycling: genomic with rapid-cycling, 10 QTL edited, and speed breeding, and genomic with rapid-cycling, 50 QTL edited, and speed breeding. The key distinction of these strategies from the base strategy is the incorporation of a maximum 90-day crop cycle for rapid-cycling, allowing for a gene editing process for 10 and 50 QTL, with posterior use of the edited plants for breeding crosses. In these strategies, the parental composition for breeding crosses was adjusted to include half (n=150) from the most recent preliminary trial and half (n=150) from individuals edited for 10 and 50 QTLs from round 1 of rapid-cycling.

Analysis and assessment

Each strategy was executed over 10 cycles following the burn-in phase (cycles 6 to 15), with each strategy repeated 50 times to capture central trends. Throughout the strategies, cycles, and repetitions, we recorded allele frequencies, genetic values, phenotypic values, estimated breeding values, genetic variance, and accuracy at each stage.

For this study, we emphasized the analysis and discussion of (i) the genetic values (in kg/ha) and (ii) the allele frequencies (percentage of loci retaining more than one allele) in lines selected for commercial deployment (n=10) and preliminary stage (n=6000). The phenotypic strategy served as the baseline for comparisons.

Data analysis and handling were enabled in R (R Core Team, 2024) and spreadsheets. Simulations were programmed using the package AlphaSimR and the code is available on https://github.com/rafaelaugvieira/Breeding-strategies-simulations.

RESULTS

The data presents a comparative analysis of genetic gains across breeding strategies in commercial deployment and preliminary stages, quantifying percentage increases over baseline values, annual genetic gains, and temporal trends (Table 1; Figures 2 and 3).

In the commercial deployment stage, the phenotypic baseline achieved an average annual genetic gain of 62 kg/ha (1.6% per year), whereas genomic strategies resulted in higher gains. The basic genomic approach increased yield by 3.1% over the average of cycles 6–15 and by 8.7% by the end of cycle 15, corresponding to an average annual genetic gain of 99 kg/ha (2.7% per year).

Integrating rapid-cycling further increased genetic gains, yielding a 4.7% increase over the average of cycles 6–15 and a 9.3% increase by the end of cycle 15, with an average annual genetic gain of 102 kg/ha. Similarly, double rapid-cycling produced a 4.7% increase in cycles 6–15 and a 9.4% increase by the end of cycle 15, maintaining an average gain of 102 kg/ha (2.7% per year).

The addition of 10 QTL editing and speed breeding resulted in an average annual genetic gain of 103 kg/ha, with increases of 4.8% in cycles 6–15 and 9.5% in cycle 15. The highest genetic improvement was achieved through the genomic strategy integrating rapid-cycling, 50 QTL editing, and speed breeding, which increases yield by 5.7% in the average of cycles 6–15 and 11.2% by cycle 15, corresponding to an average annual genetic gain of 110 kg/ha (2.9%).

In the preliminary stage, genomic strategies consistently outperformed the phenotypic approach (Table 1). The phenotypic baseline exhibited an annual genetic gain of 26 kg/ha (1.8%), whereas the basic genomic strategy increased yield by 4.5% in cycles 6–15 average and 10.9% in cycle 15, resulting in an average annual genetic gain of 69 kg/ha (3.1%).

The incorporation of rapid-cycling further slightly improved results, achieving a 5.7% increase in cycles 6–15 and 11.1% in cycle 15 over the baseline, leading to an average annual genetic gain of 70 kg/ha (3.2%). Double rapid-cycling produced similar results, with a 5.5% increase in cycles 6–15 and 10.8% in cycle 15, maintaining an annual gain of 69 kg/ha. The combination of rapid-cycling, 10 QTL editing, and speed breeding yielded equivalent results, with a 5.7% increase in cycles 6–15 over the baseline.

The combination of rapid-cycling, 10 QTL editing, and speed breeding results in a 5.7% increase in the average of cycles 6–15 and 11.1% by cycle 15, yielding an average annual genetic gain of 70 kg/ha. The most significant improvement in the preliminary stage was again achieved with the genomic strategy incorporating rapid-cycling, 50 QTL editing, and speed breeding. This approach increases yield by 6.3% in the average of cycles 6–15 and 12.6% by cycle 15, resulting in an average annual genetic gain of 76 kg/ha (3.3%).

A comparison of the average of cycles 6–15 and those at the end of cycle 15 reveals notable differences in percentage increases over the baseline. In both the preliminary and commercial deployment stages, genomic strategies yielded gains of 3.1% to 6.3% over the baseline in cycles 6–15, demonstrating their overall superiority to the baseline. By cycle 15, these gains became more pronounced, ranging from 8.7% to 12.6%, indicating that genetic improvements compound over successive cycles, leading to greater advancements in later stages.

The genomic strategy that combines rapid-cycling, 50 QTL editing, and speed breeding demonstrates superior performance, achieving genetic gains of 110 kg/ha (deployments) and 76 kg/ha (preliminary). This approach outperforms all alternative strategies, delivering an additional 7 and 11 kg/ha—representing a 10-11% increase—compared to the basic genomic strategy. Furthermore, relative to the phenotypic baseline, the strategy yields an additional 40 and 48 kg/ha (preliminary and deployments, respectively), resulting in genetic gains that are 77% and 153% higher. Relative to the rapid-cycling and double rapid-cycling strategies, it yields an 8 kg/ha (8%) increase. Similarly, it surpasses the genomic strategy incorporating rapid-cycling, 10 QTL editing, and speed breeding by 7 kg/ha (7%).

The temporal patterns in Figures 2 and 3 illustrate these trends across breeding cycles. In both commercial deployment and preliminary stages, the phenotypic baseline exhibited genetic gains in increments occurring approximately every three cycles, whereas the basic genomic strategy showed similar increases every two cycles. These increments approached 200 kg/ha, followed by extended periods of minimal change, indicating stability with limited genetic progress.

While the average genetic gain remained comparable across the genomic strategies (Table 1), the inclusion of rapid cycling and/or editing helped mitigate fluctuations, thereby ensuring more stable and consistent genetic progress (Figures

2 and 3). This stability is particularly evident when comparing the percentage improvements in the cycles 8 to 15, where genomic strategies, especially those integrating rapid-cycling, QTL editing, and speed breeding, showed pronounced gains, reinforcing their potential to enhance the reliability and stability of breeding programs over time. It quantifies the well-established importance of consistently introducing new and superior parental lines into breeding crosses, thereby preventing periods of stagnation in genetic gain.

Over time, the disparity between basic genomic and phenotypic strategies, as well as between rapid-cycling-based and basic genomic strategies, increased. After four years of deployment based on these strategies, fewer instances of overlap or convergence in genetic values were observed, highlighting the time required for their differentiation (Figure 2). Rapid-cycling, double rapid-cycling, and rapid-cycling with 10 QTL edits exhibited consistent patterns across both commercial deployments and preliminary stages (Figures 2 and 3). Notably, the genomic strategy integrating rapid-cycling, 50 QTL editing, and speed breeding distinguished from other genomic-based strategies earlier in commercial deployment (cycle 10) than in the preliminary stage (cycle 12).

The data presented in Table 2 examines the genetic diversity across various breeding strategies at different stages of commercial deployment and preliminary stages. For commercial deployments, the phenotypic baseline maintained a high percentage of loci with more than one allele within the group of individuals, averaging approximately 95% for both cycle 6-15 average and cycle 15, indicating no loss of diversity. In contrast, genomic-based strategies showed higher allele fixation percentages. Specifically, the genomic strategy had an average of 73.7% of loci with more than one allele in the average of cycle 6-15, which decreased to 63.2% by cycle 15, reflecting an average annual reduction in diversity of 3.2% since the burn-in phase (cycle 5 with 95%).

Although the exact percentages showed minor variations, similar trends were observed for the genomic with rapid-cycling, genomic with double rapid-cycling, and genomic with 10 QTL edits and speed breeding strategies (Table 2). Notably, the genomic strategy with 50 QTL edits and speed breeding demonstrated the highest allele fixation, with 56.8% of loci showing more than one allele by the end of Cycle 15, indicating an average annual reduction in diversity of 3.8% since the burn-in

phase. This pattern is further illustrated in Figure 4, where the phenotypic baseline strategy consistently showed lower loci with more than one allele throughout the breeding phase.

In the preliminary stages, where a larger number of individuals (n=6000) were used compared to the deployment stage (n=10), the phenotypic baseline preserved 100% of loci with more than one allele within the group of individuals, indicating no reduction in diversity (Table 2 and Figure 5). Similarly, other strategies exhibited high percentages of loci with more than one allele (95.1-98.4%), which remained relatively stable throughout cycle 6-15 and by the end of cycle 15, with only minor reductions observed. For instance, the genomic strategy with 50 QTL edits and speed breeding showed a slight decrease from 100.0% in cycle 5 (burn-in) to 97.6% in cycle 15, reflecting an average annual allele fixation of 0.2% (Table 2).

Unlike the genetic value chart, no significant leaps are observed when examining the loci with more than one allele. Furher, trends in Figure 5 demonstrated that most genomic strategies similarly affect the percentage of loci with more than one allele within the preliminary stage group of individuals, with the basic genomic strategy showing the lowest values. However, these values remain within acceptable thresholds, averaging approximately 95%, thereby enabling breeding opportunities that can potentially address future demands.

DISCUSSION

An extensive body of research has demonstrated the advantages of genomic approaches outperforming phenotypic selection for traits with low heritability and polygenic architecture, such as yield (Gaynor et al., 2017; Muleta et al., 2019; Ali et al., 2020; Silva et al., 2021; Lubanga et al., 2022; Chiaravallotti et al., 2023; Fritsche-Neto et al., 2024). Consistent with these findings, this study confirms the benefits of genomic strategies; however, the variety of genomic approaches explored here provides further insights into the balance between genetic gain and genetic diversity, enabling a more precise fine-tuning of this trade-off.

Simulation-based analyses frequently produce higher accuracies than observed in real-world breeding programs due to the assumption of ideal conditions, such as error-free molecular markers, the absence of epistatic interactions, and restricted germplasm exchange (Gaynor et al., 2017). Simulations also often depend on

simplified assumptions, such as the absence of genotype-environment interactions and complete linkage maps, which may not fully reflect the complexities inherent in real-world breeding. In this study, the accuracies of genomic estimated breeding values ranged between 0.20 and 0.30 (data not shown), aligning with real-world expectations by incorporating epistatic interactions and genotype-by-environment effects alongside the traditional additive effects of the simulated trait, yield. As a result, the observed annual genetic gains in commercial deployment stage were highly feasible, varying from 1.6% in the phenotypic baseline to 2.9% in the most advanced genomic-based strategy, which combined rapid cycling, gene editing, and speed breeding.

In general, genomic approaches incorporating rapid-cycling yielded more consistent genetic gains while maintaining a similar level of diversity loss compared to basic genomic strategy (without rapid-cycling). Rapid-cycling strategies facilitated steady, incremental genetic improvements rather than the abrupt leaps observed in basic genomic and phenotypic approaches. However, the increased operational complexity associated with rapid-cycling results in higher costs, which must be evaluated based on the institutional context to ensure the intended benefits are realized.

Thoughtful interpretation is essential when translating these simulation results into practical applications. Effective application of simulation insights requires empirical validation through field data, iterative refinement of models, and the integration of feedback mechanisms to enhance predictive accuracy. Notably, simulated genetic gain represents an upper theoretical bound, which is frequently diminished in practical implementation due to various constraints and selection attrition. Strategic investment decisions in breeding and R&D should prioritize the enhancement of capabilities that have the potential to facilitate the realization of these theoretical gains, such as advanced genotyping and editing capabilities, IT-driven enterprise analytics, high-capacity phenotyping, and multi-season research centers.

The depletion of genetic diversity, particularly through allele fixation at targeted loci, was more frequently observed in cohorts with smaller population sizes subjected to intense selection pressure across successive breeding stages, particularly at the commercial deployment stage (n=10). This narrowing of the

genetic base reaching the market may increase vulnerability to climatic variability and its associated risks. Conversely, preliminary testing, as the initial stage within the designed breeding framework, retained sufficient variation at both the allele level, offering potential escape routes in response to unforeseen challenges.

From a genetic gain perspective, the integration of gene editing at a rate of 10 QTL per cycle did not yield significant differentiation from single or double rapid-cycling strategies without editing. In contrast, editing 50 QTL per cycle demonstrated a clear advantage in genetic gain. Nevertheless, the necessary association with speed breeding protocols highlights that both approaches necessitate robust analytical and operational capabilities to effectively translate these benefits into breeding programs.

Emerging from the limitations of marker-assisted selection (MAS), genomic selection has become a foundational methodology in modern plant breeding. Since its introduction, computer simulations have been essential for demonstrating and optimizing genomic selection to achieve greater and more sustainable genetic gains (Bernardo, 1999; Meuwissen et al., 2001). Similarly, gene editing has emerged, gaining momentum and scale with CRISPR-Cas9. Our simulation study was the first to incorporate gene editing within a general, applicable breeding program framework, and the results proved positive.

To date, numerous applications of gene editing in plants have demonstrated practical value, primarily targeting traits controlled by a few genes with high heritability (Yao et al., 2023). Similar to how genomic selection advanced beyond marker-assisted selection, the effective application of gene editing for low-heritability traits, such as yield, could represent a breakthrough and a transformative step in plant breeding, particularly for row crops and with numerous gens being edited. A broader or decentralized gene editing approach that enables the simultaneous modification of multiple genes in a practical and cost-effective manner could redefine modern breeding practices and positively influence the development of resilient agricultural systems capable of withstanding the pressures of contemporary crop production. The results of our study are promising and informative, demonstrating that editing 50 QTLs yielded meaningful improvements, whereas editing 10 QTLs did not produce relevant effects.

Modern plant breeding have been driven by the integration of technologies such as marker-assisted selection, genomic selection, selection indexes, gene editing, artificial intelligence, process automation, data analytics, epigenomics, phenomics, enviromics, and biotechnology. As breeding programs evolve and adopt new methodologies, the demand for simulations to assess their benefits and explore alternatives remains critical in providing insights that help refine breeding pipelines and maximize genetic gains. Our study contributes to the ongoing modernization of breeding, offering empirical data and insights that underscore the importance of persistently challenging the status quo for the benefit of farmers and society.

REFERENCES

Ali, M., Zhang, L., DeLacy, I., Arief, V., Dieters, M., Pfeiffer, W. H., et al. (2020). Modeling and simulation of recurrent phenotypic and genomic selections in plant breeding under the presence of epistasis. Crop J., 8, 866–877. doi: 10.1016/j.cj.2020.04.002

Bernardo, R. (1999). Marker-assisted best linear unbiased prediction of single-cross performance. Crop Sci., 39, 1277–1282.

Chiaravallotti, I., Lin, J., Arief, V., Jahufer, Z., Osorno, J. M., McClean, P., et al. (2023). Simulations of multiple breeding strategy scenarios in common bean for assessing genomic selection accuracy and model updating. The Plant Genome, 17. doi: 10.1002/tpg2.20388

Fritsche-Neto, R., Ali, J., De Asis, E. J., Allahgholipour, M., and Labroo, M. R. (2024). Improving hybrid rice breeding programs via stochastic simulations: number of parents, number of hybrids, tester update, and genomic prediction of hybrid performance. Theor. Appl. Genet., 137. doi: 10.1007/s00122-023-04508-6

Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., et al. (2017). A two-part strategy for using genomic selection to develop inbred lines. Crop Sci., 56, 2372–2386. doi: 10.2135/cropsci2016.09.0742

Lubanga, N., Massawe, F., Mayes, S., Gorjanc, G., and Bančić, J. (2022). Genomic selection strategies to increase genetic gain in tea breeding programs. The Plant Genome, 16. doi: 10.1002/tpg2.20282

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Silva, E. D. B., Xavier, A., and Faria, M. V. (2021). Impact of genomic prediction model, selection intensity, and breeding strategy on the long-term genetic gain and genetic erosion in soybean breeding. Front. Plant Sci., 12. doi: 10.3389/fgene.2021.637133

SoyBase, the USDA-ARS soybean genetics and genomics database. Grant, D., Nelson, R. T., Cannon, S. B., Shoemaker, R. C. (2010). Nucleic Acids Research, 38(suppl_1), D843–D846. https://doi.org/10.1093/nar/gkp798

USDA (2025). Top Producing Countries - 2024/2025 Soybeans Production. https://www.fas.usda.gov/data/production/commodity/2222000

Yao, D., Zhou, J., Zhang, A., Wang, J., Liu, Y., Wang, L., Pi, W., Li, Z., Yue, W., Cai, J., Liu, H., Hao, W., Qu, X. (2023). Advances in CRISPR/Cas9-based research related to soybean [Glycine max (Linn.) Merr] molecular breeding. Front. Plant Sci., 14, 1247707. doi: 10.3389/fpls.2023.1247707

Table 1. Genetic values and gains across breeding strategies in commercial deployment and preliminary stages.

Strategy	Genetic value (kg/ha)			% Over baseline Cycle 6-15	% Over baseline Cycle 15	Genetic gain per year (%)
	Cycle 5*	Cycle 6-15	Cycle 15			
Commercial deployments						
Phenotypic (baseline)		4106	4348	100.0%	100.0%	62 kg/ha (1.6%)
Genomic		4233	4724	103.1%	108.7%	99 kg/ha (2.7%)
Genomic w/ rapid-cycling	3733	4299	4752	104.7%	109.3%	102 kg/ha (2.7%)
Genomic w/ double rapid-cycling		4300	4755	104.7%	109.4%	102 kg/ha (2.7%)
Genomic w/ rapid-cycling, 10 QTL ed. and speed bree.		4303	4762	104.8%	109.5%	103 kg/ha (2.8%)
Genomic w/ rapid-cycling, 50 QTL ed. and speed bree.		4341	4833	105.7%	111.2%	110 kg/ha (2.9%)
Preliminary stage						
Phenotypic (baseline)		3745	3989	100.0%	100.0%	26 kg/ha (1.8%)
Genomic		3913	4424	104.5%	110.9%	69 kg/ha (3.1%)
Genomic w/ rapid-cycling	3370	3957	4431	105.7%	111.1%	70 kg/ha (3.2%)
Genomic w/ double rapid-cycling		3952	4418	105.5%	110.8%	69 kg/ha (3.1%)
Genomic w/ rapid-cycling, 10 QTL ed. and speed bree.		3960	4430	105.7%	111.1%	70 kg/ha (3.1%)
Genomic w/ rapid-cycling, 50 QTL ed. and speed bree.		3982	4493	106.3%	112.6%	76 kg/ha (3.3%)

*End of burn-in phase.
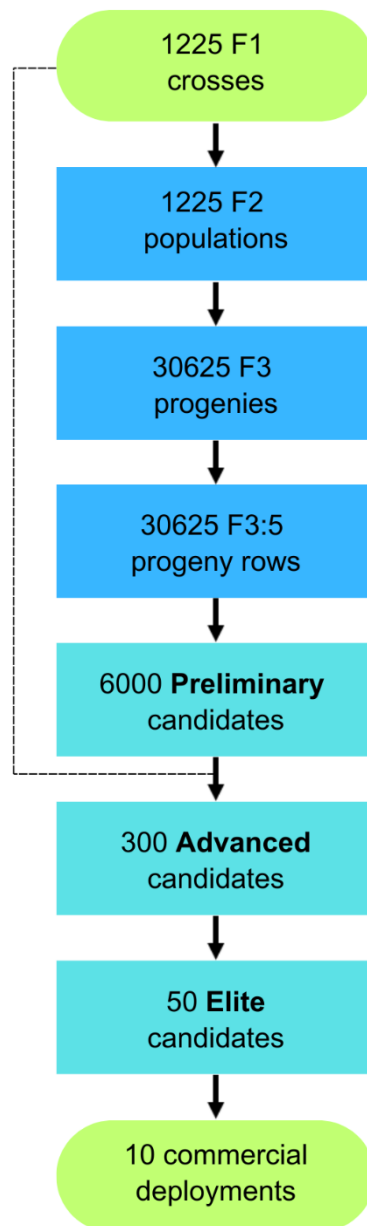
Figure 1. Flowchart demonstrating a general soybean breeding program framework, including breeding stages and parental selection.
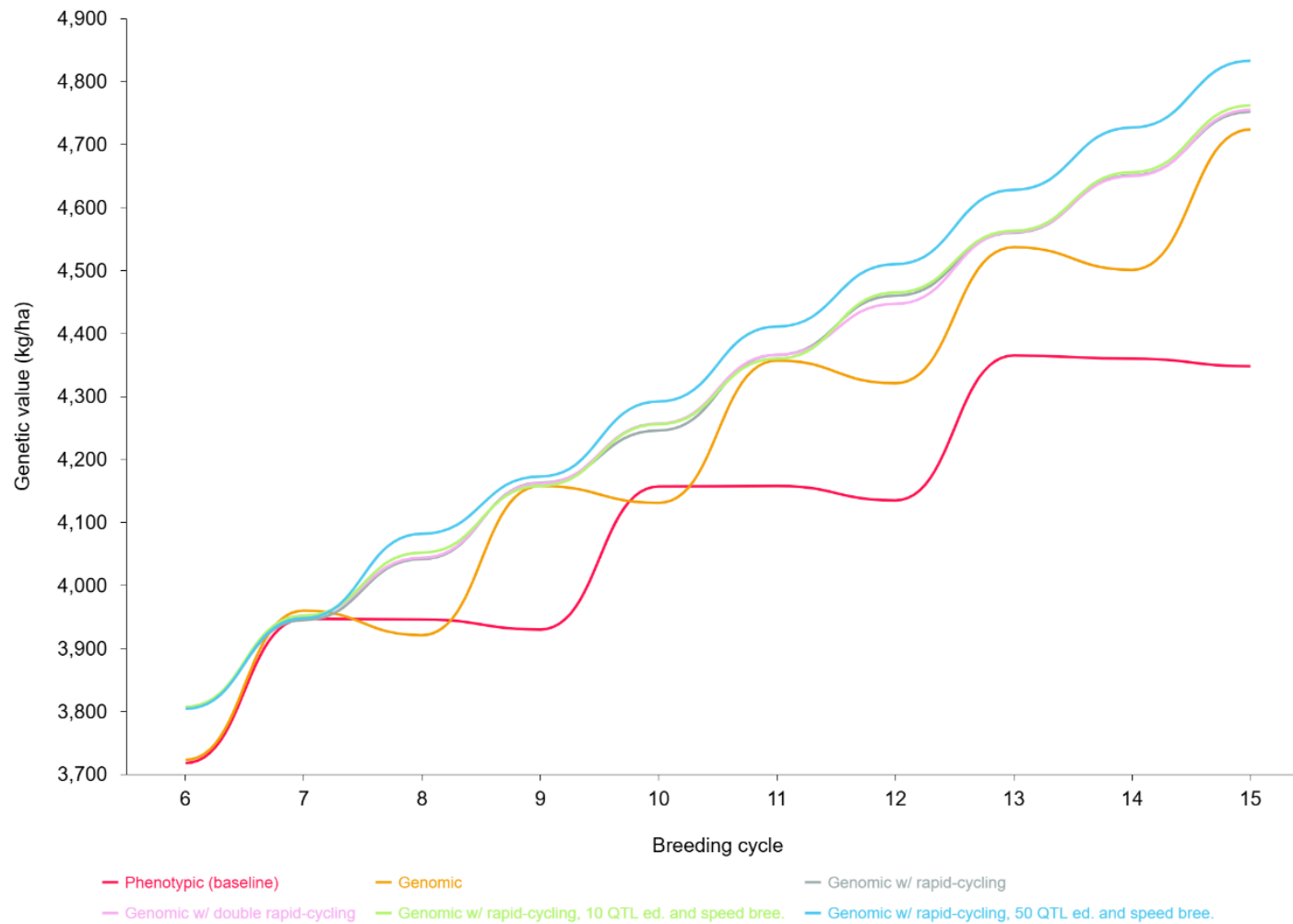
Figure 2. Temporal patterns of genetic value in commercial deployments across successive breeding cycles under different strategies.
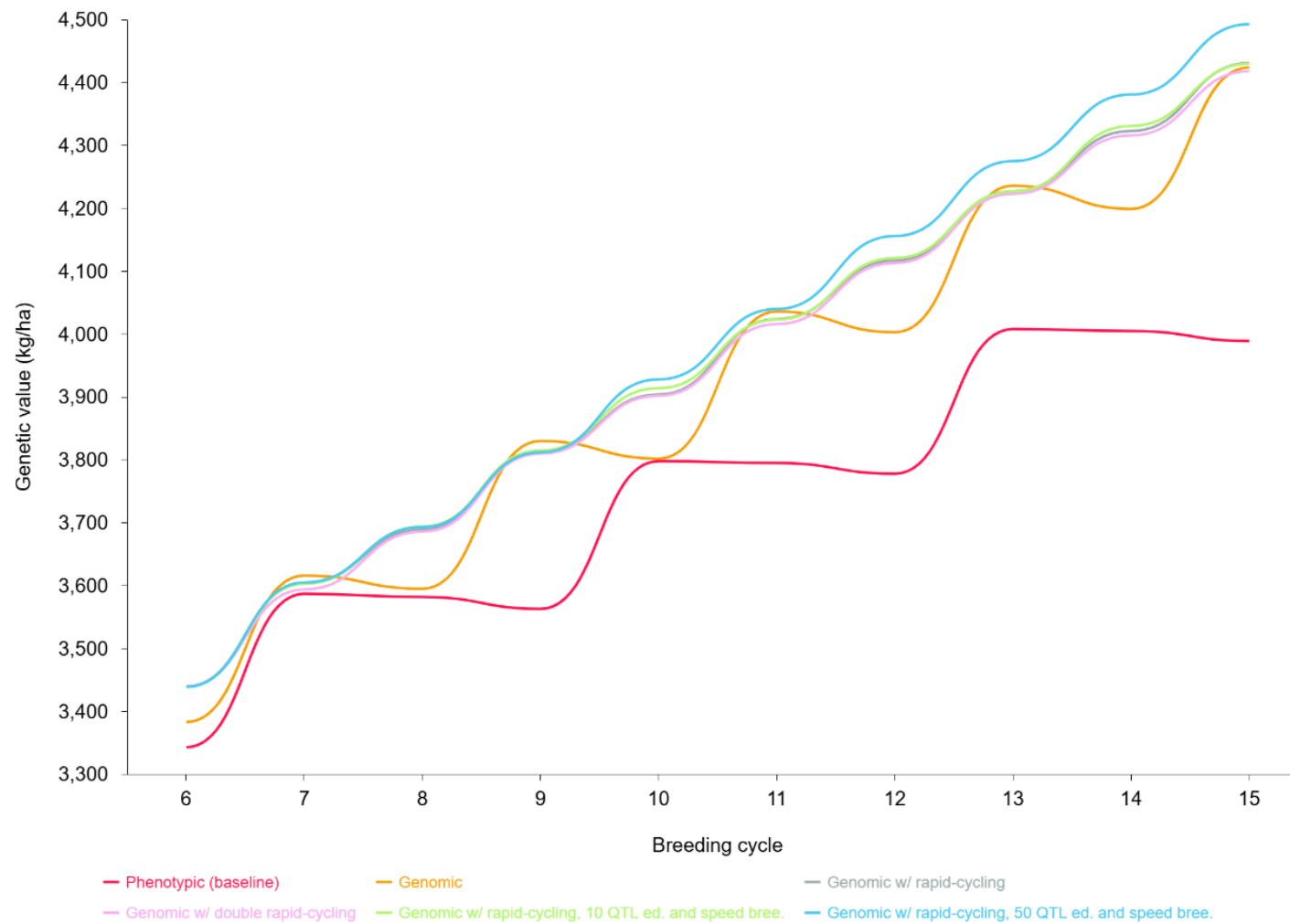
Figure 3. Temporal patterns of genetic value in preliminary stage across successive breeding cycles under different strategies.

Table 2. Genetic diversity across breeding strategies in commercial deployment and preliminary stages.

Strategy	Loci with more than one allele			Allele fixation per year (%)
	Cycle 5*	Cycle 6-15	Cycle 15	
Commercial deployments				
Phenotypic (baseline)		95.9%	95.3%	0.0%
Genomic		73.7%	63.2%	3.2%
Genomic w/ rapid-cycling	95.0%	73.4%	63.3%	3.2%
Genomic w/ double rapid-cycling		72.5%	59.9%	3.5%
Genomic w/ rapid-cycling, 10 QTL ed. and speed bree.		73.5%	65.5%	3.0%
Genomic w/ rapid-cycling, 50 QTL ed. and speed bree.		67.2%	56.8%	3.8%
Preliminary stage				
Phenotypic (baseline)		100.0%	100.0%	0.0%
Genomic		98.7%	95.1%	0.5%
Genomic w/ rapid-cycling	100.0%	99.6%	98.3%	0.2%
Genomic w/ double rapid-cycling		99.5%	98.0%	0.2%
Genomic w/ rapid-cycling, 10 QTL ed. and speed bree.		99.6%	98.4%	0.2%
Genomic w/ rapid-cycling, 50 QTL ed. and speed bree.		99.5%	97.6%	0.2%

*End of burn-in phase.
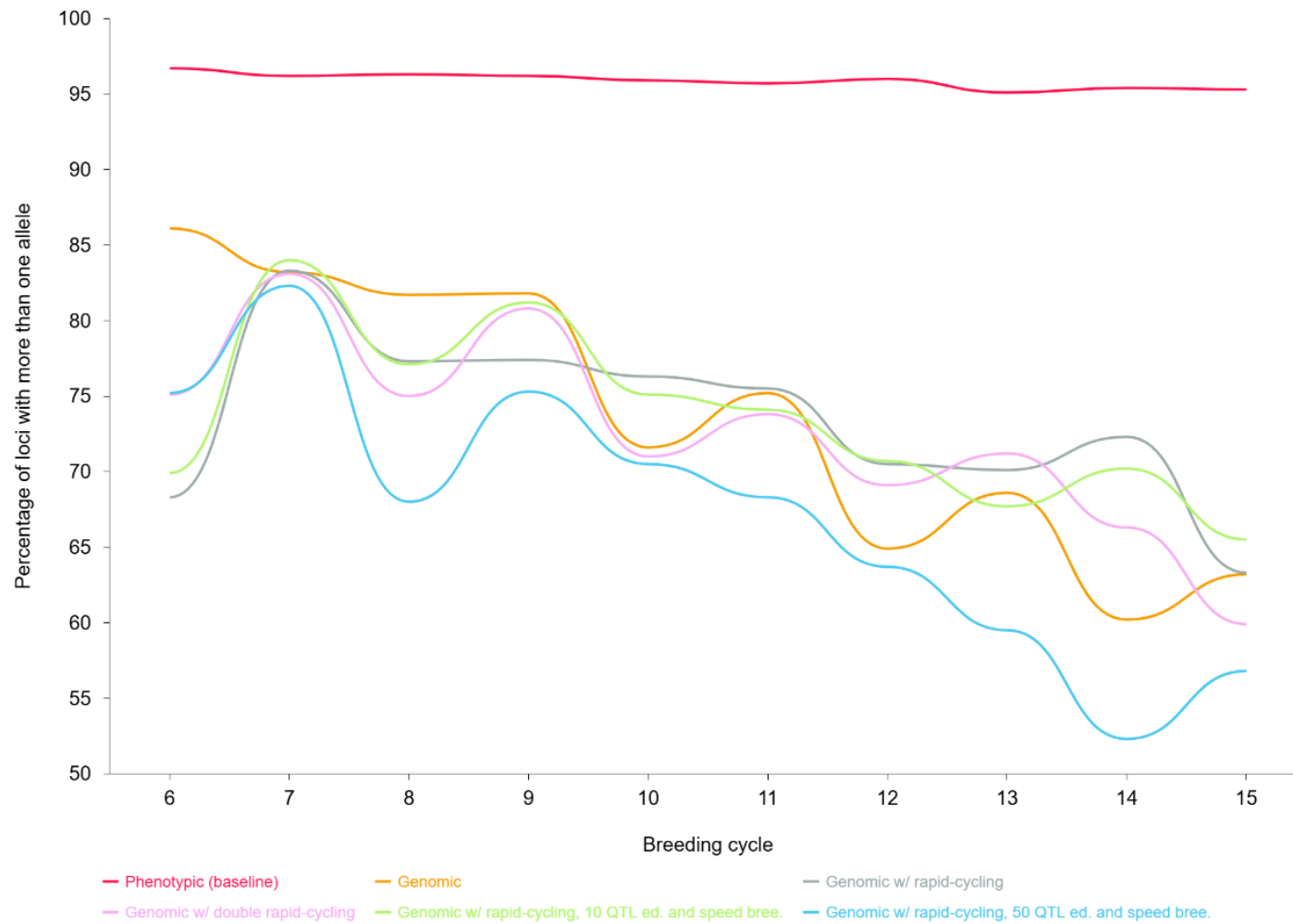
Figure 4. Temporal patterns of loci with more than one allele in commercial deployments across successive breeding cycles under different strategies.
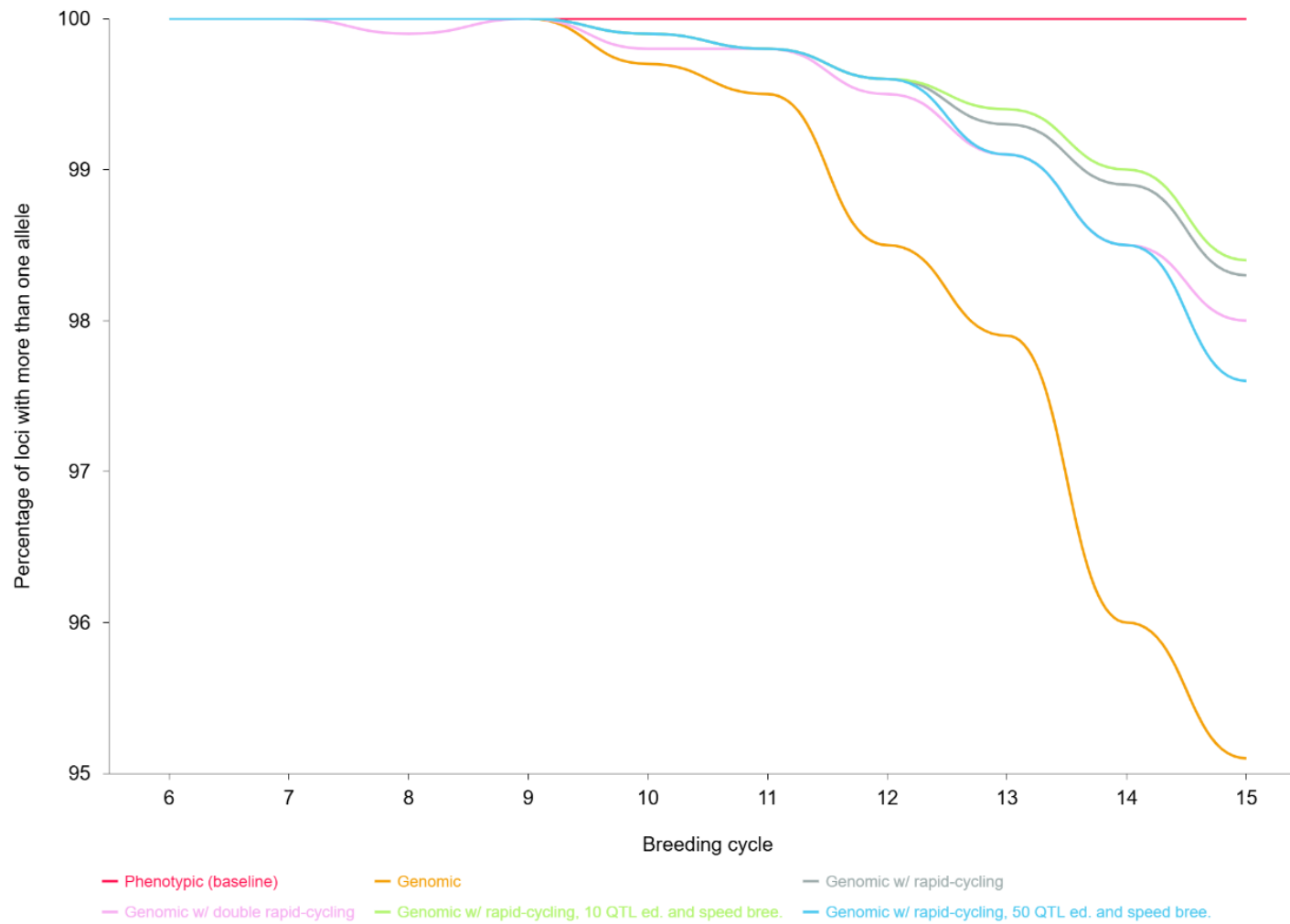
Figure 5. Temporal patterns of loci with more than one allele in preliminary stage across successive breeding cycles under different strategies.

Supplemental table 1. Genomic strategy: multi-cycle development, parental selection, and advancement timeline.

| Cycle	Parents for F1	Training for F3	year 1	year 2	year 3	year 4	year 5	year 6	year 7	year 8	year 9	year 10	year 11	year 12	year 13	year 14	year 15	year 16	year 17	year 18	year 19
1	cycle1Par	NA	F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com													
2	cycle2Par	NA		F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com												
3	cycle3Par	NA			F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com											
4	cycle1PreS	NA				F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com										
5	cycle2PreS	NA					F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com									
6	cycle3PreS	cycle3PreS						F1, F2, F3	F4, F5, Pre	Adv	Eli	Com									
7	cycle4PreS	cycle4PreS							F1, F2, F3	F4, F5, Pre	Adv	Eli	Com								
8	cycle6PreS	cycle6PreS								F1, F2, F3	F4, F5, Pre	Adv	Eli	Com							
9	cycle7PreS	cycle7PreS									F1, F2, F3	F4, F5, Pre	Adv	Eli	Com						
10	cycle8PreS	cycle8PreS										F1, F2, F3	F4, F5, Pre	Adv	Eli	Com					
11	cycle9PreS	cycle9PreS											F1, F2, F3	F4, F5, Pre	Adv	Eli	Com				
12	cycle10PreS	cycle10PreS												F1, F2, F3	F4, F5, Pre	Adv	Eli	Com			
13	cycle11PreS	cycle11PreS													F1, F2, F3	F4, F5, Pre	Adv	Eli	Com		
14	cycle12PreS	cycle12PreS														F1, F2, F3	F4, F5, Pre	Adv	Eli	Com	
15	cycle13PreS	cycle13PreS															F1, F2, F3	F4, F5, Pre	Adv	Eli	Com

Supplemental table 2. Genomic with rapid-cycling strategy: multi-cycle development, parental selection, and advancement timeline.

Cycle	Parents for F1	Training for F3	year 1	year 2	year 3	year 4	year 5	year 6	year 7	year 8	year 9	year 10	year 11	year 12	year 13	year 14	year 15	year 16	year 17	year 18	year 19
1	cycle1Par (Normal)	NA	F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com													
2	cycle2Par (Normal)	NA		F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com												
3	cycle3Par (Normal)	NA			F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com											
4	cycle1PreS (Normal) cycle2PreS (Cycling)	NA (Normal) cycle2PreS (Cycling)				F1, F2, F3	F4, F5, Pro F1RC, F2RC, F3RC	Pre	Adv	Eli	Com										
5	cycle2PreS (Normal) cycle3PreS (Cycling)	NA (Normal) cycle3PreS (Cycling)					F1, F2, F3	F4, F5, Pro F1RC, F2RC, F3RC	Pre	Adv	Eli	Com									
6	cycle3PreS, cycle4CS (Normal) cycle4PreS (Cycling)	cycle3PreS, cycle2PreS (Normal) cycle4PreS (Cycling)						F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com									
7	cycle4PreS, cycle5CS (Normal) cycle6PreS (Cycling)	cycle4PreS, cycle3PreS (Normal) cycle6PreS (Cycling)							F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com								
8	cycle6PreS, cycle6CS (Normal) cycle7PreS (Cycling)	cycle6PreS, cycle4PreS (Normal) cycle7PreS (Cycling)								F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com							
9	cycle7PreS, cycle7CS (Normal) cycle8PreS (Cycling)	cycle7PreS, cycle6PreS (Normal) cycle8PreS (Cycling)									F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com						
10	cycle8PreS, cycle8CS (Normal) cycle9PreS (Cycling)	cycle8PreS, cycle7PreS (Normal) cycle9PreS (Cycling)										F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com					
11	cycle9PreS, cycle9CS (Normal) cycle10PreS (Cycling)	cycle9PreS, cycle8PreS (Normal) cycle10PreS (Cycling)											F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com				
12	cycle10PreS, cycle10CS (Normal) cycle11PreS (Cycling)	cycle10PreS, cycle9PreS (Normal) cycle11PreS (Cycling)												F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com			
13	cycle11PreS, cycle11CS (Normal) cycle12PreS (Cycling)	cycle11PreS, cycle10PreS (Normal) cycle12PreS (Cycling)													F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com		
14	cycle12PreS, cycle12CS (Normal) cycle13PreS (Cycling)	cycle12PreS, cycle11PreS (Normal) cycle13PreS (Cycling)														F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com	
15	cycle13PreS, cycle13CS (Normal) cycle14PreS (Cycling)	cycle13PreS, cycle12PreS (Normal) cycle14PreS (Cycling)															F1, F2, F3	F4, F5, Pre F1C, F2C, F3C	Adv	Eli	Com

Supplemental table 3. Genomic with double-cycling strateg: multi-cycle development, parental selection, and advancement timeline.

Cycle	Parents for F1	Training for F3	year 1	year 2	year 3	year 4	year 5	year 6	year 7	year 8	year 9	year 10	year 11	year 12	year 13	year 14	year 15	year 16	year 17	year 18	year 19
1	cycle1Par (Normal)	NA	F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com													
2	cycle2Par (Normal)	NA		F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com												
3	cycle3Par (Normal)	NA			F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com											
4	cycle1PreS (Normal); cycle2PreS (Cycling)	NA (Normal); cycle2PreS (Cycling)					F1, F2, F3	F4, F5, Pro; F1RC, F2RC, F3RC*	Pre	Adv	Eli	Com									
5	cycle2PreS (Normal); cycle3PreS (Cycling)	NA (Normal); cycle3PreS (Cycling)						F1, F2, F3	F4, F5, Pro; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Pre	Adv	Eli	Com								
6	cycle3PreS, cycle4CS (Normal); cycle4PreS (Cycling)	cycle3PreS, cycle2PreS (Normal); cycle4PreS (Cycling)							F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com								
7	cycle4PreS, cycle5CS (Normal); cycle6PreS (Cycling)	cycle4PreS, cycle3PreS (Normal); cycle6PreS (Cycling)								F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com							
8	cycle6PreS, cycle6CS (Normal); cycle7PreS (Cycling)	cycle6PreS, cycle4PreS (Normal); cycle7PreS (Cycling)									F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com						
9	cycle7PreS, cycle7CS (Normal); cycle8PreS (Cycling)	cycle7PreS, cycle6PreS (Normal); cycle8PreS (Cycling)										F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com					
10	cycle8PreS, cycle8CS (Normal); cycle9PreS (Cycling)	cycle8PreS, cycle7PreS (Normal); cycle9PreS (Cycling)											F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com				
11	cycle9PreS, cycle9CS (Normal); cycle10PreS (Cycling)	cycle9PreS, cycle8PreS (Normal); cycle10PreS (Cycling)												F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com			
12	cycle10PreS, cycle10CS (Normal); cycle11PreS (Cycling)	cycle10PreS, cycle9PreS (Normal); cycle11PreS (Cycling)													F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com		
13	cycle11PreS, cycle11CS (Normal); cycle12PreS (Cycling)	cycle11PreS, cycle10PreS (Normal); cycle12PreS (Cycling)														F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com	
14	cycle12PreS, cycle12CS (Normal); cycle13PreS (Cycling)	cycle12PreS, cycle11PreS (Normal); cycle13PreS (Cycling)															F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com
15	cycle13PreS, cycle13CS (Normal); cycle14PreS (Cycling)	cycle13PreS, cycle12PreS (Normal); cycle14PreS (Cycling)															F1, F2, F3	F4, F5, Pre; F1RC, F2RC, F3RC*; F1RC, F2RC, F3RC**	Adv	Eli	Com

* Rapid-cycling from the given cycle. ** Rapid-cycling round 2 (initiated with F3RC from round 1) of the preceding cycle.

Supplemental table 4. Genomic with rapid-cyling, gene editing and speed breeding strategy (10 and 50 QTL edited): multi-cycle development, parental selection, and advancement timeline.

Cycle	Parents for F1	Training for F3	year 1	year 2	year 3	year 4	year 5	year 6	year 7	year 8	year 9	year 10	year 11	year 12	year 13	year 14	year 15	year 16	year 17	year 18	year 19
1	cycle1Par (Normal)	NA	F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com													
2	cycle2Par (Normal)	NA		F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com												
3	cycle3Par (Normal)	NA			F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com											
4	cycle1PreS (Normal) cycle2PreS (Cycling)	NA (Normal) cycle2PreS (Cycling)				F1, F2, F3	F4, F5, Pro F1RC, F2RC, F3RC*	Pre	Adv	Eli	Com										
5	cycle2PreS (Normal) cycle3PreS (Cycling)	NA (Normal) cycle3PreS (Cycling)					F1, F2, F3	F4, F5, Pro F1RC, F2RC, F3RC* Gene editing**	Pre	Adv	Eli	Com									
6	cycle3PreS, cycle4CS (Normal) cycle4PreS (Cycling)	cycle3PreS, cycle2PreS (Normal) cycle4PreS (Cycling)						F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com									
7	cycle4PreS, cycle5CS (Normal) cycle6PreS (Cycling)	cycle4PreS, cycle3PreS (Normal) cycle6PreS (Cycling)							F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv ---	Eli	Com								
8	cycle6PreS, cycle6CS (Normal) cycle7PreS (Cycling)	cycle6PreS, cycle4PreS (Normal) cycle7PreS (Cycling)								F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com							
9	cycle7PreS, cycle7CS (Normal) cycle8PreS (Cycling)	cycle7PreS, cycle6PreS (Normal) cycle8PreS (Cycling)									F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com						
10	cycle8PreS, cycle8CS (Normal) cycle9PreS (Cycling)	cycle8PreS, cycle7PreS (Normal) cycle9PreS (Cycling)										F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com					
11	cycle9PreS, cycle9CS (Normal) cycle10PreS (Cycling)	cycle9PreS, cycle8PreS (Normal) cycle10PreS (Cycling)											F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com				
12	cycle10PreS, cycle10CS (Normal) cycle11PreS (Cycling)	cycle10PreS, cycle9PreS (Normal) cycle11PreS (Cycling)												F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com			
13	cycle11PreS, cycle11CS (Normal) cycle12PreS (Cycling)	cycle11PreS, cycle10PreS (Normal) cycle12PreS (Cycling)													F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com		
14	cycle12PreS, cycle12CS (Normal) cycle13PreS (Cycling)	cycle12PreS, cycle11PreS (Normal) cycle13PreS (Cycling)														F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com	
15	cycle13PreS, cycle13CS (Normal) cycle14PreS (Cycling)	cycle13PreS, cycle12PreS (Normal) cycle14PreS (Cycling)															F1, F2, F3	F4, F5, Pre F1RC, F2RC, F3RC* Gene editing**	Adv	Eli	Com

* Rapid-cycling from the given cycle. ** Gene editing of 10 or 50 QTL, depending on the strategy.

Supplemental table 5. Phenotypic strategy (baseline): multi-cycle development, parental selection, and advancement timeline.

Cycle	Parents for F1	Training for F3	year 1	year 2	year 3	year 4	year 5	year 6	year 7	year 8	year 9	year 10	year 11	year 12	year 13	year 14	year 15	year 16	year 17	year 18	year 19
cycle 1	cycle1Par	-	F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com													
cycle 2	cycle2Par	-		F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com												
cycle 3	cycle3Par	-			F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com											
cycle 4	cycle1PreS	-				F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com										
cycle 5	cycle2PreS	-					F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com									
cycle 6	cycle3PreS	-						F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com								
cycle 7	cycle4PreS	-							F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli								
cycle 8	cycle6PreS	-								F1, F2, F3	F4, F5, Pro	Pre	Adv	Com							
cycle 9	cycle7PreS	-									F1, F2, F3	F4, F5, Pro	Pre	Eli	Com						
cycle 10	cycle8PreS	-										F1, F2, F3	F4, F5, Pro	Adv	Eli	Com					
cycle 11	cycle9PreS	-											F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli	Com			
cycle 12	cycle10PreS	-												F1, F2, F3	F4, F5, Pro	Pre	Adv	Eli			
cycle 13	cycle11PreS	-													F1, F2, F3	F4, F5, Pro	Pre	Adv	Com		
cycle 14	cycle12PreS	-														F1, F2, F3	F4, F5, Pro	Pre	Eli	Com	
cycle 15	cycle13PreS	-															F1, F2, F3	F4, F5, Pro	Adv	Eli	Com