frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 March 2024
DOI 10.3389/fpls.2024.1361894

# AI-assisted selection of mating pairs through simulation-based optimized progeny allocation strategies in plant breeding

Kosuke Hamazaki[†] and Hiroyoshi Iwata*

Laboratory of Biometry and Bioinformatics, Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

Emerging technologies such as genomic selection have been applied to modern plant and animal breeding to increase the speed and efficiency of variety release. However, breeding requires decisions regarding parent selection and mating pairs, which significantly impact the ultimate genetic gain of a breeding scheme. The selection of appropriate parents and mating pairs to increase genetic gain while maintaining genetic diversity is still an urgent need that breeders are facing. This study aimed to determine the best progeny allocation strategies by combining future-oriented simulations and numerical black-box optimization for an improved selection of parents and mating pairs. In this study, we focused on optimizing the allocation of progenies, and the breeding process was regarded as a black-box function whose input is a set of parameters related to the progeny allocation strategies and whose output is the ultimate genetic gain of breeding schemes. The allocation of progenies to each mating pair was parameterized according to a softmax function, whose input is a weighted sum of multiple features for the allocation, including expected genetic variance of progenies and selection criteria such as different types of breeding values, to balance genetic gains and genetic diversity optimally. The weighting parameters were then optimized by the black-box optimization algorithm called StoSOO via future-oriented breeding simulations. Simulation studies to evaluate the potential of our novel method revealed that the breeding strategy based on optimized weights attained almost 10% higher genetic gain than that with an equal allocation of progenies to all mating pairs within just four generations. Among the optimized strategies, those considering the expected genetic variance of progenies could maintain the genetic diversity throughout the breeding process, leading to a higher ultimate genetic gain than those without considering it. These results suggest that our novel method can significantly improve the speed and efficiency of variety development through optimized decisions regarding the selection of parents and mating pairs. In addition, by changing simulation settings, our future-oriented optimization framework for progeny allocation strategies can be easily implemented into general breeding schemes, contributing to accelerated plant and animal breeding with high efficiency.

KEYWORDS

breeding scheme, optimization, progeny allocation, forward simulation, mating pairs, genetic diversity, breeding values, AI breeder

OPEN ACCESS

EDITED BY
Maliheh Eftekhari,
Tarbiat Modares University, Iran

REVIEWED BY
Zitong Li,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Yong Jiang,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany
Haipeng Yu,
University of Florida, United States
Just Jensen,
Aarhus University, Denmark

*CORRESPONDENCE
Hiroyoshi Iwata
hiroiwata@g.ecc.u-tokyo.ac.jp

[†]PRESENT ADDRESS
Kosuke Hamazaki,
Molecular Informatics Team, RIKEN Center for Advanced Intelligence Project (AIP), RIKEN, Chiba, Japan

RECEIVED 27 December 2023
ACCEPTED 06 March 2024
PUBLISHED 28 March 2024

CITATION
Hamazaki K and Iwata H (2024) AI-assisted selection of mating pairs through simulation-based optimized progeny allocation strategies in plant breeding.
*Front. Plant Sci.* 15:1361894.
doi: 10.3389/fpls.2024.1361894

# 1 Introduction

To meet the increasing demand for agricultural products caused by the population explosion, developing new cultivars with desired agronomic characteristics, such as high yield, good quality, and efficient nutrient use, is urgently needed (Capper, 2011). However, in plant breeding, for instance, conventional breeding methods require several years or longer to produce new cultivars, even in annual crops, which makes it challenging to meet the increasing food demand in the near future (Eggen, 2012). Genomic selection (GS) is expected to be a critical methodology for accelerating the evaluation and selection of superior genotypes that can be implemented in conventional breeding (Meuwissen et al., 2001; Jannink et al., 2010). In GS, genotypic values of target traits are predicted using genomic prediction (GP) models based on genome-wide marker data, and the predicted values are used for selection in breeding schemes (Meuwissen et al., 2001). GS enables individual-based and crossing-based selection according to the genotypic values predicted by the models with fewer field evaluations of target traits, which leads to highly efficient and rapid breeding (Zhong and Jannink, 2007; Jannink et al., 2010; Crossa et al., 2017; Lehermeier et al., 2017). The superiority of GS over selections based on either phenotypic data or pedigree-derived prediction (pedigree BLUP) has been reported in many simulations and empirical studies (Goddard, 2009; Hayes et al., 2009; Jannink et al., 2010; Goddard et al., 2011; Crossa et al., 2013; Hickey et al., 2014; García-Ruiz et al., 2016; Rabier et al., 2016). However, even though GS has contributed to speeding up breeding during selections, the GS itself does not aim to optimize decisions in breeding. Thus, the ultimate decisions regarding parent selection and choosing mating pairs as the beginning step of the breeding scheme still largely rely on breeders' experiences because many data-based approaches have not been fully utilized by breeders due to target traits' complexity, limited resources, or implementation difficulty. In other words, the advent of GS has not yet entirely led to optimized decisions in breeding programs.

To solve the above problems and optimize the decisions regarding (1) selection, (2) mating, and (3) progeny allocation, there has been much discussion on improving the strategies in breeding schemes. First, as for the selection strategy, different selection criteria have been used to increase the efficiency of GS. While various GP models have been developed to improve the accuracy of GS (Meuwissen et al., 2001; Gianola and van Kaam, 2008; VanRaden, 2008; de los Campos et al., 2009; Jannink et al., 2010; Habier et al., 2011; de Los Campos et al., 2013; González-Camacho et al., 2016), genomic estimated breeding value (GEBV), the total sum of estimated additive marker effects, has been generally utilized for GS as a selection criterion (Meuwissen et al., 2001). Although GEBV is usually used for a short breeding period with a few cycles of selection and crossing, it does not guarantee an increased genetic gain in a long-term breeding process (Sonesson et al., 2012; Lin et al., 2017; Akdemir et al., 2019). Because genetic variance is greatly reduced due to the Bulmer effect (Bulmer, 1971; Van Grevenhof et al., 2012) and genetic diversity is also reduced by inbreeding and random drift under truncation selection (Falconer and Mackay, 1996; Li et al., 2008), the genetic gain rapidly reaches a local, rather than global, optimum, i.e., a plateau, in GEBV-based GS cycles (Moeinizade et al., 2019; Zhang and Wang, 2022). Other selection criteria have been proposed to solve the issue of GEBV-based GS by maintaining the genetic diversity and further increasing the genetic gain expected in the long-term breeding program. Weighted genomic estimated breeding value (WGEBV) is a criterion that utilizes marker allele frequencies to emphasize the marker effects of rare alleles (Goddard, 2009; Jannink, 2010). Compared to GEBV, WGEBV can maintain genetic diversity in breeding populations over the long term by suppressing the Bulmer effect (Goddard, 2009; Jannink, 2010). Other selection criteria, such as optimal haploid value (Daetwyler et al., 2015), with an expected maximum haploid breeding value (Müller et al., 2018), and optimal population value (Goiffon et al., 2017), have been introduced to maintain genetic diversity by including haplotype information and have indeed improved the final genetic gains at the end of the breeding scheme in long-term breeding programs compared to those from GEBV-based GS. Although these selection criteria are helpful in selecting parental candidates, they do not directly enable the automatic optimization of decisions regarding the selection of parents. Therefore, breeders must make ultimate decisions based on these criteria by themselves. Also, an optimal contribution selection method, which aims to optimize decisions regarding parent selection, was proposed to help breeders decide on more appropriate breeding strategies rather than just use selection criteria (Meuwissen, 1997; Grundy et al., 1998). The optimal contribution selection aims to optimize the expected contribution vector, i.e., the number of times each individual is used as a parent for the next generation, by maximizing the genetic gains while constraining the inbreeding in the next generation. However, the optimal contribution selection still cannot assist in identifying optimal mating pairs even for the next generation. Second, as for the problems regarding the selection of appropriate mating pairs, i.e., the mating design problem, many studies in animal breeding have long been conducted to select optimal pairs while considering mating constraints (Jansen and Wilton, 1985; Kinghorn, 2011), but they mostly lacked a long-term perspective. An optimal cross-selection (OCS), an extension of the optimal contribution selection, aims to select suitable mating pairs by establishing a connection between the mating pairs and the expected contribution vector (Gorjanc et al., 2018; Allier et al., 2019). Although the OCS considers genetic diversity in breeding populations by constraining inbreeding, as in the optimal contribution selection, it does not directly optimize the long-term response. Then, a look-ahead selection (LAS) approach has been proposed as a method for simultaneously optimizing decisions regarding the selection of appropriate parents and mating pairs in plant breeding (Moeinizade et al., 2019). LAS can optimize whether each pair of individuals should be mated to maximize the final genetic gain via look-ahead simulations under two simplified assumptions: one progeny is obtained from each mating pair, and random mating occurs after the first selection and mating cycle. Although this approach achieved higher genetic gains than the previous selection criteria in the final generation by considering optimal mating pairs, the assumptions were too simple to estimate the final genetic gain in practical breeding programs, and LAS could not optimize the number of progenies allocated to each mating pair, i.e., progeny allocation strategy.

Thirdly, as for the progeny allocation strategy, only one study has attempted to optimize the allocation strategy (Hunter and McClosky, 2016) although it is known to have a significant impact on the ultimate genetic gain (Wang et al., 2018). In their study, the number of progenies allocated to each pair was optimized by searching for the Pareto surface in the plane formed by the mean and variance of progenies in the next generation (Hunter and McClosky, 2016), utilizing the multi-objective optimal computing budget allocation method (Lee et al., 2010). However, this method lacks the future-oriented perspectives required for the mid-term/long-term breeding process. Moreover, there is no study on the allocation optimization of progenies for the final genetic gain under multiple selection and crossing cycles, although a couple of previous studies tried to optimize breeding program decisions from other viewpoints (Amini et al., 2021; Diot and Iwata, 2022; Moeinizade et al., 2022).

As described so far, regardless of the duration of breeding schemes, i.e., short-term or long-term, how decisions are made significantly impacts the ultimate outcome of the schemes when parent selection and mating are repeated multiple times. Thus, if the optimization framework regarding the selection of parents and mating pairs is achieved, it can be widely applied to various types of breeding programs, i.e., small-scale and large-scale breeding programs. As for small-scale breeding programs with limited resources, particularly in plant breeding, since many minor crops, including neglected and underutilized species (NUS), have attracted much attention as breeding targets, there is increasing demand for realizing more efficient breeding in small-scale programs for minor crops (Padulosi et al., 2013; Kamenya et al., 2021). Even if NUS has attracted little attention and has been entirely ignored by plant breeders so far, it is now expected to have the potential to diversify agriculture and address climate change. Another example of small-scale breeding is breeding schemes promoted by small and medium enterprises (Zambon et al., 2019). Since there are many agricultural small and medium enterprises in Asian countries, optimizing decision-making in small-scale breeding will lead to increased profits in the agricultural market. From these viewpoints, our study, which helps breeders' decisions on hybridization, is essential to improve breeding efficiency not only in large-scale but also small-scale breeding programs.

The overall objective of the study was to go beyond the conventional selection criterion-based GS by optimizing a progeny allocation strategy to maximize the ultimate genetic gains after multiple cycles of selection of parents and mating pairs. There are two significant issues in this optimization. The first issue is that it is challenging to describe the breeding process explicitly because the future state of a breeding population is unknown. The second issue is that breeding is a stochastic rather than deterministic process since progenies are randomly produced in each generation through meiosis. Therefore, the specific objectives of this study were (1) to evaluate the final genetic gain via future-oriented breeding simulations and (2) to achieve an optimal mating pair selection via a numerical optimization approach conducted by an artificial intelligence (AI) breeder while addressing the two issues above. First, GS combined with future-oriented breeding simulations enables us to evaluate the final genetic gain in a breeding scheme with multiple cycles, leading to a solution to the first issue. Second, the AI breeder in this study referred to a virtual breeder in our computer who can optimize the decisions regarding the progeny allocation strategy by using both the breeding simulations and some optimization algorithm. In this study, to design this AI breeder, we applied black-box optimization by considering the breeding process achieved by the above simulations as a black-box function whose input is a set of parameters for the allocation strategy and whose output is the ultimate genetic gain. We employed a black-box optimization algorithm, StoSOO (Valko et al., 2013), which can also optimize a stochastic black-box function, such as a breeding process, presenting a solution to the second issue. Multiple features representing some goodness of mating pairs are used to allocate progenies, and the optimized progeny allocation can automatically determine the optimal combination of features and weights to be used in the allocation. Further, the optimized progeny allocation can help breeders select the appropriate mating pairs because the AI breeder can offer information on which cross combination is prospective by showing the optimized number of progenies in continuous values. In this study, we mainly focus on optimizing the progeny allocation strategy as a proof of concept. Still, by changing the simulation settings and the parameters to be optimized, we can easily extend our framework to assist breeders in making decisions in different types of hybridization, such as parental selection in bi-parental/multi-parental populations.

## 2 Materials and methods

### 2.1 System design and overview of the methods

We assumed that breeders proceed with a breeding scheme according to the progeny allocation strategy proposed by an "AI breeder", a novel decision-making system in this study (Figure 1). The breeders can obtain a set of the optimal weighting parameters, $\left\{\hat{\mathbf{h}}^{(\tau)}\right\}_{\tau=0,1,2,3}$ ($\tau \in \{0,1,2,3\}$ is a current generation number in simulations), which is related to the allocation of progenies as in Figure 2, by passing the information on a parent panel, i.e., marker genotype, genetic marker effects, and recombination rates between genetic markers, to the AI breeder. Here, we briefly describe the overview of the framework in which the AI breeder optimizes the strategy for the resource allocation of progenies.

After receiving the parent panel information, the AI breeder attempts to simulate breeding schemes by defining the allocation strategy (Figure 1). At each generation $\tau$ in the simulated schemes, the breeder first computes the weighted breeding values (WBV) to select the parent candidates from the current generation. Then, after determining mating pairs by the diallel crossing between the parent candidates, given the parameter $\mathrm{h}^{(\tau)}$, the breeder automatically determines the number of progenies allocated to each pair $\mathbf{b}^{(\tau)}$. This step is realized by utilizing the matrix $\mathbf{\Omega}^{(\tau)}$ consisting of features representing some goodness of mating pairs, such as

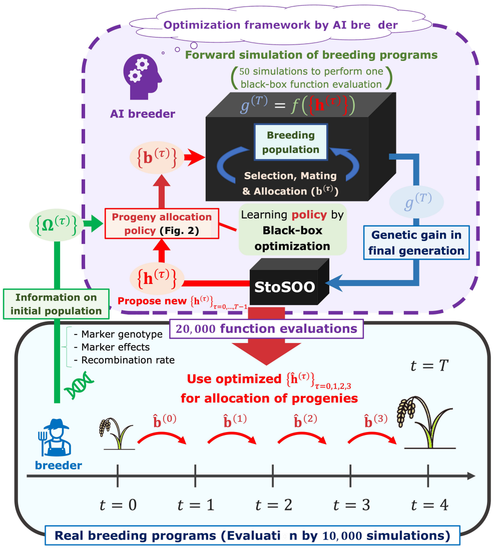
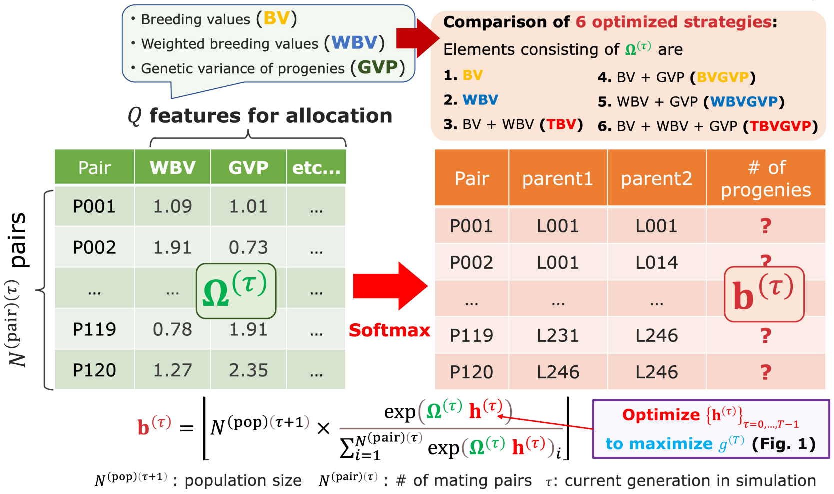

Optimization framework by AI bre der
Forward simulation of breeding programs
50 simulations to perform one black-box function evaluation
AI breeder
$g^{(T)} = f(\{h^{(\tau)}\})$
Breeding population
Selection, Mating & Allocation ($b^{(\tau)}$)
$\{b^{(\tau)}\}$
$g^{(T)}$
$\{\Omega^{(\tau)}\}$
Progeny allocation policy (Fig. 2)
Learning policy by Black-box optimization
Genetic gain in final generation
$\{h^{(\tau)}\}$
StoSOO
Propose new $\{h^{(\tau)}\}_{\tau=0,\ldots,T-1}$
Information on initial population
- Marker genotype
- Marker effects
- Recombination rate
20,000 function evaluations
Use optimized $\{\hat{h}^{(\tau)}\}_{\tau=0,1,2,3}$ for allocation of progenies
$t = T$
breeder
$\hat{b}^{(0)}$
$\hat{b}^{(1)}$
$\hat{b}^{(2)}$
$\hat{b}^{(3)}$
$t = 0$
$t = 1$
$t = 2$
$t = 3$
$t = 4$
Real breeding programs (Evaluati n by 10,000 simulations)

FIGURE 1
The framework to optimize decision-making in the breeding scheme. In our framework, the breeders can obtain a set of parameters representing the optimal allocation strategy, $\{\hat{h}^{(\tau)}\}_{\tau=0,1,2,3}$, by passing the information on a parent panel $\Omega^{(0)}$ to the AI breeder. After receiving the parent panel information, the AI breeder attempts to simulate breeding schemes by defining the allocation strategy parameterized by the parameter $h^{(\tau)}$ ($\tau \in \{0,1,2,3\}$). Breeding simulator is regarded as a black-box function whose input is a set of allocation-related parameters $\{h^{(\tau)}\}_{\tau=0,1,2,3}$ and output is an evaluated genetic gain in the final generation $g^{(T)}$. After optimizing $h^{(\tau)}$ by black-box optimization algorithm called StoSOO (Valko et al., 2013), the real breeders proceed with the breeding schemes based on the number of progenies allocated to each mating pair $\{\hat{b}^{(t)}\}_{t=0,1,2,3}$.

breeding values (BV), WBV, and expected genetic variance of progenies (GVP), which are computed from the marker genotype and effects (Figure 2). In other words, we assumed that $b^{(\tau)}$ is determined by applying the softmax function to the weighted features for the allocation, $\Omega^{(\tau)}$ $h^{(\tau)}$, as shown in Equation 1.

$$\mathbf{b}^{(\tau)} = \left\lfloor N^{(\mathrm{pop})(\tau+1)} \times \frac{\exp(\boldsymbol{\Omega}^{(\tau)}\ \ \mathbf{h}^{(\tau)})}{\sum_{i=1}^{N^{(\mathrm{pair})(\tau)}} \exp(\boldsymbol{\Omega}^{(\tau)}\ \ \mathbf{h}^{(\tau)})_i} \right\rfloor, \tag{1}$$

where is the population size in the next generation $\tau + 1$, and $x$ is the floor function of an arbitrary numeric $x$ (i.e., the maximum integer that is equal to or lower than $x$). The softmax function is used to determine the number of progenies allocated to each mating pair in proportion to probabilities reflecting the weighted sum of the multiple features. Here, $h^{(\tau)}$ can be interpreted as the weighting parameters that correspond to the importance of each feature. More detailed notations will be provided in the Progenies allocation strategies subsection. After determining $b^{(\tau)}$, the AI breeder can simulate the next generation based on the principle of meiosis. Then, by repeating the parent selection and mating steps toward the final generation of the breeding scheme, the AI breeder evaluates the performance of the allocation strategy with $h^{(\tau)}$. Thus, the goal of the AI breeder is to estimate $\{h^{(\tau)}\}_{\tau=0,1,2,3}$ that maximizes the final genetic gain via simulations of the breeding process. To optimize the weighting parameters, a black-box optimization algorithm called StoSOO (Valko et al., 2013) was used in this study (Figure 1). The StoSOO algorithm is an extension of the SOO algorithm (Munos, 2011; Preux et al., 2014), which performs global optimization of deterministic black-box functions based on a tree-based search and can be applied to stochastic black-box functions. Using the StoSOO algorithm with a finite number of function evaluations can provide a global quasi-optimal solution of an


• Breeding values (BV)
• Weighted breeding values (WBV)
• Genetic variance of progenies (GVP)
$Q$ features for allocation
$N^{(\mathrm{pair})(\tau)}$ pairs
$\Omega^{(\tau)}$
Softmax
$b^{(\tau)}$
Optimize $\{h^{(\tau)}\}_{\tau=0,\ldots,T-1}$ to maximize $g^{(T)}$ (Fig. 1)

Comparison of 6 optimized strategies:
Elements consisting of $\Omega^{(\tau)}$ are

| | |
|---|---|
| 1. BV | 4. BV + GVP (BVGVP) |
| 2. WBV | 5. WBV + GVP (WBVGVP) |
| 3. BV + WBV (TBV) | 6. BV + WBV + GVP (TBVGVP) |

| Pair | WBV | GVP | etc… |
|---|---|---|---|
| P001 | 1.09 | 1.01 | … |
| P002 | 1.91 | 0.73 | … |
| … | … | | … |
| P119 | 0.78 | 1.91 | … |
| P120 | 1.27 | 2.35 | … |

| Pair | parent1 | parent2 | # of progenies |
|---|---|---|---|
| P001 | L001 | L001 | ? |
| P002 | L001 | L014 | ? |
| … | … | … | |
| P119 | L231 | L246 | ? |
| P120 | L246 | L246 | ? |

$$\mathbf{b}^{(\tau)} = \left\lfloor N^{(\mathrm{pop})(\tau+1)} \times \frac{\exp(\boldsymbol{\Omega}^{(\tau)}\ \mathbf{h}^{(\tau)})}{\sum_{i=1}^{N^{(\mathrm{pair})(\tau)}} \exp(\boldsymbol{\Omega}^{(\tau)}\ \mathbf{h}^{(\tau)})_i} \right\rfloor$$

$N^{(\mathrm{pop})(\tau+1)}$ : population size $N^{(\mathrm{pair})(\tau)}$ : # of mating pairs $\tau$: current generation in simulation

FIGURE 2
The resource allocation strategy of progenies developed in this study. After determining mating pairs by the diallel crossing between the parent candidates, given the parameter $h^{(\tau)}$, the breeder automatically determines the number of progenies allocated to each pair $b^{(\tau)}$. This step is realized by utilizing the matrix $\Omega^{(\tau)}$ consisting of features representing some goodness of mating pairs, such as BV, WBV, and expected GVP. In other words, we assumed that $b^{(\tau)}$ is determined by applying the softmax function to the weighted sum of the multiple features, as shown in Equation 1.

implicit stochastic function within the number of evaluations. To apply StoSOO to our problem, the AI breeder performs 50 breeding simulations with one set of $\left\{\mathrm{h}^{(\tau)}\right\}_{\tau=0,1,2,3}$, and the empirical mean of the 50 final genetic gains is regarded as the objective function for StoSOO. The function evaluations are repeated 20,000 times in StoSOO, and a set of optimized parameters $\left\{\hat{\mathrm{h}}^{(\tau)}\right\}_{\tau=0,1,2,3}$ is passed from the AI breeder to the real breeders.

In this study, we evaluated this novel optimization framework by conducting simulation studies as a proof of concept. We prepared six versions of the optimized allocation strategy according to the components of $\Omega^{(\tau)}$ as follows: (a) only BV, (b) WBV, (c) both selection criteria (i.e., BV and WBV), (d) expected GVP in addition to BV, (e) expected GVP in addition to WBV, and (f) expected GVP in addition to (c) (Figure 2). We compared these six optimized strategies to (g) the equal allocation strategy that equally allocates progenies to each pair, that is, $\mathrm{h}^{(\tau)} = 0$, by conducting 10,000 breeding simulations using the optimized $\left\{\hat{\mathrm{h}}^{(\tau)}\right\}_{\tau=0,1,2,3}$ (Figure 1). In the following results, we abbreviate the above strategies as (a) BV, (b) WBV, (c) TBV (Two BVs), (d) BVGVP, (e) WBVGVP, (f) TBVGVP, and (g) EQ. Here, for the BV and WBV strategies, we only used one criterion for allocating progenies. For simplicity, we also assumed true marker effects to compute all the features for the allocation strategy. In other words, true QTL positions and effects were known throughout the study. The problem of using the true marker effects will be discussed in the Discussion section.

In this simulation study, we first simulated genome-wide QTLs for the parent panel and the corresponding QTL effects (Supplementary Figure S1). The selection of parent candidates, determination of mating pairs, and allocation of progenies to each mating pair were then repeated in sequence to move forward with a breeding scheme until the final generation. Selection and mating processes were carried out based on two selection criteria: BV and WBV, and the expected GVP. We compared two breeding strategies regarding the allocation of progenies: equal allocation and allocation optimized by a black-box optimization algorithm, StoSOO, as described above. These strategies were evaluated by simulating 10,000 breeding schemes based on each strategy.

Each element of the simulation study is described in the following subsections. First, the two selection criteria, BV and WBV, are described, followed by the simulation of genome-wide QTLs, the selection and mating processes in breeding schemes, and the evaluation of breeding schemes. Because optimized allocation requires the simulation and evaluation of breeding schemes, the method of evaluating breeding schemes is explained before describing the optimized allocation strategy. The calculation of the GVP required for optimized allocation is presented at the end of this section.

Note that, in this paper, $\tau \in \{0,1,2,3\}$ is a current generation number in simulations conducted by the AI breeder, whereas $t \in \{0,1,2,3\}$ appeared in the subsequent subsections is a current generation number in breeding schemes proceeded by real breeders.

## 2.2 Selection criteria

We used two selection criteria, BV (Meuwissen et al., 2001) and WBV (Goddard, 2009; Jannink, 2010), to select parent candidates and allocate progenies in breeding schemes since these are the two typical breeding values (Supplementary Figure S1 (1)). Here, we only used WBV for the selection step to maintain the diversity, but we used both criteria for the allocation step.

The first criterion, BV, can be expressed as in Equation 2:

$$\mathbf{u}^{(\mathrm{BV})(t)} = \sum_{m=1}^{M} \mathbf{w}_m^{(t)} \alpha_m \tag{2}$$

where $M$ is the number of markers, $N^{(\mathrm{pop})(t)}$ is the number of genotypes for a population with generation $t$ in a breeding scheme, $\mathbf{u}^{(\mathrm{BV})(t)}$ is an $N^{(\mathrm{pop})(t)} \times 1$ vector of BV for generation $t$, $\mathbf{w}_m^{(t)}$ is an $N^{(\mathrm{pop})(t)} \times 1$ vector of marker genotype at $m$ -th marker with scores of $-1, 0,$ or $1$ for generation $t$, and $\alpha_m$ is a true marker effect at the $m$ -th marker. In this study, we assumed that the true marker effect $\alpha_m$ was known, i.e., markers were identical to QTLs segregating in the population here. BV is a criterion for expressing additive genotypic values directly transmitted to the next generation and is often used in conventional breeding schemes with GS (Meuwissen et al., 2001).

The second selection criterion, WBV, can be expressed as in Equation 3:

$$\mathbf{u}^{(\mathrm{WBV})(t)} = \sum_{m=1}^{M} \mathbf{w}_m^{(t)} \alpha_m p_m^{(t)-\frac{1}{2}} \tag{3}$$

where $\mathrm{u}^{(\mathrm{WBV})(t)}$ is an $N^{(\mathrm{pop})(t)} \times 1$ vector of WBV for generation $t$, $p_m^{(t)}$ is the allele frequency at the $m$ -th marker/QTL for generation $t$, and the other terms are defined in Equation 2. Emphasizing the effects of rare alleles in each QTL, as in Equation 3, leads to the successful maintenance of genetic diversity, thus WBV can benefit the genetic gain in long-term breeding programs compared to BV (Goddard, 2009; Jannink, 2010).

## 2.3 Simulation of the parent panel and genome structure

Genome-wide QTLs for founder haplotypes were generated using the coalescent simulator GENOME (Liang et al., 2007) (Supplementary Figure S1 (2)). We assumed a diploid with the number of chromosomes ten ($2n = 20$), and that there were 500 loci, including both markers and QTLs, on each chromosome. At first, 4,000 founder haplotypes for each chromosome were independently generated using GENOME with the following parameters: "-pop 1 4000 -N 4000 -c 1 -pieces 100000 -rec 0.0001 -s 4000 -maf 0.01 -tree 0 -mut 0.00000001". Among the 4,000 loci, 500 loci whose minor allele frequency (MAF) was equal to or larger than 0.01 were randomly selected. Subsequently, 2,000 founders were generated by sampling two haplotypes from all founder haplotypes with replacement. These founders were randomly mated for one generation to simulate the pre-breeding materials of the 2,000 genotypes. These procedures for simulating the genetic resources are similar to those in the R package "BreedingSchemeLanguage" by Yabe et al (Yabe et al., 2017). From these simulated genetic resources, $N^{(\mathrm{pop})(0)} = 250$ genotypes were selected for the parent panel of a breeding scheme so that these 250 genotypes represented the simulated genetic resources in terms of genetic diversity. We applied the k-medoids method to the marker genotypes of the pre-breeding material using the "pam" function of the R package "cluster" version

2.1.2 (Maechler et al., 2021). In the k-medoids method, after clustering the pre-breeding materials into 250 groups, we selected the medoids as the representative genotypes from each group. This parent panel was regarded as generation $t = 0$ in a breeding scheme. Here, the parent panel consisted of a relatively small number of genotypes and was also genetically close to founders compared to usual breeding populations used in programs for major crops. These simulation settings, including the assumption for the parent panel, are a little apart from situations in large-scale breeding for major crops but can be justified when we assume the breeding schemes for the NUS whose breeding has not been promoted in the past.

## 2.4 Simulation of QTLs and genotypic values

In simulating QTLs and phenotypes, we assumed quantitative traits with a simple genetic architecture as target traits, as described below (Supplementary Figure S1 (3)). First, among the 500 loci simulated in the previous subsection, 2 loci per chromosome were randomly selected as QTLs (i.e., a total of $M^{(\mathbf{QTL})} = 20$ QTLs). This number of QTLs was determined assuming the situation where quantitative traits controlled by the relatively small number of QTLs still remain as target traits in small-scale breeding for the NUS. QTL effects were then sampled from the normal distribution, as shown in Equation 4.

$$p(\alpha^{(\mathrm{QTL})}) \sim \mathrm{MVN}\left(\mathbf{0}, \frac{1}{M^{(\mathrm{QTL})}}\mathbf{I}_{M^{(\mathrm{QTL})}}\right) \tag{4}$$

where $\alpha^{(\mathrm{QTL})}$ is an $M^{(\mathrm{QTL})} \times 1$ vector of QTL effects and $\mathbf{I}_{M^{(\mathrm{QTL})}}$ is an $M^{(\mathrm{QTL})} \times M^{(\mathbf{QTL})}$ identity matrix. Here, QTL effects were assumed to be additive for simplicity. The true additive genotypic values $\mathrm{u}^{(\mathrm{TGV})(0)}$ for the parent panel were then simulated using Equation 2, where each marker was replaced by each QTL ($M = M^{(\mathrm{QTL})}$, $\mathrm{w}_m = \mathrm{w}_m^{(\mathrm{QTL})(0)}$, $\alpha_m = \alpha_m^{(\mathrm{QTL})}$; $\mathrm{w}_m^{(\mathrm{QTL})(0)}$ is an $N^{(\mathrm{pop})(0)}$ $\times 1$ vector of genotype scores at $m$ -th QTL for the parent panel). Thus, $\mathrm{u}^{(\mathrm{TGV})(0)} = \mathrm{u}^{(\mathrm{BV})(0)}$ when the true marker effect $\alpha_m$ is known as the setting in this study. Since we used the true marker effects, a heritability was assumed to be 1 throughout the study.

## 2.5 AI-assisted breeding scheme

The breeding scheme in this study was carried out according to the following steps (Supplementary Figure S1 (4)). Here, we assumed small-scale plant breeding programs with recurrent selection in mind for NUS for simplicity and due to the computational limitations. Also, we did not produce inbred lines as usually done in breeding schemes for major autogamous crops. However, if the crop is vegetatively propagated, there is no need to develop inbred lines, and even for autogamous crops with seed propagation, the scheme below can be considered as a hybridization phase, and fixation can be proceeded afterwards. In any case, changing the settings of the following breeding scheme will lead to the extension of our framework for various types of breeding programs.

1. Select parent candidates (Supplementary Figure S1 (4-1))

Based on the order of increasing WBV computed using the true marker (QTL) effects $\alpha^{(\mathrm{QTL})}$ for the current generation $t$ ($t \in \{0, 1, 2, 3\}$), the top $N^{(\mathrm{sel})(t)}$ genotypes were selected as parent candidates for the next generation. In this study, we assumed $N^{(\mathrm{sel})(t)}$ was constant over generations, i.e., $N^{(\mathrm{sel})(t)} = N^{(\mathrm{sel})} = 15$. We used only WBV for the selection step in this study because selecting parents based on BV often failed to maintain the genetic diversity of the population when we chose a high selection intensity in the optimized allocation strategy.

2. Determine mating pairs for the next generation (Supplementary Figure S1 (4-2))

From the selected $N^{(\mathrm{sel})(t)}$ parent candidates, diallel crossing, including selfing, was assumed to determine the mating pairs for the next generation, that is, $N^{(pair)(t)} = \frac{N^{(\mathrm{sel})(t)}(N^{(\mathrm{sel})(t)}+1)}{2} = 120$ pairs were prepared for crossing.

3. Allocate progenies to each mating pair (Supplementary Figure S1 (4-3))

The following two strategies were used to allocate progenies to each mating pair determined in Step 2, and then a crossing table was created. Both strategies generated $N^{(\mathrm{pop})(t+1)}$ progenies for the next generation. In this study, we assumed $N^{(\mathrm{pop})(t+1)}$ was also constant over generations, i.e., $N^{(\mathrm{pop})(t+1)} = N^{(\mathrm{pop})} = 250$.

a. The optimized allocation method

One strategy was the optimized allocation method developed in this study, which applied the softmax function to the weighted sum of the multiple features for the allocation. For more details, please refer to the Progenies allocation strategies subsection.

b. Equal allocation method

We also prepared an equal allocation method that allocated $N^{(\mathrm{pop})(t+1)}$ progenies equally to the $N^{(\mathrm{pair})(t)}$ mating pairs. Thus, three progenies were allocated to each of the 10 mating pairs with the highest expected WBV of progeny, and two progenies were allocated to each of the remaining 110 mating pairs.

4. Generate progenies for the next generation (Supplementary Figure S1 (4-4))

According to the crossing table created in Step 3, two gametes were generated for each mating pair, considering the recombination between markers. Recombination rates between markers were computed using the Kosambi map function (Kosambi, 1943; Zhao and Speed, 1996) based on the linkage map generated by GENOME. These two gametes were then combined to create one new progeny, resulting in a new population (with generation $t + 1$) of $N^{(\mathrm{pop})(t+1)}$ progenies.

5. Repeat the parent selection and mating process

Steps 1-4 were repeated four times until the population reached the final generation, i.e., $t = T - 1$ where $T$ is the final generation number.

Finally, the results for each strategy were evaluated using the true additive genotypic values $\mathrm{u}^{(\mathrm{TGV})(T)}$ for the latest population with generation $T$. Here, the final generation was set as $T = 4$ in this study, which was determined assuming the situation where we needed to quickly achieve high genetic gains in small-scale breeding schemes as one example. Achieving improvements in a short-term breeding scheme holds substantial importance when considering the rapid improvement required in NUS breeding.

## 2.6 Evaluation of the outcome of breeding schemes

In one breeding scheme, the true additive genotypic value $\mathrm{u}^{(\mathrm{TGV})(t)}$ was computed for each generation. (Supplementary Figure S1 (5)). Then, the top $N^{(\mathrm{top})(t)} = 5$ genotypes were chosen in the order of increasing $\mathrm{u}^{(\mathrm{TGV})(t)}$, and the empirical mean of the $\mathrm{u}^{(\mathrm{TGV})(t)}$ for these $N^{(\mathrm{top})(t)}$ genotypes, $u^{(\mathrm{top})(t)}$, was computed to evaluate the population maximum for the breeding strategy. Here, we chose $N^{(\mathrm{top})(t)} = 5$ genotypes for the evaluation to represent the top individuals for developing a new variety and to control the stochastic variation between simulations to some extent. For a given simulation dataset of the parent panel, we conducted 10,000 different breeding schemes for each strategy and evaluated the empirical mean of the $u^{(\mathrm{top})(t)}$ using these 10,000 simulation results. The simulation and its evaluation were repeated for 10 replicates of the phenotype simulation with different QTL positions and effects.

## 2.7 Progenies allocation strategies

In this study, we developed a novel breeding strategy to quasi-optimize each feature regarding the allocation of progenies (Supplementary Figure S1 (6) and Figure 2). In other words, we determined the number of progenies allocated to each mating pair at generation $\tau$, $\mathrm{b}^{(\tau)}$, by applying the softmax function to the weighted sum of the multiple features instead of the equal allocation strategy described above, as in Equation 1. Again, $\mathrm{b}^{(\tau)}$ is an $N^{(\mathbf{pair})(\tau)} \times 1$ vector representing the number of progenies allocated to each mating pair at generation $\tau$, $\boldsymbol{\Omega}^{(\tau)}$ is an $N^{(\mathbf{pair})(\tau)} \times Q$ matrix of $Q$ different multiple features for an average progeny of each mating pair (e.g., BV, WBV, or the genetic variance of progenies in a subsequent generation (GVP) for each pair), and $x$ is the floor function (i.e., the maximum integer that is equal to or lower than $x$). In the softmax function, we can convert a vector of some weighted goodness for each mating pair, $\boldsymbol{\Omega}^{(\tau)}\ \ \mathrm{h}^{(\tau)}$, to a vector of probabilities reflecting the relative value of each element by allocating more progenies to better pairs. Here, $\mathrm{h}^{(\tau)}$ is a $Q \times 1$ vector required for determining $\mathrm{b}^{(\tau)}$, which corresponds to how much importance is given to each feature. Since the final genetic gain, $g^{(T)}$, can be represented as the output of the black-box function $f$ whose input is a set of $\mathrm{h}^{(\tau)}$, i.e., $g^{(T)} = f(\{\mathrm{h}^{(\tau)}\}_{\tau=0,1,2,3})$, the main goal of this strategy is to estimate $\mathrm{h}^{(\tau)}$ that maximizes $g^{(T)}$ via breeding simulations and the black-box optimization algorithm. When $\mathrm{h}^{(\tau)} = 0$, the softmax allocation strategy based on Equation 1 is the same as the equal allocation strategy. The final genetic gain was defined as the improvement in the true genotypic values of the top $N^{(\mathrm{top})(t)} = 5$ genotypes of the final population compared with those of the parent panel, that is, $g^{(T)} = u^{(\mathrm{top})(T)} - u^{(\mathrm{top})(0)}$.

## 2.8 Optimization of $\mathrm{h}^{(\tau)}$ to maximize the final genetic gain

In this study, we used the StoSOO algorithm (Valko et al., 2013) to optimize the parameter vector $\mathrm{h}^{(\tau)}$ (Supplementary Figure S1 (7)). The StoSOO algorithm is an extension of the SOO algorithm (Munos, 2011; Preux et al., 2014), which performs global optimization of deterministic black-box functions based on a tree-based search and can be applied to stochastic black-box functions. The StoSOO algorithm with a finite number of function evaluations can provide a global quasi-optimal solution of an implicit stochastic function within the number of evaluations. To apply StoSOO in our study, we performed 50 simulations using one set of $\{\mathrm{h}^{(\tau)}\}_{\tau=0,\ 1,2,3}$, and the empirical mean of the final genetic gains $g^{(T)}$ for these 50 simulations was used as the objective function for StoSOO. Here, we defined the domain of definition as $[0,\ 2]$ and set 1 as an initial parameter for each element of $\{\mathrm{h}^{(\tau)}\}_{\tau=0,\ 1,2,3}$. The function evaluations were repeated 20,000 times, giving us quasi-optimal solutions, $\left\{\hat{\mathrm{h}}^{(\tau)}\right\}_{\tau=0,\ 1,2,3}$, for this optimization problem. We then performed 10,000 simulations based on the estimated $\left\{\hat{\mathrm{h}}^{(\tau)}\right\}_{\tau=0,\ 1,2,3}$ to evaluate each optimized strategy developed in this study, as described in the Evaluation of the outcome of breeding schemes subsection.

## 2.9 Selection criteria and genetic diversity of progenies as the candidates for $\boldsymbol{\Omega}^{(\tau)}$

As the candidates for $\boldsymbol{\Omega}^{(\tau)}$ to compute $\mathrm{b}^{(\tau)}$, we used the two selection criteria, BV and WBV, and the expected GVP for each mating pair (Supplementary Figure S1 (8)). The BVs and WBVs of the parent candidates were computed using Equations 2 and 3, and the mean of the selection criteria of the two parents for each mating pair was used as one column vector of $\boldsymbol{\Omega}^{(\tau)}$. This vector, representing BV or WBV for each mating pair, was the same as that for the average progeny of each mating pair in this study because we only assumed additive QTL effects. To compute the GVP, we used Equation 5, based on the idea proposed by Lynch and Walsh (Lynch and Walsh, 1998; Zhong and Jannink, 2007; Lehermeier et al., 2017):

$$v_n^{(\tau)} = \frac{1}{4}\left\{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{C}^{(n_{\mathrm{M}})}\boldsymbol{\alpha} + \boldsymbol{\alpha}^{\mathrm{T}}\mathrm{C}^{(n_{\mathrm{P}})}\boldsymbol{\alpha}\right\} \tag{5}$$

where $v_n^{(\tau)}$ is an expected GVP for mating pair $n$ with generation $\tau$, $\alpha$ is a vector of the true marker effects, $\mathbf{C}^{(n_{\mathrm{M}})}$ and $\mathbf{C}^{(n_{\mathrm{P}})}$ are $M \times M$ covariance matrices caused by the segregation of maternal parent $n_{\mathrm{M}}$ and paternal parent $n_P$ of mating pair $n$, respectively. A diagonal $(m, m)$ element of $\mathrm{C}^{(n_{\mathrm{M}})}$, $C_{mm}^{(n_{\mathrm{M}})}$, is 0 if an $m$ -th marker of $n_{\mathrm{M}}$ is homozygote, and 1 if otherwise (i.e., heterozygote), and a non-diagonal $(m_1, m_2)$ element of $\mathrm{C}^{(n_{\mathrm{M}})}$ , $C_{m_1 m_2}^{(n_{\mathrm{M}})}$ , is 0 if the $m_1$ -th or $m_2$ -th marker of $n_{\mathrm{M}}$ is homozygote, $1 - 2r_{m_1 m_2}$ if both $m_1$ -th and $m_2$ -th markers are heterozygote without recombination, and. if otherwise (i.e., heterozygote with recombination). Here, $r_{m_1 m_2}$ is the recombination rate between the $m_1$ -th and $m_2$ -th markers. In this study, after computing the expected GVP for all mating pairs, $\mathrm{v}^{(\tau)}$ was used as one column vector of $\boldsymbol{\Omega}^{(\tau)}$.

We prepared six different strategies for the components of $\boldsymbol{\Omega}^{(\tau)}$ and compared them to the non-optimized strategy as written in the System Design subsection, i.e., (a) BV, (b) WBV, (c) TBV, (d) BVGVP, (e) WBVGVP, (f) TBVGVP, and (g) EQ. Even when we used only one criterion for the allocation of progenies in BV and

WBV strategies, we optimized the weighting parameter $\{h^{(\tau)}\}_{\tau=0,\ 1,2,3}$ by focusing on the optimization of the gradient of the number of progenies allocated to each pair against the relative value of that criterion.

# 3 Results

## 3.1 Evaluation of convergence conditions of the StoSOO algorithm

For the optimization of the allocation strategy to work, it was necessary to confirm that the solution obtained from the algorithm used for optimization, the StoSOO algorithm, converged stably. Thus, we first evaluated the convergence of the solution using the StoSOO algorithm for each strategy. As described above, the AI breeder conducted 50 simulations of the breeding scheme for one function evaluation and repeated 20,000 function evaluations to obtain parameter sets that were quasi-optimized by the StoSOO algorithm. We recorded the quasi-optimized parameter sets for each evaluation step of the objective function and evaluated the corresponding function values. For each of the six optimized allocation strategies, the change in the function values was plotted and compared with the function value based on the equal allocation strategy (Supplementary Figure S2).

For all strategies on the components of $\Omega^{(\tau)}$, the function value increased almost monotonically and reached the quasi-optimized function value after the 20,000 evaluations, which was much larger than that under the equal allocation strategy (Supplementary Figure S2). These results indicated that StoSOO successfully improved the expected final genetic gain evaluated by the simulations and optimized the parameter set $\{h^{(\tau)}\}_{\tau=0,1,2,3}$, resulting in the superiority of the optimized allocation strategy over the equal allocation strategy.

## 3.2 Genetic gains over four generations

For all strategies on the components of $\Omega^{(\tau)}$, we compared the optimized and equal allocation strategies by plotting the change in the genetic gain for each generation (Figure 3A). We then defined genetic gain in each generation as the difference between the empirical means of the true genotypic values of the top $N^{(top)(t)} = 5$ genotypes in the current generation (generation $t$) and the parent panel (generation 0). Using the optimized allocation strategies obtained after 20,000 function evaluations, we evaluated the expected genetic gains of the seven strategies, including the six optimized and the equal allocation strategies, based on 10,000 breeding simulation results using one replication for the quantitative phenotype simulation.

In the final generation, $t = 4$, all the strategies with optimized allocations showed higher genetic gains than the equal allocation strategy (Figure 3A). Strategies that considered GVP outperformed those without GVP in the final generation, while the strategies that only considered the selection criteria BV and WBV showed little difference in genetic gains.

In the first generation, $t = 1$, the strategies that considered GVP showed lower genetic gains than those without GVP (Figure 3A). After the second generation, $t = 2$, strategies with GVP overtook those without GVP, increasing their advantage over the other strategies through the final generation.

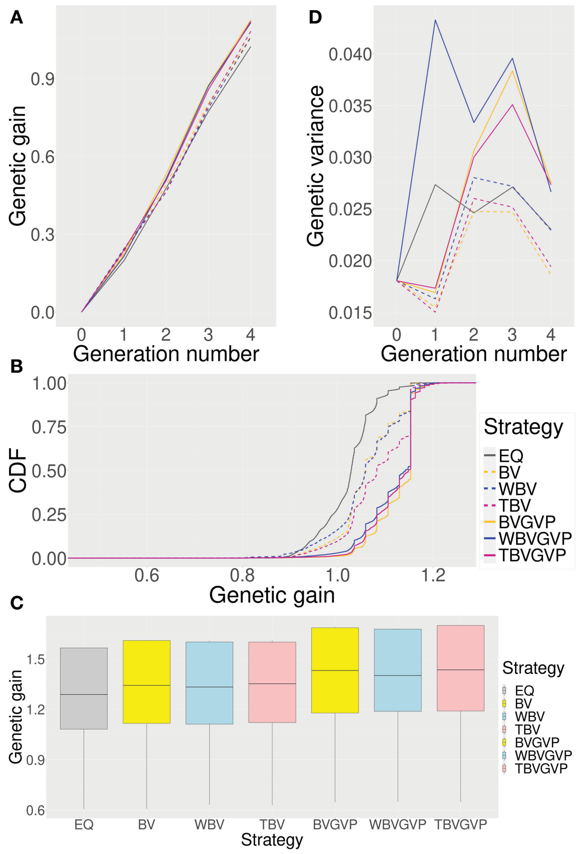


FIGURE 3
Comparison of the seven strategies (six optimized and one non-optimized) in four ways. **(A)** Change in the expected genetic gains over four generations. **(B)** Genetic gains across different simulation repetitions. **(C)** Genetic gains in the final generation using ten replications for the phenotype simulation. **(D)** Change in the genetic variances over four generations. In all Figures, we compared the seven strategies: EQ: equal allocation (ABD: black solid, C: grey), BV: optimized allocation based on BV (ABD: orange dashed, C: yellow), WBV: optimized allocation based on WBV (ABD: blue dashed, C: light blue), TBV: optimized allocation based on BV and WBV (ABD: red dashed, C: light pink), BVGVP: optimized allocation based on BV and GVP (ABD: orange solid, C: yellow), WBVGVP: optimized allocation based on WBV and GVP (ABD: blue solid, C: light blue), and TBVGVP: optimized allocation based on BV, WBV, and GVP (ABD: red solid, C: light pink). In **(A-D)**, generations 0 and 4 correspond to the populations in the initial and final generations, respectively. In addition, generations 1–3 correspond to populations in the middle of the breeding scheme. In Figure **(B)**, the horizontal axis shows the final genetic gain of an individual, whereas the vertical axis represents the percentile of the simulation repetitions. As shown in **(C)**, for each optimized strategy, the weighting parameter was optimized based on 4,000 function evaluations.

## 3.3 Genetic gains across different simulation repetitions

To assess the characteristics of each strategy in more detail, we computed the cumulative distribution functions (CDFs) of genetic gain in the final generation, $t = 4$, for each strategy based on 10,000 simulation repetitions. In Figure 3B, the horizontal axis represents the final genetic gain of an individual, whereas the vertical axis represents the percentile of simulation repetitions. Because the 1st and 99th percentiles correspond to the worst and best performances within the 10,000 simulation repetitions, respectively, the more a CDF curve of a strategy moves toward the right and downward in the graph, the better the performance of the strategy is.

Again, the optimized strategies outperformed the equal allocation strategy; among the optimized strategies, those with GVP showed better performance than those without GVP in terms of the CDFs (Figure 3B). The CDF curves for the optimized strategies were parallel to the vertical axis at the location where the genetic gain was approximately 1.15, suggesting a local optimum within the simulation repetitions. In the optimized strategies with GVP, about 50% of the breeding simulations achieved genetic gains above this local optimum (BVGVP, 55.1%, WBVGVP, 47.6%, TBVGVP, 49.4%), which were much higher than those in the equal allocation strategy (EQ, 1.85%) and strategies without GVP (BV, 16.0%, WBV, 16.4%, TBV, 30.1%).

## 3.4 Genetic gains in the final generation using ten replications of the phenotype simulation

We also evaluated the genetic gains in the final generation, $t = 4$ , using ten replications of the phenotype simulation to confirm that the optimized strategies showed better results than the equal allocation strategy for the traits with different quantitative trait loci (QTL) positions and effects (Figure 3C). When simulating phenotypes for different traits, we changed QTL positions and their effects, but the number of QTLs and distribution of QTL effects were fixed. Also, the heritabilities were assumed to be 1 for all ten replications since the true marker effects were utilized for the optimization.

To reduce the computational time for the optimization in evaluating ten replications of the phenotype simulation, optimized (or quasi-optimized) allocation strategies were determined after 4,000 function evaluations. The number of function evaluations was determined to ensure that the function value reached a certain level, although it did not converge, and to reduce the computation time based on the results in the subsection that evaluated the convergence conditions of the StoSOO. Here, the StoSOO algorithm does not necessarily have to be converged since the evaluation for the ten replicates of phenotype simulation was conducted to prove that the optimization algorithm could improve the genetic gain to some extent compared to the non-optimized strategy.

Even when the QTL positions and effects of the target trait were changed, the optimized strategies consistently outperformed the equal allocation strategy, and the strategies with GVP outperformed those without GVP (Figure 3C). When we evaluated the improvement rate in the final genetic gain of each optimized strategy compared to the equal allocation strategy using ten replications of the phenotype simulation, a similar trend was observed between the strategies with and without GVP (Table 1). In addition, the optimized strategies at least improved the genetic gain compared to the equal allocation for all ten replications. In particular, the final genetic gains for TBVGVP improved by 7.22–12.4% compared to those for the equal allocation strategy within just four generations (Table 1).

## 3.5 Genetic diversity over four generations

We compared the optimized and equal allocation strategies in genetic diversity across generations (Figure 3D). The genetic diversity of a population in each generation was calculated as the genetic variance of the true genotypic values in the population.

First, in $t = 1$, all the optimized strategies, except WBVGVP, showed lower genetic variances than those in the parent panel, $t = 0$ (Figure 3D). From $t = 2$ to $t = 4$, genetic diversity was preserved in all strategies compared to $t = 0$. In these generations, strategies with GVP maintained higher genetic variance than those without GVP and the equal allocation strategy. The difference between the strategies with or without GVP was most observed at $t = 3$, followed by a sharp decrease in genetic variance for the former strategies through the final generation, $t = 4$.

In contrast, strategies using different selection criteria (BV or WBV) showed little difference in genetic variance (Figure 3D). The strategy using only WBV as a feature for the allocation appeared to maintain a higher genetic variance than those using BV or both BV and WBV. However, the difference was much smaller than between strategies with or without GVP.

## 3.6 Optimized weighting parameters for each strategy

After the 20,000 function evaluations, we obtained a set of optimized weighting parameters $\left\{ \mathrm{h}^{(\tau)} \right\}_{\tau=0,\ 1,2,3}$ for each allocation strategy (Table 2; Supplementary Tables S1, S2). A large value of $\mathrm{h}^{(\tau)}$ indicates a large weight of the feature for the allocation.

As for the strategies without GVP, the weighting parameters in generation $\tau = 1$, $\mathrm{h}^{(1)}$, showed the smallest values among all generations, and the weights in the latter generations increased

TABLE 1 Minimum, mean, and maximum improvement rate in the final genetic gain of the optimized strategies compared to equal allocation strategy among ten different target traits (%).

| Method | Min (%) | Mean (%) | Max (%) |
|---|---|---|---|
| BV | 0.00 | 2.97 | 4.89 |
| WBV | 2.07 | 2.88 | 3.89 |
| TBV | 2.11 | 3.52 | 5.86 |
| BVGVP | 6.40 | 8.63 | 12.4 |
| WBVGVP | 6.99 | 8.28 | 9.95 |
| TBVGVP | 7.22 | 9.33 | 12.4 |

TABLE 2 Optimized weighting parameters $h^{(\tau)}$ in different generations $\tau$ / 0‹...‹$T$ 1 for BV and BVGVP strategies.

| Generation ($\tau$) | BV | | BVGVP | |
|---|---|---|---|---|
| | BV | GVP | BV | GVP |
| **0** | 1.00 | – | 1.00 | 0.33 |
| **1** | 0.33 | – | 1.00 | 1.00 |
| **2** | 1.67 | – | 1.67 | 1.67 |
| **3** | 1.89 | – | 1.67 | 1.67 |

through the final generation (Table 2, Supplementary Tables S1, S2). Using strategies with GVP, the weighting parameters for the selection criteria (BV and WBV) showed higher values in generation $\tau = 1$ than in other generations. The tendency of the selection criteria to increase in later generations was unchanged in the strategies with GVP, and the weights for GVP were similar to those of other selection criteria across generations.

## 3.7 Optimized progeny allocation in generation $t$ = 0

Finally, we present the optimized number of progenies allocated to each mating pair in generation $t = 0$, $b^{(0)}$. Here, we compared $b^{(0)}$ in breeding schemes based on the equal and the six optimized allocation strategies (Supplementary Table S3). The trend of which cross is more preferred was almost the same between the six optimized strategies, but the trend of how progenies are allocated to crosses was quite different. For example, the optimized strategies without GVP, such as WBV and TBV, tended to allocate more progenies to the best pair (G0_0669 x G0_0669). In contrast, the strategies with GVP, such as BVGVP and WBVGVP, tended to allocate progenies to many pairs so that the genetic diversity was maintained. From the results, we can see which mating pair was more emphasized in each strategy in not discrete but continuous ways. Thus, even if it is not practical for real breeders to carry out the exact allocation proposed by the AI breeder, the information on the optimized continuous progeny allocation will help breeders select appropriate mating pairs.

# 4 Discussion

In this study, we utilized the StoSOO algorithm to optimize the allocation of progenies and obtained quasi-optimal solutions for the set of weighting parameters $h^{(\tau)}$ given a finite number of function evaluations. Utilizing these quasi-optimized weighting parameters for the allocation strategy can be justified because the true genetic gain corresponding to each set of parameters almost increased as StoSOO conducted more function evaluations (Supplementary Figure S2). Thus, when applying this novel optimization approach to actual breeding schemes, the appropriate number of function evaluations will be determined according to the computational resources, and within that number of evaluations, the quasi-optimized parameters obtained from StoSOO will be used to allocate progenies.

After optimizing the weighting parameters using StoSOO, we compared the optimized allocation strategies with equal allocation strategies. The simulation results showed that the final genetic gain was higher in the following order: optimized strategies with GVP (BVGVP, WBVGVP, TBVGVP), optimized strategies without GVP (BV, WBV, TBV), and non-optimized strategy (EQ) (Figures 3A, C). These results suggest that the developed optimization framework for the progeny allocation is efficient in conventional GS and that considering the genetic diversity for later generations in the optimization framework can further improve the final genetic gains even in mid-term breeding programs ($T = 4$). In particular, the superiority of the optimized strategies with GVP was remarkable because those strategies guaranteed a certain level of the final genetic gains compared to other strategies, even when random factors, such as recombination between markers and segregation of alleles, are not realized in a real breeding scheme as they are in the optimization process (Figure 3B). Thus, because the optimized strategies with GVP are expected to consistently produce better genotypes than conventional GS, breeders can easily adopt our novel optimization framework in actual breeding schemes. Also, even when breeders find it difficult to implement the progeny allocation strategy proposed by the AI breeder as it is in actual breeding schemes, they can use progeny allocation results to select mating pairs as the results provide the information on which cross is likely to contribute to the production of better genotypes in the future generation (Supplementary Table S3).

On the other hand, the impact on the final genetic gain caused by the type of criteria (BV, WBV, or both) used for allocation was much smaller than that caused by the existence of GVP (Figures 3A, C). Because WBV was always utilized in the selection step in this study, we successfully maintained the genetic diversity throughout the breeding scheme regardless of the allocation strategy (Figure 3D), which may reduce the superiority of WBV in the allocation step against BV. We do not show the results here, but in our preliminary simulation experiments with different settings, when BV was utilized in the selection step, the final genetic gains drastically decreased compared to when WBV was utilized for selection, regardless of the allocation strategy. Thus, since the choice of selection criteria in the selection step has a more significant impact on the final genetic gain than in the allocation step, it is strongly recommended that WBV instead of BV be used in the selection step when applying the developed method to a breeding scheme.

When focusing on the genetic gains in $t = 1$, the optimized strategies with GVP showed lower genetic gains than those without GVP (Figure 3A). In contrast, the optimized strategies with GVP showed larger genetic variances than those without GVP from $t = 1$ (Figure 3D). These results suggest that the optimized strategies with GVP were able to attain higher genetic gains in the final generation by their success in maintaining genetic diversity throughout the schemes and converting it into genetic gain toward the final generation, even in mid-term breeding programs ($T = 4$). When focusing on the optimized parameters for the strategies with GVP, the optimized weights became larger after $t = 2$, which meant that the selection intensity automatically became higher in later generations (Table 2). In addition, when focusing on the optimized weighting parameters for the strategies without GVP, the optimized weight for BV in $t = 1$ was smaller than that for WBV. This may be because the larger weight of BV in the earlier generation led to less diversity, and the optimizer controlled the selection intensity by adjusting the weight to avoid such a situation. Thus, the developed framework to optimize the allocation strategy of progenies can automatically adjust the weights in each generation in an interpretable form for breeders and is expected to deal with various situations, such as breeding programs with varying deadlines $T$, by estimating proper weights for each feature for the allocation. In other words, although we assumed the situation close to small-scale plant breeding schemes with a relatively small number of breeding cycles and a small population size as one example of the simulation settings in this study, we can easily apply our novel framework to larger-scale plant breeding programs, and further animal breeding programs, by considering the assumed conditions such as mating constraints and changing the simulation settings proceeded by the AI breeder. Also, although we focused on the allocation strategies to prove that our future-oriented simulations can optimize the decision-making in breeding, we can further extend our framework to consider the optimization of other strategies, such as selection intensity or selection criteria used in the selection step by introducing parameters regarding those strategies to the simulation settings.

Next, we discuss the validity of the optimization method from the viewpoint of the genetic variance (Figure 3D). Since we assumed a genetic architecture with a relatively small number of QTLs instead of the infinitesimal model, the reduction in genetic variance was mainly influenced by the fixation of each QTL. In this study, common alleles were fixed in earlier generations, while the allele frequencies of rare positive alleles increased and reached around 0.5 from $t = 2$ to $t = 3$ (Supplementary Figure S3), resulting in an increase in genetic variance until $t = 3$ (Figure 3D). Subsequently, particularly in the optimized strategies with GVP, fixing these rare alleles from $t = 3$ to the final generation $t = 4$ led to a great increase in genetic gains compared to the other strategies (Supplementary Figures S3D–F). This result seems to suggest that the optimization algorithm aimed to solve the combinatorial problem of maximizing the probability of obtaining the ideal genotype with more positive alleles in the final generation. However, the black-box optimization algorithm itself did not specifically address the combinatorial problem and simply focused on maximizing the final genetic gain, resulting in only the above phenomenon of a rapid increase in genetic gain towards the final generation. Therefore, our framework is expected to be applicable to traits under the infinitesimal model that cannot be addressed as a combinatorial problem.

The developed future-oriented simulation-based framework achieved much higher final genetic gains than the GS method without optimized allocation of progenies. In particular, the final genetic gains for TBVGVP improved by more than 9% on average compared with those for the non-optimized equal allocation strategy (Table 1). This improvement was quite large for the genetic gains in just four generations because we compared the developed framework to the GS based on WBV (Goddard, 2009; Jannink, 2010), one of the most promising selection strategies in mid- and long-term breeding programs to date, not phenotypic selection or conventional GS based on BV (Meuwissen et al., 2001). Also, since the optimization algorithm was forced to terminate in the middle for the results of the ten replicates for the phenotype simulation due to the computational limitation, we may still underestimate our optimized results and can expect higher improvement than that obtained in Table 1. However, it is important to note that we used true marker effects to compute the selection criteria and GVP in this study. Although breeding strategies are often proposed based on the true QTL effects to simply provide a key idea (Kemper et al., 2012; Daetwyler et al., 2015; Goiffon et al., 2017; Müller et al., 2018; Moeinizade et al., 2019), in actual breeding schemes, we cannot know such true marker effects and must use estimated marker effects based on GP models. Because the optimization algorithm developed in this study can be largely influenced by the estimation accuracy of GP models, developing a robust method to optimize the allocation strategy in such cases is crucial in future studies. The other important factor when assuming real breeding schemes is the timing of GP model updates. If we can appropriately update the GP model, the estimation accuracy of the marker effects and optimization performance of the allocation strategies will be improved, which will directly lead to further improvement in the final genetic gains. In addition, we can deal with long breeding schemes by conducting appropriate model updates. Thus, we will further investigate the potential of our optimization framework by assuming real breeding schemes with the factors discussed above.

Genomic selection is a promising technique that contributes to the acceleration of breeding. In this study, we introduced and developed a novel framework that can upgrade conventional GS in breeding schemes by optimizing the allocation strategy of progenies via future-oriented simulations. To optimize the allocation strategy, we parameterized the strategy via softmax conversion by utilizing the selection criteria and expected genetic variance of progenies. From the simulation results, our novel framework with the optimized allocation greatly outperformed the non-optimized strategies, especially when we added the genetic variance of progenies to the algorithm. Thus, this future-oriented optimization framework for breeding schemes can contribute to the acceleration, high efficiency, and optimization of plant breeding at the hybridization stage, which will lead to the optimization of various kinds of decision-making in plant and animal breeding.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: All data were simulated in this study. Because we performed many breeding simulations, we cannot share all datasets themselves. Instead, we share our scripts, including the simulation of data. The scripts used in this study are available from the "KosukeHamazaki/SCOBS" repository on GitHub, https://github.com/KosukeHamazaki/SCOBS. In this study, we utilized the R package "myBreedSimulatR" to simulate and optimize breeding schemes. "myBreedSimulatR" is available from the "KosukeHamazaki/myBreedSimulatR" repository on GitHub, https://github.com/KosukeHamazaki/myBreedSimulatR. This package is a variant of "breedSimulatR", which is available from the "ut-biomet/breedSimulatR" repository on GitHub, https://github.com/ut-biomet/breedSimulatR.

## Author contributions

KH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. HI: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP 22H02306. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

## Acknowledgments

We are grateful to Mr. Julien Diot, a developer of the "breedSimulatR" package, for fruitful discussions on the implementation of breeding simulations. We would like to thank Dr. Yohei Akimoto for fruitful discussions on how to optimize decision-making in a breeding scheme. We are also grateful to Dr. Bo Zhang for her valuable comments when revising our manuscript. We would like to thank the Japan Society for the Promotion of Science (JSPS) for funding our study. We would also like to thank Editage (www.editage.com) for English language editing.

## Conflict of interest

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1361894/full#supplementary-material

## 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. *Heredity* 122, 672–683. doi: 10.1038/s41437-018-0147-1

Allier, A., Lehermeier, C., Charcosset, A., Moreau, L., and Teyssèdre, S. (2019). Improving short- and long-term genetic gain by accounting for within-family variance in optimal cross-selection. *Front. Genet.* 10, 1006. doi: 10.3389/fgene.2019.01006

Amini, F., Franco, F. R., Hu, G., and Wang, L. (2021). The look ahead trace back optimizer for genomic selection under transparent and opaque simulators. *Sci. Rep.* 11, 4124. doi: 10.1038/s41598-021-83567-5

Bulmer, M. G. (1971). The effect of selection on genetic variability. *Am. Nat.* 105, 201–211. doi: 10.1086/282718

Capper, J. L. (2011). Replacing rose-tinted spectacles with a high-powered microscope: The historical versus modern carbon footprint of animal agriculture. *Anim. Front.* 1, 26–32. doi: 10.2527/af.2011-0009

Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3* 3, 1903–1926. doi: 10.1534/g3.113.008227

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de Los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011

Daetwyler, H. D., Hayden, M. J., Spangenberg, G. C., and Hayes, B. J. (2015). Selection on optimal haploid value increases genetic gain and preserves more genetic diversity relative to genomic selection. *Genetics* 200, 1341–1348. doi: 10.1534/genetics.115.178038

de Los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. doi: 10.1534/genetics.112.143313

de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182, 375–385. doi: 10.1534/genetics.109.101501

Diot, J., and Iwata, H. (2022). Bayesian optimisation for breeding schemes. *Front. Plant Sci.* 13, 1050198. doi: 10.3389/fpls.2022.1050198

Eggen, A. (2012). The development and application of genomic selection as a new breeding paradigm. *Anim. Front.* 2, 10–15. doi: 10.2527/af.2011-0027

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics* (Harlow, Essex, UK: Longmans Green).

García-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggans, G. R., Ruiz-López, F. J., and Van Tassell, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U. S. A.* 113, E3995–E4004. doi: 10.1073/pnas.1519061113

Gianola, D., and van Kaam, J. B. C. H. M. (2008). Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178, 2289–2303. doi: 10.1534/genetics.107.084285

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi: 10.1007/s10709-008-9308-0

Goddard, M. E., Hayes, B. J., and Meuwissen, T. H. E. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* 128, 409–421. doi: 10.1111/j.1439-0388.2011.00964.x

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

González-Camacho, J. M., Crossa, J., Pérez-Rodríguez, P., Ornella, L., and Gianola, D. (2016). Genome-enabled prediction using probabilistic neural network classifiers. *BMC Genomics* 17, 208. doi: 10.1186/s12864-016-2553-1

Gorjanc, G., Gaynor, R. C., 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, 1953–1966. doi: 10.1007/s00122-018-3125-3

Grundy, B., Villanueva, B., and Woolliams, J. A. (1998). Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet. Res.* 72, 159–168. doi: 10.1017/S0016672398003474

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinf.* 12, 186. doi: 10.1186/1471-2105-12-186

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* 92, 433–443. doi: 10.3168/jds.2008-1646

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

Hunter, S. R., and McClosky, B. (2016). Maximizing quantitative traits in the mating design problem *via* simulation-based Pareto estimation. *IIE Trans.* 48, 565–578. doi: 10.1080/0740817X.2015.1096430

Jannink, J.-L. (2010). Dynamics of long-term genomic selection. *Genet. Sel. Evol.* 42, 35. doi: 10.1186/1297-9686-42-35

Jannink, J.-L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* 9, 166–177. doi: 10.1093/bfgp/elq001

Jansen, G. B., and Wilton, J. W. (1985). Selecting mating pairs with linear programming techniques. *J. Dairy Sci.* 68, 1302–1305. doi: 10.3168/jds.S0022-0302(85)80961-9

Kamenya, S. N., Mikwa, E. O., Song, B., and Odeny, D. A. (2021). Genetics and breeding for climate change in Orphan crops. *Theor. Appl. Genet.* 134, 1787–1815. doi: 10.1007/s00122-020-03755-1

Kemper, K. E., Bowman, P. J., Pryce, J. E., Hayes, B. J., and Goddard, M. E. (2012). Long-term selection strategies for complex traits using high-density genetic markers. *J. Dairy Sci.* 95, 4646–4656. doi: 10.3168/jds.2011-5289

Kinghorn, B. P. (2011). An algorithm for efficient constrained mate selection. *Genet. Sel. Evol.* 43, 4. doi: 10.1186/1297-9686-43-4

Kosambi, D. D. (1943). The estimation of map distances from recombination values. *Ann. Eugen.* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x

Lee, L. H., Chew, E. P., Teng, S., and Goldsman, D. (2010). Finding the non-dominated Pareto set for multi-objective simulation models. *IIE Trans.* 42, 656–674. doi: 10.1080/07408171003705367

Lehermeier, C., Teyssèdre, S., and Schön, C.-C. (2017). Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. *Genetics* 207, 1651–1661. doi: 10.1534/genetics.117.300403

Li, Y., Kadarmideen, H. N., and Dekkers, J. C. M. (2008). Selection on multiple QTL with control of gene diversity and inbreeding for long-term benefit. *J. Anim. Breed. Genet.* 125, 320–329. doi: 10.1111/j.1439-0388.2007.00717.x

Liang, L., Zöllner, S., and Abecasis, G. R. (2007). GENOME: a rapid coalescent-based whole genome simulator. *Bioinformatics* 23, 1565–1567. doi: 10.1093/bioinformatics/btm138

Lin, Z., Shi, F., Hayes, B. J., and Daetwyler, H. D. (2017). Mitigation of inbreeding while preserving genetic gain in genomic breeding programs for outbred plants. *Theor. Appl. Genet.* 130, 969–980. doi: 10.1007/s00122-017-2863-y

Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits* (Sunderland: Sinauer).

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2021) *cluster: Cluster Analysis Basics and Extensions*. Available online at: https://CRAN.R-project.org/package=cluster.

Meuwissen, T. H. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *J. Anim. Sci.* 75, 934–940. doi: 10.2527/1997.754934x

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

Moeinizade, S., Hu, G., and Wang, L. (2022). A reinforcement Learning approach to resource allocation in genomic selection. *Intelligent Syst. Appl.* 14, 200076. doi: 10.1016/j.iswa.2022.200076

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

Müller, D., Schopp, P., and Melchinger, A. E. (2018). Selection on expected maximum haploid breeding values can increase genetic gain in recurrent genomic selection. *G3* 8, 1173–1181. doi: 10.1534/g3.118.200091

Munos, R. (2011). "Optimistic Optimization of a Deterministic Function without the Knowledge of its Smoothness," in *Advances in Neural Information Processing Systems*. Eds. J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira and K. Q. Weinberger (New York: Curran Associates, Inc). Available at: https://proceedings.neurips.cc/paper/2011/file/7e889fb76e0e07c11733550f2a6c7a5a-Paper.pdf.

Padulosi, S.Bioversity International, , Thompson, J., Rudebjer, P. G. (2013). *Fighting poverty, hunger and malnutrition with neglected and underutilized species: needs, challenges and the way forward* (Rome: Bioversity International).

Preux, P., Munos, R., and Valko, M. (2014). "Bandits attack function optimization," in Beijing: *2014 IEEE Congress on Evolutionary Computation (CEC)*. 2245–2252.

Rabier, C.-E., Barre, P., Asp, T., Charmet, G., and Mangin, B. (2016). On the accuracy of genomic selection. *PloS One* 11, e0156086. doi: 10.1371/journal.pone.0156086

Sonesson, A. K., Woolliams, J. A., and Meuwissen, T. H. E. (2012). Genomic selection requires genomic control of inbreeding. *Genet. Sel. Evol.* 44, 27. doi: 10.1186/1297-9686-44-27

Valko, M., Carpentier, A., and Munos, R. (2013). "Stochastic Simultaneous Optimistic Optimization," in Proceedings of the 30th International Conference on Machine Learning *Proceedings of Machine Learning Research*. Eds. S. Dasgupta and D. McAllester (PMLR, Atlanta, Georgia, USA), 19–27.

Van Grevenhof, E. M., Van Arendonk, J. A. M., and Bijma, P. (2012). Response to genomic selection: the Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. *Genet. Sel. Evol.* 44, 26. doi: 10.1186/1297-9686-44-26

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Wang, L., Zhu, G., Johnson, W., and Kher, M. (2018). Three new approaches to genomic selection. *Plant Breed.* 137, 673–681. doi: 10.1111/pbr.12640

Yabe, S., Iwata, H., and Jannink, J.-L. (2017). A simple package to script and simulate breeding schemes: The breeding scheme language. *Crop Sci.* 57, 1347–1354. doi: 10.2135/cropsci2016.06.0538

Zambon, I., Cecchini, M., Egidi, G., Saporito, M. G., and Colantoni, A. (2019). Revolution 4.0: industry vs. Agriculture in a future development for SMEs. *Processes* 7, 36. doi: 10.3390/pr7010036

Zhang, Z., and Wang, L. (2022). A look-ahead approach to maximizing present value of genetic gains in genomic selection. *G3* 12, 1–8. doi: 10.1093/g3journal/jkac136

Zhao, H., and Speed, T. P. (1996). On genetic map functions. *Genetics* 142, 1369–1377. doi: 10.1093/genetics/142.4.1369

Zhong, S., and Jannink, J.-L. (2007). Using quantitative trait loci results to discriminate among crosses on the basis of their progeny mean and variance. *Genetics* 177, 567–576. doi: 10.1534/genetics.107.075358