Simulation-based future-oriented strategy can optimize decision making in a breeding program

Kosuke Hamazaki[1] and Hiroyoshi Iwata[1*]

[1]Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-Ku, Tokyo 113-8657, Japan.

Corresponding author:

Hiroyoshi Iwata

Postal address: 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

Phone & Fax: +81-03-5841-5068

E-mail: hiroiwata@g.ecc.u-tokyo.ac.jp

Abstract

Recently, genomic selection, that is, selection based on the genetic ability predicted by genome-wide markers, has begun to contribute to rapid and efficient plant breeding. The speed and efficiency of breeding depend not only on the accuracy of genomic selection, but also on decisions regarding selection and crossing, which have a significant impact on the ultimate genetic gain of a breeding program. However, methods for optimizing these decisions have not yet been fully established. This paper proposes a data-scientific framework for optimizing decisions via numerical black-box optimization. The allocation of progenies to each mating pair was parameterized according to a softmax function of multiple selection criteria. Different selection criteria are weighted in the function to optimally balance genetic gains and genetic diversity. The weighting parameters were then optimized via breeding simulations. Simulation studies to evaluate the potential of the proposed method revealed that breeding based on optimized weights attained almost 10 % higher genetic gain than breeding with equal allocation of progenies to all mating pairs. These results suggest that the proposed method can significantly improve the speed and efficiency of breeding through optimized decisions regarding selection and crossing.

Introduction

To meet the increasing demand for agricultural products caused by population explosion in a sustainable system, the development of new cultivars with excellent agronomic characteristics, such as high yield, quality, and nutrient use efficiency, must be achieved more quickly and efficiently[1]. However, conventional breeding methods require at least several years to produce new cultivars, even in annual crops, which will make it difficult to meet the increasing food demand expected in the near feature[2]. Genomic selection (GS) is expected to be a key technology for accelerating the evaluation and selection of superior genotypes, which has been an obstacle to conventional breeding efficiency[3,4]. In GS, genotypic values of target traits are predicted using genomic prediction (GP) models based on genome-wide marker data, and the predicted values obtained are used for selection in breeding programs[3]. GS enables individual-based selection according to the genetic ability predicted by the models without field evaluation of target traits, which leads to efficient and rapid breeding[4,5]. In fact, the superiority of GS over conventional phenotypic selection and selection based on pedigree-derived prediction (pedigree BLUP) has been reported in many simulations and empirical studies[4,6–11] and has also been confirmed in an actual dairy cattle breeding program[12]. However, even though GS has contributed to speeding up breeding in the selection step, decisions regarding selection and crossing in breeding programs must still rely on the intuition and experience of breeders. In other words, the advent of GS has not yet entirely led to optimized decisions in breeding programs.

One method for making decisions related to the efficiency of GS breeding is the choice of criteria for selecting parent candidates. Among the various GP models proposed to

improve the accuracy of GS[3,4,13–18], the total sum of estimated additive marker effects, called genomic estimated breeding value (GEBV), is generally utilized for GS as a selection criterion[3]. Although GEBV is a good selection criterion for a short-term breeding program with a few cycles of selection and crossing, this strategy does not guarantee increased genetic gain in a long-term program[19–21]. Because genetic diversity is greatly reduced owing to the Bulmer effect[22] in GEBV-based GS cycles[23], the genetic gain obtained in a breeding program rapidly reaches a local, rather than global, optimum. To solve the issue of GEBV-based GS, other selection criteria have been proposed to improve the genetic gain expected in long-term breeding programs. Weighted genomic estimated breeding value (WGEBV) is a criterion that utilizes marker allele frequencies to emphasize the marker effects of rare alleles[6,24]. Compared to GEBV, WGEBV can maintain genetic diversity in a breeding population over the long term by suppressing the Bulmer effect[6,24]. Other selection criteria, such as optimal haploid value[25], expected maximum haploid breeding value[26], and optimal population value[27], have been proposed to maintain genetic diversity by considering haplotype information and have surely improved the final genetic gains in long-term breeding programs compared to those from GEBV-based GS. Although these selection criteria are useful for selecting parental candidates, they do not directly enable automatic optimization of decisions regarding selection and crossing. Therefore, breeders must make final decisions based on these criteria.

To overcome the issues of selection criteria, a look-ahead selection (LAS) approach has been proposed as a method for simultaneously optimizing decisions regarding selection and crossing[28]. LAS can optimize whether each pair of individuals should be mated on

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 cycle[28]. Although this approach achieved better results than the previous selection criteria, the assumptions were too simple to estimate the final genetic gain, and LAS could not optimize the number of progenies allocated to each mating pair. Although the impact of the resource allocation of progenies is quite significant[29] to increase the final gain, only one study has attempted to optimize the allocation[30]. 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[30], utilizing the multi-objective optimal computing budget allocation method[31]. However, this method lacks the future-oriented perspectives required for long-term breeding programs. Moreover, there have been no studies on the allocation optimization of progenies for the final genetic gain under multiple selection and crossing cycles, although there have been other previous studies that tried to optimize breeding program decisions from other viewpoints[32,33].

In this paper, we propose a novel framework to replace selection criterion-based GS in long-term breeding programs by optimizing a progeny allocation strategy for the final genetic gains after multiple cycles of selection and crossing. There are two major difficulties in this optimization. One is the difficulty of explicitly describing the breeding process, because the future state of a breeding population is unknown. The second is that breeding is a stochastic rather than deterministic process, since progenies are randomly produced in each generation through meiosis. To address the first difficulty, we focused on the fact that GS combined with future-oriented simulations enabled us to evaluate the future final genetic gain in a long-term breeding program. In addition, we applied black-

box optimization by considering the breeding process as a black-box function whose input is a set of parameters for the allocation strategy and whose output is the final genetic gain resulting from the breeding program. We employed a black-box optimization algorithm, StoSOO[34], which can also optimize a stochastic black-box function such as a breeding process, and which presented a solution to the second difficulty. Multiple selection criteria are used in the allocation of progenies, and the optimization of progeny allocation can automatically determine the optimal combination of criteria and weights to be used in the allocation.

Results

System design assumed in this study

In this study, we assumed that breeders proceed with a breeding program according to the resource allocation strategy proposed by an "AI breeder," which is a decision-making system based on our proposed method (Fig. 1a). The breeders can obtain a set of parameters representing the optimal allocation strategy, $\{\hat{\mathbf{h}}^{(\tau)}\}_{\tau=0,1,2,3}$, by passing the information on an initial population, i.e., marker genotype, marker effects, and recombination rates between 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 initial population information, the AI breeder attempts to simulate breeding programs by defining the allocation strategy (Fig. 1a). At each generation τ in the simulated programs ($\tau \in \{0,1,2,3\}$), the breeder first computes the weighted breeding values (WBV) to select the parent candidates from the population. Then, after determining mating pairs by the diallel crossing between the parent candidates, given the parameter $\mathbf{h}^{(\tau)}$, the breeder automatically determines the number of progenies allocated to each pair $\mathbf{b}^{(\tau)}$, by utilizing the information on selection criteria $\mathbf{\Omega}^{(\tau)}$ computed from the marker genotype and effects, such as breeding values (BV), WBV, and expected genetic variance of progenies (GVP) (Fig. 1b). In other words, we assumed that $\mathbf{b}^{(\tau)}$ is determined by applying the softmax function to the weighted sum of the selection criteria, as shown in equation (1).

$$\mathbf{b}^{(\tau)} = \left[N^{(\mathrm{pop})(\tau+1)} \times \frac{\exp\left(\mathbf{\Omega}^{(\tau)}\, \mathbf{h}^{(\tau)}\right)}{\sum_Q \exp(\mathbf{\Omega}^{(\tau)}\, \mathbf{h}^{(\tau)})} \right], \tag{1}$$

where $N^{(\mathrm{pop})(\tau+1)}$ is the population size in the next generation $\tau+1$, and $[x]$ is the Gauss symbol of x (i.e., the maximum integer that is equal to or lower than x). Here, $\mathbf{h}^{(\tau)}$ can be interpreted as the weighting parameters that correspond to the importance of each selection criterion. After determining $\mathbf{b}^{(\tau)}$, the AI breeder can simulate the next generation based on the principle of meiosis. Then, by repeating the selection and crossing steps toward the final generation of the breeding program, the AI breeder evaluates the performance of the allocation strategy with $\mathbf{h}^{(\tau)}$. Thus, the goal of the AI breeder is to estimate $\{\mathbf{h}^{(\tau)}\}_{\tau=0,1,2,3}$ that maximizes the final genetic gain via simulations of breeding programs. To optimize the weighting parameters, a black-box optimization algorithm called StoSOO[34] was used in this study (Fig. 1A). The StoSOO algorithm is an extension of the SOO algorithm[35,36], 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 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 $\{\mathbf{h}^{(\tau)}\}_{\tau=0,1,2,3}$, 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 $\{\hat{\mathbf{h}}^{(\tau)}\}_{\tau=0,1,2,3}$ is passed from the AI breeder to the real breeders.

In this study, as a proof of concept, we evaluated this proposed framework by conducting simulation studies. We prepared six variations of the proposed allocation strategy according to the components of $\mathbf{\Omega}^{(\tau)}$ 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) (Fig. 1b). We compared these six optimized strategies to (g) the equal allocation strategy that equally allocates progenies to each pair, that is, $\mathbf{h}^{(\tau)} = \mathbf{0}$, by conducting 10,000 breeding simulations using the optimized $\{\hat{\mathbf{h}}^{(\tau)}\}_{\tau=0,1,2,3}$ (Fig. 1a). In the following results, we abbreviate the above strategies as (a) BV, (b) WBV, (c) TBV, (d) BVGVP, (e) WBVGVP, (f) TBVGVP, and (g) EQ. Here, for the BV and WBV strategies, we used only one criterion for the allocation of progenies. For simplicity, we also assumed true marker effects to compute the selection criteria.

For more details on the system design and its evaluation in this study, please refer to the Online Methods.

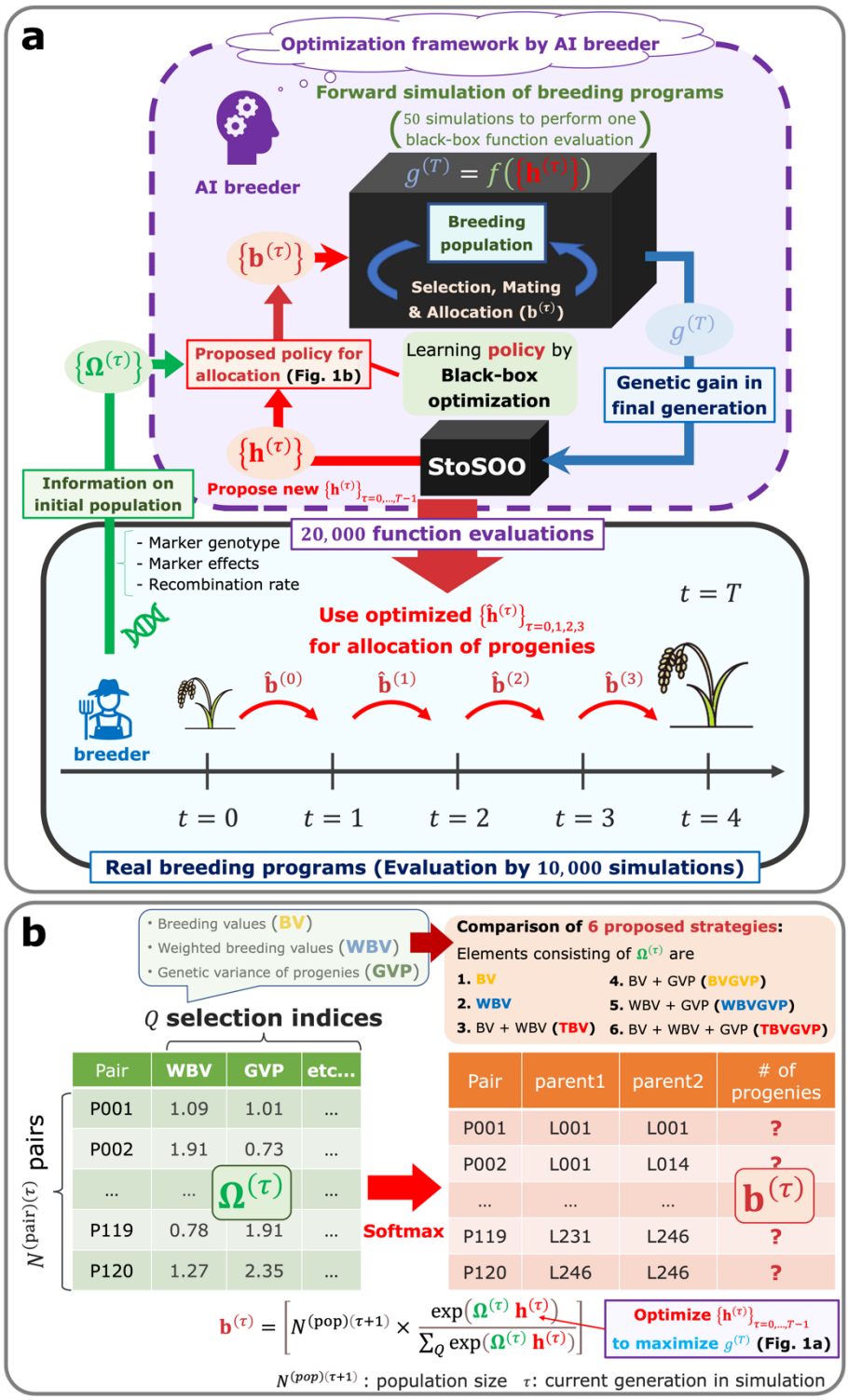

Fig. 1. **System design assumed in this study.** (a) The proposed framework to optimize decision-making in the long-term breeding program. (b) The resource allocation strategy of progenies proposed in this study.

Evaluation of convergence conditions of the StoSOO algorithm

For the optimization of the allocation strategy proposed in this study 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 derived using the StoSOO algorithm for each strategy. As described above, the AI breeder conducted 50 simulations of the breeding program 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 proposed allocation strategies, the change in the function values was plotted and compared with the function value based on the equal allocation strategy (Fig. S1).

For all strategies on the components of $\mathbf{\Omega}^{(\tau)}$, the function value increased almost monotonically and reached the quasi-optimized function value after the 20,000 evaluations, which was much larger than that expected under the equal allocation strategy (Fig. S1). These results indicated that StoSOO successfully improved the expected final genetic gain evaluated by the simulations and optimized the parameter set $\{\mathbf{h}^{(\tau)}\}_{\tau=0,1,2,3}$, resulting in the superiority of the optimized allocation strategy over the equal allocation strategy.

Genetic gains over four generations

For all strategies on the components of $\mathbf{\Omega}^{(\tau)}$, we compared the optimized and equal allocation strategies by plotting the change in the genetic gain for each generation (Fig.

2a). We 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 t) and initial (generation 0) populations. Using the optimized allocation strategies obtained after 20,000 function evaluations, we evaluated the expected genetic gains of the seven strategies based on 10,000 breeding simulation results using one replication for the phenotype simulation.

In the final generation, $t = 4$, all the strategies with optimized allocations showed higher genetic gains than the equal allocation strategy (Fig. 2a). Strategies that considered GVP outperformed those that did not consider 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 the strategies that did not consider GVP (Fig. 2a). After the second generation, $t = 2$, strategies that considered GVP overtook those that did not consider GVP, increasing their advantage over the other strategies through the final generation.

Genetic diversity over four generations

We compared the optimized and equal allocation strategies for changes in genetic diversity across generations (Fig. 2b). The genetic diversity of a population in each generation was evaluated on the basis of the genetic variance of the true genotypic values in the population.

First, in $t = 1$, all the proposed strategies, except WBVGVP, showed lower genetic

variances than those in the initial population, $t = 0$ (Fig. 2b). From $t = 2$ to $t = 4$, genetic diversity was preserved in all strategies compared to $t = 0$. In these generations, strategies that considered GVP maintained higher genetic variance than those that did not consider GVP and the equal allocation strategy. The difference between the strategies that considered GVP and those that did not was most pronounced 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 (Fig. 2b). The strategies considering only WBV as a selection criterion appeared to maintain a higher genetic variance than the strategies considering BV or both BV and WBV. However, the difference was much smaller than that between strategies that considered GVP and those that did not.

Optimized weighting parameters for each strategy

After the 20,000 function evaluations, we obtained a set of optimized weighting parameters $\{\mathbf{h}^{(\tau)}\}_{\tau=0,1,2,3}$ for each allocation strategy (Tables 1, S1, and S2). A large value of $\mathbf{h}^{(\tau)}$ indicates that the weight of the corresponding criterion is large.

As for the strategies that did not consider GVP, the weighting parameters in generation $\tau = 1$, $\mathbf{h}^{(1)}$, showed the smallest values among all generations, and the weights in the latter generations increased through the final generation (Tables 1, S1, and S2). When strategies considered GVP, the weighting parameters for the selection criteria (BV and/or WBV) did not show smaller values in generation $\tau = 1$ than in other generations. The tendency of the selection criteria to increase in later generations did not change with the strategies considering GVP, and the weights for GVP were similar to the weights of the

selection criteria across generations.

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 2c, 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 figure, the better the performance of the strategy.

Again, the proposed strategies outperformed the equal allocation strategy; among the proposed strategies, those that considered GVP showed better performance than those that did not in terms of the CDFs (Fig. 2c). The CDF curves for the proposed 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 proposed strategies with GVP, approximately 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 %).

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

Finally, we evaluated the genetic gains in the final generation, $t = 4$, using ten replications of the phenotype simulation to confirm that the proposed strategies showed better results than the equal allocation strategy for traits with different quantitative trait loci (QTL) positions and effects (Fig. 2d). When simulating different traits, we changed QTL positions and their effects, whereas we fixed the number of QTLs and distribution of QTL effects.

To reduce the computational time, optimized allocation strategies were determined after 4,000 function evaluations for each proposed strategy. The number of function evaluations was determined to ensure that the function value reached a certain level and to consider the computation time based on the results in the "**Evaluation of convergence conditions of the StoSOO algorithm**" subsection.

Even when the QTL positions and effects of the target trait were changed, the proposed strategies stably outperformed the equal allocation strategy, and the strategies that considered GVP outperformed those that did not (Fig. 2d). When we evaluated the improvement rate in the final genetic gain of each proposed strategy compared to the equal allocation strategy using ten replications of the phenotype simulation, a similar trend was observed between the strategies that considered and did not consider GVP (Table 2). In addition, the proposed 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 (Table 2).

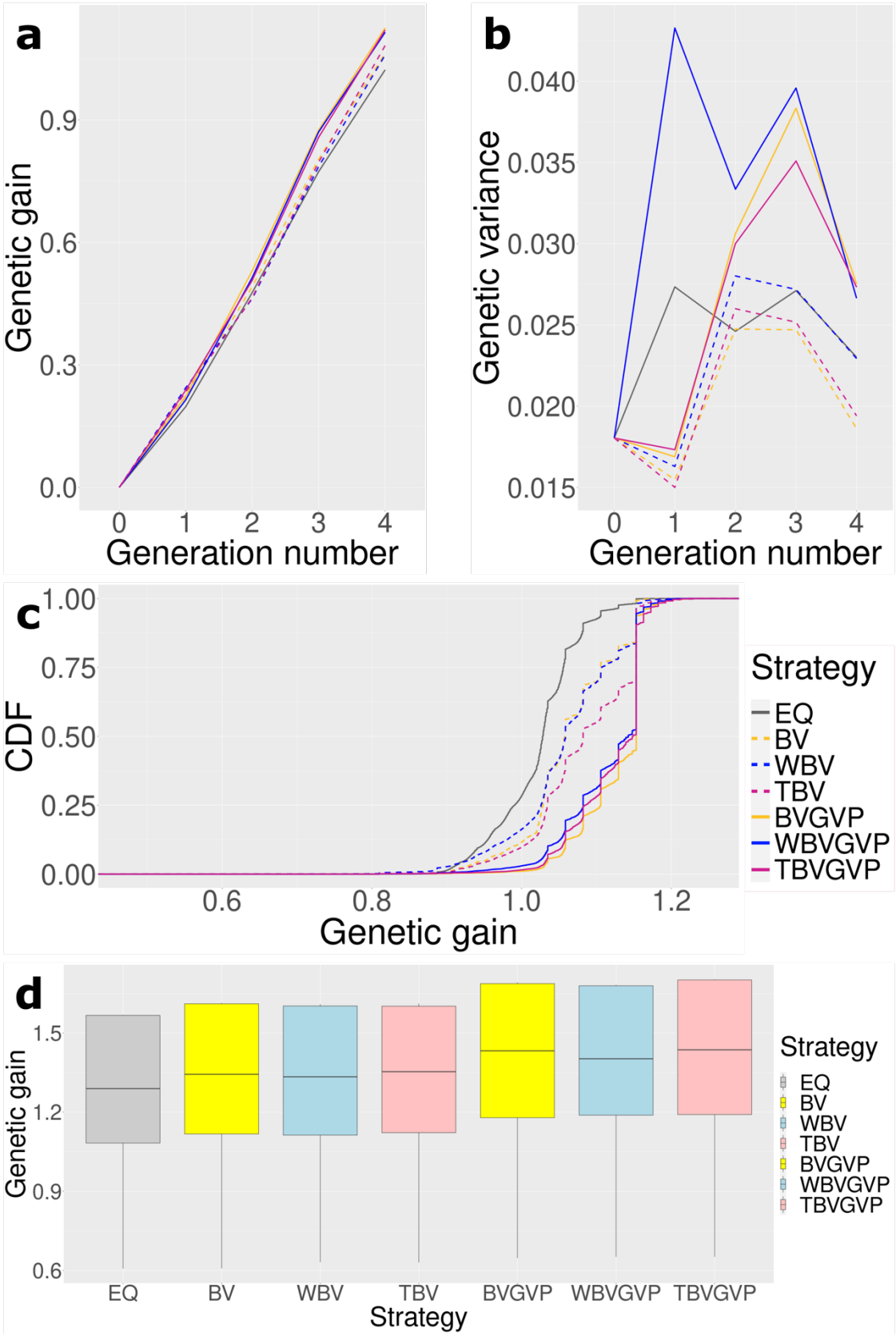

Fig. 2. **Comparison of the seven strategies (six optimized and one non-optimized) in four ways**

a: Change in the expected genetic gains over four generations.

b: Change in the genetic variances over four generations.

c: Genetic gains across different simulation repetitions.

d: Genetic gains in the final generation using ten replications for the phenotype simulation.

In all figures, we compared the seven strategies: EQ: equal allocation (**abc**: black solid, **d**: grey), BV: optimized allocation based on BV (**abc**: orange dashed, **d**: yellow), WBV: optimized allocation based on WBV (**abc**: blue dashed, **d**: light blue), TBV: optimized allocation based on BV and WBV (**abc**: red dashed, **d**: light pink), BVGVP: optimized allocation based on BV and GVP (**abc**: orange solid, **d**: yellow), WBVGVP: optimized allocation based on WBV and GVP (**abc**: blue solid, **d**: light blue), and TBVGVP: optimized allocation based on BV, WBV, and GVP (**abc**: red solid, **d**: light pink). In figures **a** and **b**, 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 program. In figure **c**, 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 figure **d**, for each proposed strategy, the weighting parameter was optimized based on $4{,}000$ function evaluations.

Discussion

Genomic selection is a promising technique that contributes to the acceleration of breeding. In this paper, we propose a novel framework that can upgrade conventional GS in breeding programs 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.

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 $\mathbf{h}^{(\tau)}$ 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 increased almost monotonically as StoSOO conducted more function evaluations (Fig. S1). Thus, when applying this novel optimization approach to actual breeding programs, 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 for the allocation of progenies.

After the optimization of 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 that considered GVP (BVGVP, WBVGVP, TBVGVP), optimized strategies that did not consider GVP (BV, WBV, TBV), and non-optimized strategy (EQ) (Fig. 2a, d). These results suggest that the proposed optimization framework for the allocation of progenies is efficient in conventional GS, and that considering the genetic diversity for later

generations in the proposed framework can further improve the final genetic gains in mid- and long-term breeding programs. In particular, the superiority of the proposed strategies considering GVP was owing to the fact that those strategies guaranteed a certain level of the final genetic gains compared to the other strategies, even when random factors, such as recombination between markers and segregation of alleles, are not realized in a real breeding program as they are in the optimization process (Fig. 2c). Thus, because the proposed optimized strategies considering GVP are expected to consistently produce better genotypes than conventional GS, breeders can easily adopt the proposed framework in real breeding programs.

However, 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 (Fig. 2a, d). Because WBV was always utilized in the selection step in this study, we were able to maintain the genetic diversity throughout the breeding program regardless of the allocation strategy (Fig. 2b), which may reduce the superiority of WBV in the allocation step against BV. We do not show the results here, but when BV was utilized in the selection step, the final genetic gains largely 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 the choice of criteria in the allocation step, it is strongly recommended that WBV be used instead of BV in the selection step when applying the proposed method to a breeding program.

When focusing on the genetic gains in $t = 1$, the proposed strategies that considered GVP showed lower genetic gains than those that did not (Fig. 2a). In contrast, the proposed strategies that considered GVP showed larger genetic variances than those that

did not from $t = 1$ (Fig. 2b). These results suggest that the proposed strategies considering GVP were able to attain higher genetic gains in the final generation by their success in maintaining genetic diversity throughout the programs and converting it into genetic gain toward the final generation. 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 stronger in later generations (Table 1). In addition, when focusing on the optimized weighting parameters for the strategies that did not consider 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 proposed 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, by estimating proper weights for each selection criterion.

The proposed future-oriented simulation-based framework achieved much higher final genetic gains than the GS method with non-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 2). This improvement is quite significant because we compared the proposed framework to the GS based on WBV[6,24], one of the most promising methods in long-term breeding programs to date, not phenotypic selection or conventional GS based on BV[3]. However, it is important to note that we used true marker effects to compute the selection criteria and GVP in this study. In real breeding programs, we cannot know such true marker

effects, and instead use estimated marker effects based on GP models. Because the proposed optimization algorithm can be largely influenced by the estimation accuracy of GP models, developing a robust method to optimize the allocation strategy in such cases is an important future study. The other important factor when assuming real breeding programs is the timing of 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 is expected to directly lead to further improvement in the final genetic gains. In addition, we can deal with long-term breeding programs by conducting appropriate model updates. Thus, in a future study, we should further investigate the potential of our proposed framework by assuming real breeding programs with the factors discussed above. The development of a future-oriented optimization framework for breeding programs proposed in this study is expected to contribute to the acceleration, efficiency, and optimization of breeding.

Tables

Table 1. **Optimized weighting parameters $\mathbf{h}^{(\tau)}$ in different generations $\tau = 0, \ldots, T-1$ for BV and BVGVP strategies.**

	BV		**BVGVP**	
Generation ($\boldsymbol{\tau}$)	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

Table 2. **Minimum, mean, and maximum improvement rate in the final genetic gain of the proposed 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**

References

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

2. Eggen, A. The development and application of genomic selection as a new breeding paradigm. *Anim. Front.* **2**, 10–15 (2012).

3. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).

4. Jannink, J.-L., Lorenz, A. J. & Iwata, H. Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics* **9**, 166–177 (2010).

5. Crossa, J. *et al.* Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci.* **22**, 961–975 (2017).

6. Goddard, M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257 (2009).

7. Goddard, M. E., Hayes, B. J. & Meuwissen, T. H. E. Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* **128**, 409–421 (2011).

8. Crossa, J. *et al.* Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3* **3**, 1903–1926 (2013).

9. Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* **92**, 433–443 (2009).

10. Hickey, J. M. *et al.* Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Sci.* **54**, 1476–1488 (2014).

11. Rabier, C.-E., Barre, P., Asp, T., Charmet, G. & Mangin, B. On the Accuracy of Genomic Selection. *PLoS One* **11**, e0156086 (2016).

12. García-Ruiz, A. *et al.* 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**, E3995-4004 (2016).

13. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).

14. de los Campos, G. *et al.* Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* **182**, 375–385 (2009).

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

16. Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186 (2011).

17. Gianola, D. & van Kaam, J. B. C. H. M. Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* **178**, 2289–2303 (2008).

18. González-Camacho, J. M., Crossa, J., Pérez-Rodríguez, P., Ornella, L. & Gianola, D. Genome-enabled prediction using probabilistic neural network classifiers. *BMC Genomics* **17**, 208 (2016).

19. Sonesson, A. K., Woolliams, J. A. & Meuwissen, T. H. E. Genomic selection requires genomic control of inbreeding. *Genet. Sel. Evol.* **44**, 27 (2012).

20. Lin, Z., Shi, F., Hayes, B. J. & Daetwyler, H. D. Mitigation of inbreeding while preserving genetic gain in genomic breeding programs for outbred plants. *Theor. Appl. Genet.* **130**, 969–980 (2017).

21. Akdemir, D., Beavis, W., Fritsche-Neto, R., Singh, A. K. & Isidro-Sánchez, J. Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity* **122**, 672–683 (2019).

22. Bulmer, M. G. The Effect of Selection on Genetic Variability. *Am. Nat.* **105**, 201–211 (1971).

23. Van Grevenhof, E. M., Van Arendonk, J. A. M. & Bijma, P. 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 (2012).

24. Jannink, J.-L. Dynamics of long-term genomic selection. *Genet. Sel. Evol.* **42**, 35 (2010).

25. Daetwyler, H. D., Hayden, M. J., Spangenberg, G. C. & Hayes, B. J. Selection on Optimal Haploid Value Increases Genetic Gain and Preserves More Genetic Diversity Relative to Genomic Selection. *Genetics* **200**, 1341–1348 (2015).

26. Müller, D., Schopp, P. & Melchinger, A. E. Selection on Expected Maximum Haploid Breeding Values Can Increase Genetic Gain in Recurrent Genomic Selection. *G3* **8**, 1173–1181 (2018).

27. Goiffon, M., Kusmec, A., Wang, L., Hu, G. & Schnable, P. S. Improving Response in Genomic Selection with a Population-Based Selection Strategy: Optimal Population Value Selection. *Genetics* **206**, 1675–1682 (2017).

28. Moeinizade, S., Hu, G., Wang, L. & Schnable, P. S. Optimizing Selection and Mating in Genomic Selection with a Look-Ahead Approach: An Operations Research Framework. *G3* **9**, 2123–2133 (2019).

29. Wang, L., Zhu, G., Johnson, W. & Kher, M. Three new approaches to genomic

selection. *Plant Breed.* **137**, 673–681 (2018).

30. Hunter, S. R. & McClosky, B. Maximizing quantitative traits in the mating design problem via simulation-based Pareto estimation. *IIE Trans.* **48**, 565–578 (2016).

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

32. Amini, F., Franco, F. R., Hu, G. & Wang, L. The look ahead trace back optimizer for genomic selection under transparent and opaque simulators. *Sci. Rep.* **11**, 4124 (2021).

33. Moeinizade, S., Hu, G. & Wang, L. A reinforcement learning approach to resource allocation in genomic selection. *arXiv [q-bio.GN]* (2021).

34. Valko, M., Carpentier, A. & Munos, R. Stochastic Simultaneous Optimistic Optimization. in *Proceedings of the 30th International Conference on Machine Learning* (eds. Dasgupta, S. & McAllester, D.) vol. 28 19–27 (PMLR, 2013).

35. Munos, R. Optimistic Optimization of a Deterministic Function without the Knowledge of its Smoothness. in *Advances in Neural Information Processing Systems* (eds. Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F. & Weinberger, K. Q.) vol. 24 (Curran Associates, Inc., 2011).

36. Preux, P., Munos, R. & Valko, M. Bandits attack function optimization. in *2014 IEEE*

Congress on Evolutionary Computation (CEC) 2245–2252 (2014). doi:10.1109/CEC.2014.6900558.

37. Liang, L., Zöllner, S. & Abecasis, G. R. GENOME: a rapid coalescent-based whole genome simulator. *Bioinformatics* **23**, 1565–1567 (2007).

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

Methods

Overview

This section provides an overview of the method used in this simulation study (Fig. S2). We first simulated genome-wide markers and QTLs for the initial population and the corresponding QTL effects. 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 program until the final generation. Selection and mating processes were carried out based on two selection criteria: breeding values (BV) and weighted breeding values (WBV), and the expected genetic variance of progenies (GVP). We compared two breeding strategies regarding the allocation of progenies: equal allocation and allocation optimized by a black-box optimization algorithm, StoSOO. These strategies were evaluated by simulating 10,000 breeding programs 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 a description of the simulation of genome-wide markers and QTLs, the selection and mating processes in breeding programs, and the evaluation of breeding programs. Because optimized allocation requires the simulation and evaluation of breeding programs, the method of evaluating breeding programs is explained before the description of the optimized allocation strategy. The calculation of the GVP required for optimized allocation is explained at the end of this section.

Selection criteria

We used two selection criteria, breeding value (BV)[3] and weighted breeding value

(WBV)[6,24], to select parent candidates and allocate progenies in breeding programs (Fig. S2-1).

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 program, $\mathbf{u}^{(\mathrm{BV})(t)}$ 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 α_m is a true marker effect at the m-th marker. In this study, we assumed that the true marker effect α_m was known. BV is known as a criterion for expressing additive genotypic values directly transmitted to the next generation and is often used in conventional breeding programs with GS[3].

The second selection criterion, WBV, can be expressed as in equation (3):

$$\mathbf{u}^{(\mathrm{WBV})(t)} = \sum_{m=1}^{M} \left\{ \mathbf{w}_m^{(t)} \operatorname{sign}(\alpha_m) \frac{\frac{\pi}{2} - \sin^{-1}\left(\sqrt{p_m^{(t)}}\right)}{\sqrt{p_m^{(t)}\left(1 - p_m^{(t)}\right)}} \right\}, \tag{3}$$

where $\mathbf{u}^{(\mathrm{WBV})(t)}$ 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 for generation t, and the other terms are defined in equation (2). Emphasizing the effects of rare alleles, 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[6,24].

Simulation of the initial population and genome structure

Genome-wide markers and QTLs for founder haplotypes were generated using the coalescent simulator GENOME[37] (Fig. S2-2). 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 material of the 2,000 genotypes. These procedures for simulating the genetic resources are the same as those in the R package "BreedingSchemeLanguage" by Yabe et al[38]. From these simulated genetic resources, $N^{(\mathrm{pop})(0)} = 250$ genotypes were selected for the initial population of a breeding program 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 genetic resources using the "pam" function of the R package "cluster" version 2.1.2[39]. This initial population was regarded as generation $t = 0$ in a breeding program.

Simulation of QTLs and genotypic values

For each chromosome, 400 loci were randomly sampled from the 500 loci and were regarded as single nucleotide polymorphism (SNP) markers available for breeders. To

simulate QTLs and phenotypes, we assumed the simplest genetic architecture, as described below (Fig. S2-3). First, among the remaining 100 loci, 2 loci per chromosome were randomly selected as QTLs (i.e., a total of $M^{(\mathrm{QTL})} = 20$ QTLs). QTL effects were then sampled from the normal distribution, as shown in equation (4).

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

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

Breeding program

The breeding program in this study was carried out according to the following steps (Fig. S2-4).

1. Select parent candidates (Fig. S2-4-1)

   Based on the order of increasing WBV computed using the true marker (QTL) effects, $\boldsymbol{\alpha}^{(\mathrm{QTL})}$ for the current population with generation t ($t \in \{0, 1, 2, 3\}$), the top $N^{(\mathrm{sel})(t)} = 15$ genotypes were selected as parent candidates for the next generation. We used only WBV as the selection criterion in this study because 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 (Fig. S2-4-2)

   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, $N^{(\mathrm{pair})(t)} = \frac{N^{(\mathrm{sel})(t)}(N^{(\mathrm{sel})(t)}+1)}{2} = 120$ pairs were prepared for crossing.

3. Allocate progenies to each mating pair (Fig. S2-4-3)

   To allocate progenies to each mating pair determined in step 2, two strategies were used to create a crossing table. Both strategies generated $N^{(\mathrm{pop})(t+1)} = 250$ progenies for the next generation.

   a. The proposed method in this study

      One strategy was the method proposed in this study, which applied the softmax function to the weighted sum of the selection criteria computed in step 1. For more details, please refer to the "**Proposed method to optimize the allocation of progenies**" 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, 3 progenies were allocated to each of 10 mating pairs, and 2 progenies were allocated to each of the remaining 110 mating pairs.

4. Generate progenies for the next generation (Fig. S2-4-4)

   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[40,41] based on the linkage map generated by GENOME. These two gametes were then combined to generate one new progeny, resulting in a new population (with generation $t+1$) of

$N^{(\mathrm{pop})(t+1)}$ progenies.

5. Repeat the selection and mating process

   Steps 1-4 were repeated four times until the population reached the final generation $T = 4$. Finally, the results for each strategy were evaluated using the true additive genotypic values $\mathbf{u}^{(\mathrm{TGV})(T)}$ for the latest population with generation T.

Evaluation of results

In one breeding program, the true additive genotypic value $\mathbf{u}^{(\mathrm{TGV})(t)}$ was computed for each generation t (Fig. S2-5). Then, the top $N^{(\mathrm{top})(t)} = 5$ genotypes were selected in the order of increasing $\mathbf{u}^{(\mathrm{TGV})(t)}$, and the empirical mean of the $\mathbf{u}^{(\mathrm{TGV})(t)}$ for these $N^{(\mathrm{top})(t)}$ genotypes, $u^{(\mathrm{top})(t)}$, was computed to evaluate the breeding strategy. For a given simulation dataset of the initial population, we conducted 10,000 different breeding programs 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.

Proposed method to optimize the allocation of progenies

In this study, we propose a novel breeding strategy/policy to quasi-optimize each selection criterion in terms of optimizing the allocation of progenies (Fig. S2-6 and Fig. 1b). In other words, we determined the number of progenies allocated to each mating pair at generation τ, $\mathbf{b}^{(\tau)}$, not by the equal allocation strategy described above, but by applying the softmax function to the weighted sum of the selection criteria, as in equation (1).

$$\mathbf{b}^{(\tau)} = \left[N^{(\mathrm{pop})(\tau+1)} \times \frac{\exp\left(\mathbf{\Omega}^{(\tau)}\, \mathbf{h}^{(\tau)}\right)}{\sum_{Q} \exp\left(\mathbf{\Omega}^{(\tau)}\, \mathbf{h}^{(\tau)}\right)} \right], \tag{1}$$

where τ is the current generation number in the breeding simulations, $\mathbf{b}^{(\tau)}$ is an

$N^{(\mathrm{pair})(\tau)} \times 1$ vector representing the number of progenies allocated to each mating pair at generation τ, $\mathbf{\Omega}^{(\tau)}$ is an $N^{(\mathrm{pair})(\tau)} \times Q$ matrix of Q different selection criteria 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 Gauss symbol (i.e., the maximum integer that is equal to or lower than x). Here, $\mathbf{h}^{(\tau)}$ is a $Q \times 1$ vector required for the determination of $\mathbf{b}^{(\tau)}$, which corresponds to how much importance is given to each selection criterion, and estimating $\mathbf{h}^{(\tau)}$ that maximizes the final genetic gain, $g^{(T)}$, via simulations of breeding programs is the main goal of this strategy. When $\mathbf{h}^{(\tau)} = \mathbf{0}$, the proposed 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 initial population, that is, $g^{(T)} = u^{(\mathrm{top})(T)} - u^{(\mathrm{top})(1)}$.

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

In this study, we used the StoSOO algorithm[34] to optimize the parameter vector $\mathbf{h}^{(\tau)}$ (Fig. S2-7). The StoSOO algorithm is an extension of the SOO algorithm[35,36], 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 implicit stochastic function within the number of evaluations. To apply StoSOO to our problem, we performed 50 simulations using one set of $\{\mathbf{h}^{(\tau)}\}_{\tau=0,1,2,3}$, and the empirical mean of the final genetic gains $g^{(T)}$ for these 50 simulations was regarded 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 $\{\mathbf{h}^{(\tau)}\}_{\tau=0,1,2,3}$. The function evaluations were repeated 20,000 times, giving us quasi-optimal solutions, $\{\hat{\mathbf{h}}^{(\tau)}\}_{\tau=0,1,2,3}$, for this optimization problem. We then performed 10,000 simulations based on the estimated $\{\hat{\mathbf{h}}^{(\tau)}\}_{\tau=0,1,2,3}$ to evaluate each optimized strategy proposed in this study, as described in the "**Evaluation of results**" section.

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

As the candidates for $\boldsymbol{\Omega}^{(\tau)}$ to compute $\mathbf{b}^{(\tau)}$, we used the two selection criteria, BV and WBV, and the expected GVP for each mating pair (Fig. S2-8). 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[42–44]:

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

where $v_n^{(\tau)}$ is an expected GVP for mating pair n with generation τ, $\boldsymbol{\alpha}$ 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_{M} and paternal parent n_{P} of mating pair n, respectively. A diagonal (m, m) element of $\mathbf{C}^{(n_{\mathrm{M}})}$, $C_{mm}^{(n_{\mathrm{M}})}$, is 0 if an m-th marker of n_{M} is homozygote, and 1 if otherwise (i.e., heterozygote), and a non-diagonal (m_1, m_2) element of $\mathbf{C}^{(n_{\mathrm{M}})}$, $C_{m_1 m_2}^{(n_{\mathrm{M}})}$, is 0 if the m_1-th or m_2-th marker of n_{M} is

homozygote, $1 - 2r_{m_1 m_2}$ if both m_1-th and m_2-th markers are heterozygote without recombination, and $2r_{m_1 m_2} - 1$ 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, $\mathbf{v}^{(\tau)}$ was used as one column vector of $\mathbf{\Omega}^{(\tau)}$.

For the components of $\mathbf{\Omega}^{(\tau)}$, we prepared six different strategies as follows: (a) only BV, (b) only WBV, (c) both selection criteria (i.e., BV and WBV), (d) the expected GVP in addition to BV, (e) the expected GVP in addition to WBV, and (f) the expected GVP in addition to (c). We compared these six optimized strategies with (g) the equal allocation strategy. We abbreviated the above strategies as (a) BV, (b) WBV, (c) TBV, (d) BVGVP, (e) WBVGVP, (f) TBVGVP, and (g) EQ. Here, for the BV and WBV strategies, we used only one criterion for the allocation of progenies. However, in such cases, we optimized the weighting parameter $\{\mathbf{h}^{(\tau)}\}_{\tau=0,1,2,3}$ by focusing on the optimization of the gradient of the number of progenies allocated to each pair against the size of that criterion.

References

39. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. cluster: Cluster Analysis Basics and Extensions. (2021).

40. Kosambi, D. D. The estimation of map distances from recombination values. *Ann. Eugen.* **12**, 172–175 (1943).

41. Zhao, H. & Speed, T. P. On genetic map functions. *Genetics* **142**, 1369–1377 (1996).

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

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

44. Lehermeier, C., Teyssèdre, S. & Schön, C.-C. Genetic Gain Increases by Applying the Usefulness Criterion with Improved Variance Prediction in Selection of Crosses. *Genetics* **207**, 1651–1661 (2017).

Declarations

Data availability

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 in the following "**Code availability**."

Code availability

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 programs. "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.

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 also like to thank Dr. Yohei Akimoto for fruitful discussions on how to optimize decision-making in a breeding program. We would like to thank Editage (www.editage.com) for English language editing.

Author contributions

KH implemented the simulation of breeding programs, developed the proposed method for optimizing allocation strategies in breeding programs, and drafted the manuscript. KH and HI conceived of and designed the study. HI provided administrative support and supervised the study. All authors have read and approved the final manuscript.

Competing interests

Not applicable. The authors declare that there are no conflicts of interest.

Ethics declarations

Not applicable.