Received: 29 March 2023 | Revised: 10 May 2023 | Accepted: 14 May 2023

DOI: 10.1111/pbr.13118

ORIGINAL ARTICLE

Plant Breeding WILEY

Optimizing quantitative trait loci introgression in elite rice germplasms: Comparing methods and population sizes to develop new recipients via stochastic simulations

John Damien Platten[1] | Roberto Fritsche-Neto[2]

[1]International Rice Research Institute (IRRI), Los Baños, Philippines

[2]H. Rouse Caffey Rice Research Station, LSU AgCenter, Rayne, Louisiana, USA

Correspondence
Roberto Fritsche-Neto, H. Rouse Caffey Rice Research Station, LSU AgCenter, Rayne, LA, USA.
Email: rfneto@agcenter.lsu.edu

Funding information
Bill and Melinda Gates Foundation (BMGF, Accelerated Genetic Gain in Rice in South Asia and Africa [AGGRi] Alliance), Grant/Award Number: OPP1194889

Abstract

This study compared three strategies to develop new recipients for quantitative trait loci (QTL) introgression (background recovery [BG], selective sweep [SS] and breeding value [BV]) in a short-term rice breeding programme (over five breeding cycles). Furthermore, we evaluated two different numbers of recipients (10 and 20) in the introgression process and how they influence the population performance and the QTL fixation over cycles. Finally, we used the International Rice Research Institute (IRRI) rice breeding framework as the model to perform the stochastic simulations. Each strategy was simulated and replicated 100 times. Regardless of the selection strategy used, the QTL introgression resulted in substantial penalties in yield performance. However, introducing fewer new parents to the augmentation process minimized this effect. Conversely, the time required to achieve fixation of target QTLs showed substantial differences, with selection for BV during augmentation outperforming other methods. Overall, the BV_10 strategy (10 parents selected based on genomic estimated BV) displayed the best trade-off between reduced penalty from introducing new QTLs with a reasonable speed at which those QTLs can achieve fixation over subsequent breeding cycles.

KEYWORDS
gene introgression, genomic selection, qualitative traits, simulation study

1 | INTRODUCTION

Modern farming will be facing major challenges in the coming years. Increases in population and standards of living are projected to double the demand for grain crops by 2050. At the same time, the effects of climate change are already being felt. Temperatures have risen by 0.95°C from 2009 to 2018 (Ting & Vasel-Be-Hagh, 2022); extreme weather events such as droughts, heat waves and typhoons are predicted to increase in frequency. At the same time, crops are increasingly being grown on more marginal land due to expanding cropping areas, displacement from urban growth and degradation of existing farmland. Disease pressures shift as cropping intensity and opportunities for spread to new regions increase. In all, rice production is predicted to require an increase of 117% by 2050 to offset these various pressures (Ray et al., 2013). Current trends in productivity due to breeding are 1% or less (Khanna et al., 2022; Ladha et al., 2021), far short of the 2.4% predicted to be required to meet these demands.

A major revolution in the speed and effectiveness of breeding for major crops is required. Therefore, many initiatives have increased genetic gains to 2.4% or higher (Ladha et al., 2021; Nayak

et al., 2022). These initiatives emphasize mechanisms to reduce the breeding cycle time, increase selection accuracy and intensity and so on. Adopting modern breeding methods based on quantitative genetics, such as enabling genomic predictions, help address many of these parameters. However, with these quantitative genetics approaches, the substantial value lies in more qualitative genetics systems. For example, selection for major genes and QTLs enables rapid improvement of a particular trait, reducing the time required to produce new varieties that are substantially better than those previously available (Kumar et al., 2014). Major-gene selection is also typically applied at different stages of the breeding process, so it is at least partially decoupled from selection for traits under polygenic control, allowing gains in a wider variety of traits per breeding cycle (Hospital & Charcosset, 1997). Crucially, the relative simplicity of selection for major genes allows these to be moved around rapidly within and between breeding programmes, providing agility that will be essential in meeting the rapidly changing demands of climate change.

A key limitation in the ability of breeding programmes to leverage these benefits of major-gene marker-assisted selection is the availability of those genes in an appropriate elite germplasm (Janaki Ramayya et al., 2021; Wu et al., 2014). Studies on current breeding programmes have shown that about half of the genes and QTLs that could be useful are not found in current elite material; a further 15% or so are present at very low frequency (Cobb, Juma, et al., 2019; Juma et al., 2021). Furthermore, these genes are mostly only available from poorly performing landraces or occasionally very old breeding material. This means breeders who use these major genes will face a major decrease in the performance of the resulting material (negative genetic gains for yield). Thus, breeders must choose between the short-term genetic gains for yield and the long-term potential for agility in improving a range of traits.

The standard approach to offset this trade-off is (marker-assisted) introgression of the target gene/QTL into elite genomic backgrounds. While this takes a dedicated effort, the result is clean material in an elite genomic background that can be used to introduce the new gene into mainstream breeding efforts. Key quality measures in this process include the size of introgression (eliminating linkage drag) and recipient recovery rates (RPRs, eliminating drag from the undesirable donor genome) (Hospital & Charcosset, 1997; Koudandé et al., 2000; Wu et al., 2014). However, this approach suffers from two major drawbacks: First, introgression is often a tedious and expensive process, especially if coming from a particularly poor-quality landrace or a related wild species. Thus, the initial introgression typically focuses on a single introgression population and will only produce one or a small number of converted lines with the new gene. Breeders cannot use these too extensively in their crossing programmes lest the overall genetic diversity of their programmes becomes depleted from the repeated use of the same parent. This means the penetration of the new gene into mainstream breeding will take multiple breeding cycles, which often last four years or more.

The second major drawback of current introgression processes is that they will always take several generations to eliminate the highly undesirable genome of the original donor landrace (Koudandé et al., 2000). This means even if they start with the best, most cutting-edge material from breeding programmes as a recipient, by the time introgression is complete, the performance of the current breeding cohorts has improved. As a result, the yield performance of the converted recipient is no longer within the range desired for present parents.

In this context, there are some studies using empirical data or theoretical approaches trying to improve QTL introgression, such as Peng et al. (2014), Cameron et al. (2017) and Frisch and Melchinger (2001). The latter presented a marker-assisted backcrossing approach for efficient introgression of a recessive gene into a desirable genetic background. The method involves using molecular markers to identify and track the target gene during the backcrossing process in a relatively short period of time, with minimal loss of favourable genetic traits.

Another and most recent study aiming to overcome these drawbacks, a two-stage process, has been proposed to introduce new genes into the mainstream breeding process (Cobb, Biswas, et al., 2019; Cobb, Juma, et al., 2019). Firstly, deployment (or conversion) creates a high-quality conversion of a modern elite background. Deployment is heavily focused on quality: ensuring no undesirable genomic contribution of the original landrace donor remains in the final product. However, as mentioned, this suffers from the above two drawbacks. To overcome this, the second stage, which is described as line augmentation, aims to rapidly introgress the new gene into a wide variety of elite genomic backgrounds as quickly as possible to enable rapid and wide-ranging impact in the breeding programme. Line augmentation is thus focused primarily on quantity and speed rather than quality; it cannot address issues of breaking linkage drag, having fertility problems or eliminating the undesirable genomic background. In addition, it requires high-quality donor material from the deployment process (or existing elite donors if available).

This concept of line augmentation as a separate and distinct activity appears to be fairly new, and many questions remain on optimizing the process to achieve maximum results (Cobb, Biswas, & Platten, 2019). Besides scale and speed, other quality control parameters for line augmentation could include recovery of the (new elite) recipient background. The aim is to minimize the influence of the first elite donor on reducing genetic diversity in the elite breeding programme and to have a recombinant selection around the target genes to embed these in additional elite haplotypes, thus avoiding selective sweeps (SSs). SS refers to the process by which a new advantageous mutation eliminates or reduces variation in linked neutral sites as it increases in frequency in the population (Nielsen et al., 2005). In this context, another option for a fast line augmentation process is the application of new methods, such as genomic prediction (Ødegård et al., 2009), in order to select parents with the target gene/QTL to be introgressed in the breeding population. This selection approach disregards the similarity between recombinant lines with the donors but prioritizes their genomic breeding values (BVs). For that, marker effects estimated based on the elite breeding population would be used. Therefore, it might reduce the well-known penalty in quantitative traits, such as grain yield (GY), caused by introgressions in elite populations (Dar et al., 2018).

Based on the above, the question is which measures or methods will ensure the most rapid utilization of target genes in the mainstream breeding programme. Unfortunately, comparing all these possibilities using empirical data would be impractical, costly and time-consuming. Also, it will be difficult to detangle the specific germplasm background effect on the conclusions, not allowing extrapolation of the results to more general applications or rules of thumb. Hence, the objective of this study was to use stochastic simulations to examine the effectiveness of three methods in promoting the uptake and fixation of new genes in the mainstream breeding programme.

2 | MATERIAL AND METHODS

Our study compared three strategies to develop new recipients for QTL introgression in a short-term breeding programme. For that, we used rice (*Oryza sativa* L.) as a model of self-pollinated crop and stochastic simulations performed by the *AlphaSimR* package (Gaynor et al., 2021). Furthermore, we evaluated two different numbers of recipients in the introgression process and how the population performance and the QTL fixation over breeding cycles were influenced.

2.1 | Historical population and genetic parameters

The historical rice founder population was simulated as 3000 unique diploid inbred individuals, with 12 chromosome pairs each, using a Markovian Coalescent Simulator (MaCS) (Chen et al., 2009). For that, 1644 biallelic segregating sites were considered, uniformly distributed across chromosomes and 360 segregating loci randomly sampled as quantitative trait nucleotides (QTNs) and 994 segregating loci as single-nucleotide polymorphism (SNP) (Arbelaez et al., 2019). The genome size (cM) and chromosome sizes follow those values described by Li et al. (2008).

The target of the simulation was a quantitative trait (such as GY) and three qualitative traits (e.g., QTL controlling biotic stress resistance or abiotic stress tolerance). For the former, the genetic parameters obtained by Li et al. (2008) were used. Each QTN received randomly allocated additive and dominance effects. Genetic values for each genotype were obtained by summing all additive and dominance effects for all QTNs. Additive effects (a) were sampled from a gamma distribution with scale and shape parameters equal to 1 and randomly assigned for each QTN. Similarly, dominance effects (d) for each QTN were computed by multiplying the absolute value of its additive effect (a_i) by locus-specific dominance degree (δ_i). Dominance degrees were sampled of a Gaussian distribution with $\delta_i \sim N(\mu_\delta, \sigma^2_\delta)$, where μ_δ is the average dominance degree equal to 0.22 and σ^2_δ is the dominance variance equal to 0.50. Then, dominance effects were assigned for each QTN according to the equation below:

$$d_i = \{0, \text{if QTN is homozygous } \delta_i \times |a_i|, \text{if QTN is heterozygous}$$

Phenotypic values for the quantitative trait were obtained by adding a random error sampled of a Gaussian distribution with mean equal to 0 and variance (σ^2_e) equal to 1, which was defined by broad-sense ($H^2 = 0.53$) and narrow ($h^2 = 0.50$) heritabilities.

We consider three independent characteristics regarding the qualitative traits, each controlled exclusively by one additive QTL, with heritability equal to 1.0. Also, we assume that all the traits were independent regarding segregation and genetic correlations (Koudandé et al., 2000). No mutations were added over breeding cycles.

2.2 | Base population and burn-in phase

In order to obtain the base populations, donors and recipients, we selected two sets of 60 individuals from 3000 lines of the historical population based on their superior phenotypic values for the quantitative trait (Figure 1), before the burn-in. The first population was selected without any favourable allele for the three QTLs and the second with two copies for each favourable QTL. As a starting point (C0: breeding cycle zero; used as the reference for the downstream analysis) to consider a programme representative of current 4-year rice breeding programmes, we ran three traditional recurrent selection cycles, totalling 12 years of breeding in the burn-in stage. A breeding cycle was as follows: first, these 60 parental lines were crossed to generate 30 F_1 plants, which were selfed to produce 230 F_2 plants from each cross (Cobb, Juma, et al., 2019). Then, single-seed descent (SSD) was conducted during the line fixation stage, from the F_2 to the F_6 generations, where the best individuals were selected based on their phenotypic values to find the next breeding cycle. Finally, after three recurrent breeding cycles, we obtained the two base populations to evaluate the augmentation/introgression schemes.

Three generations are enough to remove extreme genotypes from the population and stabilize the allele frequency and performance changes with selection. Also, from a practical perspective, three breeding cycles represent about a third of a breeder's professional life.

We selected only the best line from the donor population (QTL = 6), representing a donor with an elite genetic background. Hence, from here, the donor is an elite or set of lines from the base populations that has the target QTL. On the other hand, from the recipient population, we considered two scenarios, 10 or 20 superior lines, to develop the newest recipients in the augmentation process. Before we proceeded to further analysis, we checked the similarities between base populations regarding the allele frequencies and population performance (genetic mean) to guarantee that the only significant difference between them was the QTLs.

2.3 | Line augmentation

Line augmentation aims to diversify the range of elite genomic backgrounds containing new QTLs as rapidly as possible to

introduce this as parental material in the wider breeding programme. The starting point is an elite donor line in which effects such as linkage drag and elimination of highly unfavourable genomic backgrounds have already been eliminated (Cobb, Biswas, & Platten, 2019; Platten et al., 2019). To minimize the time required for delivery of products, introgression schemes focused on a single backcross with recipient lines, with BC_1F_3 fixed lines used as the material to introduce to breeding programmes. During the augmentation framework (Figure 2), in addition to the requirement of possessing the three target QTLs, we compared three possible methods to select favourable progeny and, ultimately, the upgraded elite lines that can be used as donor parents in further ongoing breeding efforts (Figure 1).

Background recovery (BG): This method is also known as the recipient background (RPRR); the introgression workflows typically aim to maximize recovery of the elite recipient background. Thus, individuals were selected based on their genetic similarity with the original recipient parent under this scheme. Therefore, we select individuals that have the target QTLs and are more similar to the recipient.

SS: Individuals were selected to reduce SS associated with the introgressed genes, such as shrinking the genomic region surrounding the target QTL. For that, we used the four nearest SNPs near the QTL to opportunistically inform recombinant selection on either side of the target locus. Therefore, we selected individuals that have the target QTLs and the shorter region surrounding them; in other words, we minimized the segment to be introgressed.

BV: This method selects individuals that have the desired QTL profile and then the best BV, regardless of the similarities to the recipient. For that, we considered the same training set (TS) used for the breeding population in the mainstream selection pipeline, which was composed of 1536 inbred lines originated by 30 crosses, between

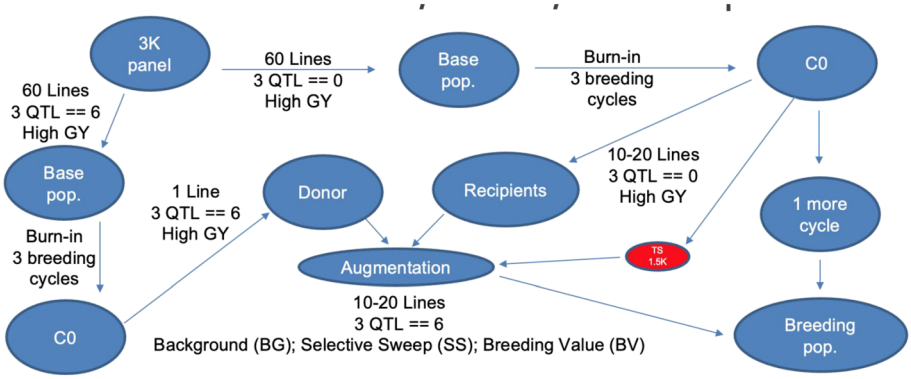

FIGURE 1 Base population (pop.) development, donors and recipients, and the methods used for augmentation. Base pop. refers to the population used before the burn-in, and C0 is the first breeding population (used as reference) after that. 3K panel, rice germplasm panel with 3000 individuals; C0, breeding cycle zero; GY, grain yield; QTL, quantitative trait loci; TS 1.5K, genomic prediction training set of 1500 individuals [Color figure can be viewed at wileyonlinelibrary.com]

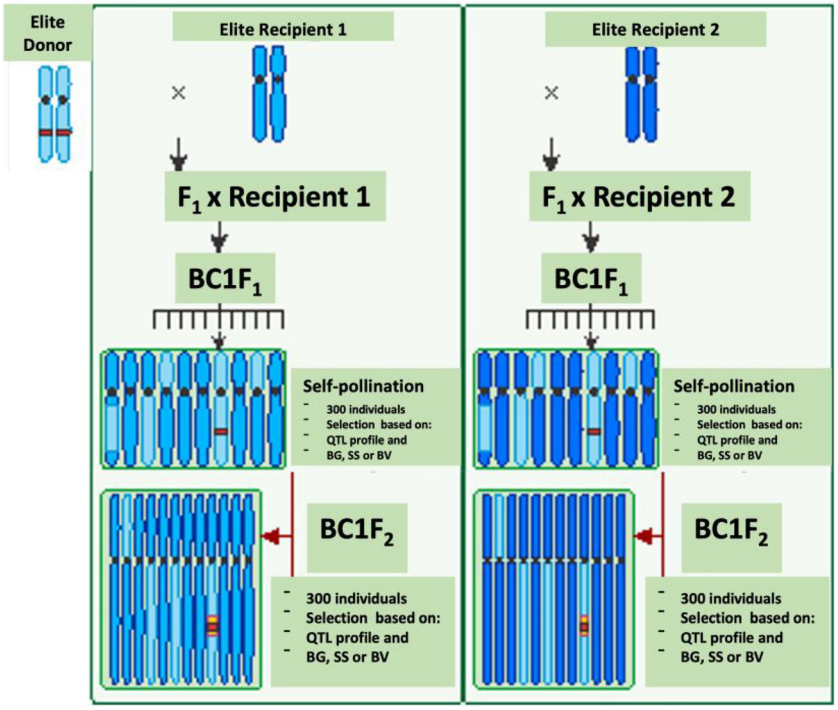

FIGURE 2 Scheme used for line augmentation across a different number of elite recipients (Cobb, Biswas, & Platten, 2019; Platten et al., 2019). In this scheme, there are two different lines (recipients) that will receive the QTL from a common donor at the same time. BG, background recovery; BV, breeding value; SS, selective sweep [Color figure can be viewed at wileyonlinelibrary.com]

60 individuals (parents), with nearly 52 plants per cross, from the base population after the burn-in stage. Marker effects were predicted using the ridge-regression best linear unbiased prediction (RRBLUP) (Endelman, 2011) according to the equation below:

$$y = 1\mu + Z_u u + \varepsilon$$

where y is the vector of individual phenotypic values from the TS; μ is the mean (intercept); u is the vector of marker effects, where $u \sim N(0, I\sigma_u^2)$; ε is the vector of random residuals; 1 is the vector of ones; and Z_u is the incidence matrix of TS genotypes for m markers. Z_u is coded as 1 for homozygous A_1A_1, -1 for homozygous A_2A_2 and 0 for heterozygous A_1A_2.

To perform the genomic selection (GS), the genomic estimated breeding value (GEBV) was estimated using the following equation: $GEBV = Mu$, where M is the incidence matrix of selection candidate genotypes and u is the vector of predicted marker effects. The GEBV was calculated for BC_1F_1 and derived BC_1F_2 material that had been preselected for the target QTLs as above. In this context, we considered the 1K-Rica SNP panel to perform the genomic predictions (Arbelaez et al., 2019).

2.4 | QTL introgression and its effects on a breeding population

After the development of the newest elite donor lines, these can be used in the mainstream breeding programme to introduce the target QTLs to the wider, mainstream breeding effort. For that, as a representative self-pollinated crop, simulations were based on the rice breeding programme structure from the International Rice Research Institute (IRRI) (Collard et al., 2019). For all scenarios, the line fixation phase was conducted by the SSD method, which collects one seed from each segregating plant to advance to the next stage until it reaches a high homozygosity level. For that, the crossing block was composed of 60 parents every breeding cycle, composing 30 crosses, in a way that each parent participated in just one cross. The parents for the next cycle were always selected based on the individual performance for the quantitative traits, non-based on the QTL profile.

The goal is to make the introgression every breeding cycle a certain number of donors as part of the crossing block. Hence, six schemes were compared: three methods and two numbers of recipients (Figure 1). Also, two methods (traditional and drift) were used as baselines. Drift is just a baseline scenario using the same framework as the traditional one, where there is no parent selection, just a random sample to compose the next generation. Therefore, we can estimate the effect of sample size on the genetic variability (drift).

In the breeding crossing block, we removed the 10 to 20 worst parents for the quantitative trait and then included the elite augmentation products as new parents into the crossing block. It is important to highlight that the augmentation products are one cycle back regarding genetic improvement due to the time spent developing them (Figure 1). Therefore, to evaluate the effect on the population, we monitor two parameters over short-term breeding: the quantitative trait's performance and the QTL frequency in the breeding population. We consider that the best control is a population without a QTL introgression because we simulate independent QTLs, in other words, without any direct genetic effect on the quantitative trait. Therefore, we understand that a breeding population following a normal pipeline versus the population with introgression provides us the 'penalty' of the residual background in the QTL introgression.

Each strategy was simulated over five breeding cycles and replicated 100 times considering the current rice breeding framework used in IRRI, which is called rapid generation advancement (RGA) (Cobb, Juma, et al., 2019; Collard et al., 2017, 2019), in other others, a variation of the SSD.

3 | RESULTS

In this study, we used stochastic simulations to examine the effectiveness of three methods in promoting the introgression and fixation of new QTLs in the mainstream breeding programme, BG, SS and genomic selection. Furthermore, we simulate the current scenario in IRRI, where the donor and recipients are elite lines (improved by breeding), less usually a landrace than the donor. For that, we used a different approach regarding the number of populations 'running in parallel' in breeding (Figure 1). This method provided elite populations for the quantitative trait with similar performances, differing only by the presence of the target QTLs.

Simulation of breeding programmes clearly showed a substantial penalty associated with introducing new material, even when this material was only one breeding cycle less advanced than the current parents (Figures 3a and 4a). There was little difference between the different selection strategies. Therefore, maximizing BG, SS or BV will produce almost identical penalties that persist over several breeding cycles, though maximizing BG was arguably slightly worse in later cycles. A difference was observed when using either 10 or 20 new parents derived from the augmentation process. Introducing 20 new parents consistently performed worse than introducing only 10 new parents. This is consistent with the size of the penalty being directly proportional to the extent of 'older' genetic material being introduced to the crossing programme. More variability in the extent of penalty was seen when introducing 10 new parents, with maximizing BG (BG_10) being substantially worse than other strategies, on a parent with SS_20. In contrast, the BV_10 and SS_10 strategies had almost half the penalty in population mean performance compared with other strategies.

In contrast to the situation with the mean performance of the breeding programme, the time required to bring new QTLs to fixation in the breeding programme was substantially faster when introducing 20 new parents rather than 10 (Figures 3b and 4b). However, substantial differences were also observed between the different

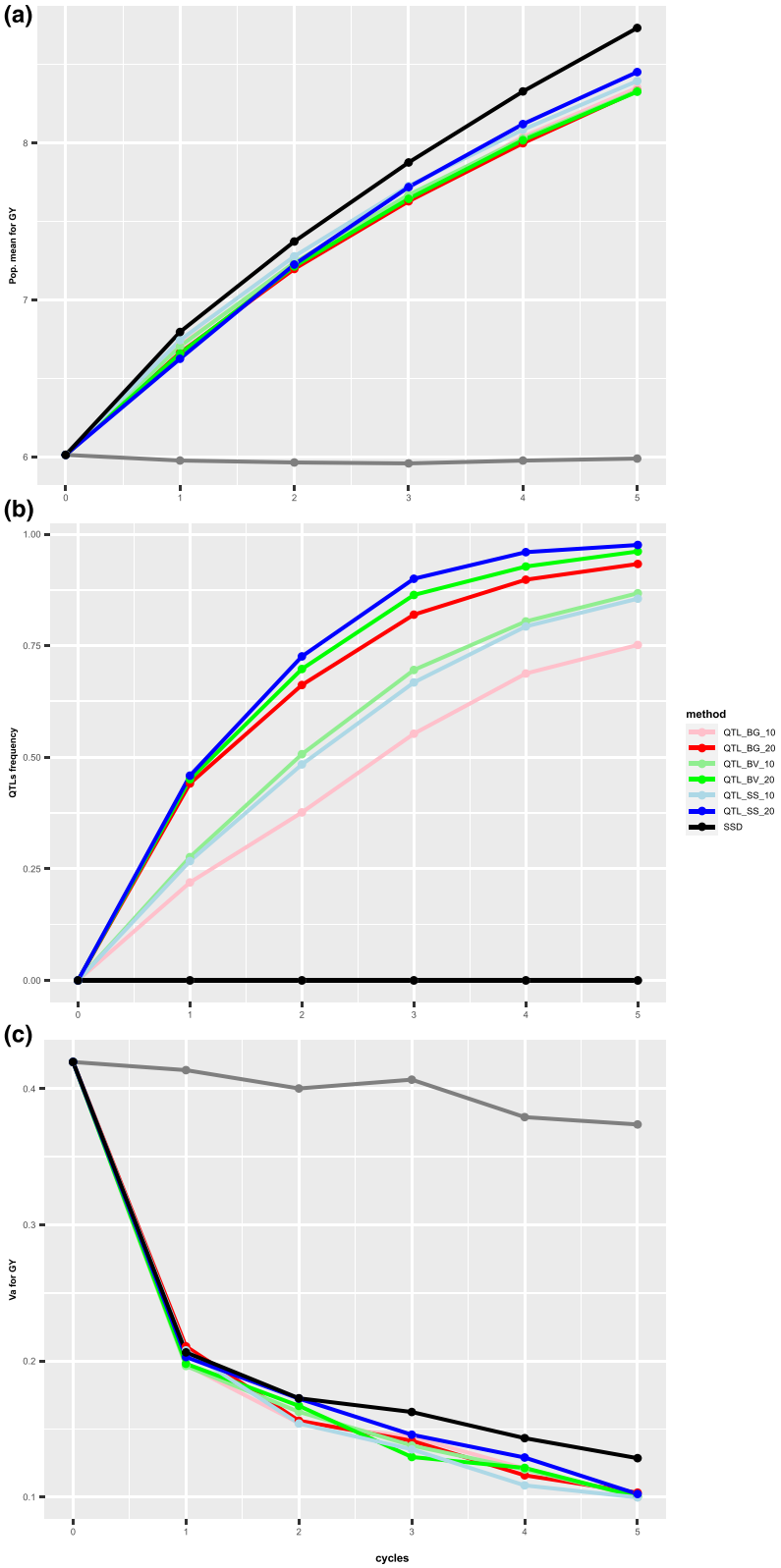

FIGURE 3 Population mean performance for yield (a), quantitative trait loci frequency (b) and genetic variability (c) over five breeding cycles. Each coloured line represents a selection method (background recovery [BG]; selective sweep [SS]; breeding value [BV]) and the number of recipients (10 or 20). GY, grain yield; SSD, single-seed descent [Color figure can be viewed at wileyonlinelibrary.com]

augmentation selection strategies. For both 10 and 20 new parents introduced, maximizing the BV of augmentation selections resulted in faster fixation of the target QTLs. At the other extreme, maximizing recovery of the recipient background produced the slowest fixation of target QTLs when using 20 parents, while minimizing SS performed the poorest when introducing 10 new parents. As a result, introducing 10 new parents always took longer than introducing 20. Still, the BV_10 strategy did not take much longer than the BG_20 strategy and achieved a frequency of approximately 70% after only three breeding cycles.

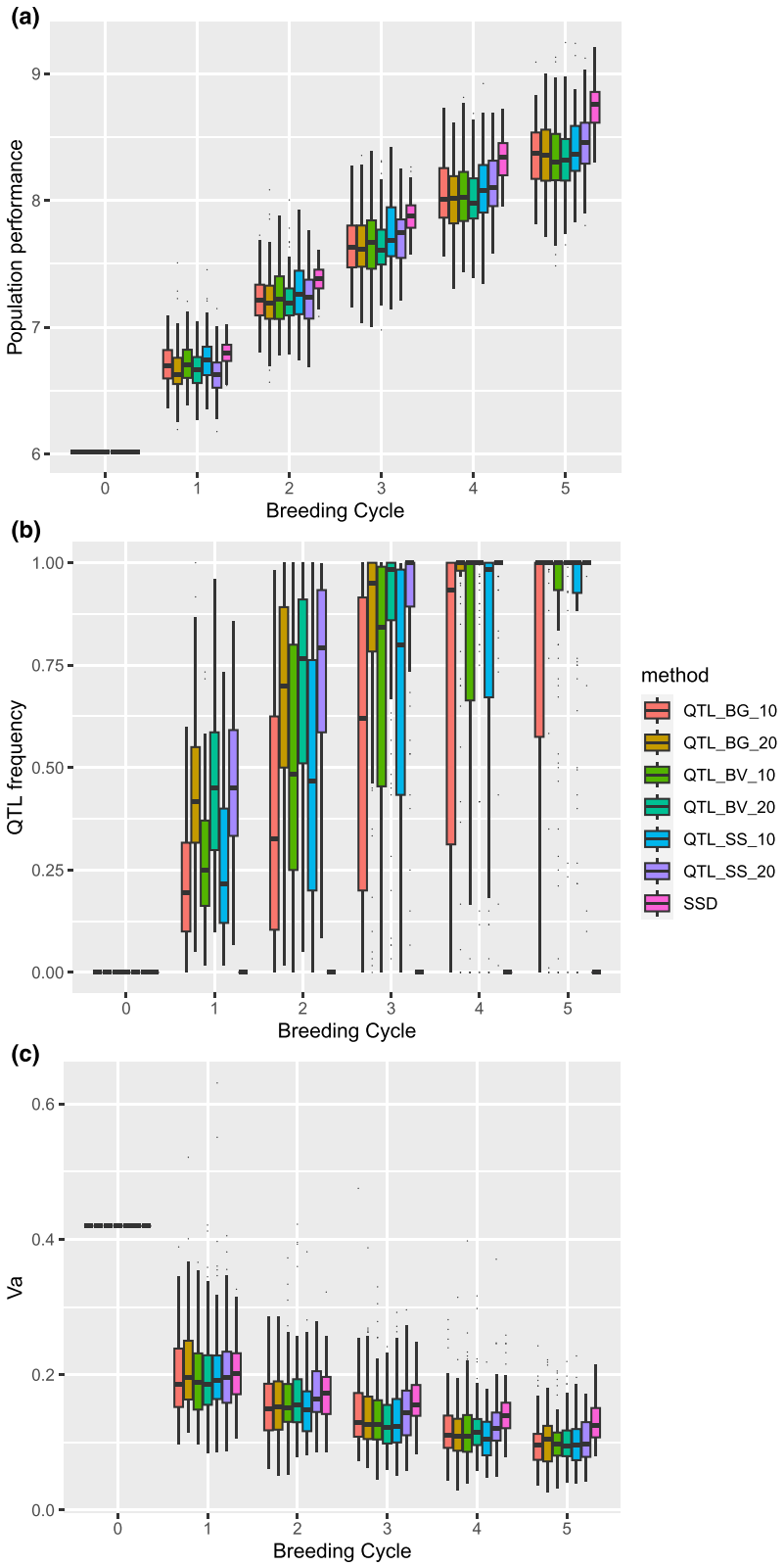

FIGURE 4 Population mean performance for yield (a), quantitative trait loci frequency (b) and genetic variability (c) over five breeding cycles. Each boxplot represents 100 replicates for the combination of the selection method (background recovery [BG]; selective sweep [SS]; breeding value [BV]) and the number of recipients (10 or 20). GY, grain yield; SSD, single-seed descent [Color figure can be viewed at wileyonlinelibrary.com]

Finally, concerning genetic variability, there is no clear difference among the tested scenarios (Figures 3b and 4b). This is mainly because we considered an elite background as the donor and not a landrace as most of the studies and real cases have used. Therefore, with the augmentation method (Cobb, Juma, et al., 2019; Cobb, Biswas, et al. 2019), it is possible to introduce new genes into the mainstream breeding process having minimal penalty in terms of performance and also without bringing 'undesirable' genetic variability to the breeding population.

4 | DISCUSSION

Modern breeding strategies focus on rapid-cycle recurrent selection achieved through intercrossing the best elite lines identified in each breeding cycle. A drawback often highlighted in this approach is that it produces (indeed relies on) a breeding programme that is effectively closed; it is difficult to introduce new genetic variation from external sources, as these typically do not have equivalent performance with current elite cohorts (Juma et al., 2021). This becomes problematic when the breeding programme needs key variations for major QTLs, such as major disease resistance or abiotic stress tolerance genes (Cobb, Biswas, & Platten, 2019). These often stem from landraces or other highly un-adapted genomic backgrounds, so introducing these to the crossing programme introduces substantial penalties in performance for other traits such as yield. The standard strategy for introducing these into the breeding programme involves the backcrossing of the target gene(s) into one or more elite backgrounds, maximizing the recovery of the elite genomic background to avoid penalties from the unfavourable donor genome and in some cases also recombinant selection around the target gene to minimize the probability of linkage drag. This one-stage process suffers a trade-off. However, it is not easy to simultaneously achieve both high-quality introgressions across a range of elite genomic backgrounds. Hence, in practice, programmes often focus on producing a single, high-quality introgression in one background or a modest number of lower quality introgressions in multiple backgrounds. To overcome this trade-off, a two-stage introgression process has been developed (Cobb, Juma, et al., 2019), whereby the initial deployment of a gene focuses on producing one high-quality elite introgression donor. This deployment product is then used as the donor in the line augmentation, rapidly introducing the new gene into various elite backgrounds. Genomic penalties and linkage drag having already been minimized, the focus of augmentation is speed and quantity of introgressions rather than quality.

There are some studies proposing QTL optimization, such as Peng et al. (2014), Cameron et al. (2017) and Frisch and Melchinger (2001), but none of them touch the topic of the donor as an elite material, different number of recipients, or even the short- to medium-term effects on a breeding population in terms of QTL frequency, population performance and genetic variability.

This study aimed to examine the optimum strategy for selecting segregants during the augmentation pipeline. The typical process of maximizing the recovery of the recipient genomic BG was contrasted with strategies to minimize SS with recombinant selection (SS) and a new approach to maximize the BV of segregants. In the last process, segregants are selected based on the most favourable genomic composition as judged by genomic predictions (Sonesson et al., 2012); priority is given to segregants displaying the highest BV irrespective of which parent (elite donor or elite recipient) contributed to any given portion of the genome.

Simulation results clearly showed that irrespective of selection strategy, substantial penalties in performance for yield were associated with introducing introgression products as new parents in the breeding programme (Figure 3). This is presumably due to introgression products being at least one breeding cycle less advanced than the most recent breeding cohort. In general, only modest differences were observed between the various selection strategies. However, there was clearly less penalty associated with introducing *fewer* new augmentation products as parents. In particular, introducing 10 augmentation products as parents using either the SS or BV selection method introduced half the penalty of introducing 20 parents or 10 parents using the traditional BG selection method. This penalty associated with introducing introgression products is almost inherent in any introgression procedure; introgression takes time, so even if the recipients represent the absolute best performers in the current cohort at the outset, the final introgression products will always be a step behind the most advanced material by the time they are finished (Hospital & Charcosset, 1997; Koudandé et al., 2000). Therefore, it is perhaps counter-intuitive that BV selection did not outperform other methods in reducing the penalty associated with introducing augmentation products. This highlights the need to minimize the time taken for introgression (augmentation in particular), thus minimizing the divergence between augmentation products and current breeding cohorts.

In contrast to the situation seen for yield performance, the time required to achieve fixation (or near fixation) of the target QTLs was minimized when introducing more parents to the target QTLs. As might be expected, introducing 20 parents with the new QTLs always resulted in faster fixation in the breeding programme than introducing only 10 new parents. For this parameter, the different augmentation selection strategies showed substantial performance differences. Of particular interest, selection based on BV always outperformed the SS and BG methods. Minimizing SS outperformed background selection when introducing 20 parents, though, with 10 parents, the opposite effect was seen. This suggests that the improved performance of augmentation segregants selected based on BV (relative to chosen segregants based on SS or BG, though not segregants of the current breeding cohort) increases their chances of being selected as parents in subsequent breeding cycles, thus increasing the QTL frequency faster than augmentation products selected based on SS or BG.

The superior outcomes when selecting via BV may be due to using marker effects estimated based on the elite breeding population. It favours those individuals genetically more related to the current population and those with the best haplotype combinations (Won et al., 2020). In other words, it reduces the penalty in GY because we can select the best BV (almost the same linkage disequilibrium [LD] and linkage phase) and accelerate the gene/QTL fixation. After all, the target ones are present and in favourable haplotypes. Furthermore, it shifts the paradigm that the donor genome is inherently and always unfavourable. In the augmentation pipeline, the donor line already has a high BV, if not as high as the most advanced elite lines. Thus, some portions of the donor genome would be expected to be favourable, even compared with the recipient genome, and crucially are also represented in the current elite breeding pool. By selecting based on overall BV, the most favourable fragments/

haplotypes are selected irrespective of whether they are contributed by the donor or recipient parents. Hence, the progeny could, in principle, outperform both parents. If this were true, it is not fundamentally opposed to the usual paradigm of maximizing recipient–parent recovery rates; in typical marker-assisted backcross (MABC) procedures, the donor parent is highly unfavourable, so in most cases, the donor parent haplotype in any given interval would decrease BV. In turn, since the BV of progeny is higher, it improves the chances that these will be selected as parents for future breeding cycles, thus speeding the fixation of target genes. Finally, stochastic simulations may help select the most important factors to be adjusted. Furthermore, it is a fast and inexpensive approach to testing various scenarios and factors (Faux et al., 2016).

In this study, we simulated the introgression of three QTLs at the same time in order to represent our reality in terms of constraints, mainly because the relationship between the number of QTLs to be introgressed and population sizes is a critical aspect. In smaller populations, introgressing multiple QTLs can be challenging due to limited genetic diversity and the potential for inbreeding. As the number of QTLs to be introgressed increases, larger populations are often necessary to maintain genetic diversity, reduce the risk of inbreeding and increase the chances of obtaining favourable recombinations. Additionally, larger populations may provide more opportunities for recombination, leading to faster progress in the introgression process. Therefore, careful consideration of population size is crucial when planning introgression programmes involving multiple QTLs, as it can significantly impact the efficiency and success of the breeding efforts.

5 | CONCLUSION

Selection strategies during line augmentation can produce substantially different outcomes as products are introduced into the breeding programme as a source of new QTLs. Substantial work has been done on optimizing many aspects of the breeding process, but this is the first attempt to quantify optimal strategies for introducing new genes and QTLs to the mainstream breeding process. In particular, introducing introgression products resulted in substantial penalties in yield performance, largely regardless of the selection strategy used. However, introducing fewer new parents to the augmentation process minimized this effect. In addition, the time required to achieve fixation of target QTLs showed substantial differences, with selection for BV during augmentation outperforming other methods. Overall, the BV_10 strategy (10 parents selected based on genomic estimated BV) displayed the best trade-off between reduced penalty from introducing new QTLs with the reasonable speed at which those QTLs can achieve fixation over subsequent breeding cycles.

AUTHOR CONTRIBUTIONS

John Damien Platten and Roberto Fritsche-Neto contributed equally to developing the hypothesis, analysing and interpreting the results, and writing. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We would like to thank Dr. Gary Atlin (Bill and Melinda Gates Foundation), Dr. Amelia Henry, the IRRI Breeders and Excellence in Breeding-CGIAR platform for the comments and suggestions.

CONFLICT OF INTEREST STATEMENT

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

DATA AVAILABILITY STATEMENT

The datasets and scripts used for this study can be found in the supporting information.

ORCID

Roberto Fritsche-Neto https://orcid.org/0000-0003-4310-0047

REFERENCES

Arbelaez, J. D., Dwiyanti, M. S., Tandayu, E., Llantada, K., Jarana, A., Ignacio, J. C., Platten, J. D., Cobb, J., Rutkoski, J. E., Thomson, M. J., & Kretzschmar, T. (2019). 1k-RiCA (1K-Rice Custom Amplicon) a novel genotyping amplicon-based SNP assay for genetics and breeding applications in rice. *Rice*, *12*, 1–15.

Cameron, J. N., Han, Y., Wang, L., & Beavis, W. D. (2017). Systematic design for trait introgression projects. *Theoretical and Applied Genetics*, *130*, 1993–2004. https://doi.org/10.1007/s00122-017-2938-9

Chen, G. K., Marjoram, P., & Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. *Genome Research*, *19*, 136–142. https://doi.org/10.1101/gr.083634.108

Cobb, J. N., Biswas, P. S., & Platten, J. D. (2019). Back to the future: Revisiting MAS as a tool for modern plant breeding. *Theoretical and Applied Genetics*, *132*, 647–667. https://doi.org/10.1007/s00122-018-3266-4

Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M., & Ng, E. H. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: Lessons from the breeder's equation. *Theoretical and Applied Genetics*, *132*, 627–645. https://doi.org/10.1007/s00122-019-03317-0

Collard, B. C. Y., Beredo, J. C., Lenaerts, B., Mendoza, R., Santelices, R., Lopena, V., Verdeprado, H., Raghavan, C., Gregorio, G. B., Vial, L., Demont, M., Biswas, P. S., Iftekharuddaula, K. M., Rahman, M. A., Cobb, J. N., & Islam, M. R. (2017). Revisiting rice breeding methods—Evaluating the use of rapid generation advance (RGA) for routine rice breeding. *Plant Production Science*, *20*, 337–352. https://doi.org/10.1080/1343943X.2017.1391705

Collard, B. C. Y., Gregorio, G. B., Thomson, M. J., Islam, R., Vergara, G. V., Laborte, A. G., Nissila, E., Kretzschmar, T., & Cobb, J. N. (2019). Transforming rice breeding: Re-designing the irrigated breeding pipeline at the International Rice Research Institute (IRRI). *Crop Breeding, Genetics and Genomics*, *1*, e190008. https://doi.org/10.20900/cbgg20190008

Dar, M. H., Zaidi, N. W., Waza, S. A., Verulkar, S. B., Ahmed, T., Singh, P. K., Roy, S. K. B., Chaudhary, B., Yadav, R., Islam, M. M., Iftekharuddaula, K. M., Roy, J. K., Kathiresan, R. M., Singh, B. N., Singh, U. S., & Ismail, A. M. (2018). No yield penalty under favorable conditions paving the way for successful adoption of flood tolerant rice. *Scientific Reports*, *8*, 9245. https://doi.org/10.1038/s41598-018-27648-y

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, *4*, 250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Faux, A., Gorjanc, G., Gaynor, R. C., Battagin, M., Edwards, S. M., Wilson, D. L., Hearne, S. J., Gonen, S., & Hickey, J. M. (2016). AlphaSim: Software for breeding program simulation. *The Plant Genome*, *9*(3), 1–14. https://doi.org/10.3835/plantgenome2016.02.0013

Frisch, M., & Melchinger, A. E. (2001). Marker-assisted backcrossing for introgression of a recessive gene. *Crop Science*, *41*, 1485–1494. https://doi.org/10.2135/cropsci2001.4151485x

Gaynor, R. C., Gorjanc, G., & Hickey, J. M. (2021). AlphaSimR: An R package for breeding program simulations. *G3 Genes|Genomes|Genetics*, *11*(2), 1–5. https://doi.org/10.1093/g3journal/jkaa017

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

Janaki Ramayya, P., Vinukonda, V. P., Singh, U. M., Alam, S., Venkateshwarlu, C., Vipparla, A. K., Dixit, S., Yadav, S., Abbai, R., Badri, J., T., R., Phani Padmakumari, A., Singh, V. K., & Kumar, A. (2021). Marker-assisted forward and backcross breeding for improvement of elite Indian rice variety Naveen for multiple biotic and abiotic stress tolerance. *PLoS ONE*, *16*, e0256721. https://doi.org/10.1371/journal.pone.0256721

Juma, R. U., Bartholomé, J., Thathapalli Prakash, P., Hussain, W., Platten, J. D., Lopena, V., Verdeprado, H., Murori, R., Ndayiragije, A., Katiyar, S. K., & Islam, M. R. (2021). Identification of an elite core panel as a key breeding resource to accelerate the rate of genetic improvement for irrigated rice. *Rice*, *14*, 92. https://doi.org/10.1186/s12284-021-00533-5

Khanna, A., Anumalla, M., Catolos, M., Bartholomé, J., Fritsche-Neto, R., Platten, J. D., Pisano, D. J., Gulles, A., Sta. Cruz, M. T., Ramos, J., Faustino, G., Bhosale, S., & Hussain, W. (2022). Genetic trends estimation in IRRIs rice drought breeding program and identification of high yielding drought-tolerant lines. *Rice*, *15*, 14. https://doi.org/10.1186/s12284-022-00559-3

Koudandé, O. D., Iraqi, F., Thomson, P. C., Teale, A. J., & van Arendonk, J. A. M. (2000). Strategies to optimize marker-assisted introgression of multiple unlinked QTL. *Mammalian Genome*, *11*, 145–150. https://doi.org/10.1007/s003350010028

Kumar, A., Dixit, S., Ram, T., Yadaw, R. B., Mishra, K. K., & Mandal, N. P. (2014). Breeding high-yielding drought-tolerant rice: Genetic variations and conventional and molecular approaches. *Journal of Experimental Botany*, *65*, 6265–6278. https://doi.org/10.1093/jxb/eru363

Ladha, J. K., Radanielson, A. M., Rutkoski, J. E., Buresh, R. J., Dobermann, A., Angeles, O., Pabuayon, I. L. B., Santos-Medellín, C., Fritsche-Neto, R., Chivenge, P., & Kohli, A. (2021). Steady agronomic and genetic interventions are essential for sustaining productivity in intensive rice cropping. *Proceedings of the National Academy of Sciences*, *118*(45), e2110807118. https://doi.org/10.1073/pnas.2110807118

Li, L., Lu, K., Chen, Z., Mu, T., Hu, Z., & Li, X. (2008). Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. *Genetics*, *180*, 1725–1742. https://doi.org/10.1534/genetics.108.091942

Nayak, S., Habib, M. A., Das, K., Islam, S., Hossain, S. M., Karmakar, B., Fritsche Neto, R., Bhosale, S., Bhardwaj, H., Singh, S., & Islam, M. R. (2022). Adoption trend of climate-resilient rice varieties in Bangladesh. *Sustainability*, *14*, 5156. https://doi.org/10.3390/su14095156

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research*, *15*, 1566–1575. https://doi.org/10.1101/gr.4252305

Ødegård, J., Sonesson, A. K., Yazdi, M. H., & Meuwissen, T. H. (2009). Introgression of a major QTL from an inferior into a superior population using genomic selection. *Genetics Selection Evolution*, *41*, 38. https://doi.org/10.1186/1297-9686-41-38

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

Platten, J. D., Cobb, J. N., & Zantua, R. E. (2019). Criteria for evaluating molecular markers: Comprehensive quality metrics to improve marker-assisted selection. *PloS One*, *14*, e0210529.

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS ONE*, *8*, e66428. https://doi.org/10.1371/journal.pone.0066428

Sonesson, A. K., Woolliams, J. A., & Meuwissen, T. H. E. (2012). Genomic selection requires genomic control of inbreeding. *Genetics Evolution*, *44*, 1–10.

Ting, D. S.-K., & Vasel-Be-Hagh, A. (Eds.). (2022). *Mitigating climate change*. Springer International Publishing.

Won, S., Park, J.-E., Son, J.-H., Lee, S. H., Park, B. H., Park, M., Park, W. C., Chai, H. H., Kim, H., Lee, J., & Lim, D. (2020). Genomic prediction accuracy using haplotypes defined by size and hierarchical clustering based on linkage disequilibrium. *Frontiers in Genetics*, *11*, 134. https://doi.org/10.3389/fgene.2020.00134

Wu, X., Li, Y., Shi, Y., Song, Y., Wang, T., Huang, Y. & Li, Y. (2014). Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. *Theoretical and Applied Genetics*, *127*, 621–631. https://doi.org/10.1007/s00122-013-2246-y

SUPPORTING INFORMATION

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

How to cite this article: Platten, J. D., & Fritsche-Neto, R. (2023). Optimizing quantitative trait loci introgression in elite rice germplasms: Comparing methods and population sizes to develop new recipients via stochastic simulations. *Plant Breeding*, *142*(4), 439–448. https://doi.org/10.1111/pbr.13118