A multi-part strategy for introgression of exotic germplasm into elite plant breeding programs using genomic selection

A multi-part strategy for introgression of exotic germplasm into elite plant breeding programs using genomic selection

Irene S. Breider
The University of Edinburgh The Roslin Institute https://orcid.org/0000-0002-5428-3734
R. Chris Gaynor
The University of Edinburgh The Roslin Institute
Gregor Gorjanc (gregor.gorjanc@roslin.ed.ac.uk)
The University of Edinburgh The Roslin Institute https://orcid.org/0000-0001-8008-2787
Steve Thorn
The University of Edinburgh The Roslin Institute
Manish K. Pandey
ICRISAT: International Crops Research Institute for the Semi-Arid Tropics
Rajeev K. Varshney
Murdoch University
John M. Hickey
The University of Edinburgh The Roslin Institute

Research Article

Keywords:

Posted Date: January 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1246254/v1

Abstract

Some of the most economically important traits in plant breeding show highly polygenic inheritance. Genetic variation is a key determinant of the rates of genetic improvement in selective breeding programs. Rapid progress in genetic improvement comes at the cost of a rapid loss of genetic variation. Germplasm available through expired Plant Variety Protection (exPVP) lines is a potential resource of variation previously lost in elite breeding programs. Introgression for polygenic traits is challenging, as many genes have a small effect on the trait of interest. Here we propose a way to overcome these challenges with a multi-part pre-breeding program that has feedback pathways to optimise recurrent genomic selection. The multi-part breeding program consists of three components, namely a bridging component, population improvement, and product development. Parameters influencing the multi-part program were optimised with the use of a grid search. Haploblock effect and origin were investigated. Results showed that the introgression of exPVP germplasm using an optimised multi-part breeding strategy resulted in 1.53 times higher genetic gain compared to a two-part breeding program. Higher gain was achieved through reducing the performance gap between exPVP and elite germplasm and breaking down linkage drag. Both first and subsequent introgression events showed to be successful. In conclusion, the multi-part breeding strategy has a potential to improve long-term genetic gain for polygenic traits and therefore, potential to contribute to global food security.

Key Message

Introgression of exotic germplasm using a multi-part breeding strategy increases long-term improvement of polygenic traits. This is achieved by creating a pre-breeding strategy with feedback pathways to optimise recurrent genomic selection.

Introduction

Some of the most economically important traits in plant breeding, such as grain yield, show highly polygenic inheritance. Selective breeding is widely used to improve grain yield in crops. Gaynor et al. (2017) developed a two-part breeding strategy to accelerate the genetic improvement that can be achieved with genomic selection (GS). This strategy splits the traditional breeding program into two components, one component for development of new commercial inbred lines and another component for the improvement of the parent population. This resulted in an optimal strategy for accelerating genetic improvement with the use of GS. However, rapid progress in genetic improvement comes at the cost of a more rapid loss of genetic variation. Gorjanc et al. (2018) showed that methods like optimal cross selection can improve the efficiency of converting genetic variation into genetic gain in the two-part program, but ultimately some genetic variation is lost.

Genetic variation is a key determinant of the rates of genetic improvement in selective breeding programs. The amount of genetic variation that is present in a breeding program is affected by the amount of genetic variation across individuals that founded the population, by the amount of new

variation that has arisen in the population due to mutations or has been introduced to the population via introgression, and by the way in which this variation has been preserved or lost from the population due to selection and drift over time.

Mikel and Dudley (2006) reported that most of the commercial maize germplasm available in the early 2000s originated from just seven ancestor lines. Due to this mostly shared background, genetic diversity between these elite breeding programs is limited (Mikel and Dudley, 2006). In the past, many of these breeding programs became closed systems (Mikel and Dudley, 2006). This combined with the effects of drift, where the gene frequencies in the population change due to random sampling of the parental gametes (Falconer and Mackay, 1996), and changes in allele frequencies amplified by small effective population sizes, contributed to a narrow genetic base for individual elite maize breeding programs.

In 1970 the Plant Variety Protection (PVP) act came into effect and in 1985 the patenting of maize inbred lines became possible. This was the onset of large-scale protection of commercial maize germplasm and caused a constraint on the germplasm available in the public domain (Mikel and Dudley, 2006). This reduction in genetic diversity in elite breeding programs combined with the reduction of germplasm available in the public domain due to the PVP act, causes a risk of sub-optimal convergence of the level of genetic gain (e.g., Vanavermaete et al, 2020). This situation can put a constraint on the genetic improvement of crop yield (amongst other traits), leading to a stagnation in profit.

Dated elite germplasm is becoming available each year through exPVP lines. This could be a useful resource for variation previously lost in elite breeding programs (Mikel and Dudley 2006). Utilizing this variation could give opportunities for crop improvement and hence increase the economic performance of these crops; even for crops which have been under intense selection in the past (Boggini et al. 1997). However, exPVP lines are 20 years behind in cycles of recombination and selection (Mikel and Dudley 2006). Introduction or reintroduction of alleles is generally challenging due to genetic lag, linkage drag, and adaptation of the germplasm to a different environment. Therefore, direct introduction will cause a decrease in genetic gain in an elite population. Linkage drag will cause unfavourable alleles to enter the elite population alongside favourable alleles. Many elite varieties are selected for specific environments and introduction of lines not adapted to this environment can reduce the performance of newly created lines. To overcome this, many have suggested the use of pre-breeding programs using recurrent selection (e.g., Hallauer and Sears, 1972; Johnson et al., 1986).

Introgression is known to be successful for monogenic traits (e.g., Stalker, 1980). However, pre-breeding for polygenic traits is more difficult (Stalker, 1980), as most polygenes only have a small effect on the trait of interest. Pre-breeding has to be implemented well to avoid a loss of favourable alleles (Stalker, 1980). Gorjanc et al. (2016) investigated how to avoid this problem by optimising the initiation of a pre-bridging germplasm that would later be used for introgression into elite breeding programs. They investigate the use of landraces, double haploids created from landraces, and test crosses between elite hybrids and landraces as the initiation point of the pre-bridging germplasm. They conclude that the source used for the initiation plays a large role in genetic variation becoming available for introgression.

Next to this, they also concluded that significant emphasis should be put on increasing frequencies of favourable alleles for polygenic traits to boost the success of introgression. However, how introgression should take place after the pre-bridging germplasm is established was outside the scope of the study by Gorjanc et al. (2016) . Yang et al. (2020) proposed the use of origin specific GS for introgression of exotic alleles to improve polygenic traits. This method partitions out the favourable alleles coming from the elite and exotic parents and optimises parental crosses based on estimates of breeding values partitioned by elite and exotic alleles. However, their method does not address how to deal with the performance gap when crossing the exotic and elite lines. Bernardo (2009) showed that pre-breeding via recurrent selection is successful in introgressing exotic material into an elite maize line for traits controlled by 100 quantitative trait loci (QTLs) when using GS. Using a simulation study, it was shown that with GS, the time required for pre-breeding can be shortened sufficiently for breeders to be able to exploit an exotic germplasm, supporting the point that increasing frequency of favourable exotic alleles is important. Crop yield is a complex polygenic trait influenced by many underlying traits, from seed number to disease resistance (Shi et al. 2009). Therefore, it is likely that many more than 100 QTLs play a role in yield performance. Allier et al. (2019b) proposed the application of usefulness criterion parental contribution (UCPC) to multi-parental crosses as a way of introgressing favourable alleles to improve polygenic traits. When selecting parents, UCPC predicts the expected progeny performance whilst considering the genome contribution of the parents to maintain diversity. Using UCPC in three-way crosses resulted in successful introgression of exotic alleles to improve performance of a trait controlled by 500 QTLs. The focus of their study was on short-term increase in genetic gain whilst maintaining diversity. They suggest the use of optimal cross-selection (Allier et al 2019a) as a complementary strategy to increase genetic gain in the long-term. Vanavermaete et al. (2021) proposed the use of deep scoping to reintroduce genetic variation and maximise genetic gain both in the short- and long-term. Deep scoping enables introgression of favourable alleles by creating selection layers through which favourable alleles can flow, due to selection of one parent based on genomic EBV (GEBV) and the other parent selected to maximise genetic variation in the offspring. This method is successful in avoiding premature convergence of genetic gain. However, only traits up to 200 QTL were investigated. Next to this, the assumption was made that the whole base population was available for use as a training population for GS. In practice, often the lines and their alleles of interest are not part of the training population and therefore not recognised by the prediction model.

We hypothesise that we can overcome the challenges of improving the long-term performance of complex polygenic traits using introgression by creating a multi-part pre-breeding strategy with feedback pathways to optimise recurrent GS and giving favourable alleles a chance to introgress successfully. The scope of this paper is to show proof of concept of such a multi-part strategy to improve quantitative traits by introgressing exotic germplasm into an elite maize breeding program using GS. We study properties of the multi-part strategy with simulation. This multi-part strategy can be extended to other plant species for achieving higher genetic gains through introgression of genetic diversity from exotic germplasms.

Materials And Methods

Stochastic simulations of breeding programs were performed using the AlphaSimR (version 0.12.2; Gaynor et al., 2020) package in R software (version 4.0.0; R Core Team, 2020) to mimic an elite maize breeding program and quantify the effects of introgression of available exPVP germplasm into this population (scripts available from https://github.com/HighlanderLab/ibreider_multi-part-breeding-strategy). To do so, whole-chromosome haplotypes were simulated, and four donor breeding programs and one elite breeding program were founded. Recent breeding was simulated using a burn-in, to ensure all scenarios of interest shared the same starting point. After the burn-in phase, baseline scenarios and introgression scenarios were simulated and compared.

Burn-in: Genome simulation

A maize-like genome consisting of 10 chromosome pairs, each with 1,000 QTLs, for 70 founding lines was simulated using the Markovian Coalescent Simulator (Chen et al., 2009) implemented in AlphaSimR. A genetic length of 2 Morgans and a physical length of $2x10^8$ bp was assigned to the chromosomes, giving a recombination rate of $2x10^8$ bp. To roughly consider the historical changes of effective population size in maize due to domestication and breeding, effective population size was set to 100 and increased with linear piecewise increases to 10,000 at 2,000 generations ago (Hickey et al., 2014; Gorjanc et al., 2016).

A single polygenic trait, such as yield, was simulated as being influenced by 1,000 QTLs per chromosome, with a trait mean of 0 and phenotypic variance of 1. This was done so that genetic gain in the final populations was expressed in standard deviations of the founder population. Genetic value of the trait for each individual was determined by summing the additive QTL allele effects. Phenotypic values were set for each doubled haploid (DH) line by adding random error to the genetic values. Error variance was set to achieve a trait narrow-sense heritability of 0.3 when simulating 1 test location and varied over field trials depending on number of test locations simulated (where error variance decreased when number of field trials increased).

Burn-in: Founder populations and recent breeding

The 70 founding lines initiated 5 independent breeding programs (Figure 1). Four breeding programs were bred forward for 30 years, to reflect maize breeding from 1970 to 2000. One breeding program was bred forward for 50 years, to reflect maize breeding from 1970 to 2020. The breeding programs bred forward for 30 years reflected exPVP germplasm that becomes available to competitors after their licence expires and can be used as a source of exotic germplasm. The breeding program continued for 50 years reflected an elite breeding program into which exotic germplasm was introgressed.

Breeding programs simulated to reflect breeding between 1970 and 2020 followed the structure of a conventional breeding program (Figure 2). One hundred F1 lines were created from 70 parental lines in the crossing block (2,415 possible combinations, parents crossed randomly), originating from the previous year's field trials. Sixty-two DH lines were produced from each F1 line. These DH lines were planted in headrows (HDRWs) to enable seed increase and visual selection. Five-hundred lines were

advanced to the preliminary yield trial (PYT) for evaluation. This evaluation represented an unreplicated trial. The best performing fifty lines were advanced to the advanced yield trial (AYT) for evaluation and were considered as parents for the next year's F1 production. This evaluation represented a small, multi-location (4 locations) replicated trial. The best performing 10 lines were advanced to the first elite yield trial (EYT1) for evaluation and were included as parents for the next year's F1 production. This evaluation represented a large, multi-location (8 locations) replicated trial. All 10 lines were advanced to the second elite yield trial (EYT2) for evaluation and were included as parents for the next year's F1 production. This was a re-evaluation of the EYT1 stage, and any lines currently included as parents for next year's F1 had their phenotypes updated. Parents for next year's F1 consisted of all lines in the AYT, EYT1, and EYT2 trials. The two best performing lines of the EYT2 were released as varieties in the following year. All selection performed was phenotypic selection.

Future breeding: general

The scope of this paper was to show proof of concept of a multi-part strategy to improve polygenic traits by introgressing exotic germplasm using GS. A known problem in using GS for introgression of exotic material is the composition of the training population (Bernardo, 2009; Gorjanc et al., 2016). However, the focus of this paper is on the structure of the breeding program including the pre-breeding strategy, rather than on solving the challenge of optimal training population composition for introgression. Therefore, initially the accuracy of GS was set to 1, to show the potential of the multi-part strategy under ideal circumstances. GS accuracy of 0.5 and 0.7 were also investigated for comparison, by adding error to the true genetic value to obtain these accuracies. After the burn-in all breeding programs were continued for 100 years. For each scenario of interest and each baseline, ten replicates were simulated.

Future breeding: Two-part breeding program without introgression

A two-part breeding program without introgression, as described by Gaynor et al. (2017) was used as a baseline trend for genetic gain and genic variance. The two-part breeding program consisted of a population improvement (PI) and a product development (PD) component. The function of the PI component was to increase the speed at which progress was made, by rapid recurrent selection on potential parents for the PD component. The function of the PD component was the development of new varieties. The PD component was structured as a conventional breeding program, as described for the burn-in, except that the F1 stage was replaced by the PI component.

For the current study, the number of breeding cycles per year in the PI component varied between 2 to 6 cycles. For comparison purposes it was assumed that there are fixed financial resources to genotype a total of 7,200 individuals per year in the PI component. Therefore, when more breeding cycles per year were performed, each cycle had less genotyped individuals available from which parents could be selected. This resulted in varied selection intensity in the PI component over scenarios, depending on the number of breeding cycles performed per year.

Starting off with 7,200 individuals in the PI component (Figure 3), individuals were divided by the number of breeding cycles per year, then randomly and equally divided into female and male parents. From these female and male parents, 50 individuals were selected each, using GS and limiting to 1 individual per half-sib family. These parents were used to create next year's cross, and at the end of each year to produce 200 individuals to be moved forward to the PD component. The PD component (Figure 3) then continued as a conventional breeding program as described for the burn-in phase; with the exception that the F1 stage was replaced by 200 individuals originating from the PD component, only 31 DH lines were produced from each line and only 1 variety was released as a commercial product. This breeding strategy was used as a baseline to evaluate the multi-part strategy against. As number of breeding cycles per year and GS accuracy play a role in the two-part strategy, this resulted in 15 baseline scenarios (3 accuracies times 5 levels of number of breeding cycles per year scenarios).

Future breeding: Multi-part breeding strategy

The multi-part strategy (Figure 4) is an extension of the two-part strategy, where exotic germplasm was introduced into the PI component using pre-breeding bridges. The function of the pre-breeding bridges is to break down linkage between favourable and unfavourable alleles and reduce the performance gap between the exPVP and elite germplasm. After the final bridge, exPVP alleles are introgressed into the PI component of the two-part strategy. From here germplasm moved to the PD component including favourable alleles can lead to a new variety. For comparison purposes it was assumed that for the pre-breeding component a total of 1,200 individuals per year were genotyped and for the PI component a total of 6,000 individuals per year were genotyped.

The exPVP breeding programs established in the burn-in were used as the source of exotic material. ExPVP breeding programs were continued as they were established during the burn-in phase (using phenotypic selection), resulting in an initial lag of 20 years. Each year the two varieties released from each exPVP breeding program were used for introgression, resulting in 8 varieties being introduced into the pre-breeding component each year (2 varieties from each of the 4 programs).

The PI component of the multi-part strategy aimed to produce the most desired parents for the PD component. The PI component was initiated in year 1 of future breeding using 500 individuals (430 best PYT lines, all 50 AYT lines, all 10 EYT1 lines and all 10 EYT2 lines) taken from last year's PD component. These individuals were randomly crossed, where the number of crosses was maximised within the budget available (total budget of genotyping for 6,000 individuals divided by number of breeding cycles per year). These individuals were randomly and equally divided into female and male parents. Fifty females and 50 males were selected, limited to 1 individual per female half-sib family. These females and males were again crossed (number of crosses limited by the budget) and their offspring was used to start the next breeding cycle. At the end of each year 200 extra crosses were performed to account for individuals moving to the PD component of the breeding program.

Pre-breeding bridges were established to break linkage drag and improve performance of the introgressed germplasm. Bridges were initiated at the beginning of the future breeding period, after the two-part

strategy was established. Bridges were initiated individually, one per year. The first bridge was initiated in two stages, starting the year after the two-part strategy was established. First, a cross was performed between the 8 exPVP varieties which became available the previous year and 48 individuals randomly selected from the previous year's last PI component. Each of the 8 exPVP varieties was crossed with 6 individuals resulting from the PI cycle, each PI individual was only used once, resulting in 48 F1 individuals. The following year these 48 F1 individuals were crossed with 52 individuals randomly selected from the previous year's last PI cycle. Consecutive bridges were initiated by randomly selecting 50 of the top individuals produced by the previous bridge and randomly selecting 50 individuals from the last PI cycle. The year after the last bridge was initiated, introgression into the PI component started. Once bridges were established, they were continued as follows (Figure 5a and b). Individuals resulting from the last breeding cycle of the previous year were randomly divided into females and males. When germplasm exchange took place, individuals moved forward from the exPVP breeding programs or a previous bridge were introduced into the female population. Germplasm was moved back from the PI component to enhance the performance in the pre-breeding component, with the aim of reducing the performance gap between the exPVP and elite germplasm. Individuals moved back from the PI component were reintroduced into the male population. Number of crosses made between females and males for each bridging-cycle was maximised within the budget available (nGenotyped, 1,200/number bridges divided/number of breeding cycles). The top individuals were selected using GS and used to start the next breeding cycle. When germplasm exchange took place, top individuals were moved forward to the next component of the breeding program (either the next bridge or the PI component).

After all bridges were established, introgression into the PI component commenced. Each cycle of the PI component started off with the maximum number of individuals available under the restrictions of the genotyping budget. These individuals were equally and randomly split into females and males. When no germplasm exchange took place 50 females and 50 males were selected. When germplasm exchange did take place 50 females, and 50 males minus the number of males to be introgressed, were selected. Females and males were crossed, and the offspring was used to start the next breeding cycle. When germplasm exchange took place, extra crosses (100 times the bridging rate (defined below) times the number of bridges) were made to account for individuals to be reintroduced into each bridge. At the end of each year another 200 extra crosses were made to account for individuals moving to the PD component of the breeding program.

The PD component followed the structure of the conventional breeding programs used, but started from the DH phase. From each of the 200 individuals received from the PI component 31 DH lines were created. Field trials were continued as described earlier for conventional breeding programs. Selection of AYT lines from the PYT was done using GS. Recycling of parents no longer took place.

Optimisation of parameters

Parameters optimised to obtain maximum genetic gain were number of bridges, number of breeding cycles in the pre-breeding and PI component, bridging rate, introgression rate, return rate, and exit rate. GS

accuracies of 0.5, 0.7, and 1 were investigated. The number of bridges used for pre-breeding varied from 1 to 3 and the number of breeding cycles per year from 1 to 6 (2 to 6 for the PI component). Bridging rate was the rate at which individuals were moved forward from one bridge to the next. This ranged from 0.05 to 0.5, with increments of 0.05. When only one bridge was used in the pre-breeding component, bridging rate was no longer applicable. Introgression rate was the rate at which individuals were moved forward from the last bridge into the PI component. This ranged from 0.1 to 0.5, with increments of 0.1. Return rate was the rate at which individuals were moved back from the PI component to the bridges. This ranged from 0.05 to 0.5, with increments of 0.05. Exit rate was the number of times per year germplasm exchange between the different components of the multi-part strategy took place. This ranged from 1 to 6 times per year and was limited by the number of breeding cycles performed per year. As a genotyping budget for 1,200 individuals was assumed for the pre-breeding component, some combinations of parameter values were not viable. A grid search was performed using a Python script. A total of 33,885 scenarios were simulated, each repeated 10 times.

Evaluation of multi-part scenarios

Performance of the multi-part strategy resulting in the highest genetic gain relative to the corresponding two-part baseline was assessed by comparing genetic gain in the EYT1 field trial of the PD component at year 100 of future breeding. A paired Welch's t-test was performed on log-transformed values, which were back-transformed to obtain 95% CI and ratio of difference (Ramsey and Schafer, 2002). To gain insight in which parameters have a major impact on the genetic gain achieved, the parameters and outcome of all scenarios were analysed using a regression tree. A linear model was fitted using R software (version 4.0.0), using genetic gain difference between the multi-part strategy and corresponding two-part strategy baseline. Accuracy of selection, number of bridges, number of breeding cycles per year and exit rate were included as a factor and introgression rate, bridging rate and return rate as continuous variables in the model. All parameters were included on their own and all interactions between pairs of variables that were significant (p=0.05) were included as well. A regression tree was then built using the "ctree()" function of the "party" package in R (Hothorn et al., 2006) and fitting a Bonferroni test control with a minimum criterion of 0.95 and a maximum tree depth of 4. For parameters showing a major impact on genetic gain difference in the regression tree analysis, the effect of the different levels of these parameters on absolute genetic gain were examined.

Haploblock analyses and QTL origin

Haploblock analyses were performed to quantify the number of haploblocks present in the exPVP breeding programs which showed higher genetic values than the corresponding haploblocks already present in the elite breeding program, at year 1 to 10, then at year 20 to 100 with increments of 10 years. This showed the potential impact of introgression on increasing genetic gain in the elite population using the optimised multi-part strategy. For the haploblock analyses, the optimised multi-part strategy was rerun, to obtain snapshots of the populations over time.

Snapshots were taken of the population consisting of 8 exPVP individuals used for introgression (exotic population), and the population at the start of the PI component (1,000 individuals, elite population), at the years of future breeding as stated above. QTLs present in each population at these time points were extracted for each individual using the pullQtlHaplo() function from AlphaSimR. QTL haplotypes for each individual were then divided into haploblocks. Analyses were performed both on haploblocks of length 1 cM (5 QTL) and 10 cM (50 QTL). Haploblock effects were obtained by multiplying QTL alleles by QTL effect and summed over QTLs within haploblocks. Average haploblock effects within population were calculated. Haploblock effects were compared between exotic and elite population to obtain the percentage exPVP haploblocks outperforming the elite haploblocks at the given years of future breeding; as well as the average, minimum, and maximum excess performance of these haploblocks. This was averaged over the 10 repeats of the optimised multi-part breeding strategy.

Origin of QTLs present in the PI component after selection (100 individuals) at year 100 of future breeding was obtained to examine the effectiveness of introgression over the years. To do so, IDs of the 8 exPVP individuals used for introgression were reset at the end of each year. These IDs were registered for cross reference after the simulation, where year 1 to 9 were treated as individual groups and year 10 to 99 were merged as groups of ten years. The identity by descent coding of the QTLs in the PI component after selection was obtained using the pullIbdHaplo() function in AlphaSimR. The identity by descent coding of alleles was then converted to ancestor IDs. By linking these IDs up with the IDs obtained for cross reference, it was possible to trace back which introgression event the QTLs originated from.

Results

Future breeding: Multi-part breeding strategy

Introgression of exPVP germplasm into an elite breeding program using the multi-part strategy resulted in the highest long-term genetic gain. Introgression rate, GS accuracy, and number of breeding cycles per year had the largest influence on the genetic gain from the multi-part strategy relative to the two-part strategy. Introgression was successful, even when GS accuracy was as low as 0.5. Introgression was only successful when the introgression rate was 0.4 or 0.5. Introgression resulted in highest long-term genetic gain when using six breeding cycles per year. Exotic haploblocks of 5 QTL carrying higher genetic value than corresponding elite haploblocks existed in the exPVP breeding programs at year 100 of future breeding. Exotic QTLs originating from different introgression events were successfully introgressed into the elite breeding program.

Introgression of exPVP germplasm into an elite breeding program using the multi-part strategy resulted in higher long-term genetic gain than other tested strategies. Figure 6 shows the genetic gain and genic variance at the EYT1 phase of the PD component resulting from 100 years of future breeding for the multi-part (blue, solid line) and two-part (black, dashed line) strategies. At year 32 of future breeding the genetic gain from the multi-part strategy started to overtake the genetic gain from the two-part strategy. At year 100 of future breeding genetic gain from the multi-part strategy was 1.53 times higher than from

the two-part strategy (95% CI: 1.40 - 1.66). Introgressed germplasm caused an increase in genic variance in the multi-part strategy from year 7 onwards. Although some of this variance was subsequently lost again, the increase in genetic gain from year 32 onwards showed that the variance originating from the exPVP, contributed to the increase in genetic gain in the elite breeding program.

GS accuracy, introgression rate, and number of breeding cycles per year had the largest influence on the genetic gain from the multi-part strategy relative to the two-part strategy. Regression tree analysis showed that an introgression rate of 0.5 generally resulted in a higher genetic gain at year 100, relative to the two-part strategy. Three or more cycles per year resulted in a higher genetic gain at year 100, relative to the two-part strategy. GS accuracy had significant interactions with many parameters and the direct effect of accuracy on genetic gain relative to the two-part strategy was not obvious from the regression tree.

Introgression of exotic germplasm was successful, even when GS accuracy was as low as 0.5. Figure 7 shows the difference in genetic gain at the EYT1 phase of the PD component obtained using the multi-part strategy when GS accuracy varied from 0.5, 0.7, to 1. When accuracy was lower, genetic gain achieved was lower and genic gain was higher at any year of future breeding. At year 100 of future breeding genetic gain resulting from the multi-part strategy with a GS accuracy of 0.5 was 1.30 times higher, with a GS accuracy of 0.7 was 1.45 times higher, and with a GS accuracy of 1 was 1.53 times higher than genetic gain from the corresponding two-part strategy. Genic variance (Figure 7) retained at year 100 of future breeding was lowest when GS accuracy was 1 and highest when accuracy was 0.5.

Introgression of exotic germplasm was only successful when the introgression rate was 0.4 or 0.5. Figure 8 shows the difference in genetic gain at the EYT1 phase of the PD component obtained using the multi-part strategy when introgression rate varied from 0.1 to 0.5. When the introgression rate was lower than 0.4, genetic gain did not differ from genetic gain obtained using the two-part strategy. When introgression rate was 0.4 the genetic gain from the multi-part strategy was 1.16 times higher than from the two-part strategy and 1.53 when introgression rate was 0.5 at year 100 of future breeding. Genic variance (Figure 8) declined for all levels of introgression rate, but this decline was much more gradual when introgression rate was 0.5.

Introgression of exotic germplasm was optimal when 6 breeding cycles per year were performed. Figure 9 shows the difference in genetic gain at the EYT1 phase of the PD component obtained using the multi-part strategy when the number of breeding cycles per year varied from 1 to 6. When the number of breeding cycles in the multi-part strategy was between 2 and 5 the genetic gain achieved at year 100 of future breeding was not significantly different and 1.22 to 1.28 times higher than resulting from the relevant two-part strategy. When the number of breeding cycles per year in the multi-part strategy was 1, the two-part strategy with 6 breeding cycles per year outperformed the multi-part strategy. No clear trend was seen in the relationship between the number of breeding cycles per year and the genic variance retained in the population.

Haploblock effects and QTL origin

Exotic haploblocks of length 5 QTLs carrying higher genetic value than corresponding elite haploblocks still exist in the exPVP breeding programs at year 100 of future breeding. Table 1 shows the percentage of exotic haploblocks that outperform the corresponding elite haploblocks for year 1 to 100 of future breeding. Although this percentage decreased as breeding progressed, at year 100 of future breeding still over 8% of exotic haploblocks of length 5 QTLs outperformed the elite haploblocks. For haploblocks of length 50 QTLs there are no longer exotic haploblocks present which outperform the existing elite haploblocks at year 40 of future breeding (Table 2).

Table 1 and 2 also show the genetic value of the exotic haploblocks outperforming the elite haploblocks, as an excess compared to the elite haploblocks. As expected, the excess genetic value of the exotic haploblocks declined as the breeding program progressed.

Exotic QTLs originating from different introgression events were successfully introgressed into the elite breeding program. Figure 10 shows from which introgression event the QTLs present in the PI component at year 100 of future breeding originate. Averaged over 10 repeats, 69% of QTLs were already present in the elite breeding program before introgression started, 3% originated from exotic germplasm introduced during the first cross between exotic and elite material. The remaining 28% originated from exotic material introgressed after year two, with the contribution declining as future breeding progressed. However, introgression continued to be successful as exotic QTLs originating from year 90 to 99 of future breeding were found in the genotypes present in the PI component at year 100. It should be noted that it takes a minimum of 8 years for exotic alleles to reach the PI component in the proposed multi-part strategy. Therefore, effectively the most recently introgressed exotic alleles can only originate from exotics produced in years 90 to 92.

Discussion

This study showed that the multi-part strategy successfully introgressed exotic germplasm to increase genetic gain of a polygenic trait. We showed that increased genetic gain with the multi-part strategy was due to the introgression of useful exotic genic variance. We showed that introgression rate, GS accuracy, and number of breeding cycles per year had a large influence on the success of the multi-part strategy. We discuss that an introgression rate maximising the number of parents carrying pre-bred germplasm was most successful. As expected, we found that higher GS accuracy resulted in faster increase and higher overall genetic gain. We discuss how GS accuracies used in our simulations relate to previous empirical studies. We discuss that maximising number of breeding cycles per year within the budget constraints resulted in highest genetic gain, due to selection being limited to once a year. We confirm that sufficient recombination events were essential to make the favourable haploblocks accessible. Finally, we discuss that both the first, as well as subsequent introgression events, contributed to the composition of the genotypes in the elite population in the final year of the breeding program.

The multi-part breeding strategy resulted in higher genetic gain

Introgression of exPVP germplasm into an elite breeding program using the multi-part strategy was successful in reintroducing genic variance, which resulted in higher and continuing long-term genetic gain of the simulated polygenic trait. The multi-part strategy was optimised for 100 years of future breeding. This was at the cost of progress in genetic gain during the first 32 years of future breeding. Optimising parameters for genetic gain at year 100 resulted in emphasis of the program shifting from closing the performance gap to retaining genic variance and breaking linkage drag. Therefore, assuring genic variance to be available for future crop improvement.

This study assumed that the exPVP breeding lines and two-part strategy baselines were closed programs. This is not the case, as breeding programs do attract and exchange germplasm from and with other programs. However, germplasm used for this purpose usually comes from a limited source (e.g., elite lines available within the company or from other countries). An additional assumption was that no mutations took place in the genotypes, for simplification purposes of the simulation. Even though these mechanisms play a role in increasing diversity in the elite breeding programs, observed rates of gain in yield potential as well as the gap between achieved yield and yield potential could indicate maize yield is currently converging to a sub-optimal level of genetic gain (Lobell and Azzari, 2017). Our simulation has shown that elite breeding programs indeed rapidly run out of genic variance when genetic improvement is accelerated using a two-part breeding strategy in line with Gaynor et al. (2017). Therefore, although these sources of genetic diversity are not considered in this study, they are not expected to significantly influence our findings.

The multi-part strategy was successful in introgressing exPVP germplasm

The multi-part strategy successfully closed the performance gap between the exotic populations and the elite breeding program. Results showed a difference in genetic gain between the two-part and multi-part strategy. The size of this difference varied with introgression rate, where higher introgression rate resulted in a larger difference in genetic gain between the two-part and multi-part strategy. This showed that the improvement in genetic gain achieved using the optimised multi-part strategy was due to introgression. Therefore, proving that exotic material has successfully progressed through the bridging phase, PI and PD. This can only have occurred if parents carrying exotic germplasm have a performance high enough to be selected as parents for the next breeding cycles.

The multi-part strategy was successful in breaking down linkage drag between favourable and unfavourable alleles. Results showed an increase in genic variance from the moment exotic material reached the EYT1 phase, relative to the two-part strategy. However, genetic gain achieved using the multi-part strategy in earlier years was lower than genetic gain achieved using the two-part strategy, indicating that initially the increase in genic variance slowed down genetic gain. Over the course of 100 years of future breeding, genic variance in the multi-part strategy decreased again whilst simultaneously genetic gain continued to increase. This indicates purging of unfavourable alleles whilst retaining or even increasing frequency of favourable alleles, hence successfully breaking down linkage drag.

The current study focuses solely on the introgression of exPVP germplasm. However, other sources with potential to improve crop yield in elite lines are available for introgression as well. Gorjanc et al. (2016) simulated an initiation of a maize pre-breeding program harnessing polygenic variation from landraces. Brauner et al. (2018) investigated the testcross performance of 89 DH lines produced from 6 landraces and found that the best DH lines approached yield levels of elite lines, thereby demonstrating the potential for using these lines for introgression into elite lines. Hölker et al. (2019) and Mayer et al. (2020) both built landrace derived DH libraries to capture genetic diversity available for the enhancement of elite lines and make this information accessible for breeding decisions. This shows that the plant breeding community is working on sources of exotic germplasm suitable for the multi-part strategy.

Introgression rate, GS accuracy, and number of breeding cycles per year have a large influence on success of the multi-part strategy

Although the optimised multi-part strategy outperformed the corresponding two-part strategy, not all multi-part scenarios did. Regression tree analyses showed that introgression rate, GS accuracy and number of breeding cycles per year influenced the achieved increase in genetic gain. It should be noted that genetic gain was optimised for year 100 of future breeding. Optimising for other time points resulted in a different combination of optimal parameter values (results not included) and parameters having the largest influence on gain achieved could differ.

Introgression rate

Introgression was most effective when introgression rate was maximised at 0.5. This meant that 100% of males in the first breeding cycle of the PI component after introgression originated from the last bridge. Therefore, parents selected for the second breeding cycle always carried germplasm originating from the bridging component. This indicates that forced introgression of pre-bred germplasm which might not immediately outperform elite germplasm (i.e. all selected male parents in the PI component originate from the last bridge) was beneficial. An introgression rate of 0.3 or lower did not result in an increase in genetic gain relative to the two-part strategy. We hypothesise this was due to the number of potential parents only carrying elite germplasm (i.e. no germplasm originating from the bridging component) available for the second breeding cycle. These parents were likely to outcompete parents carrying pre-bred germplasm, due to left-over linkage drag present in the latter. This echoes the findings of Gorjanc et al. (2016), who found that initiation of pre-bridging germplasm from elite hybrids and selected landraces resulted in reconstruction of the elite genomes. Therefore, losing the favourable variation brought in by landraces. Instead, they recommend pre-breeding germplasm to be initiated fully from landraces and to increase frequency of favourable alleles before commencing introgression. Although in the current study no pre-bridging germplasm was initiated, the bridges serve the same purpose and an introgression rate of 0.5 throughout the bridges and PI component ensured favourable variation brought in from exPVP was utilised.

Accuracy

Higher GS accuracy accelerated the rate at which genetic gain improved and resulted in higher absolute gain at year 100 of future breeding. When examining the genic variance, lower selection accuracy resulted in more genic variance being retained. However, genetic gain did not increase, indicating unfavourable alleles were able to reach the PI component, where they had a negative impact on the potential genetic gain. Results showed that although improvement in genetic gain was achieved when selection accuracy was not optimal, it took longer, and the long-term genetic gain was lower. In the current study accuracy was set rather than observed. This was done as the scope of this paper was to show proof of concept. Estimating marker effects from a training set of exotic and elite lines and their crosses is a major research question that requires further study.

Jannink et al. (2010, and references therein) reviewed GS in plant breeding and summarised that accuracies between 0.61 and 0.83 can be realistically expected in elite programs. They summarise that accuracy obtained depends on many variables, including; training population size and design, effective population size, marker density, and type of model used to estimate marker effects.

The success of introgression of exotic germplasm using GS will strongly depend on an appropriate training population and statistical model used to estimate the marker effects. When this training population exclusively consists of individuals originating from elite lines, the statistical model will not be able to capture the effect of exotic alleles on the trait of interest and therefore penalise exotic lines. When exotic and elite alleles are present in the training population simultaneously, the statistical model is expected to find stronger effects for marker alleles sitting on the elite haplotypes, because elite lines are expected to have a higher performance, and linkage between an enrichment of favourable alleles and elite marker alleles will be considerable. Therefore, the statistical model is expected to favour elite alleles over exotic alleles, meaning exotic alleles are selected against, even before they enter the elite population, as observed by Gorjanc et al. (2016). However, when only exotic alleles are in the training population, the GS model will not recognise elite alleles, which will diminish improvement in genetic gain, as exotic lines are not expected to have the same level of performance as elite lines. Further work is needed in this area.

Breeding cycles per year

Optimal genetic gain was achieved using six breeding cycles per year. Genetic gain still increased at year 100 of future breeding. Two to 5 breeding cycles per year resulted in higher genetic gain than the two-part strategy, but genetic gain reached a plateau before year 100 of future breeding when more than 2 breeding cycles per year were performed. At year 100 of future breeding the genetic gain achieved was not significantly different using 2 to 5 cycles per year, but the number of years it took to reach this level of genetic gain decreased with increasing number of cycles per year.

The difference in genetic gain achieved with 6 breeding cycles per year compared to 5 cycles per year was large. Due to the restraint on resources, an increase in number of cycles per year meant a reduction in number of individuals that could be genotyped per cycle. Budget was set to 1,200 genotypes per year for the whole bridging component. This meant that for scenarios with 6 breeding cycles per year and 2 bridges, only 100 individuals per breeding cycle could be genotyped. When no germplasm exchange took

place, 100 parents were selected for the next breeding round. Therefore, effectively no selection took place in these cycles. With an exit rate of 1 germplasm exchange only took place once a year. Therefore, 5 out of 6 breeding cycles within the bridging component did not put any selection pressure on the individuals in the bridge. This resulted in genic variance being much better retained in scenarios with 6 breeding cycles per year, as well as allowing sufficient chance for recombination to break linkage drag. As the genotyping budget was higher for the PI component, selection still took place here every breeding cycle. This suggests that the main function of the bridging component was stacking favourable alleles and the main function of the PI component was to purge unfavourable alleles.

The utilisation of the potential present in the exotic population depends on the length into which haploblocks need to be broken down

Analyses of haploblock effects have shown that exotic alleles which outperform the elite alleles existed. However, this did not always mean they were accessible for introgression. Success of introgression depended on how well linkage between favourable and unfavourable alleles was broken down. The percentage of exotic haploblocks that outperformed elite haploblocks depended on the length of the haploblocks. When haploblocks were defined as 5 QTL, the percentage of exotic haploblocks outperforming elite haploblocks was higher, than when defined as 50 QTL. Larger haploblocks are generally more accessible for introgression, as less recombinations are needed for them to become part of the elite population. However, as there are fewer large exotic haploblocks which outperform elite haploblocks, at year 40 of future breeding large favourable haploblocks were no longer present in the exotic population. Smaller haploblocks becoming accessible requires more recombinations. This study showed that with 6 breeding cycles per year in the bridging component, of which only one cycle puts selection pressure on the population, exotic haploblocks outperforming elite haploblocks were still present at year 100 of future breeding. However, the excess effect of the haploblocks decreased the further along future breeding was. This substantiates the knowledge that introgression of polygenic traits is more difficult than introgression of monogenic traits (Stalker, 1980), as the chance that haploblocks are successfully introgressed decreases with a decrease in their excess effect. The decrease in excess effect was expected as haploblocks with a larger favourable effect would be selected for in earlier years of future breeding, thereby forthgoing being classified as elite. It should be noted that the haploblock effect analyses reflect the potential that was present in the exotic breeding programs, not the actual utilisation of this potential.

Both first and subsequent introgression events contributed to the composition of the genotypes of the elite population

QTL origin analyses showed that exotic alleles with beneficial potential to the elite breeding program were successfully introgressed. Analyses also showed that introgression events continued to be successful, although their contribution sharply declined after the first introgression event (year 2 of future breeding). This declined contribution was due to on the one hand favourable alleles already having been introgressed (and therefore became classified as elite allele) and on the other hand the difference in

performance between the exPVP lines and the elite lines becoming larger. This caused the number of available exotic alleles that outperformed the elite alleles, as well as their excess effect on the trait, to decline when future breeding progressed. As exPVP breeding programs were assumed to be independent and closed breeding programs, it indeed was not expected that new alleles would arise (Hallauer and Sears, 1972) (apart from mutations, which were not considered in the simulations). Therefore, successful introgression events in later years were likely the result of the haploblock compositions of the parents having become more beneficial due to broken down linkage drag and recombination.

The establishment of the more beneficial haploblock compositions in our study occurred by chance. Bijma et al. (2020) show that Mendelian sampling variance differs among individuals and that it is possible to select parents with high gametic variance, which is analogous to the use of usefulness criterion (originally proposed in 1975 by Schnell and Utz) defined as the expected genetic mean of the selected offspring resulting from a cross. Akdemir et al. (2016) similarly used the key concepts of the usefulness criterion in their genomic mating method, where crosses are optimized by leveraging complementarity between parents. Allier et al. (2019b) extend these ideas by explicitly considering the genome contribution of the parents to diversity present in the next generation. This is a growing area of research and such techniques could be incorporated in the multi-part strategy, potentially increasing the rate at which genetic gain is achieved.

Conclusion

We showed that introgression by means of a multi-part pre-breeding strategy with feedback pathways to optimise recurrent GS is successful in improving polygenic traits. The multi-part strategy resulted in 1.53 times higher genetic gain after 100 years of future breeding, compared to the two-part strategy. This was achieved by maximising the number of parents carrying pre-bred germplasm used for introgression into the PI component. Potential of the pre-bred germplasm was highest when selection in the bridging component was only performed in one breeding cycle per year. This was due to better retaining of genic variance and optimised chance of reducing linkage drag through recombination. The reduction of linkage drag through recombination was essential for favourable haploblocks to become accessible. The first introgression event contributed most of the exotic alleles to the population at 100 years of future breeding, subsequent introgression events were successful as well. Our results suggest that the multi-part strategy has a potential to improve polygenic traits, providing a tool to avoid sub-optimal convergence of the long-term genetic gain. Therefore, this breeding strategy has a real potential to contribute to global food security. Before application in practice is possible, the challenge of training population composition for GS when introgressing exotic material needs to be addressed.

Declarations

Author contributions

JMH and RCG conceived and supervised the study. GG co-supervised the study. ISB performed the experiments and wrote the manuscript. ST contributed the script used for the grid search and gave valuable support on the use of the computer cluster. All authors were involved in valuable discussions regarding the study and reviewed and approved the manuscript.

Acknowledgements

The authors acknowledge the financial support from the BBSRC to The Roslin Institute “BBSRC BB/R019940/1” and the DBT to ICRISAT “BT/IN/UK/PORI/0I/MKP/2018-19”. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/). Authors would like to thank members of the AlphaGenes and HighlanderLab teams for their valuable comments and feedback during this study.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

Akdemir, D. and Sánchez, J. I. (2016), ‘Efficient breeding by genomic mating’, Frontiers in Genetics 7, 210. https://doi.org/10.3389/fgene.2016.00210

Allier, A., Lehermeier, C., Charcosset, A., Moreau, L., and Teyssèdre, S. (2019a). ‘Improving Short- and Long-Term Genetic Gain by Accounting for Within-Family Variance in Optimal Cross-Selection.’ Frontiers in genetics 10(1006). https://doi.org/10.3389/fgene.2019.01006

Allier, A., Moreau, L., Charcosset, A., Teyssèdre, S. and Lehermeier, C. (2019b), ‘Usefulness Criterion and Post-selection Parental Contributions in Multi-parental Crosses: Application to Polygenic Trait Introgression’, G3 Genes|Genomes|Genetics 9(5), 1469–1479. https://doi.org/10.1534/g3.119.400129

Bernardo, R. (2009), ‘Genomewide selection for rapid introgression of exotic germplasm in maize’, Crop Science 49(2), 419–425. https://doi.org/10.2135/cropsci2008.08.0452

Bijma, P., Wientjes, Y. C. J. and Calus, M. P. L. (2020), ‘Breeding Top Genotypes and Accelerating Response to Recurrent Selection by Selecting Parents with Greater Gametic Variance’, Genetics 214(1), 91–107. https://doi.org/10.1534/genetics.119.302643

Brauner, P. C., Schipprack, W., Utz, H. F., Bauer, E., Mayer, M., Schön, C-C. and Melchinger, A. E. (2019), ‘Testcross performance of doubled haploid lines from European flint maize landraces is promising for broadening the genetic base of elite germplasm.’ Theoretical and Applied Genetics 132(6), 1897-1908. https://doi.org/10.1007/s00122-019-03325-0

Chen, G. K., Marjoram, P. and Wall, J. (2009), ‘Fast and flexible simulation of DNA sequence data’, Genome research 19. https://doi.org/10.1101/gr.083634.108.

Falconer, D. and Mackay, T. (1996), Introduction to quantitative genetics, 4th edition, India: Dorling Kindersley.

Gaynor, R. C., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., Mackay, I. J. and Hickey, J. M. (2017), 'A two-part strategy for using genomic selection to develop inbred lines', Crop Science 57(5), 2372–2386. https://doi.org/10.2135/cropsci2016.09.0742

Gaynor, R. C., Gorjanc, G. and Hickey, J. M. (2020), 'AlphaSimR: an R package for breeding program simulations', G3 Genes|Genomes|Genetics 11(2). https://doi.org/10.1093/g3journal/jkaa017

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

Gorjanc, G., Jenko, J., Hearne, S. J. and Hickey, J. M. (2016), 'Initiating maize pre-breeding programs using genomic selection to harness polygenic variation from landrace populations', BMC Genomics 17(30). https://doi.org/10.1186/s12864-015-2345-z

Hallauer, A. R. and Sears, J. H. (1972), 'Integrating exotic germplasm into corn belt maize breeding programs1', Crop Science 12(2). https://doi.org/10.2135/cropsci1972.0011183X001200020016x

Hickey, J. M, Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Boddupalli M., Prasanna, B. M., Grondona, M., Zambelli, A., Windhausen, V. S, Mathews, K., Gorjanc, G. (2014), 'Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation.', Crop Science 54. https://doi.org/10.2135/cropsci2013.03.0195

Hölker, A.C., Mayer, M., Presterl, T., Bolduan, T., Bauer, E., Ordas, B., Brauner, P. C., Ouzunova, M., Melchinger, A. E. and Schön, C-C. (2019), 'European maize landraces made accessible for plant breeding and genome-based studies.' Theoretical and Applied Genetics 132, 3333–3345. https://doi.org/10.1007/s00122-019-03428-8

Hothorn, T., Hornik, K. and Zeileis, A. (2006), 'Unbiased recursive partitioning: A conditional inference framework', Journal of Computational and Graphical Statistics 15(3), 651–674. https://doi.org/10.1198/106186006X133933

Jannink, J.-L., Lorenz, A. J. and Iwata, H. (2010), 'Genomic selection in plant breeding: from theory to practice', Briefings in Functional Genomics 9(2), 166–177. https://doi.org/10.1093/bfgp/elq001

Johnson, E. C., Fischer, K. S., Edmeades, G. O. and Palmer, A. F. E. (1986), 'Recurrent selection for reduced plant height in lowland tropical maize1', Crop Science 26(2). https://doi.org/10.2135/cropsci1986.0011183X002600020008x

Lobell, D. and Azzari, G. (2017), 'Satellite detection of rising maize yield heterogeneity in the u.s. midwest', Environmental Research Letters 12(1), 014014.

https://doi.org/10.1088/1748-9326/aa5371

Mayer, M., Hölker, A. C., González-Segovia, E., Bauer, E., Presterl, T., Ouzunova, M., Melchinger, A. E and Schön, C-C. (2020), 'Discovery of beneficial haplotypes for complex traits in maize landraces', Nature Communications 11. https://doi.org/10.1038/s41467-020-18683-3

Mikel, M. and Dudley, J. (2006), 'Evolution of north American dent corn from public to proprietary germplasm', Crop Science 46(3), 1193–1205. https://doi.org/10.2135/cropsci2005.10-0371

R Core Team (2020), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Ramsey, F. and Schafer, D. (2002), The statistical sleuth: A course in methods of data analysis, 2nd edition, Duxbury Press, Pacific Grove, CA.

Schnell F. and Utz H. (1975), 'F1-Leistung und Elternwahl in der Züchtung von Selbstbefruchtern' in Berichtüber die Arbeitstagung der Vereinigung österreichischer Pflanzenzüchter (Austria: BAL Gumpenstein), 243–248.

Shi, J., Li, R., Qiu, D., Jiang, C., Long, Y., Morgan, C., Bancroft, I., Zhao, J. and Meng, J. (2009), 'Unraveling the Complex Trait of Crop Yield With Quantitative Trait Loci Mapping in Brassica napus', Genetics 182(3), 851–861. https://doi.org/10.1534/genetics.109.101642

Stalker, H. (1980), 'Utilization of wild species for crop improvement' paper no. 6188 of the journal series of the north carolina agricultural research service, raleigh, north carolina 27650.', 33, 111–147. https://doi.org/10.1016/S0065-2113(08)60165-0

Vanavermaete, D., Fostier, J., Maenhout, S. and De Baets, B. (2020), 'Preservation of Genetic Variation in a Breeding Population for Long-Term Genetic Gain', G3 Genes|Genomes|Genetics 10(8), 2753-2762. https://doi.org/10.1534/g3.120.401354

Vanavermaete, D., Fostier, J., Maenhout, S. and De Baets, B. (2021) 'Deep scoping: a breeding strategy to preserve, reintroduce and exploit genetic variation.', Theoretical and Applied Genetics. https://doi.org/10.1007/s00122-021-03932-w

Yang, C. J., Sharma, R., Gorjanc, G., Hearne, S., Powell, W. and Mackay, I. (2020), 'Origin Specific Genomic Selection: A Simple Process To Optimize the Favorable Contribution of Parents to Progeny', G3 Genes|Genomes|Genetics 10(7), 2445–2455. https://doi.org/10.1534/g3.120.401132

Tables

Table 1 Percentage of exotic haploblocks that outperform corresponding elite haploblocks (averaged over 10 repeats) for year 1 to 100 of future breeding and mean, min and max

excess performance of these haploblocks when haploblock length is 5 QTLs per bin (1 cM).

Year	Percentage	Mean	Min	Max
1	41.53	0.0017	0.0000	0.0295
2	37.97	0.0016	0.0000	0.0314
3	35.02	0.0015	0.0000	0.0288
4	32.73	0.0014	0.0000	0.0302
5	31.94	0.0013	0.0000	0.0291
6	31.13	0.0013	0.0000	0.0289
7	29.99	0.0012	0.0000	0.0276
8	29.68	0.0012	0.0000	0.0258
9	27.97	0.0011	0.0000	0.0279
10	27.31	0.0011	0.0000	0.0278
20	20.39	0.0007	0.0000	0.0230
30	15.88	0.0005	0.0000	0.0230
40	13.68	0.0004	0.0000	0.0196
50	12.00	0.0003	0.0000	0.0196
60	10.61	0.0003	0.0000	0.0199
70	9.78	0.0003	0.0000	0.0170
80	9.17	0.0002	0.0000	0.0187
90	8.58	0.0002	0.0000	0.0162
100	8.03	0.0002	0.0000	0.0171

Table 2 Percentage of exotic haploblocks that outperform corresponding elite haploblocks (averaged over 10 repeats) for year 1 to 100 of future breeding and mean, min and max excess performance of these haploblocks when haploblock length is 50 QTLs per bin (10 cM).

Year	Percentage	Mean	Min	Max
1	22.70	0.0023	0.0000	0.0384
2	13.35	0.0013	0.0000	0.0335
3	8.90	0.0008	0.0000	0.0306
4	5.30	0.0005	0.0000	0.0250
5	4.65	0.0003	0.0000	0.0195
6	3.55	0.0003	0.0000	0.0185
7	3.25	0.0003	0.0000	0.0230
8	2.05	0.0002	0.0000	0.0229
9	1.75	0.0001	0.0000	0.0145
10	1.30	0.0001	0.0000	0.0125
20	0.35	0.0000	0.0000	0.0055
30	0.10	0.0000	0.0000	0.0055
40	0.00	-	-	-
50	0.00	-	-	-
60	0.00	-	-	-
70	0.00	-	-	-
80	0.00	-	-	-
90	0.00	-	-	-
100	0.00	-	-	-

Figures

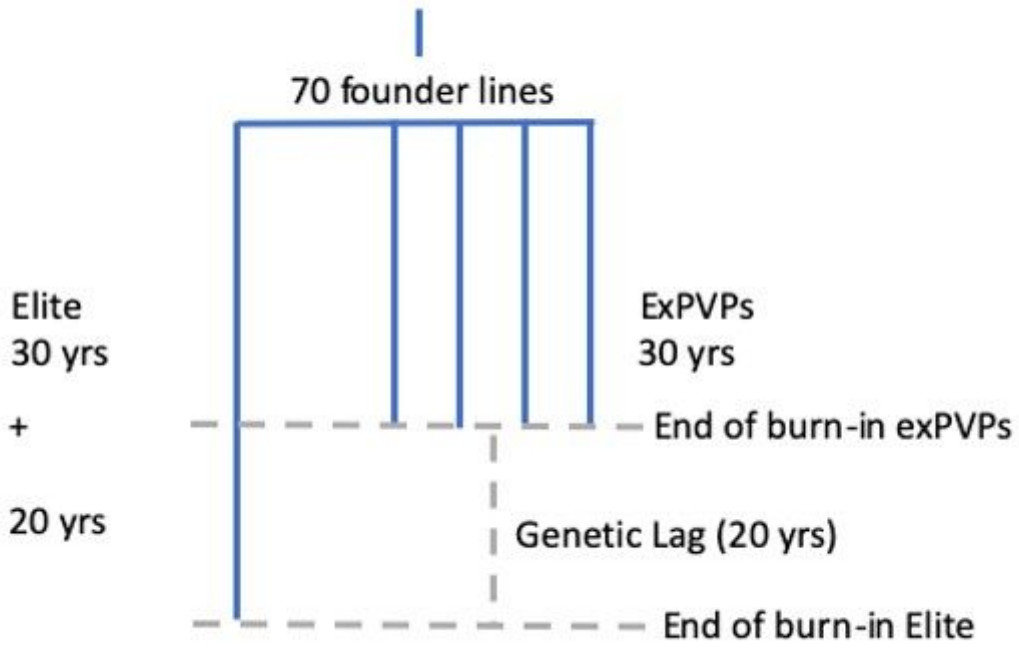


Figure 1

Founder population formation and burn-in for exPVP and elite breeding programs

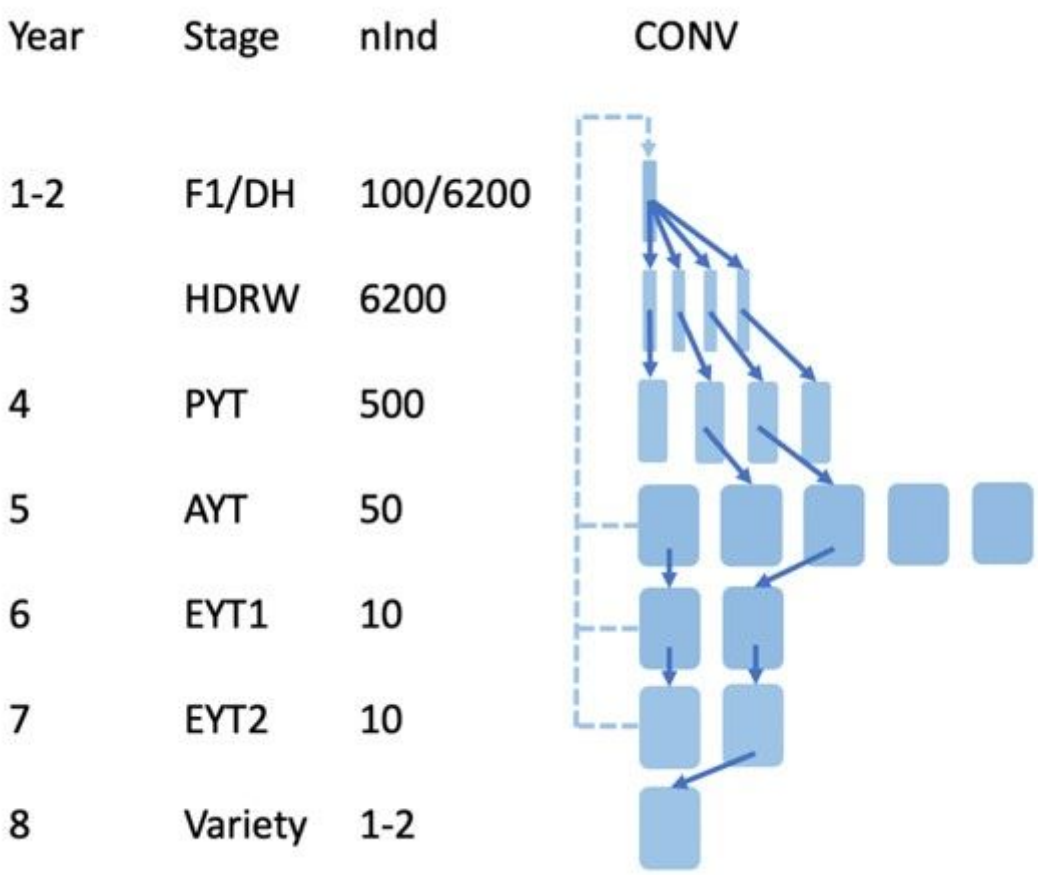


Figure 2

Structure of a conventional (CONV) breeding program used for the burn-in for ExPVP and Elite breeding program. ExPVP breeding programs were continued this way for future breeding. DH= doubled haploids; HDRW=headrows; PYT=preliminary yield trial; AYT=advanced yield trial; EYT1=elite yield trial 1; EYT2=elite yield trial 2, nInd=number of individuals in each stage

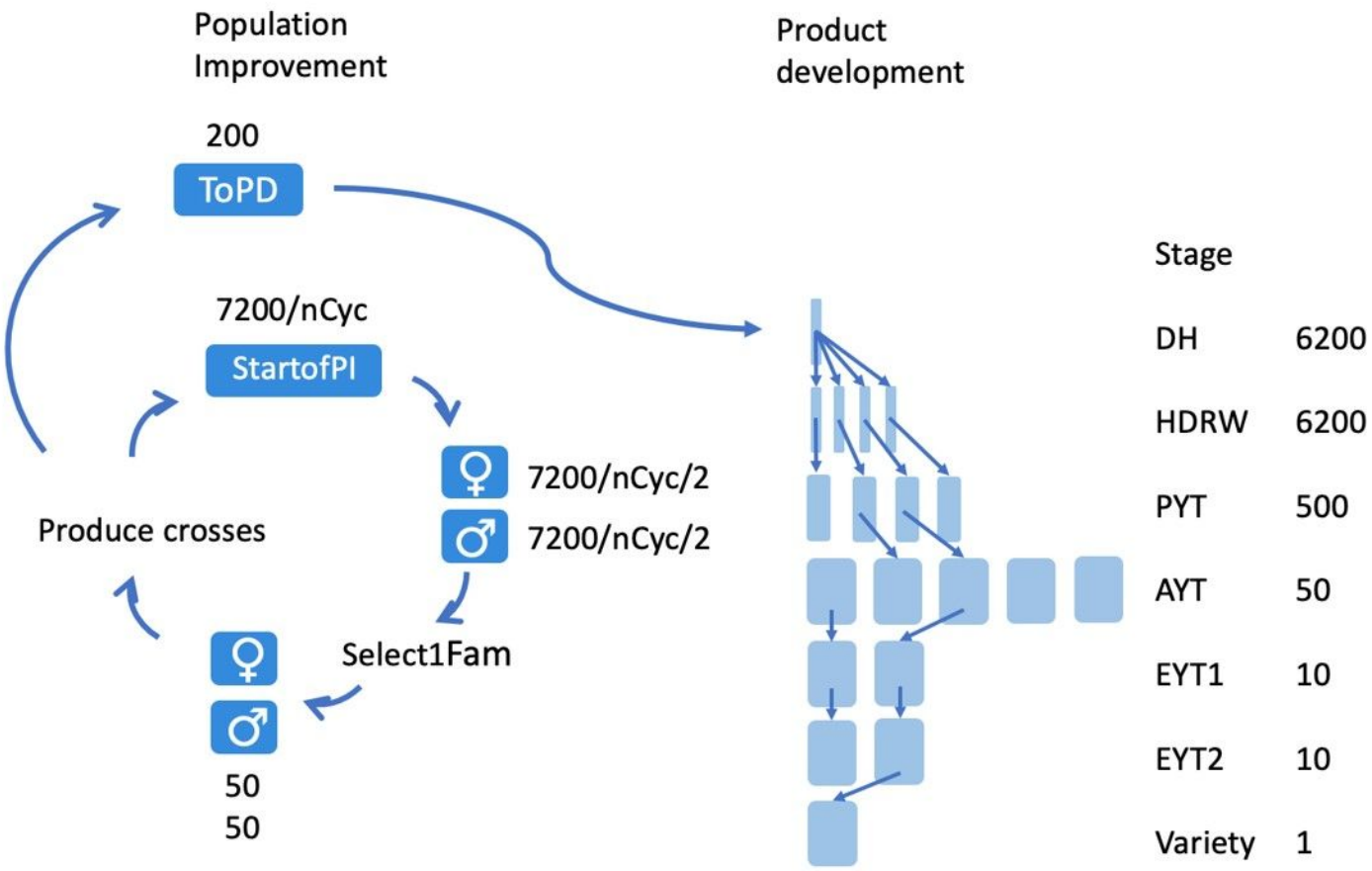


Figure 3

Structure of the two-part breeding strategy used as a baseline without introgression. Genotyping takes place at the start of the population improvement (StartofPI). Select1Fam=genomic selection of 50 parents, selection is limited to 1 individual per female half-sib family; nCyc=number of breeding cycles per year; ToPD=germplasm moving to product development, DH= doubled haploids; HDRW=headrows; PYT=preliminary yield trial; AYT=advanced yield trial; EYT1=elite yield trial 1; EYT2=elite yield trial 2

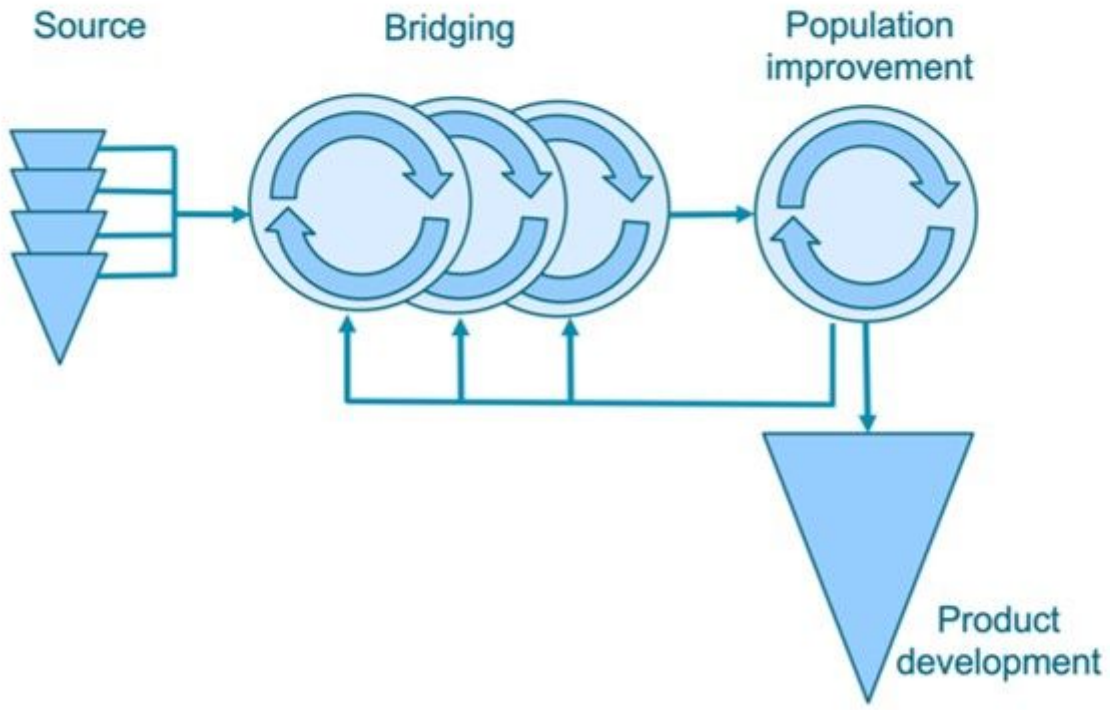
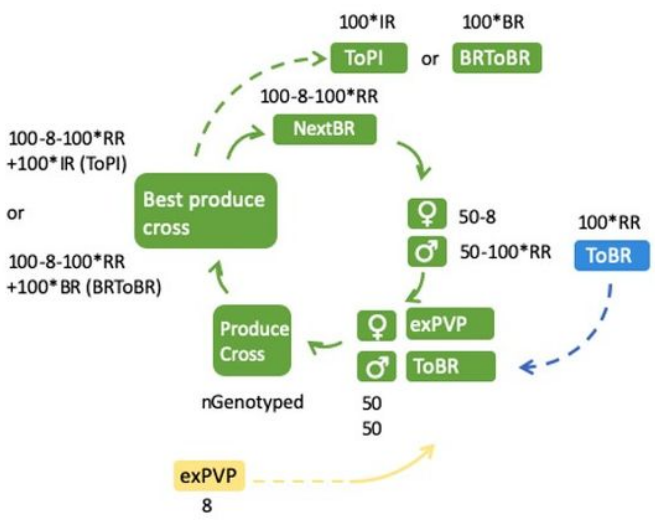


Figure 4

Structure of the multi-part breeding strategy; 4 exPVP conventional breeding programs were used as the source for exotic germplasm, the aim of the bridges used in the pre-breeding component was to break down unfavourable linkage and reduce the performance gap between the exPVP breeding programs and the elite breeding program, the aim of the population improvement component was to rapidly improve the performance of the potential parents for the product development component and the aim of the product development component was to obtain new varieties to be released. Solid lines represent germplasm flow



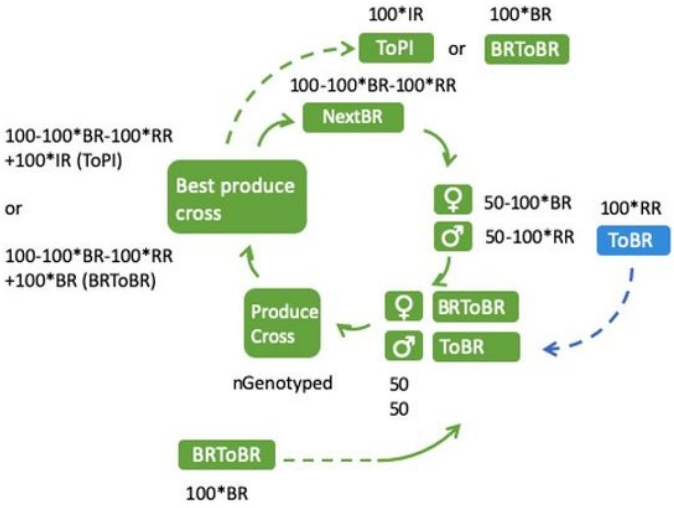


Fig. 5a, **Fig. 5b**

Figure 5

Fig.5a Structure of the first pre-breeding bridge, **Fig. 5b** structure of consecutive pre-breeding bridge(s). NextBR=start of cycle; ToBR=germplasm originating from the population improvement component passed back to the pre-breeding component; exPVP=exotic germplasm to be introgressed, nGenotyped=number of individuals genotyped each year; ToPI=germplasm to be passed on to the population improvement component; BRToBR=germplasm to be passed on to the next bridge if a next bridge exists, or received from the previous bridge if a previous bridge exists; BR=bridging rate; IR=introgression rate; RR=return rate

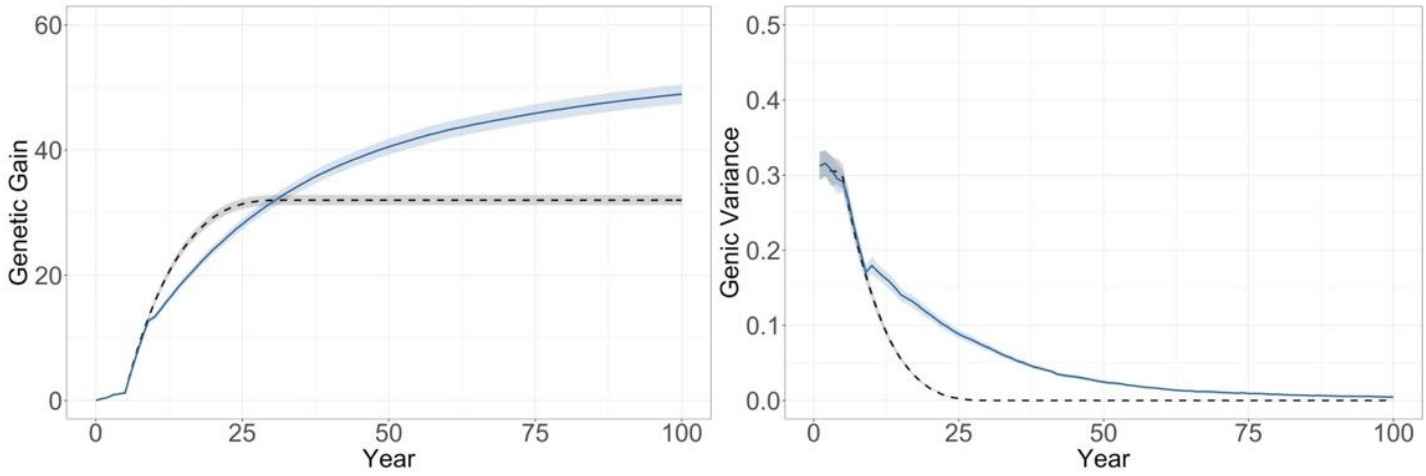


Figure 6

Change in genetic gain (left) and genic variance (right) measured at the EYT1 phase of the product development component over 100 years of forward breeding for the optimised multi-part breeding strategy (blue, solid line) and two-part breeding strategy (black, dashed line). Shaded area represents 95%-confidence interval.

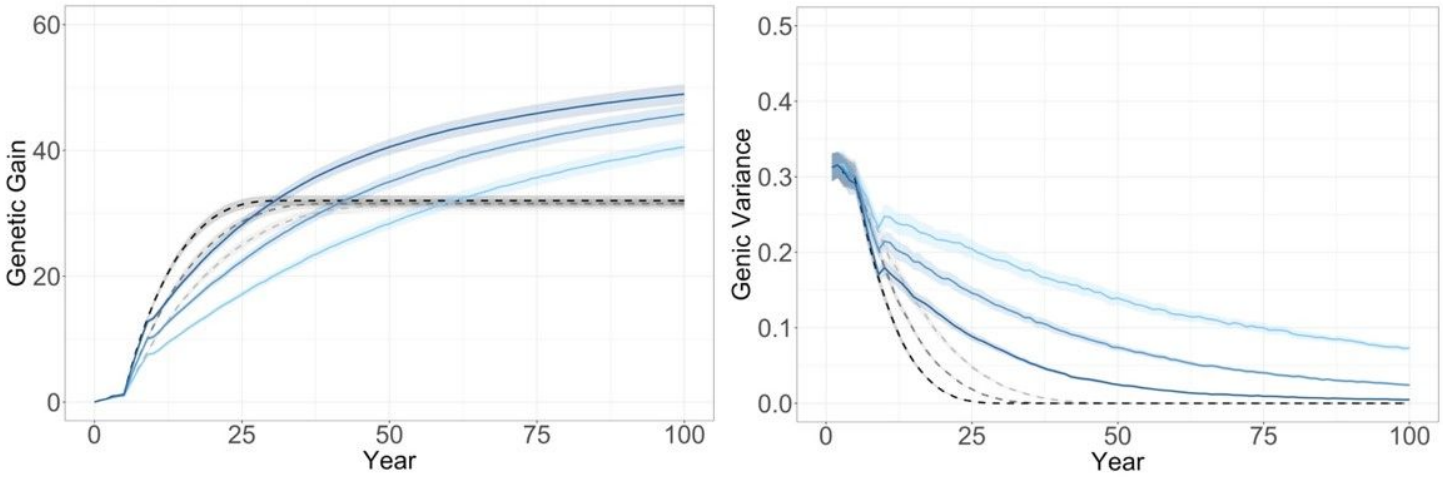


Figure 7

Change in genetic gain (left) and genic variance (right) over 100 years of forward breeding measured at the EYT1 phase of the product development component for the optimised multi-part breeding strategy with GS accuracy of 1 (dark blue, solid line), accuracy of 0.7 (middle blue, solid line), accuracy of 0.5 (light blue, solid line) and two-part breeding strategy with GS accuracy of 1 (black, dashed line), accuracy of 0.7 (dark-grey, dashed line) and accuracy of 0.5 (light-grey, dashed line). Shaded area represents 95%-confidence interval

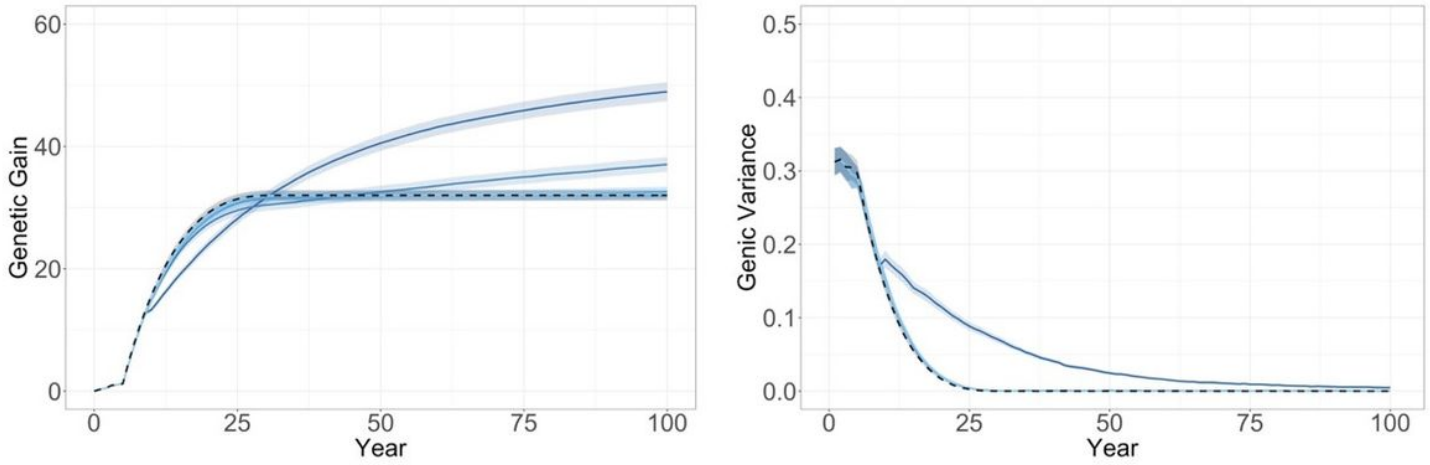


Figure 8

Change in genetic gain (left) and genic variance (right) over 100 years of forward breeding measured at the EYT1 phase of the product development component for the optimised multi-part breeding strategy varying the introgression rate from 0.1 (light blue, solid line) to 0.5 (dark blue, solid line) and two-part breeding strategy (black, dashed line, under light blue line from year 27 onwards). Shaded area represents 95%-confidence interval

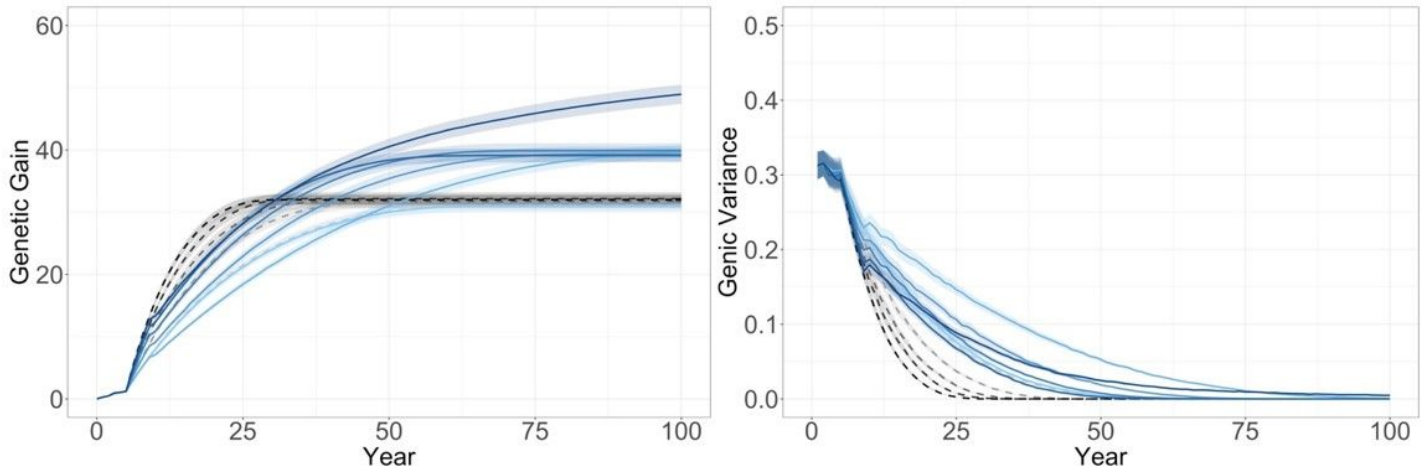


Figure 9

Change in genetic gain (left) and genic variance (right) over 100 years of forward breeding measured at the EYT1 phase of the product development component for the optimised multi-part breeding strategy with 1 (light blue, solid line) to 6 (dark blue, solid line) breeding cycles per year and two-part breeding strategy with 2 (light-grey, dashed line) to 6 (black, dashed line) breeding cycles per year. Shaded area represents 95%-confidence interval

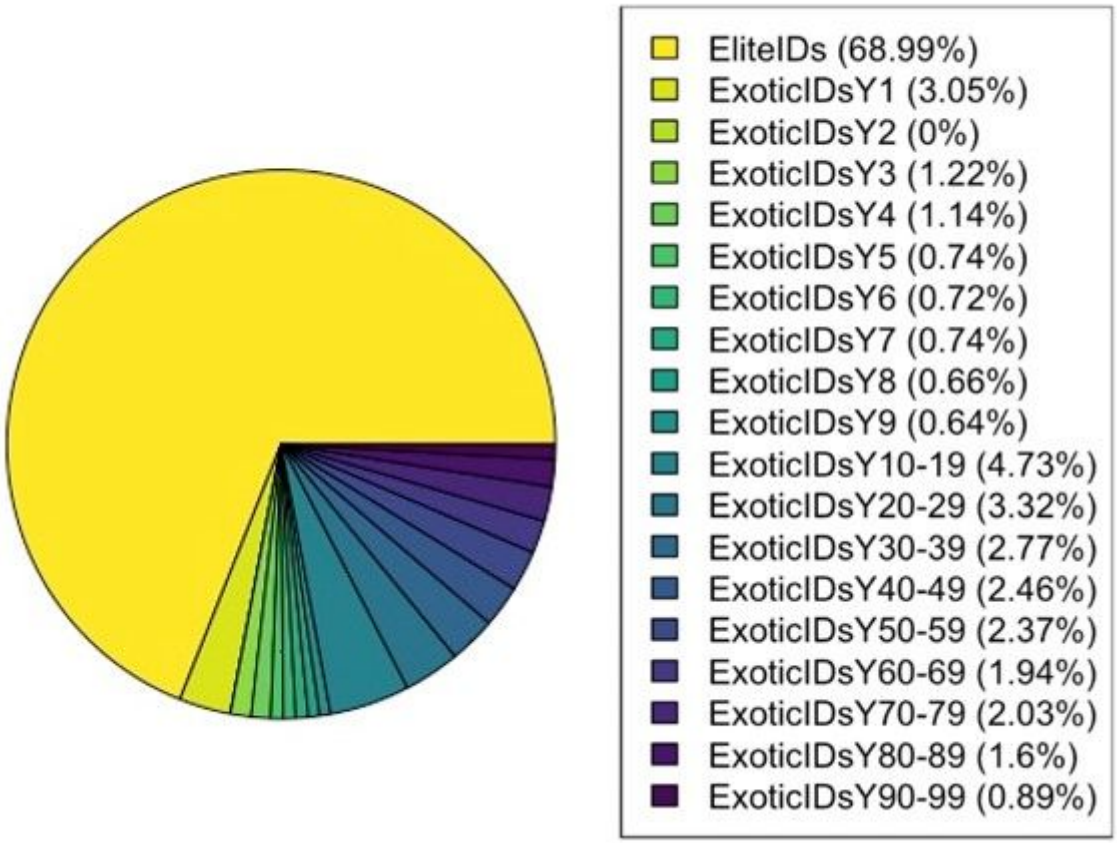


Figure 10

Percentage distribution of origin of haploblocks present in the population improvement component of the multi-part breeding strategy at year 100 of forward breeding