Euphytica (2022) 218:109
https://doi.org/10.1007/s10681-022-03065-1

Adaptive scoping: balancing short- and long-term genetic gain in plant breeding

David Vanavermaete · Jan Fostier · Steven Maenhout · Bernard De Baets

Received: 31 August 2021 / Accepted: 26 June 2022

Abstract Truncation selection is often used to rapidly achieve short-term genetic gain within a breeding program. Unfortunately, it is also associated with the loss of favorable QTL alleles in the breeding population, causing a premature convergence to sub-optimal genetic values. Parental selection strategies such as the scoping method, the population merit method, and optimal cross selection have been proposed to preserve genetic variation in the breeding population and thus maximize genetic gain in the long term. Nevertheless, for economic reasons, breeders are often interested to maximize the genetic gain in a shorter time frame. We propose a new selection strategy, named the adaptive scoping method, that aims at maximizing the genetic gain within a specific, predefined time frame. Throughout this time frame, the adaptive scoping method progressively changes its selection strategy: during the initial breeding cycles, it attempts to maximally preserve genetic variation, whereas in later breeding cycles, it prioritizes the increase of the genetic value. We demonstrate through simulation studies that the adaptive scoping method is able to maximize the genetic gain for a wide range of time frames and that it outperforms the original scoping method, both in the short and in the long term.

Keywords Genetic gain · Genetic selection · Genetic variation · Genetic value · Scoping · Plant breeding · Simulation study

D. Vanavermaete (✉) · B. De Baets
KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, B-9000 Ghent, Belgium
e-mail: david.vanavermaete@ugent.be

J. Fostier
IDLab, Department of Information Technology, Ghent University - imec, B-9052 Ghent, Belgium

S. Maenhout
Predictive Breeding, Department of Plants and Crops, Ghent University, B-9000 Ghent, Belgium

Introduction

From an economic point of view, breeders aim to maximize the genetic gain as quickly as possible. To this end, they often resort to the use of truncation selection: every generation, individuals that rank highest according to certain traits of interest are selected for breeding. When such parental selection is guided by phenotypic information, the reduction in genetic variation is limited (Piepho et al. 2008). However, when truncation selection is based on genotypic data (i.e., genomic selection), a rapid fixation of large-effect quantitative trait loci (QTL) has been observed (Clark et al. 2011; Pszczola et al. 2012; Jannink 2010). The loss of favorable QTL alleles in the breeding population reduces the maximum reachable genetic value, ultimately resulting in a premature convergence to a sub-optimal genetic value. Therefore, a successful breeding program should find a balance between genetic gain on the one hand and the

Published online: 22 July 2022

preservation of genetic variation in the breeding population on the other hand (Jannink 2010; Meuwissen 1997; Woolliams et al. 2015).

Different methods have been proposed in literature to maximize long-term genetic gain by controlling the average inbreeding coefficient of a population (Wray and Goddard 1994; Brisbane and Gibson 1995; Meuwissen 1997). The inbreeding coefficient of a diploid individual is the probability that, at any given locus, the two alleles are identical by descent (i.e., originate from the same ancestor). It is important to manage the rate at which the average inbreeding coefficient changes between consecutive breeding cycles (Woolliams et al. 2015). A high rate of inbreeding results in a quick, short-term genetic gain but a rapid loss of genetic variation, whereas a low rate of inbreeding yields a better preservation of the genetic variation at the expense of a slower genetic progress.

Different methods to predict and control the rate of inbreeding have been proposed. Wray and Thompson (1990) propose the use of the long-term genetic contribution metric, i.e., the proportion of the genes of an individual that will be passed to its descendants in the long term. Meuwissen (1997) proposed to limit the rate of inbreeding by restricting the coancestry between parents using the optimal contribution selection method. In similar approaches, the rate of inbreeding was controlled by ensuring a sufficient genetic distance between parents, thus limiting within-family selection (Sonesson et al. 2012; Allier et al. 2020b). Gorjanc et al. (2018) use the optimal cross selection method in a two-part breeding program to maximize the long-term genetic gain. Akdemir and Sánchez (2016) propose an optimal mating plan, taking into account the risk of inbreeding and the desired level of allele heterozygosity. Because it is not computationally feasible to enumerate and evaluate all possible combinations of parent individuals, optimization techniques such as genetic algorithms are often used to find a good parental population (Allier et al. 2020b; Gorjanc et al. 2018). However, such techniques tend to converge to local optima, which implies that the optimal parental population may not be found.

Vanavermaete et al. (2020) propose the scoping method as an alternative strategy to preserve the genetic variation in a breeding population and thus maximize the long-term genetic gain. The selection of parental individuals is performed in a computationally efficient manner and consists of two steps: pre-selection followed by actual parental selection. First, a certain fraction of the individuals with the highest genomic estimated breeding values (GEBVs) are pre-selected from the breeding population. This fraction is referred to as the scoping rate (SR). A low scoping rate results in the pre-selection of a small set of individuals with only the highest GEBVs, whereas a high scoping rate yields a larger, more diverse set of candidate parents. From this set, parents are selected and coupled aiming for genetic progress as well as the preservation of genetic variation. The scoping method was demonstrated to outperform parental selection methods such as truncation selection, the population merit method (Lindgren and Mullin 1997) and the maximum variance total method (Cervantes et al. 2016), with a more pronounced superiority in the long term.

For economic reasons, breeders are often interested in the maximization of genetic gain in a short time frame. To this end, we propose a modification of the scoping method that aims at optimizing the genetic gain of a breeding population within a predefined number of breeding cycles. This method, referred to as the *adaptive* scoping method, dynamically changes the scoping rate throughout the different breeding cycles: initially, high scoping rates are considered such that the preservation of genetic variation is emphasized, whereas during later breeding cycles, increasing the genetic value is gradually prioritized through lower scoping rates. The adaptive scoping method takes only a single parameter, namely the time frame t (expressing the number of breeding cycles) during which the genetic gain should be maximized. This unique feature enables breeders to balance exploration and exploitation of their breeding population.

Materials and methods

The base population and breeding scheme are adopted from Neyhart et al. (2017). The base population is constructed from two datasets of North American barley (Hordeum *vulgare*) from the University of Minnesota (UMN) and the University of North Dakota (NDSU), counting respectively 384 and 380 six-row spring inbred lines with 1590 biallelic SNP loci. The simulation study was constructed

in a similar way as reported by Vanavermaete et al. (2020), ensuring that the performance of the adaptive scoping method can be compared with that of the original scoping method.

Breeding scheme

The recurrent breeding scheme is depicted in Fig. 1 and has been described by Vanavermaete et al. (2020) as well as by Neyhart et al. (2017). In the initial breeding cycle, 50 individuals with the highest phenotypic values of the NDSU dataset are coupled with 50 individuals with the highest phenotypic values of the UMN dataset. Each couple produces 20 offspring and after 2 generations of single-seed descent, the base population is obtained containing 1000 individuals. From this point onward, the parents are selected solely based on the genomic estimated breeding values (GEBVs) to reduce the financial cost of phenotyping. The GEBVs are predicted using a linear mixed effects model that has been fitted using the base population, which contains both phenotypic and genotypic information.

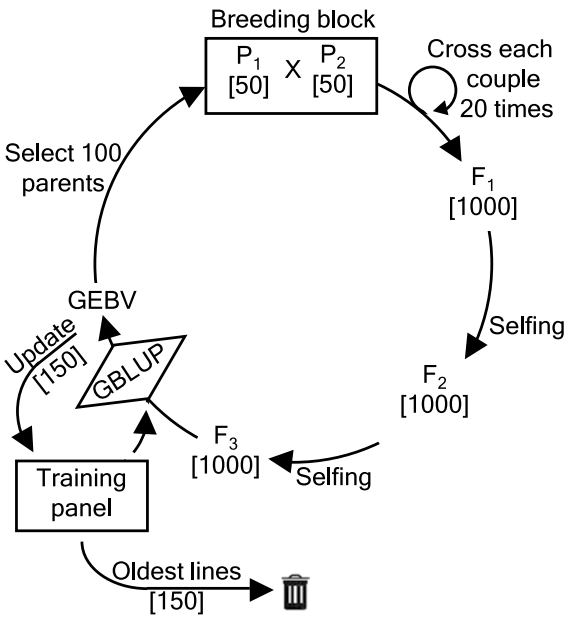


Fig. 1 Overview of the recurrent breeding scheme. First, 50 pairs of parents (P_1, P_2) each produce 20 offspring, yielding a total of 1000 F1 individuals. After two generations of single-seed descent, 1000 F3 individuals are obtained. From those F3 individuals, new parental lines are selected. In each breeding cycle, 150 individuals that have been the longest in the training panel are removed. Next, 150 individuals are phenotyped and added to the training panel according to the tails method (Neyhart et al. 2017), selecting 75 individuals with the highest GEBVs and 75 individuals with the lowest GEBVs

In each subsequent breeding cycle, 100 parents are selected and coupled according to one of the parental selection methods considered to construct a crossing block. Each couple produces 20 offspring resulting in a total of 1000 F1 individuals. After two generations of single-seed descent, 1000 F3 individuals are obtained. These individuals represent the new breeding population from which parents can again be selected. Each simulation run consists of 50 breeding cycles and all results are averaged over 100 simulation runs.

The scoping method

Vanavermaete et al. (2020) proposed the scoping method to preserve genetic variation in a breeding population and thus maximize genetic gain in the long term. The scoping method consists of two steps: pre-selection followed by parental selection. During pre-selection, individuals with the highest GEBVs are chosen. This ensures that genetic progress can be made in the next breeding cycles. The fraction of individuals that is pre-selected is controlled by the scoping rate which ranges between a minimum value SR_{min} and a maximum value SR_{max}. Here, $SR_{max} = 1$ (pre-select 100% of the individuals), while $SR_{min} = n_s/n_t$, with n_s the number of parent individuals to be selected for breeding and n_t the total population size. For example, if $n_s = 100$ parents are to be selected from a population size $n_t = 1000$, then a fraction of at least 10% ($SR_{min} = 0.1$) of the individuals must be pre-selected.

Next, from these pre-selected individuals, $n_s/2$ parental pairs are consecutively chosen as follows: the individual with the highest GEBV that has not yet been selected as a parent is selected as the P1 parent, whereas the P2 parent is selected by maximizing the F_{score} such that the genetic variation of all selected parents thus far is maximized over all markers. Mathematically, the F_{score} is given as:

$$F_{score} = \sum_{i=1}^{k} \text{var}\,(\mathbf{Z}_i) p_i\,, \tag{1}$$

with k the number of genetic markers, $\mathbf{Z}_i$ the i-th column of the $n \times k$ matrix $\mathbf{Z}$ containing the genotypes (coded as -1, 0 and 1) of the n already selected individuals and $\mathbf{p}$ a Boolean vector of length k. Initially, p_i is set to 1 for all marker positions. When both alleles

at marker i are present, p_i is set to 0. This way, the F_{score} takes into account only those markers for which both alleles are not yet present in the selected population. Once both alleles are present for all marker positions, all p_i are again set to 1 such that the variance is again maximized over all markers.

The core idea is that the P1 parents drive the genetic progress of the offspring, whereas the P2 parents ensure the preservation of genetic variation. Clearly, the scoping rate controls the trade-off between the degree in which genetic variation can be preserved on the one hand, and the rate at which genetic gain can be made on the other hand.

The adaptive scoping method

The scoping method uses a single, fixed value for the scoping rate across successive breeding cycles. In contrast, the adaptive scoping method gradually decreases the scoping rate from its maximum value $\mathrm{SR}_{\mathrm{max}}$ to its minimum value $\mathrm{SR}_{\mathrm{min}}$. The adaptive scoping method takes a single parameter t, expressed through the number of breeding cycles over which the scoping rate is varied. Specifically, at breeding cycle i, the scoping rate takes the value:

$$\mathrm{SR}(i) = \begin{cases} \dfrac{\mathrm{SR}_{\mathrm{min}} - 1}{t-1} i + \dfrac{t - \mathrm{SR}_{\mathrm{min}}}{t-1} & , \text{ if } 1 \leq i \leq t \\ \mathrm{SR}_{\mathrm{min}} & , \text{ if } i > t \end{cases} . \tag{2}$$

In other words, the scoping rate decreases linearly over t breeding cycles from $\mathrm{SR}_{\mathrm{max}} = 1$ at breeding cycle 1 to $\mathrm{SR}_{\mathrm{min}}$ at breeding cycle t (and later cycles). As a consequence, during the first breeding cycles, the adaptive scoping method pre-selects a larger number individuals, focusing on the preservation of the genetic variation (exploration). In contrast, at breeding cycles t and later, only elite individuals are pre-selected, maximizing the genetic progress (exploitation). From this set of pre-selected candidate parents, parent pairs are chosen in an identical manner as in the scoping method.

H-optimal genomic mating method

The H-optimal genomic mating (H-OGM) method was designed to compare two existing methods using the same setting as the adaptive scoping method. In other words, the existing methods are combined with a preselection using the same SR values as for the adaptive scoping method. To do so, a fraction of the individuals with the highest GEBVs is preselected. Only the preselected individuals can be used as a parent. Similar to the adaptive scoping method, the preselection is controlled by the SR allowing for both methods to use the same pool of individuals as potential parents. Next, the actual parents are selected. Individuals with the highest GEBVs are selected as P1 parents, whereas the P2 parents are selected based on the highest H-score.

The H-score is a metric proposed by Allier et al. (2020a) by which individuals are scored based on the presence of favorable haplotype segments that are not available in the elite P1 parents. A full description of the H-score was also reported by Vanavermaete et al. (2020). First, the haplotype estimated breeding value (HEBV) is calculated for each individual. In this simulation study, the genotype is decomposed into different haploid segments using a window size of 20 and a step size of five. A matrix $\mathbf{M}$ of size $k \times n_H$, with k the number of markers and n_H the number of haplotype segments, is constructed to keep track of the selected markers per haplotype segment, such that $M_{ij} = 1$ if marker i is part of the j-th haplotype segment and $M_{ij} = 0$ otherwise. Mathematically, the HEBV matrix $\mathbf{H}$ can be written as:

$$\mathbf{H} = (\mathbf{X} \circ \mathbf{1}_{2n} \boldsymbol{\beta}^T)\mathbf{M}\,, \tag{3}$$

with $\mathbf{X}$ a matrix of size $2n \times k$ containing the haplotype of n different individuals and k different markers coded as 0 and 1 (such that the haplotype of individual i is represented at rows $2i-1$ and $2i$), $\circ$ the Hadamard product operator, $\mathbf{1}_{2n}$ a vector of size $2n$ containing 1s and $\boldsymbol{\beta}$ a vector of size k with estimated marker effects. The HEBV can then be calculated as:

$$\mathrm{HEBV}(i,j) = \lambda \sum_{h=1}^{n_H} \max\left(\mathbf{H}_{2i-1,h}, \mathbf{H}_{2i,h}, \mathbf{H}_{2j-1,h}, \mathbf{H}_{2j,h}\right), \tag{4}$$

with λ a scaling parameter defined as the ratio between the step size and the window size. Finally, the H-score of individual i can be calculated as:

$$H(i) = \max_{j \in \mathrm{P1}} \mathrm{HEBV}(i,j)\,. \tag{5}$$

The crossing block is designed using the optimal genomic mating method proposed by Akdemir and

Sánchez (2016). This method includes information on the complementarity of parents, allowing for a better crossing design in the long term. The OGM method was performed using the getGaSolutionsFrontier function of the R package `GenomicMating` using default values (Akdemir and Sánchez 2016).

Prediction model

The GEBVs are predicted by fitting a linear mixed effects model:

$$\mathbf{y} = \mathbf{1}_n\beta + \mathbf{Zu} + \boldsymbol{\epsilon}\,, \tag{6}$$

with $\mathbf{y}$ a vector of phenotypic values, $\mathbf{1}_n$ a vector of size n containing ones, n the number of individuals in the training panel, β the fixed effect (phenotypic mean), $\mathbf{Z}$ the incidence matrix of the training panel with marker information, $\mathbf{u}$ the marker effects following a normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{G})$ with $\mathbf{G} = \sigma_u^2\mathbf{I}_k$ (with $\mathbf{I}_k$ the identity matrix of dimension k), k the number of markers and $\boldsymbol{\epsilon}$ the residual effects following a normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{R})$ with $\mathbf{R} = \sigma_e^2\mathbf{I}_n$. Both variance components σ_u^2 and σ_e^2 are estimated by means of restricted maximum likelihood (REML). The GEBVs of the individuals are calculated as:

$$\hat{\mathbf{g}} = \mathbf{Z}\hat{\mathbf{u}}\,, \tag{7}$$

with $\hat{\mathbf{g}}$ the GEBVs, $\mathbf{Z}$ the marker information and $\hat{\mathbf{u}}$ the predicted marker effects.

At the start of the simulation study, both the UMN and NDSU datasets are used as training population. In the subsequent breeding cycles, 150 new individuals are phenotyped and added to the training panel according to the tails method, selecting 75 individuals with the highest GEBVs and 75 individuals with the lowest GEBVs (Neyhart et al. 2017). According to Neyhart et al. (2017), this results in a (non-significantly) higher genetic gain compared to other update methods. Before updating the training panel, 150 individuals that have been the longest in the training panel are removed from the training population. This reduces the computational time without reducing the prediction accuracy (Neyhart et al. 2017).

The linear mixed effects model in Eq. (6) is fitted using the package `rrBLUP` in R (Endelman 2011). Even though it has been recommended to remove markers with low levels of polymorphism from the training panel (Chang et al. 2018), we kept all markers as this resulted in a higher prediction accuracy.

Simulation of the population

The simulation study was built upon the work of Neyhart et al. (2017), using the packages `GSSimTPUpdate` and `hypred` in R (version 3.6.3). First, the genome of barley is constructed based on marker position, allele, and chromosomal information. One hundred QTL ($L = 100$) are selected randomly from the available 1590 biallelic SNP loci. The remaining 1490 biallelic SNP loci are available as markers for prediction and selection purposes. The QTL effects are calculated according to a geometric series. At the k-th QTL, the favorable homozygote will have a value a^k, the heterozygote a value zero, and the unfavorable homozygote a value $-a^k$ with $a = (L-1)/(L+1)$. Dominance and epistatic effects were assumed to be absent. The phenotypic value is calculated over three different environments, each drawn from a normal distribution with mean 0 and a variance component σ_E^2 which is defined as eight times the genetic variance (Bernardo 2014). The phenotypic value of the i-th individual in the j-th environment (y_{ij}) is calculated as follows:

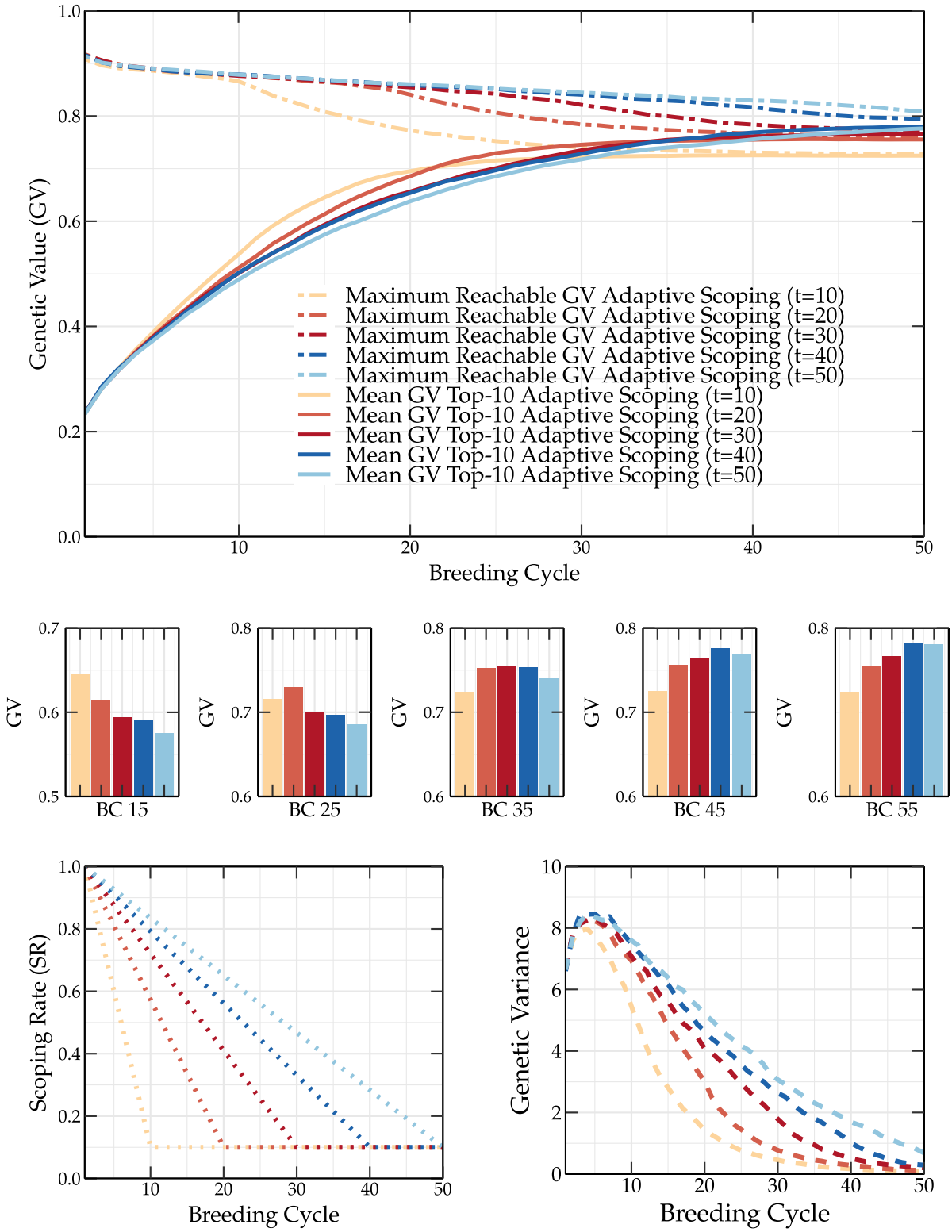
$$y_{ij} = g_i + e_j + \epsilon_{ij}\,, \tag{8}$$

with g_i the genetic value of the i-th individual, e_j the j-th environmental effect and ϵ_{ij} the residual effect of the i-th individual and the j-th environment. The residual effect is drawn from a normal distribution with mean 0 and a variance component σ_R^2, with σ_R^2 scaled to simulate a population with a heritability (h^2) of 0.5. The phenotypic value of an individual is the averaged value over the three environments. A comprehensive overview of the simulation study has been described by Vanavermaete et al. (2020).

To track the inbreeding coefficient of the population, the genetic relationship matrix $\mathbf{G}$ is required, which is calculated as follows (VanRaden 2008):

$$\mathbf{G} = \frac{\mathbf{MM'}}{2\sum_{i=1}^{k} P_i(1-P_i)}\,, \tag{9}$$

with $\mathbf{M}$ a matrix of size $n \times k$ of which each column is calculated as $\mathbf{Z}_i - \mathbf{1}_n[2(P_i - 0.5)]$, n the number of individuals in the breeding population, $\mathbf{Z}_i$ the genotype of n individuals at the i-th marker, $\mathbf{1}_n$ a vector


Genetic Value (GV)
Breeding Cycle
Maximum Reachable GV Adaptive Scoping (t=10)
Maximum Reachable GV Adaptive Scoping (t=20)
Maximum Reachable GV Adaptive Scoping (t=30)
Maximum Reachable GV Adaptive Scoping (t=40)
Maximum Reachable GV Adaptive Scoping (t=50)
Mean GV Top-10 Adaptive Scoping (t=10)
Mean GV Top-10 Adaptive Scoping (t=20)
Mean GV Top-10 Adaptive Scoping (t=30)
Mean GV Top-10 Adaptive Scoping (t=40)
Mean GV Top-10 Adaptive Scoping (t=50)
GV
BC 15
BC 25
BC 35
BC 45
BC 55
Scoping Rate (SR)
Genetic Variance

◂**Fig. 2** Top panel: the mean genetic value of the top-10 individuals and the maximum reachable genetic value for the adaptive scoping method with a value for t of respectively 10, 20, 30, 40 and 50 breeding cycles. Middle panels: the mean genetic values for the different parental selection methods at breeding cycles 15, 25, 35, 45 and 55. Bottom left panel: the SR for the different parental selection methods. Bottom right panel: the genetic variation for the different parental selection methods

of size n containing ones, k the number of markers, and P_i the frequency of the second allele at the i-th marker. Next, the averaged inbreeding coefficient F can be calculated as:

$$F = \frac{1}{n}\sum_{i=1}^{n} G_{ii} - 1\,, \tag{10}$$

with G_{ii} the diagonal elements of the genetic relationship matrix. To track the fixation of unfavorable QTL alleles, the maximum reachable genetic value is calculating as the sum of the QTL effects that are fixed (both favorable and unfavorable) and the sum of the favorable QTL effects that are not yet fixed. It represents the maximum genetic value that could still be reached, taking into account the fixation of unfavorable QTL alleles. The maximum reachable genetic value and the mean genetic value are rescaled such that the maximum reachable genetic value has a value of 1. As in Vanavermaete et al. (2020), the mean genetic value of the top-10 individuals is reported. These individuals represent the superior lines that are prime candidates for commercialization.

Data availability

The scripts, figures, datasets of the base population and supplementary data are available from the GitHub repository https://github.com/biointec/AdaptiveScoping. The dataset and the simulation of the recurrent breeding scheme have been reported and published by Neyhart et al. (2017).

Results

Performance of the adaptive scoping method

The adaptive scoping method is designed to maximize the preservation of genetic variation in the first breeding cycles and maximize the genetic progress in later cycles. This is achieved by linearly decreasing the scoping rate over the course of t breeding cycles. Fig. 2 shows the mean genetic value of the top-10 individuals of the population for different values of $t = 10$, 20, 30, 40 and 50 breeding cycles. The mean genetic value of the top-10 individuals and the maximum reachable genetic value of the adaptive scoping method and the scoping method are also reported in the supplementary material in Tables S1 and S2, respectively. For a value of $t = 10$, the preservation of variation is quickly traded off for a rapid genetic gain: from breeding cycle 10 onward, only the individuals with the highest GEBVs are considered as parents. At breeding cycle 15, this manifests itself in an at least 4 percentage points higher mean genetic value compared with the other values of t. Nevertheless, due to the rapid reduction of genetic variation, the adaptive scoping method with $t = 10$ quickly loses its ability to drive genetic progress further, yielding the worst genetic values beyond about 30 breeding cycles. The adaptive scoping method with $t = 20$ preserves the genetic variation somewhat longer before focusing on genetic gain. Around breeding cycle 25, this yields the highest mean genetic value compared with other values of t. Again, this advantage quickly degrades throughout later breeding cycles.

In general, the behavior of the adaptive scoping method can be understood as follows: the higher the value of t, the longer genetic variation is preserved before genetic gain is prioritized. As soon as t breeding cycles are completed, the adaptive scoping method resembles the behavior of truncation selection (although the pairing of parents is not random). Therefore, the adaptive scoping method will yield the highest genetic gains shortly after t breeding cycles. This is indicated in the middle panels of Fig. 2, where the results for different values of t are compared at different breeding cycles: the adaptive scoping method with $t = 10$ yields the highest gain at cycle 15, the adaptive scoping method with $t = 20$ yields the highest gain at cycle 25, etc. The only exception is the adaptive scoping method with $t = 50$. In that case, the adaptive scoping method is not able to outperform the adaptive scoping method with $t = 40$ at breeding cycle 55 but converges to the same value (see Table S1). Clearly, by choosing a particular value for t, a breeder can expect the highest gains during the breeding cycles that immediate

follow t, outperforming the adaptive scoping method with different values of t.

We also compare the adaptive scoping method with the original scoping method for a fixed scoping rate across the breeding cycles. In Vanavermaete et al. (2020), a scoping rate of 0.3 was suggested to maximize the genetic gain in the short as well as in the long term and is hence also used here. Compared with the scoping method, the adaptive scoping method uses a higher scoping rate and hence pre-selects more individuals during the first breeding cycles, allowing for a better preservation of the available genetic variation. As a consequence, during those initial breeding cycles, the adaptive scoping method suffers less from the loss of favorable QTL alleles and hence preserves a higher maximum reachable genetic value at the expense of a lower mean genetic value of its top-10 individuals (see Fig. 3).

The short-term sacrifice in genetic gain pays off in the long term. The adaptive scoping method with $t = 10$ outperforms the scoping method after 13 breeding cycles and yields higher genetic values up to breeding cycle 20. At that point, the adaptive scoping method has exploited the remaining genetic variation, quickly leading to the convergence of the genetic value from that point onward. Similarly, the adaptive scoping method with $t = 20$ outperforms the scoping method after 20 breeding cycles. Finally, the adaptive scoping method with $t = 50$ surpasses the scoping method at breeding cycle 30 and yields the highest long-term gain (a 4 percentage point increase compared to the scoping method at breeding cycle 50). At that point, the genetic value of the adaptive scoping method is even higher than the maximum reachable genetic value of the scoping method. This means that the loss of favorable QTL alleles from the population during the initial breeding cycles of the scoping method has caused an insurmountable disadvantage in the long term. We conclude that the adaptive scoping method is able to outperform the original scoping method, both in the short term (when low values of t are used) and in the long term (for high values of t).

We have also compared the adaptive scoping method to the H-OGM method, which combines the H-score and the OGM method using the same setting as the adaptive scoping method (Allier et al. 2020a; Akdemir and Sánchez 2016). The mean genetic value of the top-10 individuals and the maximum reachable genetic value of the H-OGM method are reported in the supplementary material in Tables S3 and S4, respectively. In the short term, the H-OGM method results in a lower genetic gain compared to the (adaptive) scoping method (see Fig. 4). When t increases, the short-term genetic gain of the H-OGM method decreases. However, in contrast to the adaptive scoping method, the genetic gain is not optimized after t breeding cycles. For each t value, the genetic value converges to almost the same value in the long term. It is clear that adaptive scoping is able to improve upon the scoping method, but cannot directly be used

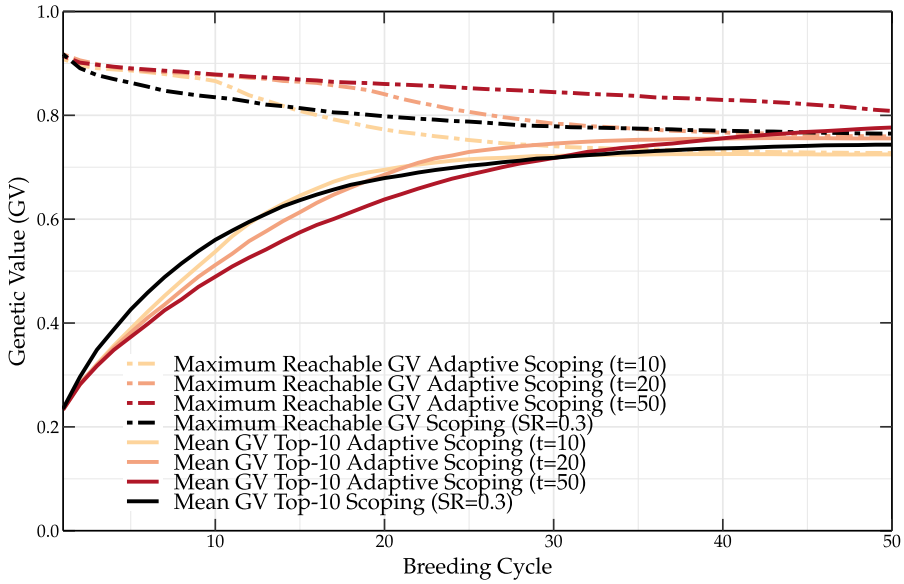


Fig. 3 The mean genetic values of the top-10 individuals using the scoping method (SR = 0.3) and the adaptive scoping method for $t = 10$, $t = 20$ and $t = 50$

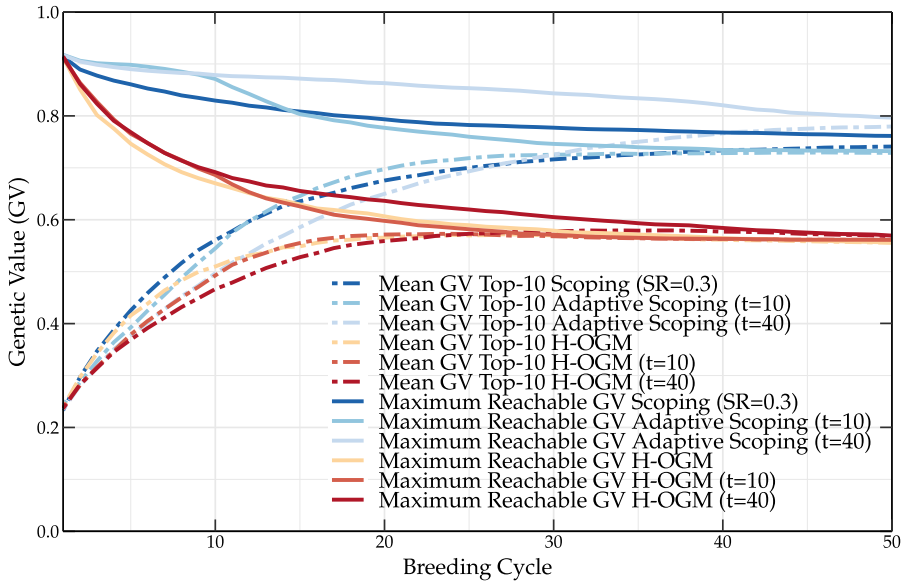


Fig. 4 The mean genetic value of the top-10 individuals and the maximum reachable genetic value for the scoping method (SR = 0.3), the adaptive scoping method ($t = 10$ and $t = 40$) and the H-OGM method (SR = 0.3, $t = 10$ and $t = 40$)

to improve other methods like the H-OGM method within a given time frame. At breeding cycle 50, a 17 and 23 percentage points higher genetic value is observed for the adaptive scoping method compared to the H-OGM method for a t value of 10 and 40, respectively.

Robustness of the adaptive scoping method

The original scoping method and the adaptive scoping method have been evaluated in different simulation settings. In each experiment, these methods were assessed using 100 different genomes such that the effects of different QTL and marker positions are averaged. The effects of the heritability and the number of QTL on the genetic gain using both methods have also been tested: simulation studies were performed using a heritability of 0.2 and 0.8 using 100 QTL, and a heritability of 0.5 using 50 and 200 QTL (see Fig. 5).

In each case, shortly after t breeding cycles, the adaptive scoping method resulted in the highest genetic value throughout a certain number of breeding cycles. For $t = 50$, the adaptive scoping method always yielded the highest long-term genetic gain. Increasing the heritability improves the prediction accuracy, resulting in higher genetic gains for all methods. Similarly, the GEBVs can be more accurately predicted when fewer QTL are present. For lower values of the heritability, the effect of the environment becomes more pronounced, making it more challenging to accurately select the best parents (based on the GEBVs). As the adaptive scoping method is better at preserving the genetic variation in the first breeding cycles, a slower but more accurate fixation of the QTL alleles is observed, resulting in higher long-term genetic gains compared to the scoping method.

Discussion

The effect of a variable scoping rate on the genetic gain

The loss in genetic variation and the resulting risk associated with truncation selection are well known (Jannink 2010; Meuwissen et al. 2001; Vanavermaete et al. 2020). In response to this, parental selection techniques such as the scoping method were developed to better preserve genetic variation and thus maximize genetic gain in the long term. To achieve this, individuals with a lower GEBV can also be considered as a parent when they contribute to the genetic diversity. To avoid an adverse effect on the rate of genetic progress, a fraction of individuals with the highest GEBVs in the breeding population is preselected. This fraction is controlled by the scoping rate. The (original) scoping method relied on a fixed scoping rate throughout the different breeding cycles.

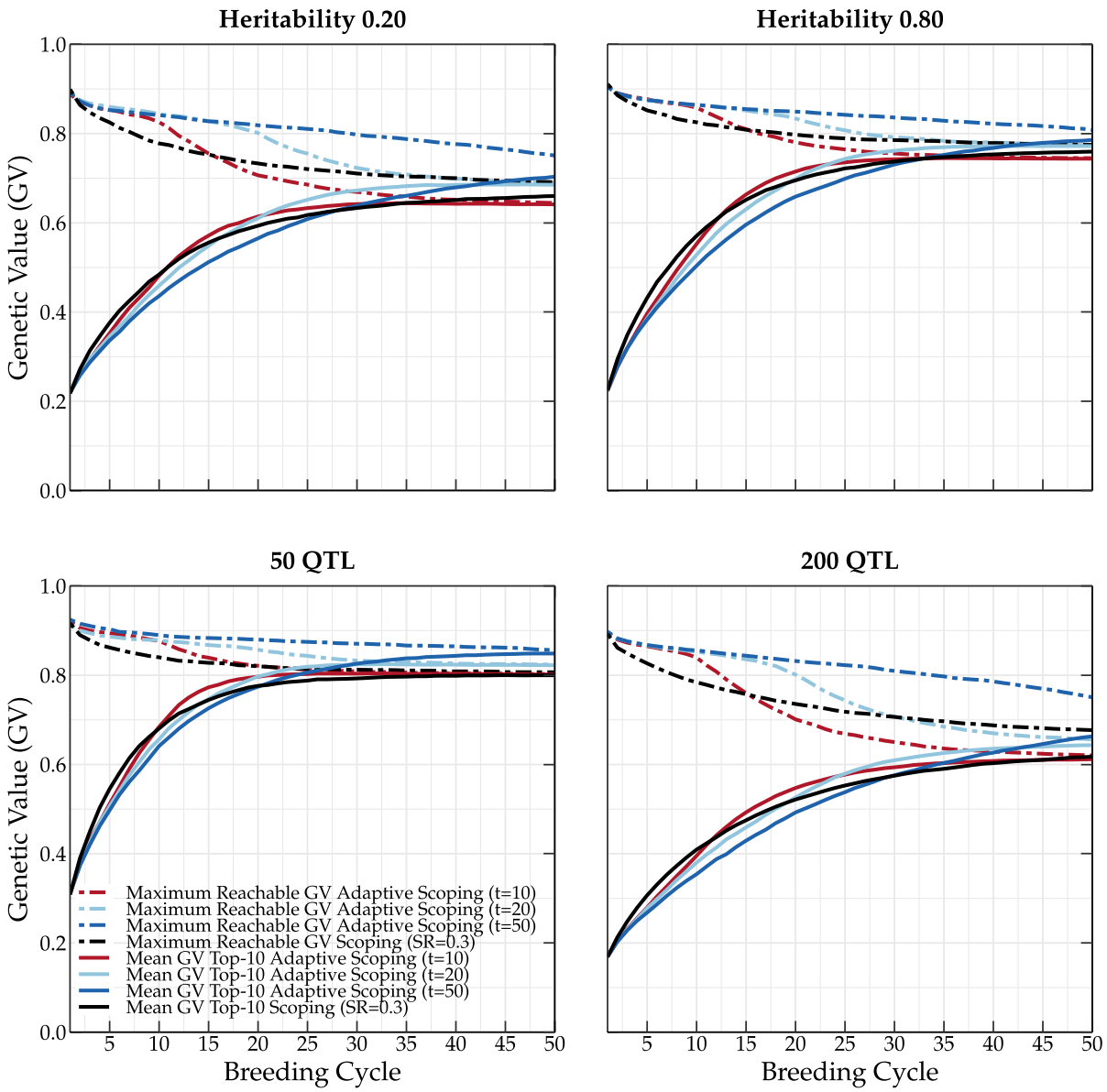


Fig. 5 Simulation results of the original and adaptive scoping methods (using $t = 10$, 20 and 50) for a heritability of 0.2 and 0.8 using 100 QTL (top) and for a heritability of 0.5 using 50 and 200 QTL (bottom). The impact of both methods on the genetic value and on the maximum reachable genetic value is reported

For a scoping rate of 0.3, the maximum reachable genetic value decreases significantly in the first breeding cycles, indicating that several favorable QTL alleles are lost from the population at the early stages of the breeding program. In principle, this could be avoided by increasing the scoping rate (and thus preselecting more parents), but this would unavoidably slow down the genetic progress and, hence, require a very large number of breeding cycles to outperform the truncation selection method.

By introducing a variable scoping rate, the trade-off between genetic gain and genetic variation can be controlled during the breeding process itself. In the first breeding cycles, a high value for the scoping rate prevents the loss of favorable QTL alleles. The scoping rate is decreased linearly, gradually prioritizing

genetic progress over preserving genetic variation. This leads to a slower, but more accurate fixation of the QTL alleles, translating into lower short-term, but higher long-term genetic gains. The parameter t represents the number of breeding cycles over which the scoping rate is varied, and can thus be used to control the time frame over which the genetic value is to be optimized. After t breeding cycles, the adaptive scoping method fully prioritizes the increase of the genetic gain, and a rapid fixation of QTL alleles is observed.

Optimizing the breeding population within a predefined time frame

The key advantage of the adaptive scoping method is that it can be used to optimize the genetic gain of a breeding population within a predefined time frame. Depending on the goals of the breeder, an appropriate choice for t can be made: a low value of t will provide fair genetic values in the short term, whereas higher values of t will lead to higher genetic values in the long term. Once t breeding cycles have been completed, genetic gain is fully prioritized and the highest genetic values will quickly be reached during the next few breeding cycles. Irrespective of the choice of t, the breeder can expect the adaptive scoping method to yield superior genetic values during a short time window that follows breeding cycle t. This is shown in Fig. 6 where at each breeding cycle, the results of the method that yields the highest genetic values are shown.

In a commercial breeding program, elite lines are selected in each breeding cycle for commercialization. Because the adaptive scoping method requires t breeding cycles before it renders elite individuals, it cannot be used in a commercial setting. The adaptive scoping method should be used in a situation where no output is expected in the first t breeding cycles. Such a scenario arises in prebreeding where individuals with a high variation and low genetic value are crossed for several breeding cycles to maximize the genetic gain before introducing these individuals into a commercial breeding population.

The adaptive scoping method proves robust even when the prediction accuracy is low. This was demonstrated in Fig. 5 by decreasing the heritability or increasing the number of QTL. In both cases, selecting the best parents based on the GEBVs becomes more tedious.

Comparison of the scoping and adaptive scoping methods

Compared to the (original) scoping method which uses a fixed scoping rate, the adaptive scoping method has two important advantages.

First, during the initial breeding cycles, the adaptive scoping method uses a higher scoping rate and thus better prevents the loss of favorable QTL.

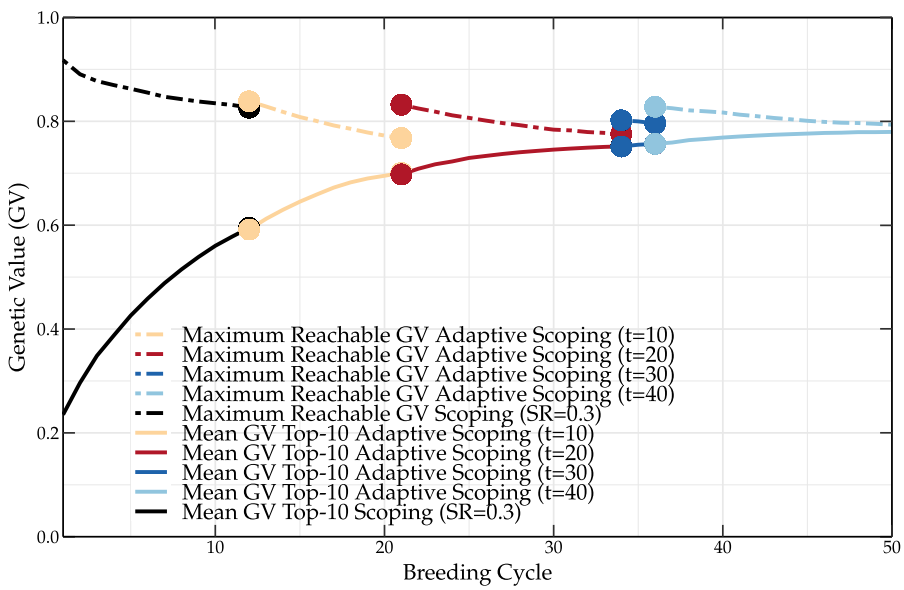


Fig. 6 Simulation results of the scoping and adaptive scoping methods. At each breeding cycle, the mean genetic value of the top-10 individuals and the maximum reachable genetic value is depicted for the method and/or value of t that yields the highest genetic value

The effect of the loss of favorable QTL is clearly observed in Figs. 3 and 5: during the first few breeding cycles, the maximum reachable genetic value of the scoping method decreases significantly whereas this is less pronounced for the adaptive scoping method.

Second, after t breeding cycles have been completed, the adaptive scoping method relies on a low scoping rate to efficiently convert the remaining genetic variation into genetic gain. From breeding cycle t onward, the scoping rate reaches its minimum value and the pre-selection procedure yields the same parental population as truncation selection (i.e., the individuals with the highest GEBVs). However, whereas truncation selection relies on a random crossing of parents, the adaptive scoping method constructs the crossing block using an identical procedure as the scoping method. The latter was demonstrated to result in an overall higher genetic gain (Vanavermaete et al. 2020). As such, the adaptive scoping method allows for a better and more accurate exploitation of the remaining genetic variation toward the end of the pre-defined time window.

Except for the case where the optimization of a breeding population in a very short period of time is desired, the adaptive scoping method outperforms the (original) scoping method. In turn, the scoping method was demonstrated to outperform parental selection methods such as truncation selection, the population merit method and the maximum variance total method in simulation studies (Vanavermaete et al. 2020). As such, the adaptive scoping method appears to be an attractive parental selection method.

Comparison of the H-OGM and the adaptive scoping method

The OGM method uses an evolutionary algorithm to optimally cross the selected parents. However, compared to the scoping method and the adaptive scoping method, the H-OGM method yields much lower genetic gains. The inbreeding coefficient for the scoping, adaptive scoping and H-OGM methods is shown in Fig. 7. A higher increase in inbreeding is observed for the (adaptive) scoping method compared to the H-OGM method. The H-OGM method crosses parents in such a way that the inbreeding coefficient is minimized in each breeding cycle. This is not the case for the scoping and adaptive scoping methods. Although the F_{score} is used to maximize the genetic variation between two crosses, over time, favorable marker alleles will accumulate in the breeding population while unfavorable alleles will still be preserved. The scoping method balances the preservation of genetic variation and the maximization of genetic gain by selecting parents that maximize the genetic gain (P1) and the genetic variation (P2). Therefore, the (adaptive) scoping method will result in a higher inbreeding coefficient compared to methods that minimize the inbreeding coefficient, but a lower inbreeding coefficient compared to methods that maximize the genetic gain (e.g. truncation selection). This also

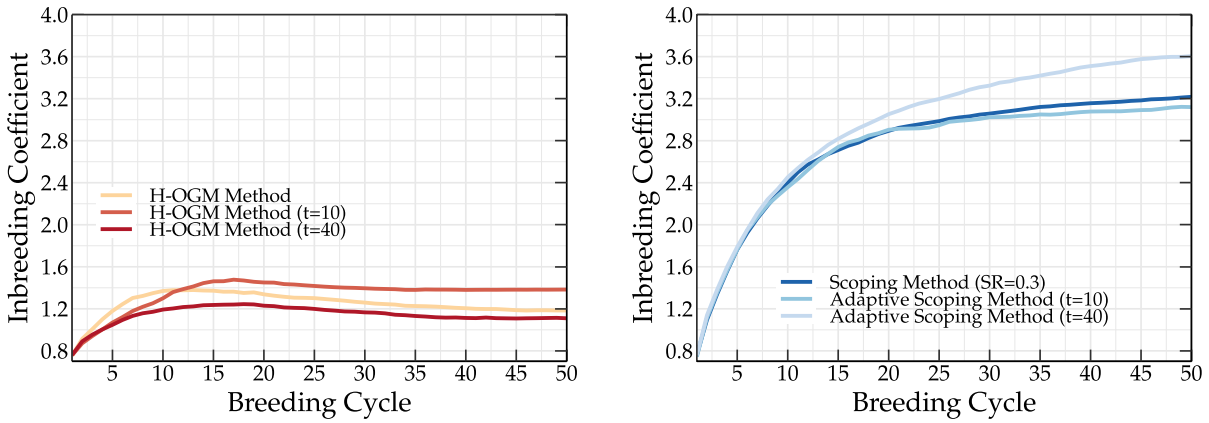


Fig. 7 Left panel, the evolution of the inbreeding coefficient in each breeding cycles for the H-OGM method (SR = 0.3, $t = 10$ and $t = 40$). Right panel, the evolution of the inbreeding coefficient in each breeding cycles for the scoping method (SR = 0.3) and the adaptive scoping method ($t = 10$ and $t = 40$)

means that the H-OGM method will result in a higher effective population size compared to both scoping methods. Nevertheless, the adaptive scoping method can reach higher genetic values in the long term and can better preserve the genetic variation compared to the H-OGM method. This indicates that the inbreeding coefficient is not sufficient as a metric to guide the selection of parental combinations. Whereas the H-OGM method minimizes the inbreeding coefficient, the adaptive scoping method can find a balance between preservation and genetic gain, allowing for a higher inbreeding coefficient but also a higher genetic progress.

Conclusion

We proposed the adaptive scoping method as an enhanced version of the original scoping method. By dynamically balancing genetic progress and genetic variation, we demonstrated its ability to maximize the genetic gain of a breeding population within a specific, predefined time frame of interest. This unique ability enables breeders to balance between exploration and exploitation of their breeding population: they can obtain fair genetic values in a relatively short term, or they can aim for the highest genetic values in the longer term. Regardless of this choice of time frame, the adaptive scoping method was shown to outperform the original scoping method.

Author Contributions DV, JF, SM, and BDB conceived and supervised the study. DV designed and performed the experiments, and wrote an early version of the manuscript. All authors reviewed and approved the manuscript.

Funding Funding was provided by Ghent University.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Akdemir D, Sánchez JI (2016) Efficient breeding by genomic mating. Front Genet 7:1–12

Allier A, Teyssèdre S, Lehermeier C, Charcosset A, Moreau L (2020) Genomic prediction with a maize collaborative panel: identification of genetic resources to enrich elite breeding programs. Theor Appl Genet 133(1):201–215

Allier A, Teyssèdre S, Lehermeier C, Moreau L, Charcosset A (2020) Optimized breeding strategies to harness genetic resources with different performance levels. BMC Genet 21(1):1–16

Bernardo R (2014) Genomewide selection of parental inbreds: Classes of loci and virtual biparental populations. Crop Sci 54(6):2586–2595

Brisbane JR, Gibson JP (1995) Balancing selection response and inbreeding by including predicted stabilised genetic contributions in selection decisions. Genet Sel Evol 27(6):541–549

Cervantes I, Gutiérrez JP, Meuwissen THE (2016) Response to selection while maximizing genetic variance in small populations. Genet Sel Evol 48(1):1–9

Chang LY, Toghiani S, Ling A, Aggrey SE, Rekaya R (2018) High density marker panels, SNPs prioritizing and accuracy of genomic selection. BMC Genet 19(1):1–10

Clark SA, Hickey JM, Van Der Werf JHJ (2011) Different models of genetic variation and their effect on genomic evaluation. Genet Sel Evol 43(1):1–9

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4(3):250–255

Gorjanc G, Gaynor RC, Hickey JM (2018) Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theor Appl Genet 131(9):1953–1966

Jannink JL (2010) Dynamics of long-term genomic selection. Genet Sel Evol 42(1):1–11

Lindgren D, Mullin TJ (1997) Balancing gain and relatedness in selection. Silvae Genet 3(2):124–129

Meuwissen THE (1997) Maximizing the respond of selection with a predifined rate of inbreeding. J Anim Sci 75:934–940

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Neyhart JL, Tiede T, Lorenz AJ, Smith KP (2017) Evaluating methods of updating training data in long-term genomewide selection. G3 7(5):1499–1510

Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161(1):209–228

Pszczola M, Strabel T, Mulder HA, Calus MPL (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. J Dairy Sci 95(1):389–400

Sonesson AK, Woolliams JA, Meuwissen THE (2012) Genomic selection requires genomic control of inbreeding. Genet Sel Evol 44(27):1–10

Vanavermaete D, Fostier J, Maenhout S, De Baets B (2020) Preservation of genetic variation in a breeding population for long-term genetic gain. G3 10(8):2753–2762

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91(11):4414–4423

Woolliams JA, Berg P, Dagnachew BS, Meuwissen THE (2015) Genetic contributions and their optimization. J Anim Breed Genet 132(2):89–99

Wray NR, Goddard ME (1994) Increasing long-term response to selection. Genet Sel Evol 26:431–451
Wray NR, Thompson R (1990) Prediction of rates of inbreeding in selected populations. Genet Res 55:41–54