Maximizing multi-trait gain and diversity with Genetic Algorithms

Guilherme F. Simiqueli[1,*] ⓘ , Rafael Tassinari Resende[2,3] ⓘ ,

Marcos Deon V. de Resende[4,5] ⓘ

Abstract
Genetic gain followed by loss of diversity is not ideal in breeding programs for several species, and most studies face this problem for single traits. Thus, we propose a selection method based on Genetic Algorithms (GA) to optimize the gains for multi-traits that have a low reduction of status number (N_S), which takes into account equal contributions from individuals as a result of practical issues in tree breeding. Real data were used to compare GA with a method based on a branch and bound algorithm (BB) for the single-trait problem. Simulated and real data were used to compare GA with a multi-trait method adapted from Mulamba and Mock (MM) (a genotypic ranking approach) through a range of selected individuals' portions. The GA reached a similar gain and N_S in a shorter processing time than BB. This shows the efficacy of GA in solving combinatorial NP-hard problems. In a selected portion of 1% and 2.5%, the GA had low reduction in the overall gain average and greater N_S than the MM. In a selection of 20%, the GA reached the same N_S as the base population and a greater gain than MM for the simulated data. The GA selected a lower number of individuals than expected at 10% and 20% selection, which contributed to a more practical breeding program that maintained the gains and without the loss of genetic diversity. Thus, GA proved to be a reliable optimization tool for multi-trait scenarios, and it can be effectively applied in tree breeding.

Keywords
Optimization — Tree breeding — Heuristic index — Status number — Evolutionary algorithm — Artificial intelligence

[1] *Corteva AgriScience, Juiz de Fora – MG, Brazil*

[2] *Plant Breeding Sector, School of Agronomy, Federal University of Goiás, 74690-900, Goiânia, Goiás, Brazil*

[3] *Forest Engineering Department, University of Brasília, 70297-400, Brasília, Federal District, Brazil*

[4] *Brazilian Agricultural Research Corporation (EMBRAPA Café), Brasília – DF, Brazil*

[5] *Universidade Federal de Viçosa (DET),Departamento de Estatística, Viçosa – MG, Brazil*

*Corresponding author*: guisimiqueli@gmail.com

1. Introduction

The selection of related individuals can be harmful to genetic diversity in a tree breeding program by enabling their crosses and, consequently, the expression of genetic load by allowing deleterious alleles in homozigose. This is likely to happen, because related individuals can share the same genes that produce the best phenotypes (Ahlinder, Mullin, & Yamashita, 2014). Thus, flexible algorithms that promote a sustainable balance between genetic gains for many traits in selection processes and genetic diversity are essential to a successful tree breeding.

In this context, detailed pedigree information is of great importance, because gives more precise information about genetic crosses and the resemblance between relatives without knowing the genetic design (El-Kassaby, Cappa, Liewlaksaneeyanawin, Klápště, & Lstibůrek, 2011; Sonesson, Woolliams, & Meuwissen, 2012). This infor-

mation is useful for estimating the genetic size of a population through group co-ancestry (θ) (Cockerham, 1967; Lindgren, Gea, & Jefferson, 1996), which represents the genes that are identical to descendants (IBD), providing a measure of genetic diversity in the base breeding population and the next selected populations.

Lindgren et al. (1996), Noiton and Alspach (1996), and Wei et al. (1998) suggest the use of the status number (N_S), which is half the inverse of the θ mean and equal to the census number when all individuals are not IBD. The θ and N_S are independent of the genetic crossing design and useful in measuring the genetic relationships between individuals. However, in finite populations, the selection of related individuals is common, which is why it is important to control the endogamy and the genetic gains that are reached (Pong-Wong & Woolliams, 2007; Sonesson et al., 2012). The selection of few individuals will likely carry a higher number of individuals per family and, consequently, result in a lower N_S. Besides
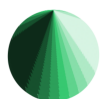
that, inbred individuals will greatly reduce the genetic diversity of a population, so the selection of large family numbers with the same number of individuals per family and a lower number of inbred individuals is needed (de Oliveira Castro et al., 2019; M. D. V. Resende, 2002).

This problem has been studied intensively in the recent past (Brisbane & Gibson, 1995; de Oliveira Castro et al., 2019; Fernández & Toro, 1999; Meuwissen, 1997; R.-P. Wei et al., 1998; R. P. Wei & Lindgren, 1995), and a method of dynamic optimization selection using Lagrange multipliers was derived (optimum contribution – OC) for animal breeding by maximizing genetic gain under a given restriction on group co-ancestry (Meuwissen, 1997). In tree breeding, an algorithm based on OC (Meuwissen, 1997) was developed because of differences in the mating system between trees and animals (Kerr, Goddard, & Jarvis, 1998) and tested in various mating schemes, including a diallel progeny trial of Scots pine (Hallander & Waldmann, 2009a). Another important algorithm is the Population Merit Selection (PMS), which maximizes genetic gain under a given weighting restriction on group co-ancestry (Lindgren & Mullin, 1997). However, the PMS is very time consuming due to the recalculation of group co-ancestry (Lindgren & Mullin, 1997) and requires weights that adjust the importance of genetic diversity, which depends on economic, biological, and other parameters (Hallander & Waldmann, 2009b). Subsequently, a software called GENCONT was developed (Meuwissen, 2002). However, GENCONT rounds negative contributions of the candidates, obtaining suboptimal solutions (Pong-Wong & Woolliams, 2007; Woolliams, Berg, Dagnachew, & Meuwissen, 2015). Therefore, semidefinite programming (SDP) (Pong-Wong & Woolliams, 2007) was developed, and the optimal solution was found with a shorter processing time than with GENCONT (Meuwissen, 1997, 2002). Currently, there is a faster version of GENCONT (Gencont2) due to the use of Gauss-Seidel method (Dagnachew & Meuwissen, 2016).

The SDP was applied in tree breeding and it obtained excellent results for clonal seed orchard selection. However, the computational cost would have been higher if a large candidate list (12,000 genotypes plus ancestors) was considered, and this would also have been very time consuming (Ahlinder et al., 2014). Thus, an alternative and faster method based on second-order cone programming (SOCP) was proposed for analyzing unequal contributions of parental genotypes to the seed orchard (Yamashita, Mullin, & Safarina, 2015), showing better results than GENCONT and SDP, mainly, for large problems in a matter of seconds. Currently, SOCP is one of the best software programs developed for unequal contribution solutions.

However, due to a lack of human and financial re-sources, tree breeding programs encounter difficulties when analyzing specific crossings with a variable family size and a larger number of parents per crossing. Consequently, many breeders prefer to cross a specific number of parents with equal contributions to the next breeding population. As a result of these practical issues, Mullin and Belotti (2016) suggested a branch and bound algorithm (BB) to solve this mixed integer quadratically constrained optimization (MIQCO) problem. Thus, objective functions and/or constraints are quadratic, and the variables to be optimized must be binary. Consequently, contributions are equal for all candidates. BB proved to be an excellent algorithm, although MIQCO is still considered an NP-hard problem. Hence, other methods should be developed because the search to find the best candidates in a small population tends to become exhaustive (Fernandez & Toro, 2001).

However, most of these algorithms used single traits and a multi-trait problem is more realistic and challenging in breeding programs (Chapuis, Pincent, & Colleau, 2016; Colleau, Tual, de Preaumont, & Regaldo, 2009; Yanchuk & Sanchez, 2011). Furthermore, antagonistic traits negatively impact the search for optimal conditions, in addition to the simple control of co-ancestry (Chapuis et al., 2016; Yanchuk & Sanchez, 2011). Thus, the selection of a group of individuals with highly desirable genetic gains for all traits without a loss of diversity is a complex problem that justifies the use of heuristic algorithms for MIQCO problems. Heuristic methods, such as simulated annealing (SA), have been applied to assigned matings (Hallander & Waldmann, 2009a; Kerr et al., 1998) to maximize genetic gains with a restriction on co-ancestry in animal (Chapuis et al., 2016; Colleau et al., 2009) and tree breeding (Fernandez & Toro, 2001; Yanchuk & Sanchez, 2011). However, these methods assign unequal contributions to their candidates.

Another algorithm that is robust and easy to implement is the Genetic Algorithm (GA). GA is based on the concept of natural selection proposed by Darwin (Holland, 1975). Although its name suggests it is a method used solely for biological problems, it is currently used in many areas of science, such as multi-depot vehicle routing problems, manufacturing systems, and 3D Truss Structures (Karakatič & Podgorelec, 2015; Mejía & Gutiérrez, 2016; Mitchell, 1998; Toğan & Daloğlu, 2016). In genetics and crop breeding, there are studies that use it to map quantitative trait loci (QTL) (Nakamichi, Ukai, Kishino, Ukai, & Nakamichi, 2001) and genomic selection (Akdemir, Sanchez, & Jannink, 2015). In animal breeding, GA has been applied to maximize average genetic gains of direct and social effects under the control of inbreeding, and it showed advantages in relation to BLUP (best linear unbiased prediction) (Wu, Jiang, Zhu, Li, & Tang, 2016).

However, the authors did not give substantial background regarding their approach. Thus, the objective of this study was proposing the use of GA to optimize genetic gain for multiple traits and genetic diversity of a breeding population at the same time.

2. Material and Methods

2.1 Data

2.1.1 Case 1: Simulated data and multi-trait index

This dataset was used to evaluate the optimization behavior of selection via GA in a simulation scenario. Here, the population structure and the genetic correlations between the traits were known. Four generations with 20 sires and 1000 dams were created with a simulation using the 16th workshop QTLMAS (Usai, Gaspa, Macciotta, Carta, & Casu, 2014). The generations did not overlap. Three traits were created relating to milk production and the genetic correlation between them. All traits relate to the female ones. Therefore, in this study, the algorithm was applied to the data of 3000 individuals in generations G1-G3. These data were chosen because the simulating quality and the resemblance to dioecious plant species. These plant species include many fruit trees, for which most phenotypes are also measured in female individuals. This is the case for the Brazilian Pine (*Araucaria angustifolia*) and Yerba Mate Tea (*Ilex paraguariensis*).

2.1.2 Case 2: Real data and single-trait index

In order to compare GA with the method reported by Mullin and Belotti (2016), the same dataset for single-traits used by these authors was examined. These data include current information regarding the pedigree and estimated breeding values (EBVs) of a breeding program in Sweden. From 50 parents, 24 breeding populations of *Pinus sylvestris* were crossed, and 242 individuals within these populations were used as the F0 to produce the F1 and F2 generations. The cycles were evaluated through open and controlled pollination. From this population, the study used the same 5250 individuals, which were examined using the method that Branch and Bound implemented in OPSEL software (Mullin & Belotti, 2016). The EBV average was compared for both methods (BB and GA) at the same restrictions and co-ancestry coefficient (Woolliams et al., 2015).

2.1.3 Case 3: Real data and multi-trait index

The *Pinus taeda* dataset was provided by Resende et al. (2012). A population of 926 individuals was formed from a partial diallel with 32 top parents. The individuals were asexually propagated as clones, which generated

experimental replicates in three different environments. Seventeen traits were evaluated by the authors, which pertained to growth, wood properties, and disease susceptibility.

In this study, the following five traits were examined: diameter at breast height (DBH), tree height (HEI), wood specific gravity (WSG), lignin content (LIG), and susceptibility assessed as gall volume (GV). All traits with correlations and heritability were in the same range as typical forest tree pedigrees, and they represented a multi-trait selection of individuals with equal contributions that maximized genetic gain with low diversity reduction.

2.2 Index formulation

The index for multi-trait selection was formulated using GA, which is considered to be a stochastic, evolutionary, metaheuristic index that is very efficient at finding the optimal solution for combinatorial NP-hard problems (Haupt & Haupt, 2004). GA is a methodology proposed by Holland (1975) and it is based on the theory of evolution by natural selection. Thus, its associated language is derived from biological terms like chromosomes, genes, alleles, population, fitness, selection, crossover, mutation, and elitism. The flow-chart (Figure 1) shows the steps followed in a GA analysis.

Generally, the algorithm randomly creates a set of solutions (population), and each solution is called a chromosome. Each chromosome is formed by many units (genes) that have discrete values (alleles) with binary representation. The algorithm either selects (allele equals one) or does not select (allele equals zero) individuals from a breeding population. Our chromosome was a binary vector yielding equal contributions. However, other codifications are also possible.

All solutions are evaluated by the fitness function (objective function), which calculates the genetic gains of many traits and the N_S of the group of individuals for each chromosome, generating a fitness value. The N_S was applied instead of group co-ancestry, because the N_S is better understood than the concept of group co-ancestry (Cockerham, 1967). This makes more sense with respect to the number of non-inbred individuals and unrelated genotypes (Lindgren & Mullin, 1997). However, Kerr et al. (1998) showed differences between the use of status number and group co-ancestry as a restriction to the optimum contribution (OC). Hence, the breeder can choose what parameter is most appropriate for the specific problem. However, further investigations must be performed to assess the differences between them. The fitness function was formulated and applied in the GA package in R software (Scrucca, 2013). Its representation (equation 1) and restrictions (equations 2, 3, and 4) are shown below:
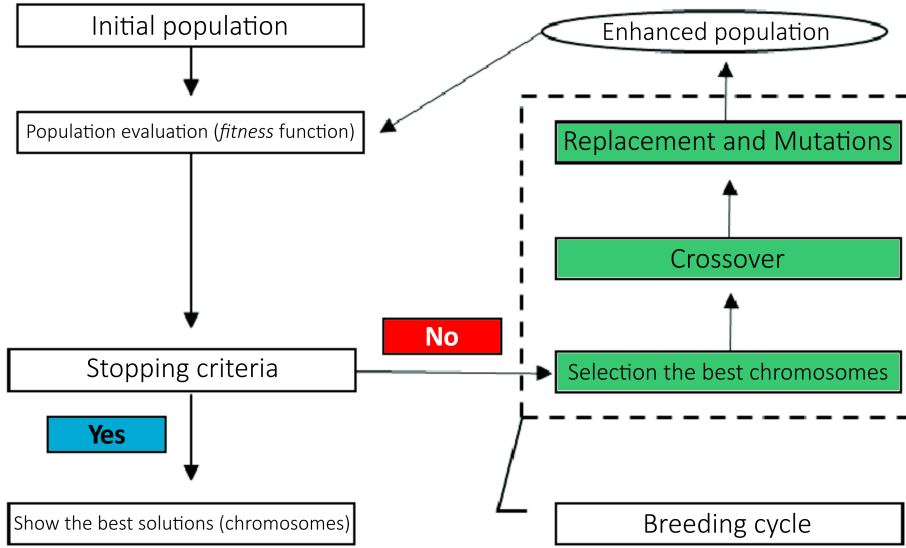
Figure 1. Flow chart showing all steps followed by the GA method. The base population is created randomly, and all of its chromosomes are evaluated according to the fitness before entering the breeding cycle. The best chromosomes are selected (exploitation) and enhanced crossovers and mutations (exploration) are performed. After this, an improved population is obtained and submitted to a new iteration (generation). The process repeats until one of the stopping criteria is achieved.

$$f(x) = \frac{1}{2}\left\{\left[Z'\left(\frac{Nx(x'x)^{-1}}{Y}-1\right)(Z'Z)^{-1}\right]+\left[\left(\frac{1}{N_{S0}}\frac{(x'x)(x'x)}{x'Ax}\right)\right]\right\}-$$
$$RZ-C \tag{1}$$

$$\frac{Nx(x'x)^{-1}}{Y}-1 > Q \tag{2}$$

$$\frac{Nx(x'x)^{-1}}{Y}-1 \leq Z \tag{3}$$

$$x'x < S_j \tag{4}$$

$$x_{(mx1)} \in [0,1], \tag{5}$$

and:

$$N_S = \frac{(x'x)(x'x)}{x'Ax} \tag{6}$$

$$N_{S0} = \frac{m}{u'Au} \tag{7}$$

where: k = number of traits; m = total of individuals; Z = vector of ones $(k \times 1)$; N = EBVs matrix $(k \times m)$; x = binary vector of the solution (chromosome) $(m \times 1)$; Y = vector of EBVs' average $(k \times 1)$; A = additive kinship matrix ($2\times$coancestry coefficients) $(m \times m)$; R = vector of penalty constant for traits $(1 \times k)$; C = penalty constant for the maximum number of selected individuals; Q = vector of zeros $(k \times 1)$; u = vector of ones $(m \times 1)$; S_j = maximum number of selected individuals according to the selected portion j; N_S = status number of selected individuals (equation 6); N_{S0} = status number of the base population with m individuals (equation 7).

The data consisting of traits represented by EBVs were normalized to vary from 0 to 1, and the mean represented the genetic gain. Normalization was necessary to avoid biased selection of individuals that have high absolute values for a specific trait, allowing us to set them at the same scale as the relation between status number of selected individuals and the base population.

The fitness function maximizes the mean between the N_S/N_{S0} ratio (the second element of equation 1) and the average genetic gain for all traits (the first element of equation 1). When the former is close to 1, this indicates

that there is a low reduction in diversity. The N_{S0} was chosen because it avoids the question of what value should be set, preventing a large variation in group co-ancestry and, consequently, population bottlenecks (Woolliams et al., 2015).

When k traits for equation 2 are not met (no gain or a negative gain), the sum of the k trait penalties (constants) is subtracted in equation 1. In this study, the value of the penalty was 2 for each trait. This penalty value was chosen because the averages of the first and second elements of equation 1 will result in a maximum value close to the unit. Hence, when any penalty is not fulfilled, the fitness function results in negative values for that chromosome (x), indicating that the chromosome is unlikely to be selected and enhanced through genetic crossover and mutation (Figure 1). Furthermore, any penalty value over this unit could be chosen.

Equation 3 limits the unit's genetic gain maximum to 100% of the gain, which is equal to double the base population's mean. This strategy balances the selection averages for many traits with respect to N_S and N_{S0} so that all averages varied from 0.0 to 1.0. This meant that a maximum of 100% of the genetic gain was limited to each trait.

Equation 4 allows the GA to identify a group of individuals that has a lower or equal value as the maximum number established by the selected portion. If unfulfilled, the penalty equals 2 (C), and this is applied in equation 1 for the same reason explained above. Thus, the selected portion in this case is the maximum condition allowed. Equation 5 refers to the GA binary codification, indicating equal contributions from all individuals selected and the size of the chromosome in the GA package (Scrucca, 2013). Equations 6 and 7 calculate the N_S

and N_{S0}, respectively. Because the binary codification of GA in equation 5, our method allows the selection of the minimum number of individuals necessary to achieve an equilibrium between the genetic gain and the N_S/N_{S0} ratio within the limits imposed by equation 4.

Therefore, this new multi-trait method will obtain the fitness values for all chromosomes by selecting individuals that have a higher EBV average for all traits with a lower reduction of status number by using a minimum number of selected individuals.

The greater the fitness value (meaning the higher genetic gain and better control of N_S), the greater the probability that these solutions will be selected. This probability is given by the linear rank method proposed by Baker (1985), which results in a slower convergence of the algorithm compared to others methods, such as the Roulette Wheel (RW) method. However, it guarantees a greater diversity of possible solutions (Mitchell, 1998; Sivanandam & Deepa, 2007). The linear rank method sets a maximum value for the best chromosome, and it sets 0 for the worst chromosome. Then, it determines the probabilities for the other chromosomes in a linear fashion. Thus, each chromosome receives a probability ranking according to its fitness value and not on its magnitude (Baker, 1985; Sivanandam & Deepa, 2007).

The RW method determines the probability of selection for each chromosome according to the magnitude of its fitness value. Therefore, if a chromosome has a very high fitness value, the others will have a low probability of being selected using the RW method. This leads to rapid convergence (Sivanandam & Deepa, 2007). After selection, chromosomes exchange genes (cross-over) and mutations occur, which are modifications to some gene values (alleles). These processes explore the solution space.

The genetic operators (crossover and mutations) were the default settings adopted by the GA package (Scrucca, 2013). They generate new chromosomes based on the best ones from the previous generation, and they are evaluated again using the fitness function. This process then repeats to identify the best chromosomes and maintain them across generations (elitism) (Haupt & Haupt, 2004; Scrucca, 2013).

The GA is robust, flexible, and it avoids local optima. This is because it has mechanisms for exploration (evaluation of new solutions in the solution space) and exploitation (the use of information from better solutions obtained previously) (Sivanandam & Deepa, 2007).

The steps involved in selection, crossover, and mutation that create the enhanced population are called breeding cycles (Mitchell, 1998; Sivanandam & Deepa, 2007). Two-stopping criteria were adopted, which included a preset maximum number of generations and

a stall generation to take into account the number of generations without an improved solution (Sivanandam & Deepa, 2007). The number of iterations was chosen according to the problem size because of the evolutionary nature of GA, which could consequently run forever (Glover & Kochenberger, 2006). Thus, the algorithm will stop when the search does not yield improved results or the maximum number of generations is reached.

Economic weights could be applied using weighting factors for which the sum should be equal to 1. In this study, all traits were proportionally equal. An important study in multi-trait selection by (Yanchuk & Sanchez, 2011) could be a method to use for tree breeding. The authors maximized the classical Smith-Hazel (SH) index under a constraint on group co-ancestry. While they did this in two steps, they first found the best individuals based on EBV through the SH index, and then they optimized the SH index under a constraint using group co-ancestry via simulated annealing. Our approach does the same in only one step.

2.3 Analysis for each case study

In cases 1 and 3, an index based on Mulamba and Mock (1978) (MM) was used as a comparison for the GA method. The MM maximizes the genetic gain of traits using the sum of their average ratings, and it selects the top-ranked individuals. However, the MM does not consider economic weights or the N_S during selection. Therefore, it is not a sufficient method for maintaining genetic diversity across breeding cycles. The N_S was calculated after the individuals had been selected using the MM method.

In case 2, the solution vectors (chromosomes) were remodeled to fit the 50 individuals that had been selected according to the restrictions on the number of individuals established by Mullin and Belotti (2016), which permitted a lower space for solutions. The value of the genes (alleles) was set as the identification number of individuals, and it ranged from 1 to 5250. Thus, each chromosome had 50 identification numbers (genes) that represented the candidates. Additionally, in the fitness function, the selection of 50 individuals was considered with restrictions on a maximum of six individuals per family and a co-ancestry coefficient less than or equal to 0.015. This was equivalent to selecting a group possessing an N_S higher than 33.33. All restrictions were equal to those used by Mullin and Belotti (2016).

The same process was performed using the co-ancestry coefficient of 0.0167 (N_S = 29.94) with a predefined processing time of 30 minutes in order to compare BB and GA in a shorter time. The dsOpt solver embedded in OPSEL software was used to implement the BB algorithm (Mullin & Belotti, 2016). All results were obtained

on a personal computer (Windows 7, 64 bits, 2.5 GHz, 6 GB RAM). GA was compared to BB because both algorithms are used to address MIQCO problems.

For case studies 1 and 3, defined selected portions were 1%, 2.5%, 5%, 10%, and 20% without a specific selected portion (free). This was performed by excluding the restriction (equation 4) and penalty constant (C). In these cases, the processing time for GA was efficient, which encouraged other researchers to improve this method and develop others that are similar.

The selection was performed using both methods (GA and MM) to cross individuals randomly and obtain a lower probability of inbreeding. Using these conditions, the next breeding population will have a smaller risk of inbreeding depression and higher genetic diversity.

3. Results

3.1 Case 1: Simulated data and multi-trait index

Table 1 shows the results for the MM and GA methods in different selected portions. This includes the number of selected individuals, the new averages for the three traits and their percentage genetic gains (T1, T2, and T3), the number of selected families (NF), the number of individuals per family (IF), the number of inbred individuals (NI), the processing time for GA, and its number of iterations.

For the MM method, the 1% selected portion had an N_S of 6.1, which is nine times lower than the N_S of the base population (54.9). Meanwhile, the GA method generated 50% of the N_S of the base population (Table 1). The MM method displayed 47% of the overall gain and the GA had 44%.

The Figure 2 shows that the iterations using GA progress up to its convergence. Figures 2a, 2b, and 2c represent the selected portions of free, 1%, and 10%, respectively. In the selected portion of 1%, it is not possible to reach higher values of N_S. In the selected portion of 10%, the final results after convergence are similar to the selected portion of the free context with respect to N_S, genetic gain, and numbers of iterations.

In the selection portion of 5%, GA reached an N_S equal to the base population, and it displayed a similar overall gain in the selected portions of 10%, 20%, and free. The overall gain for GA (0.33) was slightly lower than that of MM (0.36), for which the N_S reached 60% of the base population's N_S (Table 1). Thus, a selected portion of 5% is suitable for using the GA method to analyze these data.

For the selected portions of 10% and 20%, the GA obtained a smaller number of individuals (177 and 185, respectively) than expected (300 and 600, respectively). This was because the best N_S was found, which guar-

anteed genetic diversity and avoided over-selection of individuals that reduces the overall gain. The same phenomenon occurred with the selected portion of the free category, which had a similar number of individuals with selected portions of 10% and 20%. Obtaining fewer individuals with a high N_S and overall gain makes the selection process more practical and robust.

The GA method selected a smaller number of families and a larger number of individuals per family compared to the MM method (Table 1), which jeopardized the genetic diversity of the breeding population (Resende 2002). However, genetic diversity depends on the base population structure and the relatedness between and within families. Thus, the GA selected the best unrelated individuals from the best families.

The MM produced a significant reduction in the N_S of selection portions by 1% and 2.5% because it did not have any control over the related individuals. For this method, the N_S reduction can be controlled by using a selected portion of 10% or more. Alternatively, the same number of individuals per family can be selected with a higher number of families, which implies a lower genetic gain for both strategies.

The genetic correlations between T1 and T2 or T3 are 0.81 and -0.44, respectively. The correlation between T2 and T3 is 0.15. According to simulated data, there are positive, negative, and null correlations, which influence the way the genetic gain is set with respect to the traits. When traits are highly negatively correlated, there is an increased processing time because the search will be restricted to a few individuals that are related to each other with high breeding values.

In this context, the GA could be applied as an efficient tool in genetic breeding for combinatorial NP-hard problems because it effectively lowered processing time (from 7.55 to 39.58 minutes, Table 1).

3.2 Case 2: Real data and uni-trait index

The BB method ran for 120 minutes using a personal computer with OPSEL software and obtained a co-ancestry coefficient (θ) equal to 0.015 (N_S = 33.33) and an EBV average equal to 319.81. The GA ran until it identified an EBV average greater than or equal to 320.5, and it then selected 50 individuals from 45 families with a maximum of 6 individuals per family and obtained a co-ancestry coefficient (θ) equal to 0.015 (N_S = 33.33) and an average EBV of 320.69 in 64 minutes and 10 seconds (141059 iterations).

With a co-ancestry coefficient of 0.0167 (N_S = 29.94) generated in the same processing time (30 minutes), the GA method obtained an average EBV of 332.42, while BB in OPSEL yielded an average EBV of 332.88 with 50
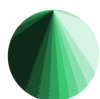
Table 1. Results obtained from both selection indexes using simulated data (Usai et al., 2014). The status number (N_S), GA processing time (Time), genetic gains for the three traits (T1, T2, and T3), and overall gain (average of the percentage genetic gain for the three traits) are shown.

Sel. Int.[a]	Sel. Ind.[b]	Ns	T1	T2	T3	Overall Gain	NF[c]	IF[d]	NI[e]	Time (min)	Iter.[f]
Mulamba and Mock											
1%	30	6.08	253.553 (0.433)[g]	22.086 (0.685)	0.028 (0.287)	0.47	13	2.3	1	–[h]	–
2.5%	75	28.1	214.352 (0.366)	19.002 (0.590)	0.026 (0.266)	0.41	65	1.2	2	–	–
5%	150	33.8	181.561 (0.310)	16.579 (0.514)	0.024 (0.248)	0.36	113	1.4	6	–	–
10%	300	40.4	172.851 (0.296)	14.806 (0.459)	0.018 (0.191)	0.32	216	1.39	14	–	–
20%	600	46.1	155.140 (0.265)	12.429 (0.386)	0.013 (0.134)	0.26	417	1.44	32	–	–
Genetic Algorithm											
1%	30	24.64	305.546 (0.522)	21.508 (0.667)	0.013 (0.136)	0.44	28	1.07	0	7.55	1217
2.5%	75	43.27	247.987 (0.424)	17.641 (0.547)	0.012 (0.120)	0.36	36	2.08	0	13.78	2189
5%	150	54.92	234.671 (0.401)	16.039 (0.498)	0.009 (0.097)	0.33	55	2.73	0	21.35	3375
10%	177	54.92	240.245 (0.411)	16.375 (0.508)	0.009 (0.093)	0.33	78	2.27	2	23.81	3680
20%	185	54.93	238.600 (0.408)	16.333 (0.507)	0.009 (0.096)	0.34	82	2.26	3	21.8	3118
Free	184	54.93	236.129 (0.404)	16.338 (0.507)	0.010 (0.101)	0.34	82	2.24	3	39.58	4641

[a] Selected portion; [b] Selected individuals; [c] Number of families selected; [d] Number of selected individuals per family; [e] Number of selected inbred individuals (i.e., inbreeding coefficient > 0); [f] Number of iterations; [g] Percentage of genetic gain (values in parentheses); [h] Instantaneous process.

selected individuals and a maximum of six individuals per family.

3.3 Case 3: Real data and multi-trait index

The Table 2 shows the results from using real data (M. F. R. Resende et al., 2012) with the MM method and the GA method at different selected portions. This table reports the number of selected individuals, the new mean of the five traits, the number of selected families (NF), the number of individuals per family (IF), the number of inbred individuals (NI), the processing time for GA, and its number of iterations.

For the MM method, a selected portion of 1% showed an N_S of 3.0, which represents an eight-fold reduction in the base population's status number (24.1). In contrast, the GA method produced almost one-third (7.5) of the base population's status number (Table 2) and a lower gall volume (GV) with a small difference (0.07) in the overall gain between the two methods.

For GA, the best selected portion that yields high N_S with high genetic gains is 10%. Similar results are obtained with the 20% and free selected portions. For selected portions of 10% and 20%, the GA method isolated
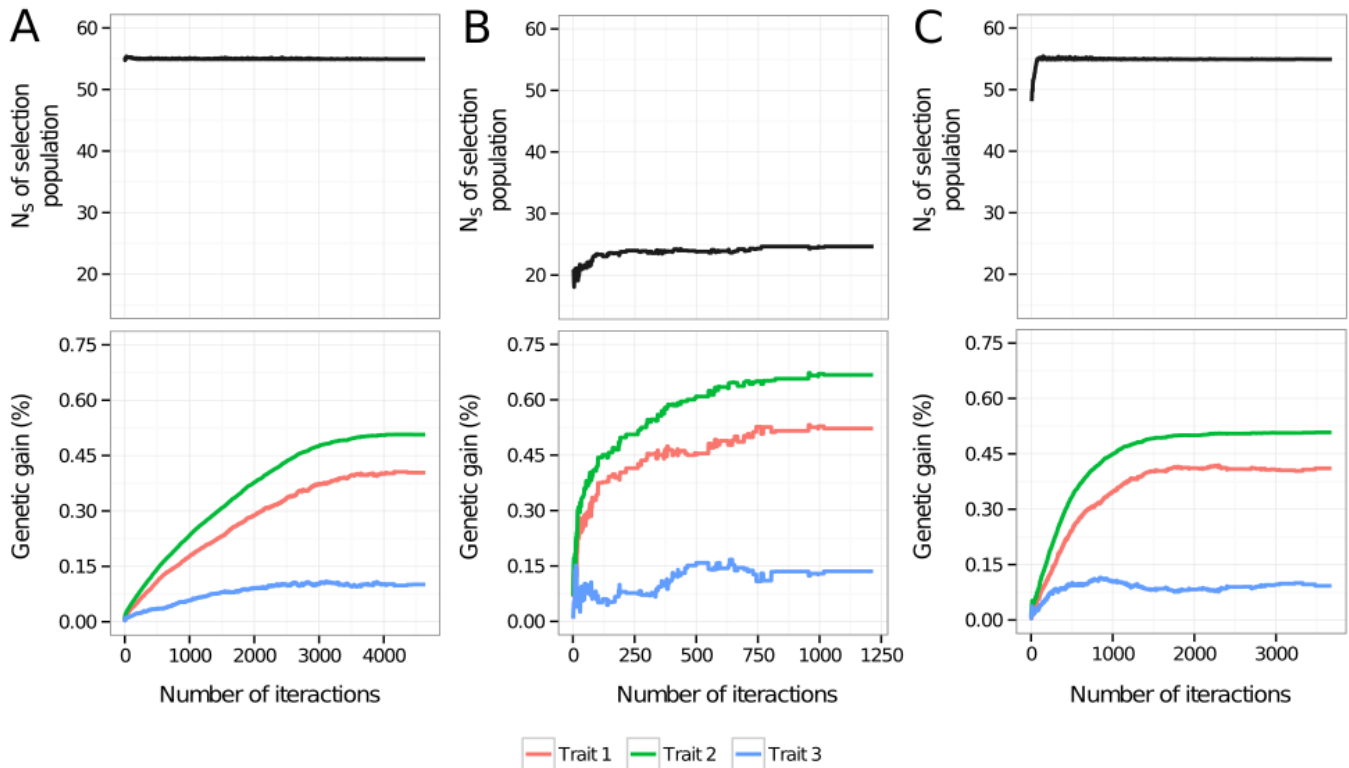
Figure 2. The GA method's convergence progress for case 2. (a) The selected portion is free (Free) to maximize the N_S and the genetic gain without restricting the number of selected individuals. (b) The selected portion of 1% (maximum of 30 selected individuals). (c) The selected portion of 10% (maximum of 300 selected individuals). The three charts on the top show the maximization of genetic diversity.

a lower number of individuals (91 and 122, respectively) than expected (93 and 185, respectively) because the best N_S was identified, guaranteeing genetic diversity and avoiding over-selection of individuals that can reduce the overall gain. The same result was obtained for the selected free portion, which had a similar number of individuals as the 20% selection.

Using the GA method, the genetic gain values for the selected portion of 5% are lower than those for the 10% portion because the latter accommodates a greater number of individuals with higher genetic values for the five traits. For instance, the HEI had 67.4% of selected individuals with genetic values higher than the average in the 5% selected portion. Meanwhile, in the 10% selected portion, 76.9% of the individuals had genetic values higher than the average.

The GA method selected a larger number of families and fewer individuals per family than the MM (Table 2), which preserves the genetic diversity of the breeding population (Resende 2002) because it selects the best unrelated individuals from the best families.

The difference in the overall gain between the se-

lected portion of 1% and 2.5% is lower using the MM method and higher using the GA method. This is because of the base population's structure, which has individuals with higher genetic values in only a few families. This explains the smaller number of selected families obtained with the MM method than by the GA method for these selected portions (Table 2). For this real data, the GA had a shorter processing time (ranging from 2.54 to 12.38 minutes), which may be due to the absence of highly negative correlations between the traits.

Controlling selection by imposing a restriction of a maximum of three individuals per family (REST method) for the MM method (results not shown) increased the number of selected families and improved diversity (high N_S). By doing this, the MM method had a higher N_S for all selected portions, but this was still inferior to the GA method. The overall gain of the MM was similar to that of GA for a selected portion of 10%. However, the gain was smaller for the MM with respect to the 20% portion compared to GA. Showing that the strategy for number of individuals controlling per family slightly improved genetic diversity and enhanced the desired genetic gains.
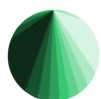
Table 2. Results obtained from both selection indexes using real data (M. F. R. Resende et al., 2012). The status number (N_S), GA processing time (Time), and estimated breeding values relating to diameter at breast height (DBH), tree height (HEI), wood specific gravity (WSG), lignin content (LIG), and susceptibility assessed as gall volume (GV) are shown.

Sel. Int.[a]	Sel. Ind.[b]	Ns	DBH	HEI	WSG	LIG	GV	Overall Gain	NF[c]	IF[d]	NI[e]	Time (min)	Iter.[f]
Mulamba and Mock													
1%	9	2.4	1.938 (0.511)[g]	92.741 (0.412)	0.470 (0.290)	0.408 (0.562)	-0.177 (0.154)	0.39	3	3.0	0	–[h]	–
2.5%	23	4.9	1.620 (0.411)	90.875 (0.401)	0.361 (0.211)	0.373 (0.509)	-0.192 (0.165)	0.34	10	2.3	0	–	–
5%	46	8.1	1.499 (0.374)	83.143 (0.357)	0.239 (0.123)	0.303 (0.399)	-0.242 (0.204)	0.29	15	3.1	0	–	–
10%	93	10.9	1.395 (0.341)	75.154 (0.311)	0.243 (0.126)	0.251 (0.319)	-0.182 (0.158)	0.25	26	3.6	0	–	–
20%	185	13.7	1.158 (0.267)	62.134 (0.237)	0.178 (0.079)	0.218 (0.269)	-0.165 (0.144)	0.20	38	4.9	0	–	–
Genetic Algorithm													
1%	9	7.5	1.588 (0.401)	86.064 (0.374)	0.313 (0.176)	0.363 (0.493)	-0.180 (0.156)	0.32	8	1.1	0	2.54	514
2.5%	23	17.5	0.752 (0.139)	49.324 (0.164)	0.135 (0.048)	0.134 (0.139)	-0.121 (0.115)	0.12	23	1.0	0	3.15	672
5%	46	24.1	0.762 (0.143)	41.624 (0.120)	0.079 (0.008)	0.097 (0.082)	-0.057 (0.061)	0.08	38	1.2	0	9.92	2230
10%	91	24.1	0.926 (0.194)	49.520 (0.165)	0.181 (0.082)	0.120 (0.117)	-0.098 (0.093)	0.13	53	1.8	0	8.73	1922
20%	122	24.1	0.904 (0.187)	51.224 (0.175)	0.180 (0.08)	0.126 (0.126)	-0.108 (0.100)	0.13	59	2.1	0	12.01	2317
Free	122	24.1	0.929 (0.195)	50.286 (0.169)	0.171 (0.074)	0.130 (0.132)	-0.106 (0.099)	0.13	58	2.1	0	12.38	2282

[a] Selected portion; [b] Selected individuals; [c] Number of families selected; [d] Number of selected individuals per family; [e] Number of selected inbred individuals (i.e., inbreeding coefficient > 0); [f] Number of iterations; [g] Percentage of genetic gain (values in parentheses); [h] Instantaneous process.

4. Discussion

Genetic breeding programs may produce low genetic variability and have difficulty selecting other traits because they do not maintain population diversity (Noiton & Alspach, 1996). Thus, the selection of many traits using co-ancestry guarantees genetic gain and diversity for more than one trait (Chapuis et al., 2016; Colleau et al., 2009; Yanchuk & Sanchez, 2011). However, the balance between genetic gain and genetic diversity (co-ancestry or N_S) is typically divergent (R.-P. Wei et al., 1998; Woolliams et al., 2015) and many studies have developed efficient algorithms to solve this type of problem, like algorithms based on Lagrange multipliers (Dagnachew & Meuwissen, 2016; Meuwissen, 1997, 2002), SDP (Pong-Wong & Woolliams, 2007), the branch and bound algorithm (Mullin & Belotti, 2016), SA (Chapuis et al., 2016; Colleau et al., 2009; Fernandez & Toro, 2001; Yanchuk & Sanchez, 2011) and SOCP (Yamashita et al., 2015). Except for branch and bound, these algorithms address unequal genetic contributions from animal and tree breeding, which is considered ideal for long-term genetic gain (Hallander & Waldmann, 2009a; Meuwissen, 1997; Rodríguez, 2000). Despite of unequal genetic contribution benefits, many tree breeders prefer to cross a specific number of parents with equal contributions to the next breeding population due to the lack of human and financial resources.

Moreover, most of these studies concerned the genetic gain of one trait, although the genetic breeding of many traits was suggested to be extremely important by Wei et al. (1998). Nevertheless, if the genetic breeding of many negatively correlated traits is desired, this will negatively impact the search for an optimal gain in-

stead of simply controlling co-ancestry (Chapuis et al., 2016; Yanchuk & Sanchez, 2011). In this regard, the GA method was efficient when applied to a MIQCO multi-trait problem with a shorter processing time (a maximum of 40 minutes) for simulated data (Table 1). In this case, the correlations between two traits were -0.44, and the GA increased the correlation after selection (results not shown). For real data with correlations and heritabilities in the range for typical forest tree pedigrees (Table 2), the same result was observed, yielding a maximum processing time of 12.38 minutes.

Some authors applied Simulated Annealing (SA), another artificial intelligence algorithm, to maximize the gain for one or more traits using a restriction on co-ancestry in animal (Chapuis et al., 2016; Colleau et al., 2009) and tree breeding (Fernandez & Toro, 2001; Yanchuk & Sanchez, 2011). SA is an excellent heuristic method, but it has been criticized for its slow speed of convergence (Glover & Kochenberger, 2006). While it circumvents the problem of becoming stuck in local optima in the first steps of the algorithm at high temperature, SA can become stuck if a low temperature is chosen (Fernandez & Toro, 2001; Press, Flannery, Teukolsky, & Vetterling, 1989). In these studies, SA was applied to address unequal contributions, and it was not utilized to solve the MIQCO problem in a multi-trait scenario. For this reason, this study did not compare SA with GA. However, the selection of different algorithms to solve optimization problems should be considered carefully due to the no-free-lunch theorem, which posits that the average performance of all algorithms is equal to all types of problems. In other words, while the SA or GA can function better than other algorithms for certain problems, they are not necessarily the best analytical tools for all problems (Wolpert & Macready, 1997). Thus, future work should focus on comparing both methodologies because certain algorithms are more appropriate than others for solving particular problems. As instance, the SOCP method can be considered a specific algorithm (Yamashita et al., 2015), but it should still be tested for its ability to solve multi-trait problems, MIQCO problems, or both.

Pong-Wong and Woolliams (2007) argued that the GA method does not guarantee the same solution since it is a stochastic process and its convergence is slower when the number of restrictions increase. However, GA incorporates the elitism parameter, which involves the maintenance of the best solutions (chromosomes) across iterations (generations), making GA extremely efficient. In addition, the GA has two important mechanisms of exploitation and exploration (Figure 1), which ensure the acquisition of a global optimum and escape from the local optima, even though GA does not guarantee an exact solution (Haupt & Haupt, 2004; Scrucca, 2013; Sivanandam

& Deepa, 2007). For combinatorial NP-hard problems, many local optima are necessary when using methods that converge to a global optimum, which is similar to GA. Therefore, the use of GA is very relevant because it guides genetic breeding programs to a global optimum despite its stochastic nature (Figure 1). Moreover, the performance of the algorithm dominates the analysis of practical applications when the size of the optimization problems increases, becoming the exact solution irrelevant (Kerr et al., 1998). Thus, the processing time is only important when very large candidate populations are considered, but even a small number of candidates results in an exhaustive search (Fernandez & Toro, 2001). Generally, when there are many non-flexible restrictions, it is not possible to find a feasible solution. This increases the processing time because the program searches for a nonexistent solution. Therefore, optimization methods should be used as guides during the decision-making process such that other solutions are considered.

To meet all the restrictions imposed and find an optimal solution may result in not identifying an optimal solution close to a breeder's objective, and this may require a longer processing time. This was shown in case 2 where both methodologies (BB and GA) required a processing time longer than one hour to find the same EBV at a specific co-ancestry coefficient. Furthermore, the efficiency of single-trait GA was demonstrated by the fact that similar results were obtained when compared to BB for two different scenarios. In the first scenario (θ equals 0.015), GA obtained an EBV average slightly higher (320.69) than BB (319.81) in a shorter processing time. In the second scenario (θ equals 0.0167 with a processing time of 30 minutes), GA obtained an EBV average slightly lower (332.42) than BB (332.88). Although GA is a stochastic algorithm and it can identify different EBV averages, it converged towards a global optimum and avoided local optima as much as possible. Therefore, both algorithms are excellent for solving MIQCO problems.

The GA algorithm could identify the number of individuals, which guaranteed an optimal solution with respect to the genetic gains and the diversity with or without (free) a restriction in the number of selected individuals (Table 1 and 2, and Figure 2). For instance, in case 1, the selected portion of 10% (with a maximum of 300 individuals) showed that the selection of more than 177 individuals negatively influenced the genetic gains and the genetic diversity. This happened because over-selection reduced the genetic gain and genetic diversity due to the selection of related and inbred individuals (Table 1). The convergence process for the selected portion (free) is similar to 10% (Figure 2) and 20% of selection (Table 1 and 2), and only a slight difference in processing time occurred.

The establishment of a specific number of individuals per family, which is called REST by Brisbane and Gibson (1995), is other simple way to control de loss of diversity. However, it slightly enhances genetic diversity and reduces the desired genetic gains (Case 3), because important combinations of individuals that ensure a better N_S and genetic gain for many traits will be lost.

Inbreeding depression may be a more severe problem in the short- and mid-term than the loss of genetic variance (Brisbane & Gibson, 1995; Kerr et al., 1998; Woolliams et al., 2015). In this context, the minimal losses observed with respect to genetic gain (Table 1 and 2) justify the improvement in genetic diversity that was obtained with the GA index, which avoided a reduction in the genetic variance and inbreeding depression. The latter can put an end to genetic breeding programs in the short- and mid-term (Meuwissen, 1997; Mullin & Belotti, 2016; M. D. V. Resende, 2002; Sonesson et al., 2012). For tree breeding, another strategy used to avoid inbreeding depression is optimizing the assignments of mating pairs (Hallander & Waldmann, 2009a; Kerr et al., 1998) that can be also combined with the GA index.

In addition, some breeding programs aim to maintain the Germplasm Bank's diversity (Pong-Wong & Woolliams, 2007) or introduce a new Germplasm Bank (Noiton & Alspach, 1996) and GA is able to address this need. It is also possible to define specific levels of diversity and gains for each trait while including individual accuracy as a weight by changing the fitness function or give different weights to genetic gain and $\frac{N_S}{N_{S0}}$ in equation 1, instead of a weight of 0.5 for both. This demonstrate the extremely flexibility of GA algorithm. Moreover, GA can be applied to select individuals with unequal contributions and to assign matings in one step instead of two, as demonstrated by Kerr et al. (1998) and Hallander and Waldmann (2009a) in tree breeding.

In this study, the control of individual accuracy and breeding values from the previous generations was not applied. If it had been, there would be a multi-generation and multi-trait index that controlled the genetic diversity and individual accuracy, which is a complex and nontrivial problem (Brisbane & Gibson, 1995; R.-P. Wei et al., 1998; R. P. Wei & Lindgren, 1995). In addition, this study was not applied in more than one generation, which is a limitation that will be fulfilled in future works.

The GA is easy to implement in software like R (R Core Team, 2022) through the GA package (Scrucca, 2013) and the fitness function, which can be modified according to the breeder's objective by including other factors that favor genetic breeding strategies, guarantee diversity and genetic gain.

While individual accuracy was not considered in this study, the data suggest that methods for obtaining EBVs should have a high level of accuracy, ensuring reliability with respect to the individuals selected. Furthermore, phenotypic information combined with molecular markers obtained through genome-wide selection approaches provide higher selective accuracies and greater control of the individual endogamy rate than the use of phenotypic selection alone (Daetwyler, Villanueva, Bijma, & Woolliams, 2007; Sonesson et al., 2012). Therefore, the GA method may be used in conjunction with genome-wide selection (GWS) to initially select less-related individuals, which could ensure more control over the N_S and higher genetic gains across the generations for many traits.

5. Conclusions

The GA method proved efficient for finding a global optimum for the genetic diversity and genetic gain of many traits in a convenient processing time, like other methods found in the literature, by reproducing similar results as the branch and bound method for one trait in a mixed integer, quadratically constrained optimization problem. Additionally, the fitness function's flexibility allows for the inclusion of different restrictions, which are adjusted according to the breeder's objective. Thus, GA should be the first approach used to when considering more than one trait and controlling the status number in tree breeding.

Data Archiving Statement

We follow the standard policies for Tree Genetics and Genomes. All data are available at `figshare.com/articles/QTLMASXVI_xlsx/4265129`. The data were uploaded and there are no access numbers. All data in this work are public and available online. Data for case 1 are available at `qtl-mas-2012.kassiopeagroup.com/en/dataset.php`. The section data is in the file QTLMASXVI.dat [117.2 KB] and the pedigree information is in the file QTLMASXVI.gnl [100.1 KB]. These data relate to the pedigree and EBVs of the simulated data relating to the milk production case study. Data for case 2 are in a public repository available at `datadryad.org/resource/doi:10.5061/dryad.4r1f0`. These data relate to the pedigree and EBVs for the Scots pine case study, and they are reported at the following paper: Mullin TJ, Belotti P (2016) Using branch-and-bound algorithms to optimize selection of a fixed-size breeding. Tree Genetics & Genomes 12: 4. `dx.doi.org/10.1007/s11295-015-0961-z`. Data for case 3 are available in File S2 and File S3, and can be found at `www.genetics.org/content/190/4/1503/suppl/DC1`. These data relate to the pedigree and EBVs for the Loblolly pine case study,
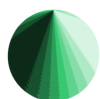
and are reported in the following paper: Resende MFR, Muñoz P, Resende MDV, et al (2012) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). Genetics 190:1503–1510.

Acknowledgments

References

Ahlinder, J., Mullin, T. J., & Yamashita, M. (2014). Using semidefinite programming to optimize unequal deployment of genotypes to a clonal seed orchard. *Tree Genetics & Genomes*, 10(1), 27–34.

Akdemir, D., Sanchez, J. I., & Jannink, J.-L. (2015). Optimization of genomic selection training populations with a genetic algorithm. *Genetics Selection Evolution*, 47(1), 1. doi.org/10.1186/s12711-015-0116-6

Baker, J. E. (1985). Adaptive selection methods for genetic algorithms. In *Proceedings of an International Conference on Genetic Algorithms and their applications* (pp. 101–111). Hillsdale, New Jersey.

Brisbane, J. R., & Gibson, J. P. (1995). Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. *Theoretical and Applied Genetics*, 91(3), 421–431.

Chapuis, H., Pincent, C., & Colleau, J. J. (2016). Optimizing selection with several constraints in poultry breeding. *Journal of Animal Breeding and Genetics*, 133(1), 3–12. doi.org/10.1111/jbg.12178

Cockerham, C. C. (1967). Group inbreeding and coancestry. *Genetics*, 56(1), 89.

Colleau, J.-J., Tual, K., de Preaumont, H., & Regaldo, D. (2009). A mating method accounting for inbreeding and multi-trait selection in dairy cattle populations. *Genetics, Selection, Evolution : GSE*, 41, 7. doi.org/10.1186/1297-9686-41-7

Daetwyler, H. D., Villanueva, B., Bijma, P., & Woolliams, J. A. (2007). Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics*, 124(6), 369–376.

Dagnachew, B. S., & Meuwissen, T. H. E. (2016). A fast Newton–Raphson based iterative algorithm for large scale optimal contribution selection. *Genetics Selection Evolution*, 48(1), 70.

de Oliveira Castro, C. A., Nunes, A. C. P., Roque, J. V, Teófilo, R. F., Santos, O. P., Santos, G. A., . . . Resende, M. D. V. (2019). Optimization of Eucalyptus benthamii progeny test based on Near-Infrared Spectroscopy approach and volumetric production. *Industrial Crops and Products*, 141, 111786.

El-Kassaby, Y. A., Cappa, E. P., Liewlaksaneeyanawin, C., Klápště, J., & Lstibůrek, M. (2011). Breeding without breeding: Is a complete pedigree necessary for efficient breeding? *PLoS ONE*, 6(10). doi.org/10.1371/journal.pone.0025737

Fernández, B. J., & Toro, M. A. (1999). The use of mathematical programming to control inbreeding in selection schemes. *Journal of Animal Breeding and Genetics*, 116(6), 447–466.

Fernandez, J., & Toro, M. A. (2001). Controlling genetic variability by mathematical programming in a selection scheme on an open-pollinated population in Eucalyptus globulus. *Theoretical and Applied Genetics*, 102(6–7), 1056–1064.

Glover, F. W., & Kochenberger, G. A. (2006). *Handbook of metaheuristics* (Vol. 57). Springer Science & Business Media.

Hallander, J., & Waldmann, P. (2009a). Optimization of selection contribution and mate allocations in monoecious tree breeding populations. *BMC Genetics*, 10(1), 70.

Hallander, J., & Waldmann, P. (2009b). Optimum contribution selection in large general tree breeding populations with an application to Scots pine. *Theoretical and Applied Genetics*, 118(6), 1133–1142.

Haupt, R. L., & Haupt, S. E. (2004). *Practical genetic algorithms with CD-Rom*. www.amazon.com/gp/product/0471455652/ref=oh_details_o08_s01_i01?ie=UTF8&psc=1

Holland, J. H. (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* U Michigan Press.
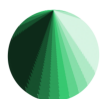
Karakatič, S., & Podgorelec, V. (2015). A survey of genetic algorithms for solving multi depot vehicle routing problem. *Applied Soft Computing*, 27, 519–532.

Kerr, R. J., Goddard, M. E., & Jarvis, S. F. (1998). Maximising genetic response in tree breeding with constraints on group coancestry. *Silvae Genetica*, 47(2), 165–173.

Lindgren, D., Gea, L., & Jefferson, P. (1996). Loss of genetic diversity monitored by status number. *Silvae Genetica*, 45(1), 52–58.

Lindgren, D., & Mullin, T. J. (1997). Balancing gain and relatedness in selection. *Silvae Genetica*, 46(2–3), 124–129.

Mejía, G., & Gutiérrez, E. (2016). Scheduling Complex Manufacturing Systems Using a Genetic Algorithm. In *Metaheuristics for Production Systems* (pp. 223–240). Springer.

Meuwissen, T. H. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science*, 75(4), 934–940.

Meuwissen, T. H. (2002). GENCONT: an operational tool for controlling inbreeding in selection and conservation schemes. In *Proceedings of the 7th Congress on Genetics Applied to Livestock Production* (pp. 19–23).

Mitchell, M. (1998). An Introduction to Genetic Algorithms, 209. doi.org/10.1016/S0898-1221(96)90227-8

Mulamba, N. N., & Mock, J. J. (1978). Improvement of yield potential of the ETO blanco maize (Zea mays L.) population by breeding for plant traits [Mexico]. *Egyptian Journal of Genetics and Cytology*.

Mullin, T. J., & Belotti, P. (2016). Using branch-and-bound algorithms to optimize selection of a fixed-size breeding population under a relatedness constraint. *Tree Genetics and Genomes*, 12(1), 1–12. doi.org/10.1007/s11295-015-0961-z

Nakamichi, R., Ukai, Y., Kishino, H., Ukai, Y., & Nakamichi, R. (2001). Detection of Closely Linked Multiple Quantitative Trait Loci Using a Genetic Algorithm. *Genetics*, 158(1), 463–475. www.genetics.org/cgi/content/abstract/158/1/463%5Cnwww.genetics.org/cgi/reprint/158/1/463.pdf

Noiton, D. A. M., & Alspach, P. A. (1996). Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. *Journal of the American Society for Horticultural Science*, 121(5), 773–782.

Pong-Wong, R., & Woolliams, J. A. (2007). Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming (Open Access publication). *Genetics Selection Evolution*, 39(1), 1.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). Numerical Recipes, the Arts of Scientific Computing Fortran Version. *Cambrigdge University Pess, Cambridge, UK*.

R Core Team. (2022). R: a language and environment for statistical computing. Vienna, Austria. www.r-project.org/

Resende, M. D. V. (2002). *Genética biométrica e estatística no melhoramento de plantas perenes*. (Embrapa,Ed.). Brasília: Embrapa: Informação Tecnológica.

Resende, M. F. R., Muñoz, P., Resende, M. D. V., Garrick, D. J., Fernando, R. L., Davis, J. M., . . . Kirst, M. (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics*, 190(4), 1503–1510.

Rodríguez, L. S. (2000). Balanced vs. slightly unbalanced selection. *Heredity*, 84(6), 685–691.

Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 53(4), 1–37. doi.org/10.1359/JBMR.0301229

Sivanandam, S. N., & Deepa, S. N. (2007). *Introduction to genetic algorithms*. Springer Science & Business Media.

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

Toğan, V., & Daloğlu, A. T. (2016). Genetic Algorithms for Optimization of 3D Truss Structures. In *Metaheuristics and Optimization in Civil Engineering* (pp. 115–134). Springer.

Usai, M. G., Gaspa, G., Macciotta, N. P. P., Carta, A., & Casu, S. (2014). XVI th QTLMAS: simulated dataset and comparative analysis of submitted results for QTL mapping and genomic evaluation. In *BMC proceedings* (Vol. 8, p. 1). BioMed Central.
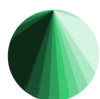
Wei, R.-P., Hansen, C. R., Dhir, N. K., & Yeh, F. C. (1998). Genetic gain with desired status number in breeding programs: a study on selection effects. *Canadian Journal of Forest Research*, 28(12), 1861–1869.

Wei, R. P., & Lindgren, D. (1995). Optimal family contributions and a linear approximation. *Theoretical Population Biology*, 48(3), 318–332.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82. doi.org/ 10.1109/4235.585893

Woolliams, J. A., Berg, P., Dagnachew, B. S., & Meuwissen, T. H. E. (2015). Genetic contributions and their optimization. *Journal of Animal Breeding and Genetics*, 132(2), 89–99.

Wu, P., Jiang, Y., Zhu, L., Li, X., & Tang, G. (2016). Optimizing the gain of social genetic effect under the control of inbreeding using genetic algorithm. *Livestock Science*, 190, 70–75.

Yamashita, M., Mullin, T. J., & Safarina, S. (2015). An efficient second-order cone programming approach for optimal selection in tree breeding. *ArXiv Preprint ArXiv:1506.04487*.

Yanchuk, A. D., & Sanchez, L. (2011). Multivariate selection under adverse genetic correlations: impacts of population sizes and selection strategies on gains and coancestry in forest tree breeding. *Tree Genetics & Genomes*, 7(6), 1169–1183.