University of São Paulo
"Luiz de Queiroz" College of Agriculture

Long-term selection of biparental crosses: a comparison among genomic methods and phenotypic selection

Rafael Storto Nalin

Thesis presented to obtain the degree of Doctor in Science.
Area: Genetics and Plant Breeding

Piracicaba
2019

Rafael Storto Nalin
Bachelor in Agronomy

Long-term selection of biparental crosses: a comparison among genomic methods and phenotypic selection

Advisor:
Prof. Dr. **ANTONIO AUGUSTO FRANCO GARCIA**

Thesis presented to obtain the degree of Doctor in Science.
Area: Genetics and Plant Breeding

Piracicaba
2019

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Nalin, Rafael Storto

Long-term selection of biparental crosses: a comparison among genomic methods and phenotypic selection / Rafael Storto Nalin. – – Piracicaba, 2019.

69 p.

Tese (Doutorado) – – USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Predição de cruzamentos 2. Simulação computacional 3. Seleção à longo prazo 4. *Triticum spp.* L. I. Título

DEDICATORY

To my beloved parents, Julio Cesar and Rosana,
and my brothers Rodrigo and Renan

To Professor Roland Vencovsky
(*in memoriam*)

ACKNOWLEDGEMENTS

I would like to start thanking the one who was with me at the begining of this long (but pleasurable) journey. During the first one year and a half I had the biggest academic honor of my life working with Professor Roland Vencovsky. There are no words to express how much I learned from him, not only about statistics, genetics, and breeding during his excelent explanations, but also as a human being. I feel very fortunate for the opportunity to be by his side daily, learning from him on how a scientist must be. Thank you for everything, I will always take your examples with me, whenever I am. I would like to extend this words to his wife, Maria Olavia Vencovsky and his family.

My sincere thanks also goes to Prof. Antonio Augusto Franco Garcia, for welcome me in our lab during my second year and on, and for provide me all the support that I needed during my studies. Thank you for all the patience, pleasurable discussions, and insights that you gave to me, not only about my project but also about science, in general. Additionally, I would like to thanks the Statistical Genetics Lab (ESALQ/USP): Cris, Camila, Elaine, Gabriel, Getúlio, João Bachega, Jhonathan, Kaio, Letícia, Luís Felipe, Marianella, Matheus, Rodrigo, and many others that were in our lab during these four years.

Special acknowledgement goes to Prof. Lucia Gutierrez (University of Wisconsin-Madison), for all the support during my time in Madison. Thank you for put so much efforts to take me into our CBQG Lab and also for all the priceless weekly discussions that we had. This project definitely would not be the same without your support. I'm also very thankful to the crew of the Cereal Breeding and Quantitative Genetics Lab (UW-Madison) for the great time of learning and friendship that I had in Madison: Chris, Gabi, Inês, Madhav, Maria, Pablo G., Pablo S., Raegan, Sarah, and Valentina.

I also would like to thanks all the staff and professors of the Genetics and Statistics Departments for their valuable time during the classes and also for all the discussions that helped me to clarify several topics on statistics, programming, genetics, and breeding. Many thanks goes also to all the professors present during my undergraduation (IFMT) and Masters (Federal Univesity of Lavras) degree. Special thanks goes to Professor Magno Ramalho (Federal University of Lavras), for the great time of learning that I had during my masters, as his student, and also for all the discussions about my project during my Ph.D., especially during my first year.

To Martin Quincke (INIA, Uruguay), for provide me the data set from the wheat breeding program of INIA, used on this project. Also, for his patience and availability to talk about the data and the results whenever I need it.

Many thanks to Fernando Henrique Toledo (CIMMYT, Mexico) and Julio Bueno Filho (Federal University of Lavras) for all the discussions about Monte Carlo simulation and for help me on my own simulations, turning into a more biological meaningful and realistic one.

Special thanks also goes to all my friends that were with me during this journey, especially: Dayane and Lucas (also for the support and friendship during my stay in Madison), Jéssica, Mario, and Vitor.

To the "Luiz de Queiroz"College of Agriculture, University of São Paulo (ESALQ/USP) for provide me such a great structure, physical and intellectual, to the development of this thesis. To the University of Wisconsin, Madison, for having me during my sandwich program. Also, to the Center for High Throughput Computing (CHTC, UW-Madison) for all the technical support.

To the Coordination of Superior Level Staff Improvement (CAPES) and National Council for Scientific and Technological Development (CNPq) for the scholarship during my Ph.D..

Last, but not least, special thanks to my parents, Julio Cesar and Rosana, and my brothers, Rodrigo and Renan, for being the best family that I could have. I would not be able to get this far without all your support (Amo vocês!).

EPIGRAPH

"***There is nothing more practical than a good theory***"

Kurt Lewin (1952)

SUMMARY

RESUMO

Seleção de cruzamentos biparentais em longo prazo: Uma comparação entre métodos genômicos e seleção fenotípica

A seleção de cruzamentos é parte importante de um programa de melhoramento e o uso de uma estratégia adequada é crucial. Uma boa estratégia deve balancear a seleção dos melhores indivíduos e a manutenção da diversidade genética ao longo dos ciclos de seleção, visando ganhos a longo prazo. Dentre os métodos propostos na literatura podemos destacar a predição genômica com progênies simuladas, que pode ser utilizada para se estimar a média e a variância genética de cada combinação de parentais candidatos, provendo valiosa informação para o melhorista. No entanto, não há relatos sobre como esse método se comporta em um processo de seleção a longo prazo. Portanto, o objetivo desse trabalho foi avaliar a performance do método de predição genômica com progênies simuladas em relação aos métodos tradicionais de seleção fenotípica, ao longo de dez ciclos de melhoramento. Simulações *In silico* e utilizando um conjunto de dados foram utilizadas para investigar essas metodologias em relação ao ganho genético e diversos outros parâmetros relacionados a diversidade genética. Simulou-se um programa de melhoramento de trigo com capacidade para avaliar 1000 genótipos a cada ciclo. Diferentes cenários para herdabilidade e a combinação número de populações e número de progênies por população foram avaliados. Um conjunto de dados reais de 1465 linhagens de trigo também foi utilizado com o objetivo de proceder com uma simulação baseada em dados reais. Nesse caso, marcas foram aleatoriamente designadas como genes. Os resultados indicam que o melhor método é dependente da herdabilidade da característica, da estratégia adotada pelo melhorista quanto ao número de cruzamentos realizado e também se o objetivo de melhoramento é a obtenção de ganhos genéticos a curto ou à longo prazo. No geral, os métodos envolvendo seleção genômica, especialmente o que faz uso de progênies simuladas, apresentaram melhores resultados quando a herdabilidade é baixa e o número de populações é alta, tanto a curto como à longo prazo. No entanto, embora a conversão de variabilidade genética em ganhos genéticos seja mais rápida com essa estratégia, a perda de variabilidade é mais acentuada, sendo interessante a reposição de novas fontes de diversidade com o avançar dos ciclos de melhoramento. A adoção de uma restrição no número de vezes que um genótipo atua como genitor é, também, de fundamental importância para a obtenção de ganhos à longo prazo.

Palavras-chave: Predição de cruzamentos, Simulação computacional, Seleção à longo prazo, *Triticum spp.* L.

ABSTRACT

Long-term selection of biparental crosses: a comparison among genomic methods and phenotypic selection

The selection of crosses is a fundamental part of a breeding program, and the use of an adequate strategy is crucial. A good strategy should balance the selection of the best individuals and maintenance of genetic diversity throughout cycles of breeding, aiming for long-term genetic gains. Among the methods proposed in the literature, we can highlight the genomic prediction with simulated offsprings, which can be used to estimate the mean and genetic variance of each combination of candidate parents, providing useful information to the breeder. However, as far as we know, there are no reports on how this method performs concerning long-term genetic gain. Thus, the goal of this study was to evaluate how genomic prediction with simulated offsprings performs compared with the traditional phenotypic selection across five cycles of breeding. In silico and data-based simulation was used to investigate these approaches in terms of genetic gain and several other parameters related to the genetic diversity. We simulated an *In silico* standard wheat breeding program, with a capacity to evaluate 1000 lines per cycle. We considered different scenarios for the heritability, number of populations and the number of offspring per population. A real dataset of 1465 wheat inbred lines was also used to perform simulations. In this case, markers were randomly assigned to be genes. The results indicated that the best method is dependent of the heritability of the trait under consideration, the breeder's strategy about how many crosses will be done and also if the breeding goal is to have short or long-term genetic gains. In general, the genomic methods, especially the genomic prediction with simulated progenies, presented the best results under scenarios of low heritability and high number of population, either on short or long-term. However, even though the conversion of genetic variability into genetic gains is faster than any other strategy, the losses of variability are also higher, being interesting to bring new sources of variability with the advance of the cycles of breeding. The adoption of the restriction on the number of times a genotype is a parent in crosses is also of fundamental importance for obtaining long-term genetic gains.

Keywords: Prediction of crosses, Computer simulation, Long-term selection, *Triticum spp.* L.

LIST OF FIGURES

1 PREFACE

The selection of superior individuals have been a common practice since the begining of the human race. However, with much less knowledge and science involved, our anscestors started the first breeding process by selecting individuals that were phenotypically more atractive. It was just on the 1950's, several years after the the rediscover of Mendel's laws of heredity, the foundation of the Biometrical school (mostly based on Francis Galton's approach), and posterior development on quantitative genetics, embracing both approaches (largely accomplished by Fisher, Haldane and Wright) that breeding started to be applied using a solid theoretical basis, on both, animal and plant species (Lynch and Walsh, 1998).

The impact of adoption of quantitative genetics theory on plant breeding can be seen on several crop species. Nowadays, the selection of individuals rely on large data sets that allows breeders to estimate the breeding and genotypic values from the candidate individuals with much more accuracy than before, increasing the genetic gain (Bernardo, 2010). Also, a whole new set of techniques became widely available with the development of molecular markers for a reasonable price. Genomic selection, proposed by Meuwissen *et al.* (2001) is one of these tools and refers to the use of molecular markers to predict accurately, the genetic merit of a set of individuals. This method takes the advantage of the high number of markers that we can have nowadays and also the fact that some of these markers will be on linkage disequilibrium with the causal variants of the trait of interest. Using these information, the breeder can calculate the level of association of each marker with the trait, by fitting a random regression model with all the markers, at the same time, as covariables. With all the marker effects, it is possible to calculate the merit of each individual of intereset, by summing up the genetic values of each marker, even the ones without any phenotypic information.

The use of genomic selection to predict the genetic merit of individuals is revolutionizing breeding schemes (Jannink, 2010). Its use allows the breeder to select individuals on earlier cycles of breeding than traditional methods (Heffner *et al.*, 2010). A recurrent selection scheme, for example, involves three different steps: progeny development, evaluation and finally, the recombination of the selected individuals, composing the ones for the next cycle (Hallauer *et al.*, 2010). However, as usually the evaluation step requires the synthesis and evaluation of progenies of each individual on a experimental trial, this leads to a low genetic gain per unit of time, as it might take too much time to accomplish a cycle. But, with genomic selection it is possible to combine evaluation and recombination into a single step, where the genomic estimated breeding value is used to select the individuals and cross them, without the need to evaluate their progenies (Schwantes *et al.*, 2018).

Moreover, genomic selection can increase the rates of genetic gain not only shortening the breeding cycles, but also improving testing efficiency in field evaluations (Bernardo and Yu, 2007; Heffner *et al.*, 2010), connecting genotypes on different environments by modelling Genotype $\times$ Environment interaction (Crossa *et al.*, 2017).

Despite all the use of genomic selection since the first application on plants (Bernardo and Yu, 2007), the use of genome wide markers applied to the selection of crosses, instead of individuals, started just on 2014 (Bernardo, 2014). Until this date, the most traditional method to choose what genotypes to cross was based just on the mean breeding values of the parents (Lado *et al.*, 2017). However, not taking into account the variability that a cross have, might lead to wrong decisions on the selection of crosses. The variance within a cross, which is a statistical measure of variability, is a function of the divergence on allele frequencies of each locus in the parental lines (Lynch and Walsh, 1998). Besides the average breeding values, these divergences allow the breeder to find progenies that are way superior to the parents. However, it is not rare to have crosses with high mean but with very low variability, not being possible to select superior progenies from these ones. Thus, the selection of crosses

should take into account the average breeding value and also the possible variability within a cross, and the use of genomic selection can be very helpful to accomplish this task. It is also important to mention that the possible genetic gain, in this application of genomic selection, it is not based on the fact that we can shorten cycles, but relies simply on the selection of better crosses than the average breeding values of the parents.

This procedure was applied in a short-term selection on a real data set by LADO *et al.* (2017). The results were promising, as the crosses selected based on the genomic selection approach were different than the common approach used by breeders; however, no genetic gain was calculated. While short-term responses sometimes can be reasonably predicted from the base-population variance components alone, long-term response depends on the genetic architecture (number of genes, allelic effects and frequencies) (WALSH and LYNCH, 2018). The long-term response can be studied using a deterministic approach, where the theoretical formulas are derived. Real experiments of long-term selection can also be found in the literature. The most iconic of these is the Illinois long-term corn selection experiment, started by the chemist Cyril Hopkins in 1896 and since then, more than 90 generations of selection were performed, with no apparent selection limit, acchieving a total response of 22 additive standard deviation (σ_A) (GOLDMAN, 2004; DUDLEY, 2007). While, the deterministic approach requires several assumptions, simulation procedures, like *Monte Carlo* method, can be more flexible, relaxing some of these assumptions and it is also much less time consuming than a real experiment, being a good alternative for this kind of study.

Simulations like the *Monte Carlo* procedure can be very useful to model variables that follows some distribution of probability. Because of that, this approach can be very useful on plant breeding experiments, as mostly of the phenomena are random, like meiosis. It is possible, for example, to model the meiosis procedure using a set of probability distributions. We can use a Poisson distribution (λ = Total size in Morgans) to sample values related on how many crossing-overs we might have in a chromosome, followed by an uniform distribution to model where each crossing-over will be ($U \sim$ (0, Chrom. size)). However, there are several models that aims to mimic the meiosis procedure in real life. We will detailed this topic and others in the literature review section, with the objective to present in details most of the fundamental topics related to the paper, also in this manuscript.

The paper presented on this thesis aimed to study, using two simulation approaches, how different methods of selection of crosses perform on a long-term selection. Different values for parameters, like heritability (h^2) and different ratio between number of crosses (N_{POP}) and number of progenies per cross (N_{Prog})were considered. As far as we know, there is no studies using and comparing this approach with different methods of cross selection in a long-term way. The format of the paper express our intention to submit it to *Genetics* journal. The partial results were presented in a poster format on the PAG XXVII (Plant and Animal Genome Conference), San Diego, California, USA.

2 LITERATURE REVIEW

2.1 Selection of crosses in plant breeding

The selection of crosses is one of the fundamental steps in every breeding program (Fehr, 1987). By selecting the best crosses, the breeder expects to find progenies with superior performance than the parents. However, this is not an easy task as the germplasm usually counts with hundreds or even thousands of genotypes and the number of population that can be created and evaluated are limited to just a few hundred, as the breeding programs are limited on the number of progenies or plots per season (Ramalho *et al.*, 2012b; Bernardo, 2010). Considering only biparental population, we can have $\frac{n(n-1)}{2}$ combinations from n parents, so this number grows fast and every cross that does not produce interesting progenies it is a waste of time and financial resources. Thus, selecting the best crosses is important to support the breeding program and achieve genetic gains throughout cycles of selection.

Quantitative genetics has been an important source of information, providing knowledge to the breeder to guide the decisions on a breeding program since its development (Bernardo, 2010; Lynch and Walsh, 1998; Walsh and Lynch, 2018). This is also true when we look to the theory about the selection of crosses. Several studies can be found related to the development of population in a breeding program, like proposing methods of cross selection (Hallauer *et al.*, 2010; Bernardo, 2014), studying the trade-off between the number of biparental population and number of progenies per population (Wricke and Weber, 1986; Fouilloux and Bannerot, 1988; Bernardo, 2003), comparing biparental crosses against other classes of crosses (Fehr, 1987; Ramalho *et al.*, 2012a), etc.

About the methods used to select crosses, we can classify them in two categories, according to Baenziger and Peterson (1991): the first involves the procedures that use information from the parents only and the second refers to all the methods that use the performance of the progenies, demanding the breeder to produce the crosses and evaluate these populations. However, due to the usual high number of candidate lines to be used as parents, it is unfeasible to produce all the possible crosses and evaluate them, being important the adoption of methods that allow the choice, *a priori* of the best crosses (Witcombe *et al.*, 2013). Among these methods, we can highlight the use of parental means, coancestry coefficient, genetic distances and more recently the genomic prediction with simulated offsprings.

In the parental means method, we use just the *per se* performance of the candidate genotypes to select which ones should be used as parents. The problem with this is that we do not consider the complementarity between two genotypes, being not possible to predict how much variability the breeder will have in the cross to apply selection (Ramalho *et al.*, 2012a). This is the most popular method used in breeding programs as it is easy to apply and also provide good results, although do not take into account the possible variability generated within a cross (Bernardo, 2010, 2014; Beckett *et al.*, 2019). Using just classical quantitative genetics, we can predict the mean of a population (Bernardo, 2010); however, the genetic variance is difficult to predict. The use of coancestry coefficient and genetic distances, on the other hand, with pedigree, molecular markers or multivariate analysis using morphologic traits to define what crosses will provide higher genetic variability. Several papers can be found using these methodologies to select crosses (Ghaderi *et al.*, 1984; Souza and Sorrells, 1991; Bohn *et al.*, 1999; Carpentieri-Pípolo *et al.*, 2000; Tambasco-Talhari *et al.*, 2005)

Until a few years ago, the applications of genomic prediction, proposed by Meuwissen *et al.* (2001), was restricted to the prediction of genotypic and breeding values of unphenotyped individuals. Just with the paper from Bernardo (2014) the method also became useful for the selection of crosses. In this method, with the adjusted means and the marker data set, the breeder fits a regression model to estimate the marker effects (e.g., Ridge regression BLUP or any of the different Bayesian methods), and with a genetic linkage map, simulates several progenies from every possible cross from a list of

candidate parents. The computer codes that simulate these progenies involves the reproduction of the meiosis phenomenon, crossing-over, gamete formation, and fertilization. With the simulated progenies, it is possible to calculate different parameters, as the mean and variance, providing useful information to the breeder. The probability of success in obtaining superior genotypes in a cross is a function of the breeding value of the parents and also the variance that comes from the complementarity between parents (Bernardo, 2014). The genomic prediction procedure concatenates both aspects.

Further studies, like Mohammadi *et al.* (2015) and Tiede *et al.* (2015), also evaluated the usefulness of this method and also facilitated its implementation in a breeding program through a package in the *R* environment (R CORE TEAM, 2015). The method was studied in a short-term manner by Lado *et al.* (2017) using a wheat dataset. Also comparing this method with the phenotypic selection (mean breeding values of the parents), the authors showed that there are no relevant differences on what crosses should be selected with the GS adoption when the yield was used. However, major differences were reported for grain protein content, which is a higher heritability trait. Some concerns with this method were also described by Lehermeier *et al.* (2017). The use of a random regression model shrinks the marker effects, and these effects are used to estimate the breeding values of the simulated progenies. The variance is then calculated as the variance of the breeding values, which are shrunk, as it is also based on the marker effects, being the variance underestimated, especially when the shrunk effect is large (when traits present low heritability). Thus, it is important to find ways to circumvent this issue and obtain a better estimation of the variance of each cross, being possible to select crosses more accurately. Otherwise, with an underestimation of the variance, the selected crosses tend to be the same as those selected based only on the mean breeding values of the parents. Also, Osthushenrich *et al.* (2017) presented expressions to calculate the mean (μ), superior progeny mean (μ_{sp}) and the genetic variance (V_G) of a cross, being the results highly correlated with the ones from the PopVar package.

Using a training population of two-row barley lines, and calculating the superior progeny mean, from the mean and genetic variance, Neyhart and Smith (2019) predicted 330,078 possible parent combinations for *Fusarium* head blight (FHB) severity, heading date and plant height. Twenty-seven of these combinations were done, and these populations were then evaluated for the same traits. They find out that mean and superior progeny mean are predictable with moderate to high accuracy ($r_\mu = 0.46 - 0.69$), but the progeny variance can present low accuracy in some cases ($r_{V_g} = 0.01 - 0.4$) and this accuracy is a function of the heritability of the trait.

Beckett *et al.* (2019) also applied this method, using the PopVar package from Mohammadi *et al.* (2015), in a maize population. Results show that combining germplasm from different sources/companies in new breeding crosses can produce inbreds with improved performance in a hybrid test cross. Also, the authors presented results about the use of genetic distance between two parents of a breeding population as presenting little correlation with the predicted genetic variance.

2.2 Computer simulation to plant and animal breeding

Since the initial development of molecular markers, we tremendously improved our knowledge about the mechanisms underlying the genetic control of a quantitative trait. Before this, our decisions relied on the information provided from the rediscovery of Mendel's law, in the twentieth century, and the outstanding work made by Sir Ronald Fisher, early 20th century, combining the Mendelian theory with the biometric school (Falconer and Mackay, 1996; Lynch and Walsh, 1998; Visscher and Goddard, 2019), showing that a quantitative trait can be modeled considering multiple genes contributing to the variation, each of them following Mendel's law.

Especially with the development of SNP (*Single Nucleotide Polimorphism*) markers, that recently became widely available due to the advancement of genotyping technology (Poland and Rife,

2012), a huge amount of molecular data started to be available to the breeders, helping to support their decisions. However, with this fast development, we clearly started to see a gap between the huge amount of information that we have and the breeding practices (Li *et al.*, 2012). With these new tools in the box, several new questions that we did not though before, arised, like: How can we use these markers to optimize our resource allocation (should we invest more on genotyping than phenotyping)? (Lorenz, 2013; Endelman *et al.*, 2014); What is the best method and approach to apply this information on my breeding program? (Crossa *et al.*, 2017); How can I redesign my breeding program, turning into a more efficient one? (Heffner *et al.*, 2010; Robertsen *et al.*, 2019) How can we deal with Genotype × Environment in this whole new scenario? (Lado *et al.*, 2016; Monteverde *et al.*, 2018).

To find out answers to these questions and integrate the knowledge that we have gained in a breeding program, converting it into genetic gain, we can use different approaches, including computer simulation. Today, computer simulation can be considered the third component of the scientific research triad, along with theoretical and experimental approaches (Li *et al.*, 2012). Especially with this explosion of knowledge, computer simulation became very useful to address these questions in a fast way, as we do not have to wait for a long time to collect empirical data.

Thus, computer simulation is an important tool to solve problems in genetics and breeding, addressing answers to several questions. Like a bridge between theory and experimentation, computer simulation has become a powerful tool in scientific research, providing not only preliminary validation of theory but also guiding the user on the adoption of new techniques and technologies (Li *et al.*, 2012). In computer simulation, parameters and restrictions are set up in a way that we can isolate the effects of certain factors and conveniently control them (Kemphthorne, 1988; Wang *et al.*, 2007; Cruz *et al.*, 2013). In this way, simulation can be seen as a system that mimics a certain reality, aiming to find out what could possibly happen in a real system if changes were made (Dachs, 1988).

According to Li *et al.* (2012), the major applications of computer simulation in genetics can be defined as follow: (1) breeding method comparison: finding the best breeding scheme taking into account the multiple factors; (2) gene mapping: validating the effectiveness of new mapping methods, calculating *LOD* score and confidence interval (CI); (3) plant breeding platforms: integrative simulation programs that can simulate the whole plant-breeding process; and (4) crop modeling: combining crop models, genetic architecture of traits, and environmental information to fill the gap between genotype and phenotype.

The first papers of simulation using computer resources in genetics and plant breeding were presented by Fraser (1957a,b). In these two papers, the authors evaluated the influence of genetic linkage on the genetic gains when mass selection is applied. Since then, several papers came along using computer simulation techniques aiming to answer fundamental questions of plant breeding programs. Among these questions we can highlight: measure the efficiency of methods of conduction of segregant population in autogamous species (Casali and Tigchelaar, 1975); efficiency of marker assisted selection (MAS) applied to breeding programs (Edwards and Page, 1994; Lande and Thompson, 1990; Zhang and Smith, 1992, 1993).

Several examples of the usefulness of computer simulation can be found in the literature. Bernardo (2009), for example, used a simulation procedure to study right after the boom of the double haploid (DH) technique on corn, if DH lines should be induced among F_1 or F_2 plants, providing useful results in a short period of time. His results showed that plants should be induced from F_2 plants rather than F_1, because of the increase in the amount of genetic variance explored when F_2 plants are induced. Also, in the same year of the groundbreaking paper from Meuwissen *et al.* (2001), which also used computer simulation to present the theory of genomic selection and how useful it could be if we have enough markers covering the genome, Bernardo (2001) presented the paper "What if we knew all the genes for a quantitative trait in hybrid crops?", to study how useful gene information would be in scenarios

where a few or a high number of loci control the trait of interest. The same author studied the parental selection, the optimum number of breeding population and size of each breeding population (Bernardo, 2003), using the same simulation procedure, as presented on the section 2.1. The results showed that parental selection is more important than the trade-off between the number of population and size of each population. Using computer simulation, Abreu *et al.* (2010) presented results about the efficiency of stratification in mass selection under different scenarios of heritability and allelic interactions. The same authors studied different recombination strategies in an intrapopulation recurrent selection scheme in *Eucalyptus* species (Abreu *et al.*, 2013), also considering different scenarios for heritability, allelic frequencies, and allelic interactions.

Papers comparing breeding methods using computer simulation can also be found. Casali and Tigchelaar (1975), for example, studied the effectiveness of the classical methods of selection in autogamous species: pedigree method, bulk, and single seed descent. Wang *et al.* (2003) compared the modified pedigree/bulk selection method (*MODPED*) against the selected bulk method (*SELBLK*) in terms of genetic gain. The authors found out that *SELBLK* method provides slightly higher genetic gain and significant improvements in cost-effectiveness.

In the genomic era, there are also several papers providing useful information to guide decisions on how to implement this new set of techniques in plant breeding programs. In the context of the traditional marker assisted selection, Zhang and Smith (1992) compared the value of this technique, which use the linkage disequilibrium between markers and QTLs, against the traditional selection using the *Best Linear Unbiased Prediction* (BLUP) procedure, not taking into account molecular data, and a method that combine both sources of information. The results showed that scenarios with high LD (F_2 population) provided higher gains than a population with low LD (F_{20} population) when only MAS were applied, probably due the few number of markers covering the simulated genome. Also, the BLUP procedure provided higher gains than MAS, but the combined method overcame both individual methods. Also, Edwards and Page (1994) studied the application of MAS in a recurrent selection applied to an allogamous annual species context, such as maize. Again the results showed that a good cover of the genome is a fundamental condition in MAS. Also, markers were most effective in fully exploiting the genetic potential when fewer QTLs controlled the trait. Besides that, MAS offers the advantage of enabling two selection cycles per year, while the phenotypic selection required two years per cycle.

The pioneering paper on genomic selection, the most recent bandwagon (Bernardo, 2016), is also an important example of the usefulness of computer simulation. Starting in 2001, Meuwissen *et al.* (2001) presented the methodology and how useful the method could be if we had enough markers covering the genome, which was not the case by the time, as we did not have high throughput methods to genotype a large number of genotypes with a good amount of molecular markers covering the genome. Thus, the authors used computer simulation to present their ideas. It was only late 2000s that genomic selection became an object of study. The first simulation of genomic selection in a plant breeding program was published just on 2007 (Bernardo and Yu, 2007), showing that the method is superior to MAS (18% to 43%, depending on the scenario) if molecular markers started to be cheaper than it was.

Also using computer simulation, Jannink (2010), studied how genomic selection performs in a long-term manner. The choice for a selection procedure should not only take into account the short-term response but also what is the response in a long-term, as the maintenance of genetic diversity along several generations of breeding allow us to have future gains with the selection. In his paper, the author described how recombination, population size, heritability influence the prediction accuracy. The results also showed that losing favorable alleles that are in weak linkage disequilibrium with markers is perhaps inevitable when using genomic selection. Thus, some strategies need to be adopted to minimize this rate of loss.

Lorenz (2013) and Endelman *et al.* (2014) studied the resource allocation when genomic

selection is applied in a plant breeding program. The first author simulated, under different scenarios, the trade-off between population size and the number of replications, considering a single biparental population of double haploid lines. The results showed that a high degree of flexibility was observed when *RR-BLUP* method was applied, meaning that different scenarios of this trade-off provide similar results in terms of genetic gain. Endelman *et al.* (2014), on the other hand, simulated the trade-off between population size, number of replications and number of locations, comparing methods that take into account only the phenotypic information, only the genotypic information and both. These questions about resource allocation are very important, due to the fact that now we have additional costs of genotyping, but using the same budget as before. In this way, we need to figure out what is the best choice under different scenarios considering all these trade-offs that a breeder works on a daily basis.

Among the methods of simulation, we can highlight the *Monte Carlo* method, which takes into account all the categories of systems that has a probabilistic or stochastic base (Ferreira, 2001; Suess and Trumbo, 2008). This method has great importance in agriculture, especially in genetics and plant breeding, as most of the phenomena has a random behavior, like meiosis phenomenon. A *Monte Carlo* procedure involves the use of probability distributions to describe the behavior of a particular phenomenon. As we will see below in section 2.2.1, to simulate a counting-location model of meiosis, we need to use several generators of random numbers of some probability distribution, like Poisson and Uniform. Also, to simulate other procedures on breeding, we need to create random numbers that follow some normal and exponential distribution, as we will see in section 2.2.2.

Obtaining estimates of genetic parameters is a fundamental part of a breeding program as several decisions can be made based on these information. Thus, a good estimate of these parameters is crucial to make a correct decision. To obtain an estimate we can use several estimators, so we need to figure out which one is the best to use (Ferreira, 2001). *Monte Carlo* simulation is also a very useful tool to test different estimators, as the researcher defines the parametric value of the parameter, being this a known value. The simulation process can be replicated thousands of times, considering different scenarios, and with this, study the bias, variance, and consistency of the estimator. Zeng and Cockerham (1991), for example, evaluated the estimator proposed by Wright to calculate the number of genes of a quantitative trait using computer simulation, showing that the estimator suffers most seriously from linkage. Also, with reasonable assumptions about unequal allelic effects and initial allele frequencies, the values, not considering linkage, are likely to be on the order of one-third of the real number of genes.

To simulate the different breeding procedures, the breeder needs to build a set of computer code in some programming language (e.g., *R*, *C++*, Fortran, Python), aiming to mimic the different biological phenomena that happens in real life. The adoption of a specific language impacts directly on the simplicity and performance of the code. Languages like *C++* or Fortran sometimes are called low-level languages as the programmer needs to define features of the system by itself. However, even though it is harder to write and understand, its performance highly overcome languages like *R*, usually called high-level language, as it is written in a form that is close to our human language, using English words, enabling the programmer to focus just on the problem being solved. However, environments like *R* also allow the user to incorporate pieces of *C++* code, for example, to speed up some part that is struggling in time. Packages like *Rcpp* (Eddelbuettel, 2013) have been very helpful to easily integrate these two languages, overcoming problems related to not only time but memory use.

2.2.1 Simulating the meiosis

It is important to mention that some procedures in a simulation process need to be replicated millions of times. Meiosis, for example, is present in almost every single step of a simulation procedure. Meiosis must occur when we self plants from one generation to another, or crossing individuals or even

when we generate double haploid lines. Because of that, we need a function that mimics meiosis and also is very efficient computationally. As mention in section 2.2, the use of a low-level language, like *C++* sometimes is the best option to optimize the time that a simulation procedure needs to run. Nowadays, there are some very efficient packages that help the researcher to implement models that mimic meiosis and support the simulation of common tasks in breeding, like crossing, selfing, and creation of double haploid lines (See section 2.2.3).

There are different models to simulate meiosis. The most simple model considers unlinked loci, which is a very strong assumption about the correlation that occurs between loci that are close to each other due linkage. Even though it is a simple model, sometimes can be sufficient to provide important information, like the ones presented by Veiga *et al.* (2000) and Fernandes (2016). In this case, each locus follows the first Mendel's law, independently, which can be simulated easily using an uniform distribution to define what class (AA, Aa, or aa) an offspring will be from a heterozygous genotype and loci. However, when the goal is to use molecular markers, this model becomes useless as we need to simulate the linkage disequilibrium (LD) that occurs between markers and between markers and genes.

The count-location model is an alternative to the unlinked loci model and a much better approximation of reality. To simulate meiosis we need to define two things: (1) How many crossing-overs will occur? (2) Where these crossing-overs will going to be located in the chromosome?. To address the first question, the count-location model uses a poisson distribution in which the parameter of the distribution refers to the size of the chromosome ($\lambda = L$) (Speed and Waterman, 1996). The disadvantage of this method occurs when we need to define where these crossing-overs will occur as this method uses an uniform distribution ($U \sim (0, L)$), in which, L refers to the size of the chromosome. By assuming this, we also assume that there is no interference between crossing-overs, meaning that the probability of two crossing overs occur close to each other is the same of occurrence far away in the chromosome (Browning, 2000). However, this method also has the advantage of easily incorporate obligate chiasma, which is common in biology (Falque *et al.*, 2007).

When the count-location model is applied, the pairwise recombination frequency between loci follows the Haldane's linkage function (Haldane, 1919), which also assumes that the distance between two loci relates to their recombination frequency following a Poisson distribution, as presented on equation 2.1.

$$m_{ij} = -\frac{1}{2} log(1 - 2r_{ij}) \tag{2.1}$$

m_{ij} represents the genetic distance, measured in centiMorgans (cM), between markers i and j; and r_{ij} refers to the recombination frequency between markers i and j, measured in centiMorgans (cM). The relation between the count-location model and Haldane linkage function can also be seen in figure 2.1. In this case, 100 markers were distributed along four Morgans, and several gametes were generated using the count-location model. Their pairwise recombination frequency was then calculated. The red line represents the relation between genetic distance and recombination frequency, according to Haldane's model.

However, in some cases, modeling the ratio of interfering crossing-overs to noninterfering crossing-overs differs between the species in which it has been studied (Falque *et al.*, 2007)

The model of Lande and Stahl (1993), on the other hand, is a stationary renewal process and allows to take into account the interference between crossing-overs(Speed and Waterman, 1996). However, one of the disadvantages of this model initially was that it did not allow to consider obligate chiasmas. Just with the paper from Falque *et al.* (2007) a model that also takes into account this phenomenon was proposed, turning the modeling process into a more realistic one.

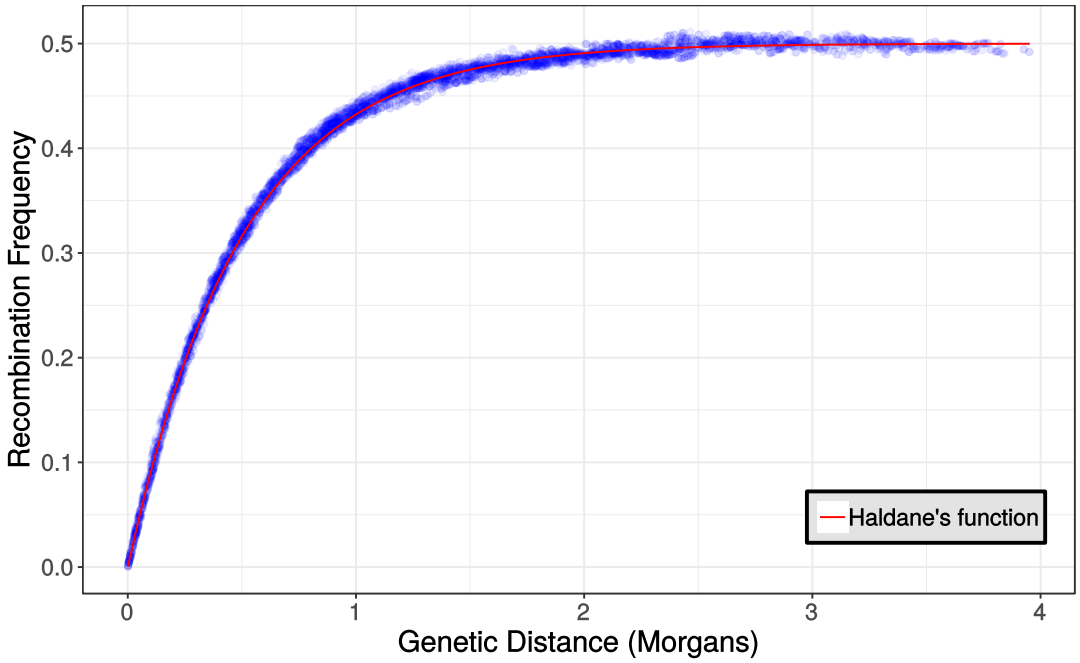

Figure 2.1. Pairwise recombination frequency of 100 markers in a linkage group of four Morgans using the count-location model. The function line represents the recombination frequency on genetic distance, according to Haldane's linkage function (Equation 2.1)

2.2.2 Simulating other procedures in plant breeding

To simulate a breeding program, we need to have a set of codes besides the one to simulate the meiosis phenomena, discussed in section 2.2.1. Especially when we are interesting to evaluate the effects of selection, we need to define the genotypic values of the genotypes under consideration. To do that, we can define a genotypic value for each loci using some probability distribution. An exponential distribution, for example, has been used on several papers (Bernardo, 2001; Santos *et al.*, 2002) as it is a good approximation of a quantitative trait (Goddard, 2009). Using this distribution, we assume that many loci have a small effect and a few loci have large effects. As in the past, some programming languages did not provide good generators of random numbers for several distributions, a few papers also present a geometric series in order to reproduce the same shape of an exponential distribution (Bernardo, 2003; Bernardo and Yu, 2007; Bernardo, 2009). However, we can also find papers that simulate the gene effects by also using a normal distribution, centered in zero (Zeng and Cockerham, 1991). In this case, most of the genes present effects close to zero and a few of them have large effects, negative and positive. One way to calculate the genotypic values is to multiply the realizations from the probability distributions by the allele dose (e.g., 0, 1, and 2, for a diploid species), and, finally, add all the loci up.

So far we have discussed only about the genetic part of the simulation. It turns out that some times we are interested in attribute an error effect into our observations, to compose a phenotypic value, which is the sum of the error effect and the true genotypic effect, that comes from the genotype (allele dose) and the gene effects. In this case, the most usual way to do it is control the magnitude of this variance component by fixing the heritability of interest and calculating the genotypic/genetic variance directly from the true value (Bernardo, 2001, 2009; Cruz, 2001), according to the expression below:

$$h^2_A = \frac{\sigma^2_G}{\sigma^2_G + \sigma^2_\varepsilon} \leftrightarrow h^2_A(\sigma^2_G + \sigma^2_\varepsilon) = \sigma^2_G \leftrightarrow \sigma^2_\varepsilon = \frac{1 - h^2_A}{h^2_A}\sigma^2_G \tag{2.2}$$

h^2 refers to the broad heritability sense, if we consider the genotypic variance (variance of the additive

and non-additive effects) or narrow sense, if we consider the genetic variance (only the variance of additive effects); σ^2_G refers to the genotypic or genetic variance, depending on the nature of the effects, and can be calculated as the variance of the genotypic/genetic values; finally, σ^2_e refers to the error variance.

The error effects usually are modeled using a normal distribution $\varepsilon \sim N(0, \sigma^2_e)$, and the realizations of the distribution are added up with the genotypic values. Also, we can simulate the spatial dependency of these errors by using some correlation structure (e.g., autoregressive structure), sampling values from a multivariate normal distribution, as presented by Mramba *et al.* (2018).

It is important to mention that the procedures described here, refers to an *In silico* simulation, in which the researcher entirely create the data set, from the meiosis and genetic effects to the error attribution to compose the phenotypic values. However, there are several other ways to build a simulation and investigate what the researcher is looking for. As presented in section 3, we also used a data-based simulation in which we started the simulation based on a real data set, using the estimated marker effects as true genetic effects for each locus and also using the population structure present on the real data to start the simulation process. This is also an interesting approach since we can use real data set to create a simulation, sometimes, closer to one reality, and also create different scenarios (e.g., different parameter values or methods) for testing. We can also found some references on literature that uses this approach (Moehring *et al.*, 2014; Borges *et al.*, 2019).

2.2.3 Software available for simulation in plant breeding

Several software are currently available to simulate breeding schemes and test new techniques, helping the breeder to preliminarily investigate the efficiency of adoption of these procedures. As mention in section 2.2, simulation is especially useful to test procedures that were recently proposed, or it takes a lot of resources (money and time) to test. One of these software is QU-GENE, presented with more details and exemplified its application in a computer simulation review published by Wang *et al.* (2007). The software has been used to compare two selection strategies (Wang *et al.*, 2003), to study the effects on selection of dominance and epistasis (Wang *et al.*, 2004), and to optimize marker assisted selection to efficient pyramid multiple genes (Wang *et al.*, 2007).

Also, the *R* environment (R CORE TEAM, 2015) counts with a few packages that help the researcher to build breeding schemes. The *Meiosis* package (Mueller and Broman, 2017), for example, allows the researcher to easily produce crosses, self individuals, extract gametes from an individual, and produce double haploid lines. However, the backbone of this package, especially the function that mimics the meiosis process, uses another package, *simcross* (Broman, 2019), which is written in *C++*, being much faster than a code written directly in *R* language. However, before start using the features cited above, the researcher needs to define some parameters related to the meiosis process, for example: the size of each chromosome, the interference parameter, which can assume values between 0 and 1, if the parameter is equal to zero, the probability of two crossing-overs happens close to each other is the same of happening far distant in the chromosome. In this case, the model is called count-location model, and we have the scenario presented in figure 2.1. If we consider some value between 0 and 1, the model will allow less frequently the occurrence of crossing-over close to each other. This model is called Stahn model (Copenhaver *et al.*, 2002) and the count-location model can be considered a special case of the Stahn model. We had used this package to simulate the meiosis phenomenon in the manuscript presented on section 3. Also, the package *isqg* (Toledo and CIMMYT, 2018), presents similar functions, but its implementation in *C++* and the *S4* class, allow the package to have an incredible performance. The package *isqg* also has an interesting feature of enabling the researcher to create your own meiosis model and easily implement inside *isqg*, running using the *C++* and *S4* machinery of the package.

The package *BreedingSchemeLanguage* (*BSL*) (Yabe *et al.*, 2017) not only present the base

functions similar to the *Meiosis* and *isqg*, but also several others related to the creation automatically of population and progenies, the attribution of an experimental error based on the heritability and number of replication, etc.

Recently, the package *AlphaSimR* (Gaynor, 2018) was also released on CRAN (the official web server where we can download packages directly from *R*). This package also has some functions beyond the basic functions for modeling common tasks in an animal and plant breeding program, like a function to perform genome editing, compute additive and dominance variance, make designed crosses, etc. Also, *AlphaSimR* uses a Markovian Coalescent Simulator (MaCS) for fast simulation of biallelic sequences according to a population demographic history (Chen *et al.*, 2009).

It is important to mention that the main advantage of using this kind of package inside an *R* environment is that we can also use several other packages to run statistical models, manipulate data, summarize results, produce plots, and several other procedures, all in the same place. This helps a lot, as we can automate all the simulation procedure also using different programming languages to speed up some part of code that is struggling in time and/or memory.

3 LONG-TERM SELECTION OF BIPARENTAL CROSSES: A COMPARISON AMONG GENOMIC AND PHENOTYPIC METHODS

3.1 Introduction

Plant breeding is one of the main responsible for the supply of food and other essential products for humankind. The cultivars obtained from private and public breeding programs significantly increased yield, adaptation, and quality (Vencovsky *et al.*, 2012). However, for most crops, breeding is an expensive and costly process, as it can take a lot of resources and several years to develop a new cultivar (Watson *et al.*, 2018). Also, the most important traits in breeding are complex quantitative traits, thus controlled by many genes and highly influenced by the environment (Falconer and Mackay, 1996; Lynch and Walsh, 1998). In this way, it is extremely important to adopt strategies that maximize the probability of finding genotypes that are better, considering several agronomic traits, than the previously released cultivars.

An essential step in every breeding program refers to the selection of parents and crosses (Fehr, 1987; Ramalho *et al.*, 2012b). The methods used to select crosses can be classified in two categories: the first involves the procedures that use information from the parents only and the second refers to all the methods that use the performance of the progenies, demanding the breeder to produce the crosses and evaluate these populations (Baenziger and Peterson, 1991). However, due to the usual high number of candidate lines to be used as parents, it is unfeasible to produce all the possible crosses and evaluate them, being important the adoption of methods that allow the choice, *a priori* of the best crosses (Witcombe *et al.*, 2013). Among these methods, we can highlight the use of parental means, coancestry coefficient, genetic distances and more recently the genomic prediction with simulated offsprings. Until a few years ago, the applications of genomic prediction, proposed by Meuwissen *et al.* (2001), was restricted to the prediction of genotypic and breeding values of unphenotyped individuals. Just with the paper from Bernardo (2014) the method also became useful for the selection of crosses.

Further studies, like Mohammadi *et al.* (2015) and Tiede *et al.* (2015), also evaluated the usefulness of this method and also facilitated its implementation in a breeding program through a package in the *R* environment (R CORE TEAM, 2015). The method was studied in a short-term manner by Lado *et al.* (2017) using a wheat dataset. Also comparing this method with the phenotypic selection (mean breeding values of the parents), the authors showed that there are no relevant differences on what crosses should be selected with the GS adoption when the yield was used. However, significant differences were reported for grain protein content, which is a higher heritability trait. However, some concerns with this method were also described by Lehermeier *et al.* (2017). The use of a random regression model shrinks the marker effects, and these effects are used to estimate the breeding values of the simulated progenies. The variance is then calculated as the variance of the breeding values, which are shrunk, as it is also based on the marker effects, being the variance underestimated, especially when the shrunk effect is large (when traits present low heritability). Thus, it is important to find ways to circumvent this issue and obtain a better estimation of the variance of each cross, being possible to select crosses more accurately. Otherwise, with an underestimation of the variance, the selected crosses tend to be the same as those selected based only on the mean breeding values of the parents.

In this method, with the adjusted means and the marker data set, the breeder fits a regression model to estimate the marker effects (e.g., Ridge regression BLUP or any of the different Bayesian methods), and with a genetic linkage map, simulates several progenies from every possible cross from a list of candidate parents. The computer codes simulate these progenies involves the reproduction of the meiosis phenomenon, crossing-over, gamete formation, and fertilization. With the simulated progenies, it is possible to calculate different parameters, as the mean and variance, providing useful information

to the breeder. The probability of success in obtaining superior genotypes in a cross is a function of the breeding value of the parents and also the variance that comes from the complementarity between parents (Bernardo, 2014). The genomic prediction procedure concatenates both aspects.

The standard application of genomic selection in a long-term was studied by Jannink (2010). In his paper, the author described how recombination, population size, heritability influence the prediction accuracy. The results also showed that losing favorable alleles that are in weak linkage disequilibrium with markers is perhaps inevitable when using genomic selection. Thus, some strategies need to be adopted to minimize this rate of loss. Jacobson *et al.* (2015), on the other hand, compared the impact of genomic selection within a biparental cross instead of the standard phenotypic selection, commonly used by breeders. The results presented by these authors showed that genomic selection can significantly increase the similarity among lines when a high intensity of selection is used ($i = 5\%$), but not with other values for intensity, meaning that genomic selection causes only a minimal loss of genetic diversity within biparental populations. Even though this is a promising method to select crosses in breeding programs, as far as we know, there are no studies about the performance of this method in a long-term selection process. It is important to mention that an optimum method should balance selection and maintenance of genetic diversity, in a way that future gains can also be achieved (Gorjanc *et al.*, 2018). Especially for methods routinely used, cycle after cycle, in a breeding program, its impacts in a long-term way should be studied prior to its implementation.

For this kind of study, that takes several years to elucidate, the use of computer simulation can be considered a potential tool to speed up the process to obtain information. Like a bridge between the theory and experimentation has become a powerful way to achieve results in scientific research, providing not only preliminary validation of theories but also orientation about the adoption of technologies and techniques (Li *et al.*, 2012). Using this tool, parameters and restrictions are set up in a way that the desired effects can be studied conveniently in a specific model species (Cruz *et al.*, 2013). Thus, simulation consists of a system that mimics the reality, aiming to study what would happen in the real system if changes in some specific factor were made (Dachs, 1988).

The goals of this study were to evaluate, under different scenarios, how genomic prediction with simulated offsprings performs compared with the traditional phenotypic selection across several cycles of breeding, characterizing long-term selection. Also, different aspects of the adoption of this method were evaluated. We used *in silico* and data-based simulation.

3.2 Material and Methods

In this study, we used two different approaches to compare, in a long-term manner, how genomic approaches perform in relation to the standard phenotypic procedure used by breeders. First, we simulated a fully *In silico* wheat breeding program applying different strategies to select crosses. Also, we performed simulations with the same scenarios but using a real wheat dataset with 1465 lines from the National Agricultural Research Institute (INIA), from Uruguay. Both approaches were simulated during ten cycles of selection. Statistics related to the genetic gain and maintenance of genetic diversity were calculated.

3.2.1 Genetic model and *In silico* wheat breeding program

We simulated a wheat breeding program with the capacity to evaluate 1000 lines per cycle of a quantitative trait controlled by 100 genes (g). Also, 5,000 thousand molecular markers (m) were simulated. All the loci ($g + m$) were randomly distributed along a simulated wheat genome of size 31.58 Morgans and 21 chromosomes based on the linkage map presented by Lado *et al.* (2017).

The allelic effect of each gene (α_j) was sampled from an exponential distribution of rate 1 ($\lambda_{exp} = 1.0$). The use of this distribution allows us to have a few QTL (*Quantitative trait Loci*) with large effect and several with small effects, mimicking a regular quantitative trait (Bernardo, 2003, 2014; Bloom *et al.*, 2015; Kearsey and Farquhar, 1998). Non-additive effects were not considered in this study. Thus, the breeding value of the *i-th* individual was calculated as the sum of the genotypic effect of each gene.

The *In silico* simulation started with a cross between two completely contrasting inbred lines. Their resulting F_2 generation, with one thousand individuals, were then randomly crossed for 35 generations, achieving the linkage disequilibrium (LD) between markers found in Laidò *et al.* (2014), in a real dataset (Figure S9)

Double haploid (DH) lines were then generated from this random crossed population. This population of DH will be referred to as base population. To form the cycle zero of breeding, DH lines were randomly sampled from the base population to be the parents of n_{POP} populations. For each population, n_{Prog} progenies were also obtained as double haploid lines and as Recombinant inbred lines (RILs) in one of the Phenotypic methods. The following scenarios were considered for n_{POP} and n_{Prog} (in parenthesis): 5(200), 20(50), 50(20) and 100(10).

By the end of the cycle zero, each strategy of cross selection was applied. Three phenotypic procedures were used. The first, in each cycle, 10% of the best individuals were selected, and pairs of parents were randomly sampled from the selected individuals to form n_{POP} populations. The selection of individuals was based on the phenotypic value. The second (using double haploid lines) and third (using RILs) method used a chain selection of parents scheme, in this case, after order the genotypes based on their phenotypic value, crosses were made between the first and second, second and third best genotypes, and so on, until n_{POP} crosses.

The phenotypic value was calculated as the sum of the genotypic value and a random error sampled from a normal distribution, $\varepsilon \sim N(0, \sigma^2_\varepsilon)$, in which the σ^2_ε were based on the heritability of the trait, following the expression:

$$h^2_A = \frac{\sigma^2_G}{\sigma^2_G + \sigma^2_\varepsilon} \leftrightarrow h^2_A(\sigma^2_G + \sigma^2_\varepsilon) = \sigma^2_G \leftrightarrow \sigma^2_\varepsilon = \frac{1 - h^2_A}{h^2_A}\sigma^2_G \tag{3.1}$$

h^2_A is the narrow sense heritability; σ^2_G is the genetic additive variance, calculated as the variance of the breeding values; σ^2_ε is the error variance. We considered the following values for heritability: 0.3, 0.5, and 0.8.

For the genomic prediction using simulated offsprings, on other hand, every possible combination among the candidate parents was predicted. The prediction was performed using the package PopVar (Mohammadi *et al.*, 2015) that follows the method proposed by Bernardo (2014). In this method, simulated offsprings are generated for each possible combination of parents, and with these simulated offsprings and marker effects, we can calculate the breeding value of each offspring. This information allows us to have a prediction of the mean and genetic variance of each population and select crosses based on this. To calculate the marker effects, a ridge regression BLUP (RR-BLUP) were adjusted as described by Meuwissen *et al.* (2001) and Bernardo and Yu (2007), as it is faster than Bayesian methods and it usually provides similar prediction accuracies in plants (Thavamanikumar *et al.*, 2015; Xavier *et al.*, 2016). The model can be defined as follows:

$$y = 1\mu + Xm + \epsilon \tag{3.2}$$

y refers to the vector $N \times 1$ of phenotypic observations of the N lines; 1 is a vector $N \times 1$ with 1's; μ is the intercept; m refers to the vector of random markers effects, in which $m \sim N(0, \sigma^2)$; X refers to the incidence matrix that relates y and their genotypes.

The best populations were then selected based on the mean of the 10% superior breeding values of a population. The scheme on Figure 3.1 shows each step of the simulation:

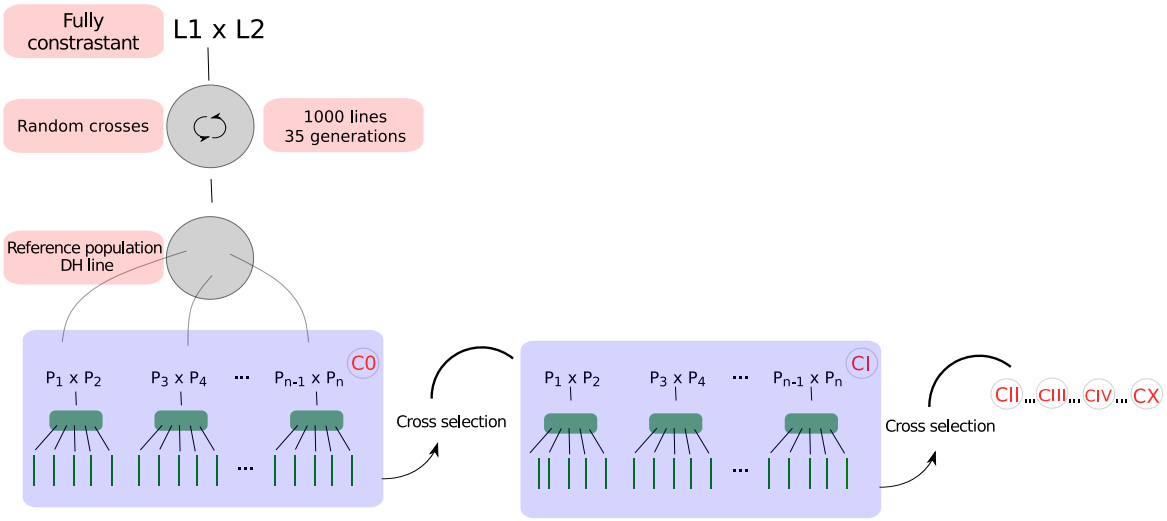

Figure 3.1. Scheme of the *In Silico* simulation.

To avoid the higher losses of variability we also simulated a scenario with a restriction on the number of times a genotypes can be a parent on the selected crosses. For this, we consider that some genotype can be a parent in no more than three crosses. According to the scenario where we did not applied this restriction, several crosses with a common parent can be selected due the performance of some genotype. This can cause loss of alleles due the bottleneck effect, reducing the gain of selection in a long-term. Also, a random selection of parents were simulated aiming to measure the loss of variability due to genetic drift.

The simulations were performed using the package Meiosis (Mueller and Broman, 2017) from *R* (R CORE TEAM, 2015). The package is based on the Stahl model (Copenhaver *et al.*, 2002), with the interference parameter being an integer. However, a count-location model was used to simulate the meiosis phenomenon. In this model, the number of crossing overs in each chromosome can be defined as Poisson distributed random variable with the average number of crossing-overs equals to the size of the chromosome ($\lambda_{poisson}$ =*Chromosome size*). Also, the position of crossing overs is sampled from a uniform distribution; therefore, the crossing overs are randomly distributed and do not take into account a possible interference (Sun *et al.*, 2011).

For each scenario (combination among the values of each parameter) we ran 20 simulations to have stable estimates as postulated by the Law of Large numbers (Suess and Trumbo, 2008).

3.2.2 Data-based simulations

Aiming to have a population structure as close as a real breeding program we also performed simulations using a real wheat dataset with 1465 inbred lines from the wheat breeding program conducted by the National Agricultural Research Institute (INIA), in Uruguay. These lines were phenotyped for grain yield during five years (2010-2014) in four locations (Dolores, Durazno, La Estanzuela, Ruta2, and Young) and four different sowing dates in La Estanzuela, as described in Lado *et al.* (2016), totalizing 35 environments. These lines represent the advanced inbred lines coming from the breeding program, as they belong to Preliminary Yield Trials (F_7 generation), Advanced Yield Trials (F_8 generation) and Elite Yield Trials (F_9 generation).

The following model was adjusted to estimate the genotypic least square means:

$$y_{ijklm} = \mu + g_i + e_j + t_{k(j)} + r_{l(jk)} + b_{m(ljk)} + \epsilon_{ijklm} \tag{3.3}$$

where μ is the intercept; g_i is the fixed effect of i^{th}; e_j is the fixed effect of j^{th} year–location combination; $t_{k(j)}$ is the random effect of the k^{th} trial nested within the jth year–location, $t_{k(j)} \sim N(0, \sigma_t^2)$; $r_{l(jk)}$ is the random effect of the l^{th} replicate nested within the k^{th} trial and j^{th} year–location, $r_{l(jk)} \sim N(0, \sigma_r^2)$; $b_{m(ljk)}$ is the random effect of the m^{th} incomplete block nested within the l^{th} replicate, k^{th} trial, and the j^{th} year-location, $b_{m(ljk)} \sim N(0, \sigma_b^2)$; and ϵ_{ijklm} is the residual error. The best linear unbiased estimations were estimated using the nlme package (Pinheiro *et al.*, 2018) in R statistical software (R CORE TEAM, 2015).

The genotypic information for these lines was obtained using the GBS method. The SNP calling was conducted using Tassel pipeline as described in Lado *et al.* (2013). Markers with minor allele frequency smaller than 0.01 and call rate more than 80% were removed from the data set. A total of 81,999 SNPs were keeped after this process. Of these, 3,844 SNPs were mapped.

In this simulation approach, we considered the real inbred lines as being the reference population. Based on the yield and molecular markers of these genotypes, we adjusted an RR-BLUP model to calculate the marker effects. After that, g markers were randomly picked to be considered as genes and their estimated effects as true genetic effects. The following on Figure 3.2 shows how the reference population and the simulation were set up:

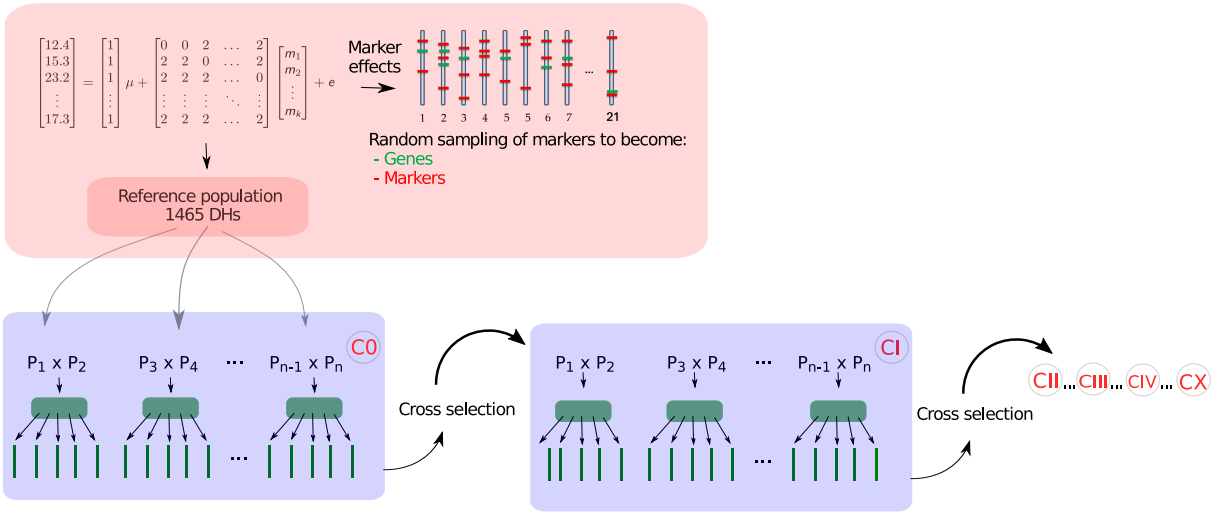

Figure 3.2. Scheme of the data-based simulation.

The process after the creation of the reference population it is the same as presented in the *In Silico* simulation and also goes from cycle one to five. But it is important to highlight that the reference population is recreated for each Monte Carlo simulation in the *In Silico* simulation, differently from the Data-based simulation, which is based on a single real panel of genotypes; however, in each simulation, we ran a new randomization process to set what markers will be considered as genes.

3.2.3 Evaluation criteria

Some criteria were used to measure the performance of the different methods in different scenarios. The first refers to mean breeding value of the inbred lines derived by the end of each cycle.

The changes in genetic diversity throughout the cycles were measured using a few criteria. The first is the Nei's diversity coefficient (Nei, 1987), calculated based on $H_e = 1 - p^2 - (1-p)^2$, that measures the expected heterozygosity if the individuals were randomly crossed and it is based entirely on the allele

frequencies p and q of each locus. The value recorded was the average of all loci. The second criterion refers to the number of loci that lose the benefit alleles from the entire population, in each cycle. We considered the benefit allele as the one responsible to increase the breeding value of the individuals. The alternative allele, in another hand, responsible to decrease the breeding value, were considered the non-benefit allele. Also, the total number of alleles lost were calculated based on the number of monomorphic loci. Fourth, the average distance between biparental populations was calculated based on the F_{ST}, proposed by Wright (1949) and detailed by Weir and Cockerham (1984), according to the following estimator:

$$F_{ST} = \frac{H_T - \bar{H}_S}{H_T} \tag{3.4}$$

In which: H_T represents the expected heterozygosity if all subpopulations were mixed in a single one and their individuals were randomly crossed; $\bar{H}_S$ refers to the average expected level of heterozygosity if the individuals of each subpopulation were randomly crossed.

We also calculated the efficiency of conversion of genetic diversity into genetic gain, according to Gorjanc *et al.* (2018). This is based on a linear regression between the genetic gain ($y_t = \frac{\mu_{a_t} - \mu_{a_5}}{\sigma_{a_5}}$) on the lost genetic diversity ($x_t = 1 - \frac{\sigma_{a_t}}{\sigma_{a_5}}$), i.e. $y_t = a + bx_t + e_t$, where b is the slope of the regression and the efficiency of conversion.

These measures were used with the objective to describe changes in the populations under different aspects throughout cycles of breeding. It is important to highlight that after run all the simulations for each scenario and method, we calculated the mean and the quantiles 5% and 95% of the Monte Carlo distribution for each criterion, having a better idea on how is the dispersion of the estimates.

Finally, aiming to understand how much of what was lost is important to improve the genetic value of the general trait, we also calculated the variance proportion for the wheat data-based simulation. For each cycle of breeding we calculated the square of the average effect of each loci, then separate these loci into three categories: loci that fixed the benefit allele (the one that increase the breeding value of the genotype), loci that fixed the non-benefit allele (the one that decrease the breeding value of the genotype), and loci that are still segregating.

3.3 Results

3.3.1 *In silico* simulation

The *In silico* simulation process can be divided into two parts: the creation of an evolutionary process of the species under consideration and second, the breeding program by itself. To mimic this evolutionary process, we started by crossing two constrasting inbred lines and making several random crosses among one thousand individuals. We did an *ad hoc* simulation to find out the number of generations of random crosses needed to achieve a level of linkage disequilibrium (LD) similar to the panel of wheat genotypes presented in Laidò *et al.* (2014). Figure S9 shows that 35 generations are sufficient to have a similar amount of LD, compared to the panel from the paper cited above. Also, the decrease in the LD levels starts to be very small after 35 generations of random crossing. Similar to what was presented by Laidò *et al.* (2014) and Breseghello and Sorrells (2006) we used the percentile 95 of the distribution of Unlinked-r^2, beyond which LD is caused by genetic linkage, as a measure of the level of LD in the population.

After the definition of the evolutionary process, the *In silico* breeding program started and went on for ten cycles of breeding. The results showed that the best method depends on the heritability, number of population, and what cycle we are taking into consideration. Considering a scenario of low heritability ($h^2 = 0.2$) and high number of population ($N_{POP} = 100$), the genomic methods outperform

the phenotypics in a long-term selection (after ten cycles of breeding), in terms of average breeding value of the progenies. The best method, in this case, was the genomic with restriction on the number of times a genotype is a parent (Genomic-SO-Const), being it superior in 0.9%, 5.38%, and 7.37% to Genomic-SO, Genomic-GEBV, and Phenot-Ch-DH, respectively, as can be seen in Figure 3.3 (Top figure).

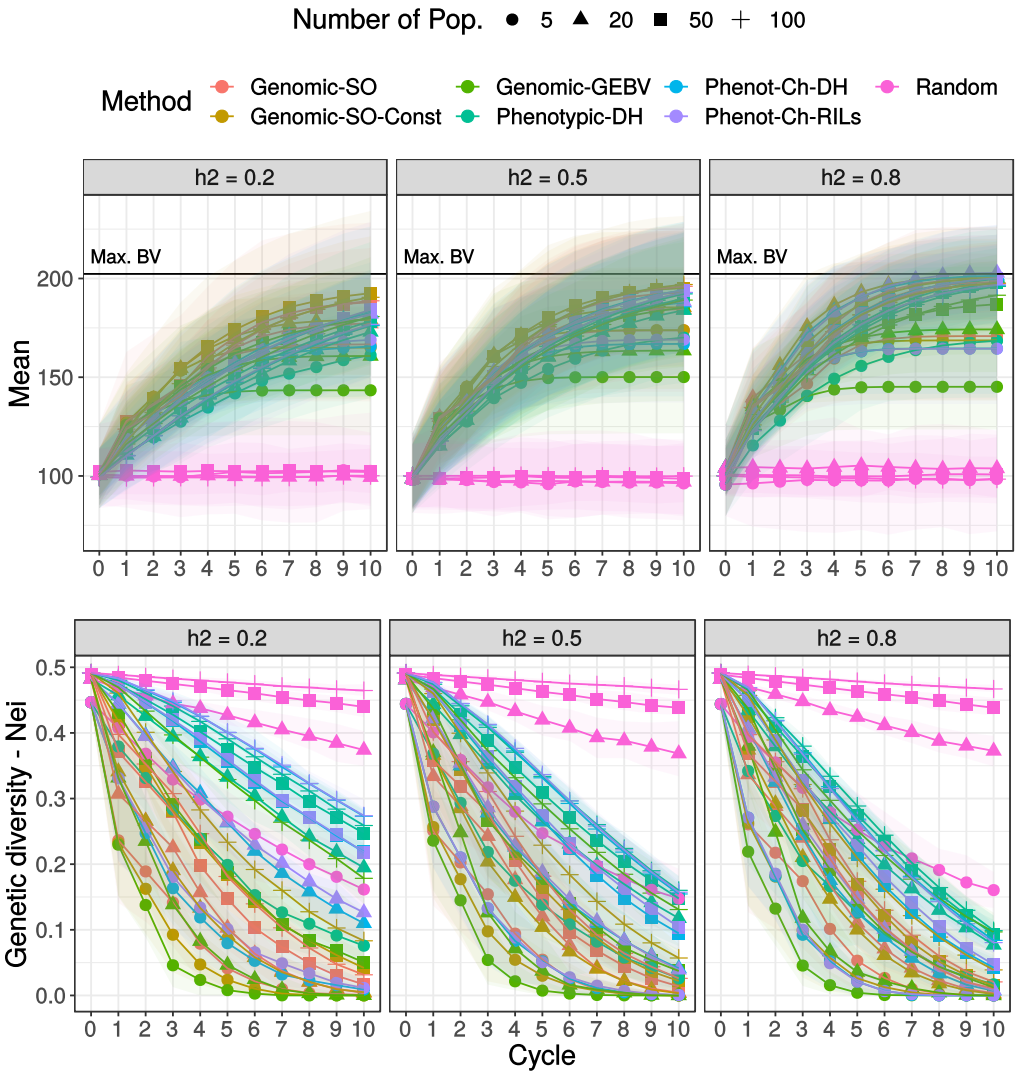

Figure 3.3. Mean breeding values and genetic diversity of the *In silico* simulation during ten cycles of selection of crosses using the following methods: genomic prediction with simulated offspring (Genomic-SO), genomic prediction with simulated offspring and restriction (Genomic-SO-Const), mid-parent GBLUP (Genomic-GEBV), phenotypic selection (Phenotypic-DH), phenotypic chain selection using double haploid lines (Phenot-Ch-DH), phenotypic chain selection using recombinant inbred lines (Phenot-Ch-RILs), and random crosses. Different scenarios for heritability (h^2) and number of population (N_{POP}) were considered. The points represents the average value of 20 simulations, while the shadow of different colors represent the quantile 5% and 95%. The horizontal line on top figure represents the maximum breeding on average, across 20 simulations.

For the scenario of high heritability ($h^2 = 0.8$) and again high number of population and considering a long-term selection (after ten cycles), the best method is still the Genomic-SO-Const, but the superiority and the rank changes. In this case, Genomic-SO-Const is superior in 0.18%, 0.47%, and 0.59% compared to Genomic-SO, Phenot-Ch-RILs, and Phenot-Ch-DH, respectively. The Genomic-GEBV was the worst method of selection, being ahead just of the random method and 3.22% lower than Genomic-SO-Const.

Considering a short-term selection (after one cycle of breeding) and the scenario of low heritability ($h^2 = 0.2$) and high number of population ($N_{POP} = 100$), the differences are more pronounced, being the Genomic-SO the best method and superior to Genomic-SO-Const (5.1%), Genomic-GEBV (6.4%), and Phenotypic-DH (12.0%). For a high heritability scenario ($h^2 = 0.8$), the rank is the same but with slight differences in the magnitude of the performance among the methods. In this case, the Genomic-SO was superior on 7.0%, 8.9%, and 9.3% to the Genomic-SO-Const, Genomic-GEBV, and Phenotypic-DH, respectively.

When we take a closer look at a medium number of population ($N_{POP} = 20$), we start to notice several differences on the rank of the methods. Considering the last scenario of heritability, for example, the best method starts to be the Phenot-Ch-RILs, followed by Phenot-Ch-DH, and just in the third position, the Genomic-SO-Const. However, they perform similarly, being the Phenot-Ch-RILs superior on 0.7% and 1.3% to the two other methods. For a low heritability scenario ($h^2 = 0.2$), the ranking is the same, but with a few differences on the magnitude. In this case, the Phenot-Ch-DH overperform in 0.2% and 1.9% the Phenot-Ch-RILs and Genomic-SO-Const.

Also, as presented in Figure 3.3, not surprisingly higher heritability provides higher genetic gains after ten cycles of breeding. Considering that the perfect individual (homozygous for the benefit allele in all loci), on average across simulations, presented a breeding value of 199.65, the Genomic-SO-Const provided individuals, also on average, with breeding values 190.46, meaning that 95.3% of the maximum breeding value was achieved. Genomic-SO, Genomic-GEBV, and Phenot-Ch-RILs were capable of producing individuals with 188.7 (94.5%), 180.7 (90.5%), and 176.8 (88.5%) on average, compared with the perfect individual.

Visually, it is easy to notice that the Genomic-GEBV method provides the lowest average breeding value in a long-term selection and this is also clearly due to the big loss of variability right in the first five cycles of breeding. For the extreme inferior case of number of population, after ten cycles, the genetic diversity was 17.9%, 37.1%, 79.3%, and 85.7% lower in this method compared to Genomic-SO-Const, Genomic-SO, Phenot-Ch-DH, and Phenot-Ch-RILs, respectively.

Still considering aspects related to the genetic diversity (Figure 3.3), genetic variance, and F_{ST} (Figure S12), the number of population also has a large impact, especially on the scenarios of low heritability. Considering only the genetic diversity, estimated according to Nei's expression, the genomic methods provides a faster decrease than the phenotypic methods, according also to Figure 3.3 (Bottom figure), which is expected as the selection tends to increase frequencies of the benefit allele towards the selection direction, decreasing the genetic diversity.

Another important aspect to consider is the number of genes that become monomorphic throughout the cycles of breeding and also, how many of these genes lost, by drift or linkage drag, the benefit allele. Figure S10 presents these results for each method of selection, heritability, and number of population. The random method presents how much alleles we are losing by drift only, while the other methods also take selection into account.

Similar to the results presented on Figure 3.3 and Figure S12, a small number of populations causes a high number of alleles lost and the main reason for this is drift, as can be seen in the random method. For higher number of populations, drift it is almost inexistent, being the losses caused only by selection forces. Heritability does not play an important role in the number of monomorphic genes when genomic selection is applied. When a high number of population is used ($N_{POP} = 100$) a very low number of benefit alleles is lost on both methods, genomic and phenotypic.

Figure 3.4 presents the efficiency of conversion of genetic standard deviation (the square root of genetic variance) into genetic gain, calculated according to (Gorjanc *et al.*, 2018). The angle formed by the arrows is the efficiency of conversion. The ideal scenario is a long arrow with high angle, meaning that a high mean was achieved without compromise the genetic variance for next generations of breeding. We

can see that the phenotypic methods provided higher efficiency of conversion, even though a lower average breeding value at the end of cycle ten was achieved, the low lost of $\sqrt{\sigma^2_G}$, compensates the process. ..

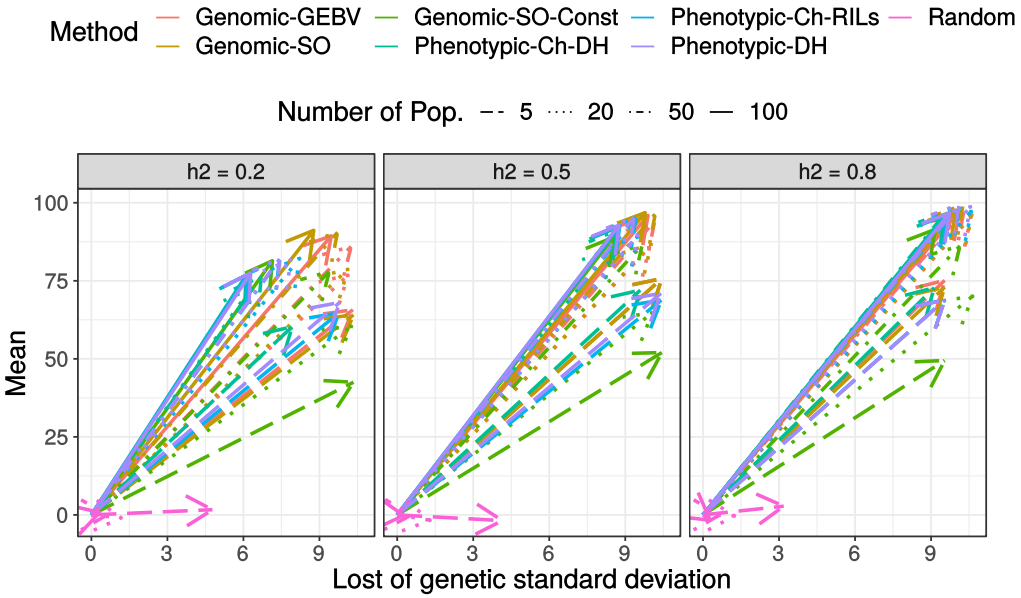

Figure 3.4. Efficiency of conversion of genetic standard deviation ($\sqrt{\sigma^2_G}$) into genetic gain according to Gorjanc *et al.* (2018) for the *In silico* simulation for all combinations of heritability (h^2), number of population, and method of selection. The higher the angle of the lines, higher is the conversion

In this procedure, no significant differences were found when we consider short-term and long-term selection, looking to different generations of breeding. In other words, considering only the average breeding value, except for minor changes, the best method on early cycles of breeding (e.g., cycle 1) is still the best method on late cycles of breeding (e.g., cycle 10). However, with higher heritability, we can see that the differences among the mean are shrunk with the advance of the cycles. On cycle 1, for example, Genomic-SO-Const was the best method followed by Genomic-SO, Phenot-Ch-RILs, and Phenot-Ch-DH, being superior by 7.0%, 8.9%, and 9.2% respectively. However, looking to cycle ten, we see that the Genomic-SO-Const provided higher mean, followed by Genomic-SO, Phenot-Ch-RILs, and Phenot-Ch-DH, being superior in 0.1%, 0.4%, and 0.5%.

3.3.2 Wheat data-based simulation

Figure 3.5, presented below, shows the distribution pattern of the square of the correlation between markers (r^2). The red horizontal line indicates the 95% percentile of the distribution of r^2 (0.1746), showing that the level of LD present in this population is slightly superior to the average LD level across simulation in the *In Silico* approach (LD (r^2)= 0.12). This result might be due to the structure of population that these genotypes come from. The genotypes from Laidò *et al.* (2014) belongs to a wide panel of 230 genotypes, consisting of 128 durum wheat varieties plus 102 wild and domesticated accessions. In the other hand, the 1465 lines present on the panel used on this research came from a wheat breeding program from the National Agricultural Research Institute (INIA), from Uruguay. Thus larger diversity was expected on the Laidò *et al.* (2014) panel than the one from INIA.

The results from the wheat data set were similar to the *In Silico* simulation. A high number of population ($N_{POP} = 100$) again provided higher gains, as we can see on top of Figure 3.6. As can be notice it on the bottom part of Figure 3.6, the loss of diversity is much higher when we use a few crosses. As the random crosses method also presented the same pattern, the cause of this loss of diversity is drift.

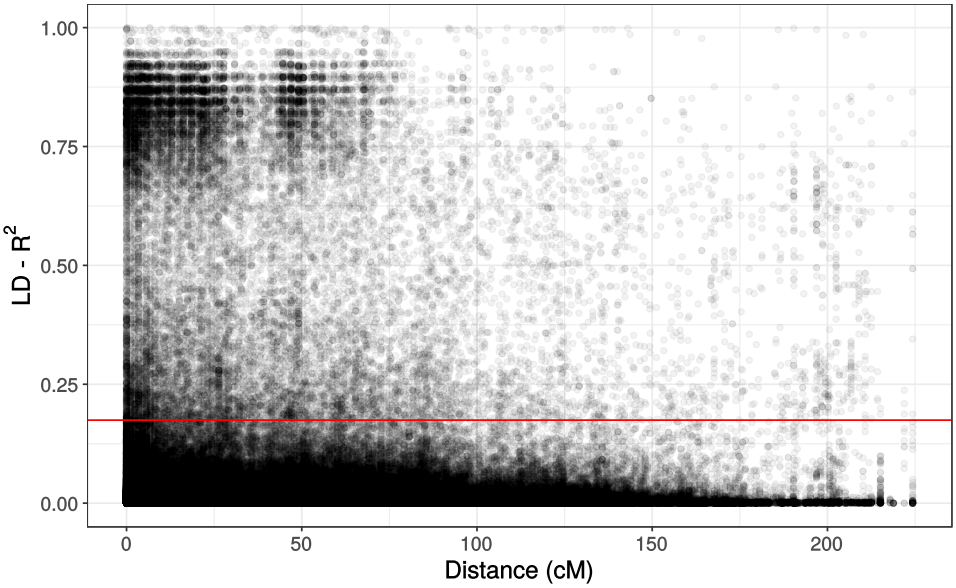

Figure 3.5. Overview of the distribution of the LD parameter r^2 of the intrachromosomal pairs in the whole collection, according to the distance in centiMorgans. The horizontal line indicates the 95% percentile of the distribution of r^2.

In this method no selection is applied, so the genotypes are picked by chance and thus all the loss of diversity is caused by this phenomena. The higher loss of benefit alleles (Top of Figure S11) and total number of alleles (Bottom of Figure S11) also show this.

The heritability (h^2) also impacts significantly the genetic gain and other aspects related to diversity. As expected, higher heritability provides higher gains, but also higher loss of variability. However, when selection is applied, the loss of variability is due not only by drift but also because of the conversion of genetic variance into genetic gains.

It is important to notice the interactions that might occur among all the variables under consideration. In a short-term selection, both methods that use genomic selection are way more efficient than phenotypic selection to select the best crosses. Considering cycle 1, the highest heritability ($h^2 = 0.8$), and the highest number of population ($N_{POP} = 100$), genomic selection using simulated offspring (Genomic-SO) presented the highest mean (91.51), being 19.2% 25.3%, 26.9%, 27.1%, and 27.3% superior compared to the selection using Genomic-SO-Const, Genomic-GEBV, Phenotypic-DH, Phenot-Ch-DH, and Phenot-Ch-RILs, respectively. However, this scenario changes when we consider a long-term selection. If we look at the cycle 5, for example, the order from the best to the worst method in terms of average breeding values are: Genomic-SO, Genomic-SO-Const, Phenot-Ch-RILs, Phenot-Ch-DH, Genomic-GEBV, and Phenotypic-DH. Additionally, if we consider the cycle 10, the order becomes: Phenot-Ch-RILs, Phenot-Ch-DH, Phenotypic-DH, Genomic-SO-Const, Genomic-GEBV, and Genomic-SO, being the Phenot-Ch-RILs 5.9% superior to the Genomic-SO.

However, when we consider a scenario of low heritability ($h^2 = 0.2$), the methods that use genomic selection prevails either on a short-term selection (after one cycle of breeding) and long-term selection (after ten cycles). For the short-term selection, Genomic-SO prevails being the best method and superior on the magnitude of 36.79% and around 91% compared to Genomic-GEBV and the other methods but again considering only the average breeding values of the progenies. In a long-term selection, the Genomic-SO-Const starts to be the best method, in front of Genomic-GEBV (2.54%), Genomic-SO (6.9%), and Phenot-Ch-RILs (13.7%)

Considering the genetic diversity (Bottom of Figure 3.6), we can see that higher number of

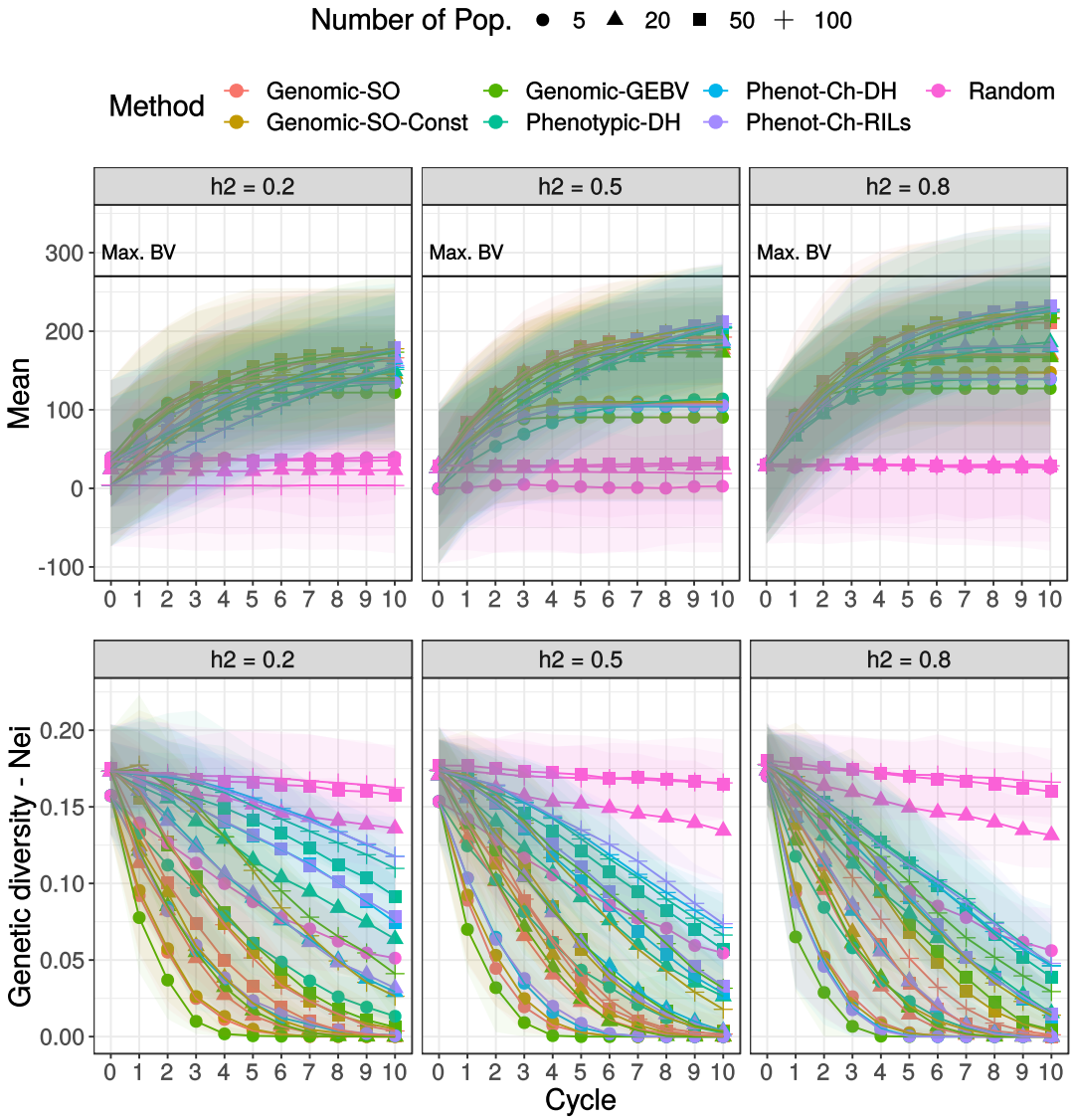

Figure 3.6. Mean breeding values and genetic diversity of the data-based simulation during ten cycles of selection of crosses using the folloowing methods: genomic prediction with simulated offspring (Genomic-SO), genomic prediction with simulated offspring with restriction (Genomic-SO-Const), mid-parent GBLUP (Genomic-GEBV), phenotypic selection (Phenotypic-DH), phenotypic chain selection using double haploid lines (Phenot-Ch-DH), phenotypic chain selection using recombinant inbred lines (Phenot-Ch-RILs), and random crosses. Different scenarios for heritability (h^2) and number of population (N_{POP}) were considered. The points represents the average value of 20 simulations, while the shadow of different colors represent the quantile 5% and 95%. The horizontal line on top figure represents the maximum breeding on average, across 20 simulations.

population contributes to the maintenance of genetic gains in a long-term process. The heritability also plays an important role, as higher heritability ($h^2 = 0.8$) allows the breeder to go much further with the selection, depleting the existing genetic variability. Again we can see that the Genomic-GEBV method causes the highest losses of genetic variability.

The efficiency of conversion for the data-based simulation, using the wheat dataset is present in Figure 3.7. According to this figure, even though the conversion of genetic variance into genetic gain is lower in the phenotypic methods (Phenotypic-Ch-DH, Phenotypic-Ch-RILs, and Phenotypic-DH), their low loss of genetic variability compensates and increase the conversion in a long-term. These differences

are more pronounced in low heritability scenario ($h^2 = 0.2$). Also, the number of population impacts tremendously as a higher number of population contribute to the maintenance of genetic diversity.

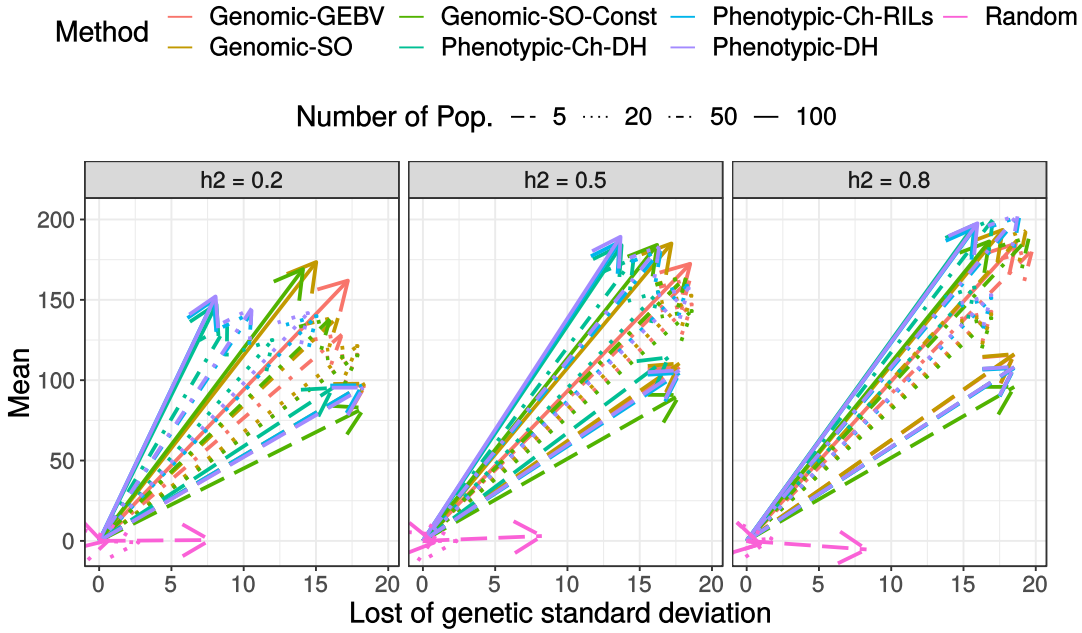

Figure 3.7. Efficiency of conversion of genetic variance into genetic gain according to Gorjanc *et al.* (2018) for the data-based simulation for all combinations of heritability (h^2), number of population, and method of selection. The higher the angle of the lines, higher is the conversion.

Even though plots like Figure 3.6 and S11 are informative in terms of the diversity that we have throughout cycles of breeding, showing how many loci became monomorphic and how many we lose the benefit allele, it is important to have an idea of what was the importance of the loci that we lost in this process. Figure 3.8 presents the variance proportion for each of the following types: Loci that fixed the benefit allele, loci that fixed the non-benefit allele and loci that are still segregating. This measure was calculated as the average of the square of the genetic effect for each loci and each individual. These effects were them added up and calculated as proportions. We can see that a big part of the losses is due to the low number of population, as can be noticed on the random method. Also, even though the genomic methods present higher mean, they also tend to lose more than the phenotypic methods, even in the case that we control for the number of times that a genotype is a parent (Genomic-SO-Const).

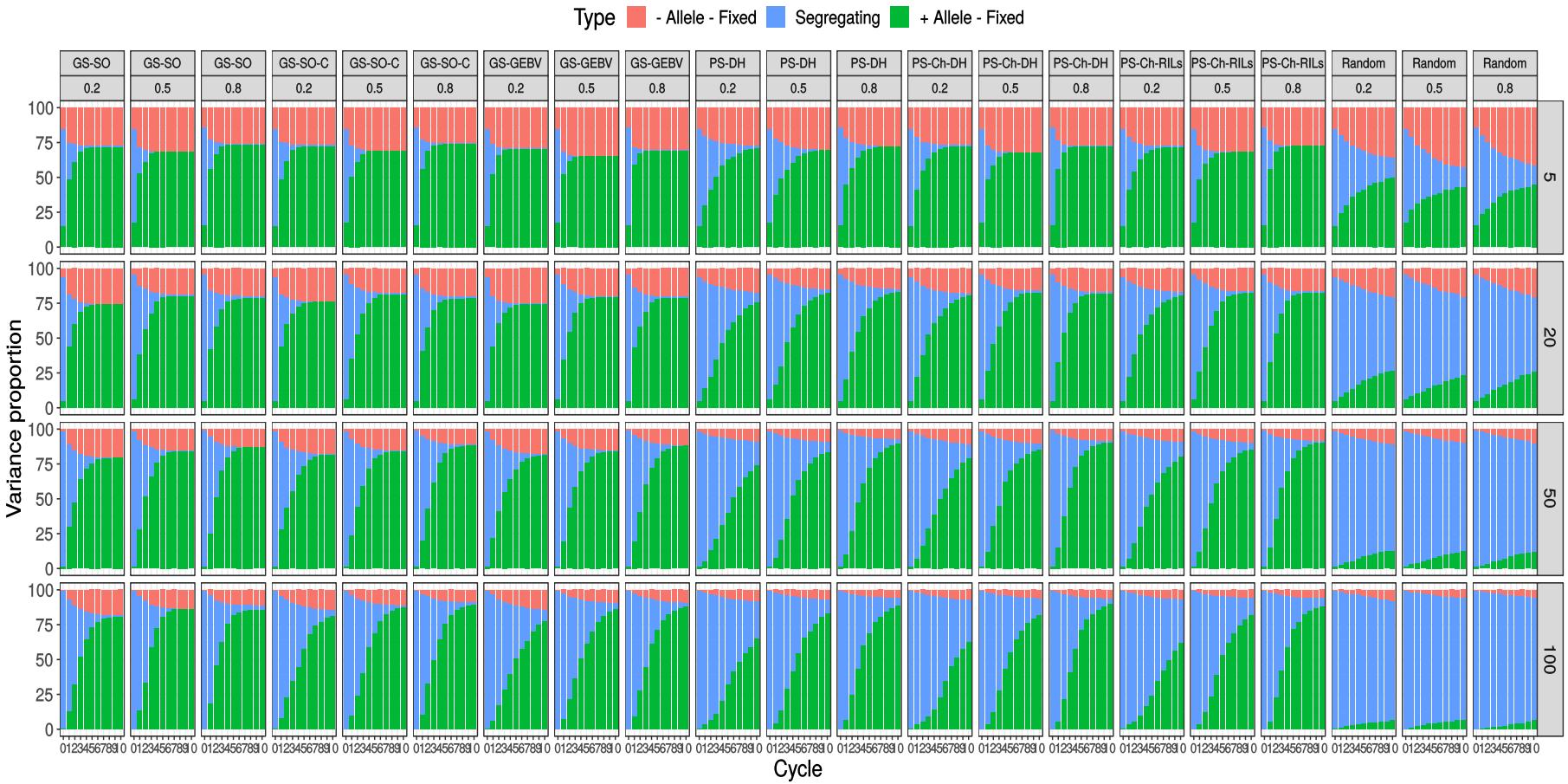

Figure 3.8. Proportion of genetic variance explained by the following parts: genes that are still segregating, genes that have the non-benefit allele fixed (-Allele-Fixed), and genes that have the benefit allele fixed (+Allele-Fixed).

3.4 Discussion

The choice for the right crosses has been a major concern on animal and plant breeding since we migrate from selecting the most beautiful, colorful, and tasteful plants on a natural pollinated crop to select individuals to compose some controled pollination scheme, directing what individuals must be crossed, with the goal to have the best offsprings. Even though quantitative genetics theory has been an important supporter to the breeder's decision in different aspects (Lynch and Walsh, 1998; Walsh and Lynch, 2018), the selection of crosses has been done mostly based on the average breeding value or genotypic mean of the candidate parents, not taking into account how much variability we can have within a cross (Bernardo, 2014; Ramalho *et al.*, 2012a). According to Baenziger and Peterson (1991), this method, called expected mid-parent, can be classified on the class of methods that use information from the parents only. Methods that use coancestry coefficient and genetic distance to select divergent crosses, can also be classified in the same class of methods.

Expected mid-parent is the most popular method used in breeding programs as it is easy to apply and also provides good results, although do not take into account the possible variability generated within a cross (Bernardo, 2010, 2014; Beckett *et al.*, 2019). Using just classical quantitative genetics, we can predict the mean of a population, however, the genetic variance is difficult to predict (Bernardo, 2010). Methods that use also the information of the offspring can be very useful to select which crosses should be produced. However, the need to produce the crosses and evaluate these populations, limits its application due to the usual high number of candidate lines to be used as parents, being important the adoption of methods that allow the choice, *a priori*, of the best crosses (Witcombe *et al.*, 2013).

The use of molecular markers to predict the breeding value of individuals was proposed on 2001 by Meuwissen *et al.* (2001). By the time, the applications of genomic prediction, was restricted to the prediction of genotypic and breeding values of unphenotyped individuals. Just with the paper from Bernardo (2014) the method also became useful for the selection of crosses, being possible to predict not only the mean but also the variance that might occur in a cross. Even though methods that use genetic distance and coancetry coefficient has the objective to take the possible variability within a cross into account, there is no significant correlation between these methods and the actual predicted genetic variance, as proposed by Bernardo (2014), according to Beckett *et al.* (2019). In this method, with the adjusted means and the marker data set, the breeder fits a regression model to estimate the marker effects (e.g., Ridge regression BLUP or any other marker model), and with a genetic linkage map, simulates several progenies from all possible crosses from a list of candidate parents. The computer codes that simulate these progenies involves the reproduction of the meiosis phenomenon, crossing-over, gamete formation, and fertilization. With the simulated progenies, it is possible to calculate different parameters, like the mean and variance, providing useful information to the breeder.

Even though the method proposed by Bernardo (2014) has been studied by Lado *et al.* (2017), Osthushenrich *et al.* (2017), Beckett *et al.* (2019), and Neyhart and Smith (2019), as far as we know, there were no studies of long-term selection using and comparing this method with the traditional methods used by breeders. For that, we used two simulation approaches, an *In silico* simulation, and also a data-based simulation, using a large panel of wheat genotypes. Long-term studies usually are very time consuming and laborious, as it may take decades to have meaningful results, especially in plants. Considering that in our simulations we conducted a breeding program for ten cycles of breeding, this corresponds around to 40 years for the scenarios that use double haploid lines, as it may take four year for each cycle (Cobb *et al.*, 2019). Scenario that uses recombinant inbred lines (RILs) this time can be even higher, demanding several decades to perform a study like this.

One of the most famous long-term experiment is the Illinois long-term selection for oil and protein in corn, initiated in 1896 by C.G. Hopkins (Goldman, 2004). Since them, more than 100

generations of selection was done. The results were reported and updated several times in literature, and surprisingly there are still genetic gain after many cycles of breeding. However, as can be noticed, this is long and time consuming process. For this kind of study, which aims to investigate the impacts of artificial selection in a long-term process, simulation can be especially useful, as it can provide valuable and useful information to the breeder, helping to decide what methods or also what techniques should be used in order to have higher genetic gains. Also, the choice for a selection procedure should not only take into account the short-term response but also what is the response in a long-term, as the maintenance of genetic diversity across several generations of breeding allow us to have future gains with the selection (Gorjanc *et al.*, 2018).

According to Jannink (2010) in a long-term selection study, also using computer simulation, is perhaps inevitable to not lose favorable alleles that are in weak linkage disequilibrium with markers when genomic selection is applied. The results found in the present study agree with the ones from Jannink (2010), even though the situation on which genomic selection is applied is different, as we are selecting crosses and not unphenotyped individuals. Jacobson *et al.* (2015), using computer simulation, on other hand, compared the impact of genomic selection within a biparental cross instead of the standard phenotypic selection, commonly used by breeders. The results presented by these authors showed that genomic selection can significantly increase the similarity among lines when a high intensity of selection is used ($i = 5\%$), but not with other values for intensity, meaning that genomic selection causes a minimal loss of genetic diversity within biparental populations. In general, either, genomic selection with simulated offspring or even the method that uses GBLUP (VanRaden, 2008), caused higher losses of genetic diversity than the phenotypic methods. It is important to notice that even when we applied a restriction on the number of times a genotype is a parent, the losses are still reasonable, especially for scenarios of low number of population. The proportion of genetic variance is a good way to visualize the efficiency of selection and also how much we are losing in terms of possible increase on genetic value, as presented for the wheat data set. Even though genomic selection speeds up the gains with selection, the fixation of non-benefit alleles with valuable effects leads to the loss of a possible higher genetic value. It is important to notice that the losses are not only due to drift, which obviously occurs as we can see on the three columns of the variance proportion plot related to the random method, but also are due to the fixation of loci of small and negative effect that are close to a locus of big and positive effect, leading to a drag effect (Walsh and Lynch, 2018).

Lado *et al.* (2016) described the differences on what crosses would be selected if we consider mid-parent value against genomic prediction with simulated offsprings, and also not taking into account the LD among markers, using the expression presented by Endelman (2011) in two wheat datasets. Similar to what was found in Lado *et al.* (2016) the crosses selected when genomic prediction is applied would be different from the phenotypic methods. In our case, with a simulation approach, we are able to not only look at the differences on what crosses would be selected but also what selection method would provide higher genetic gains. In the case of the *In silico* simulation, for example, considering a low heritability scenario, the genomic methods outperformed the phenotypics, providing higher average breeding value after ten cycles of breeding. The results described by Lado *et al.* (2016), however, showed that no big differences would be found considering a trait like grain yield, but in a trait like grain protein content, the crosses selected would change.

Using a two-row barley dataset, Neyhart and Smith (2019) also investigated the applicability of genomic prediction to select crosses. In this case, the authors predicted 330,078 crosses and selected 27 of these to develop biparental populations and evaluate the progeny performance, measuring the mean and variance. The authors found a very good correlation of the population's mean, but the prediction of the genetic variance was less accurated. Also, the authors observed a strong negative bias in the variance prediction. This was previously reported by Lehermeier *et al.* (2017), showing that due to

the shrinkage present on models like *RR-BLUP*, the markers effects are shrunk towards the mean and as we use these marker effects to calculate the breeding values of the simulated progenies, the genetic variance is underestimated. Lehermeier *et al.* (2017) also presents an alternative solution using a Bayesian model, aiming to mitigate this underestimation of the genetic variance. Even though we used an *RR-BLUP* model to estimate the marker effects on each cycle of breeding, genomic methods (Genomic-SO, Genomic-SO-Const, and Genomic-GEBV), provided higher genetic gains, especially as presented in scenarios of high number of populations.

Besides that, the estimation of each genetic variance component, associated with each cross, is independent with each other. The consequence of this is that if we have a genotype with good breeding value, this genotype will be selected as parental for several crosses with other genotypes. By doing that, we drastically reduce the number of parents used in the crossing block, reducing also the genetic variability due to the elimination of several benefit alleles that are not in the hall of crosses. Aiming to have a long-term gain selection, Gorjanc *et al.* (2018) proposed a method of selection of crosses using the relationship matrix and the genotypic values of the candidate parents. The method jointly optimizes selection, maintenance of genetic diversity, and cross allocation with the AlphaMate software. Even though the method optimizes these aspects, it does not take into account the genetic variance that might occur within accross, like the method proposed by Bernardo (2014), applied here.

If we look at different times (breeding cycles), we might find differents methods providing the best results in terms of genetic gain, especially for the data-based simulation. This highlight the importance of long-term selection studies, as the methods for a short-term selection might not be the best ones in a long term. For medium to high heritability, for example, the genomic methods outperform the phenotypic methods in a short-term, but not in a long-term as it introduce higher levels of inbreeding in the population (Jannink, 2010; Jacobson *et al.*, 2015).

The number of population also plays an important role on the maintenance of genetic diversity. This subject was investigated before for a few authors (Wricke and Weber, 1986; Fouilloux and Bannerot, 1988; Bernardo, 2003). Bernardo (2003), for example, studied the trade-off between the number of population (N_{POP}) and the number of progenies per population (N_{PROG}). The results presented by the author showed that the method of selection is more important than the configuration of the trade-off $N_{POP} : N_{PROG}$. However, as presented here, higher number of population contributes to mitigate the drift effect that might occur when benefit alleles are losed by chance in a scenario of low number of population.

As we can notice, after ten cycles of breeding (40-50 years) the variability approximates to zero due the high number of fixed loci, especially loci fixed for the benefit allele, as we applied directional selection. Also, as mentioned before, the loci fixed for the non-benefit allele represents a very small proportion of what we can get from all the possible genetic variance. This fast decay on genetic variability is more pronounced in the scenario of high heritability, as the selection becomes more efficient because what we see is what we get from the genotype (Falconer and Mackay, 1996; Bernardo, 2010).

To clarify these results, it is important to highlihg also some major assumptions that we needed to consider in both simulation approaches. A single generic quantitative trait with 100 loci following an exponential distribution in the *In silico* was used. In the data-based simulation, the effects came from a sample of the marker effect set, calculated using *RR-BLUP* model. This scenario is usually not realistic, as breeders usually needs to breed for several traits at the same time and very often these traits presents negative correlation with each other (Bernardo, 2010; Lynch and Walsh, 1998; Ramalho *et al.*, 2012a). This negative correlation causes a competition and turns the breeder's work much harder, reducing drastically the genetic gain, when we compared with cases where breeding is done for a single trait.

Genotype by environment interaction (G $\times$ E) is another very important phenomena that occurs

in breeding, especially in plant breeding (Falconer and Mackay, 1996; Lado *et al.*, 2016; Malosetti *et al.*, 2016). When G × E is significant, the genotypes do not perform similar accross enviroment (e.g., year and location). This can turn the breeder's life much harder, as it introduce a new source of variation, reducing the genetic gain. In our simulations, we do not took into account the interactions between genotype and enviroments. This means that there was no changes on the breeding value of the genotypes due to some different expression of the genes. All the 100 gene effects were generated right in the beginning of the simulation and presented the same alellic effects in all following breeding cycles. This simplification can lead to higher genetic gains and consequently, fast depletion of the genetic variance.

Moreover, the absence of non-additive effects in the simulation procedure is another assumption that relates with the complexity of the genomes. Epistasis, for example, is a very common effect found in plant species (Falconer and Mackay, 1996; Crossa *et al.*, 2017). The presence of this kind of effect, just like G × E interactions, can add another source of variation and confounding with additive and dominance effects, increasing the complexity of the effects estimation, and reducing the capability of selection, hampering the fast decrease on diversity.

As mention before, the process to create the reference population in the *In silico* simulation tries to reproduce a scenario of linkage disequilibrium similar to a different panel of genotypes from the literature, but with an average allele frequencies approximate to $p = q = 0.5$, as we started with two contratant inbred lines and no selection was acting on this population. The two approaches of simulation presented similar results in terms of what methods performed better in each scenario. There are several ways to start and conduct an *In silico* simulation process. Gorjanc *et al.* (2018), for example, used a Markovian Coalescent Simulator, as proposed by Chen *et al.* (2009), creating whole genome sequences from this process. After this initial process, 50 genotypes to stablish the founder genomes. Also, 20 years of breeding was used as burn-in, to eliminate some possible bias from the reproduction of the historical background of the species. The use of real data sets to compose the reference population, as presented in this paper, to start the simulation, is also a valuable approach. Moehring *et al.* (2014) used real historical data of variance components and also a relationship matrix to generate genetic values and start a simulation process to evaluate different strategies of resource allocation on experimental design.

Finally, the efficiency of conversion, calculated as described by Gorjanc *et al.* (2018) shows that even though the average breeding value achieved by the phenotypic methods (Phenotypic-Ch-DH, Phenotypic-Ch-RILs, and Phenotypic-DH) is lower than the genomic methods, their low loss of genetic variability compensates and increase the conversion in a long-term. These differences are more pronounced in the low heritability scenario. It is important to remember that this is a closed simulation procedure, with no introduction of new sources of variability and no new mutation during the breeding process. To circumvent this problem, especially in the genomic methods, the breeder should adopt strategies to mitigate this fast decrease on genetic variance. New introductions in the breeding program can be done to increase the amount of genetic variability and keep pushing the genetic gains in a long-term selection. Also, new techniques are arising to help breeders to go even further. *CRISPR* hopefully will be able help the breeder to replace non-benefit alleles to benefit alleles very precisely in genes identified by association studies, for example, increasing the gain with the selection (Bernardo, 2017).

3.5 Conclusion

The simulation process was informative to select what strategy should be used in each scenario of breeding under consideration. Also, in general, the approaches were concordant, besides the differences in the reference population.

The results of both approaches showed that the best method depends on the combination of heritability, number of population, and if the strategy of breeding is to have genetic gain in a short or

long-term. According to both types of simulation, in scenarios of low heritability ($h^2 = 0.2$) and high number of population ($N_{POP} = 50$ or 100), genomic prediction with restriction performs better than any other on both strategies. Also, it is important to restrict the number of times a genotype is a parent, aiming long-term gains with the selection.

The Genomic-GEBV method, based on GBLUP, presents a good performance in a short-term selection, but as it tends to select individuals genetically close related, it performs poorly in a long-term process, due to the introduction of population inbreeding, causing a drastic reduction of genetic variability.

In scenarios with high heritability ($h^2 = 0.8$), the phenotypic selection overperform genomic selection.

Even though the conversion of genetic variability into genetic gains is faster than any other strategy, the losses of variability are also higher. It is important to mention that both approaches of simulation under consideration here are closed ones, with no introduction of new sources of variability. Based on these results, especially with the adoption of genomic prediction, the introduction of new alleles with the advance of the cycles of breeding can be an interesting approach, aiming not to mine the genetic variability and have long-term genetic gains.

3.6 Supplementary material

Annex A

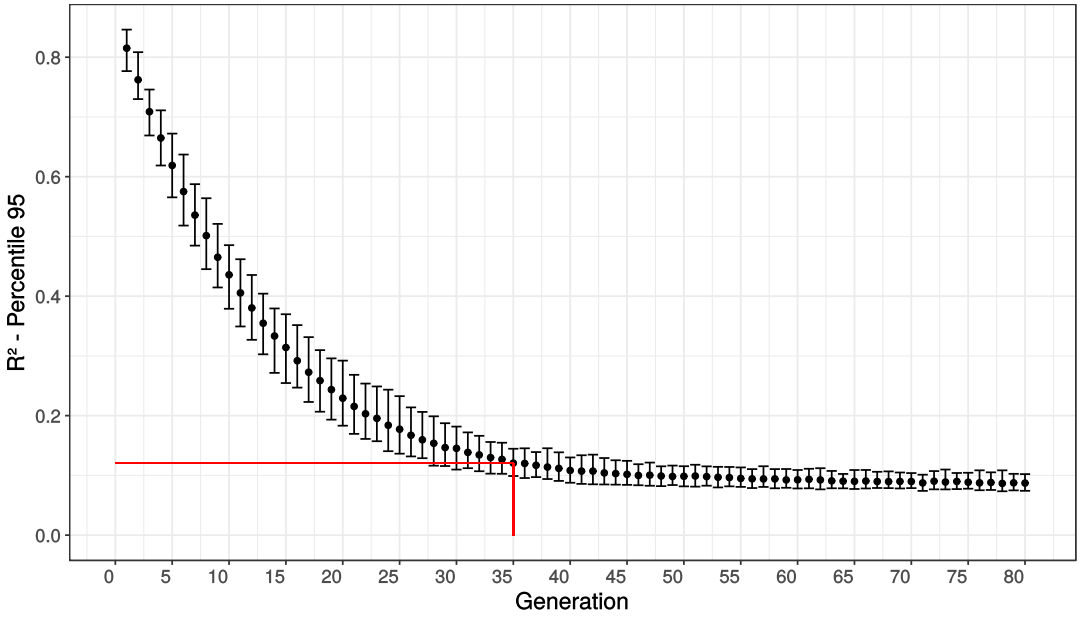

Figure S9. Levels of linkage disequilibrium (LD-R^2) after several generations of random mating among 1000 individuals of an F_2 segregating for $g + m$ loci.

Annex B

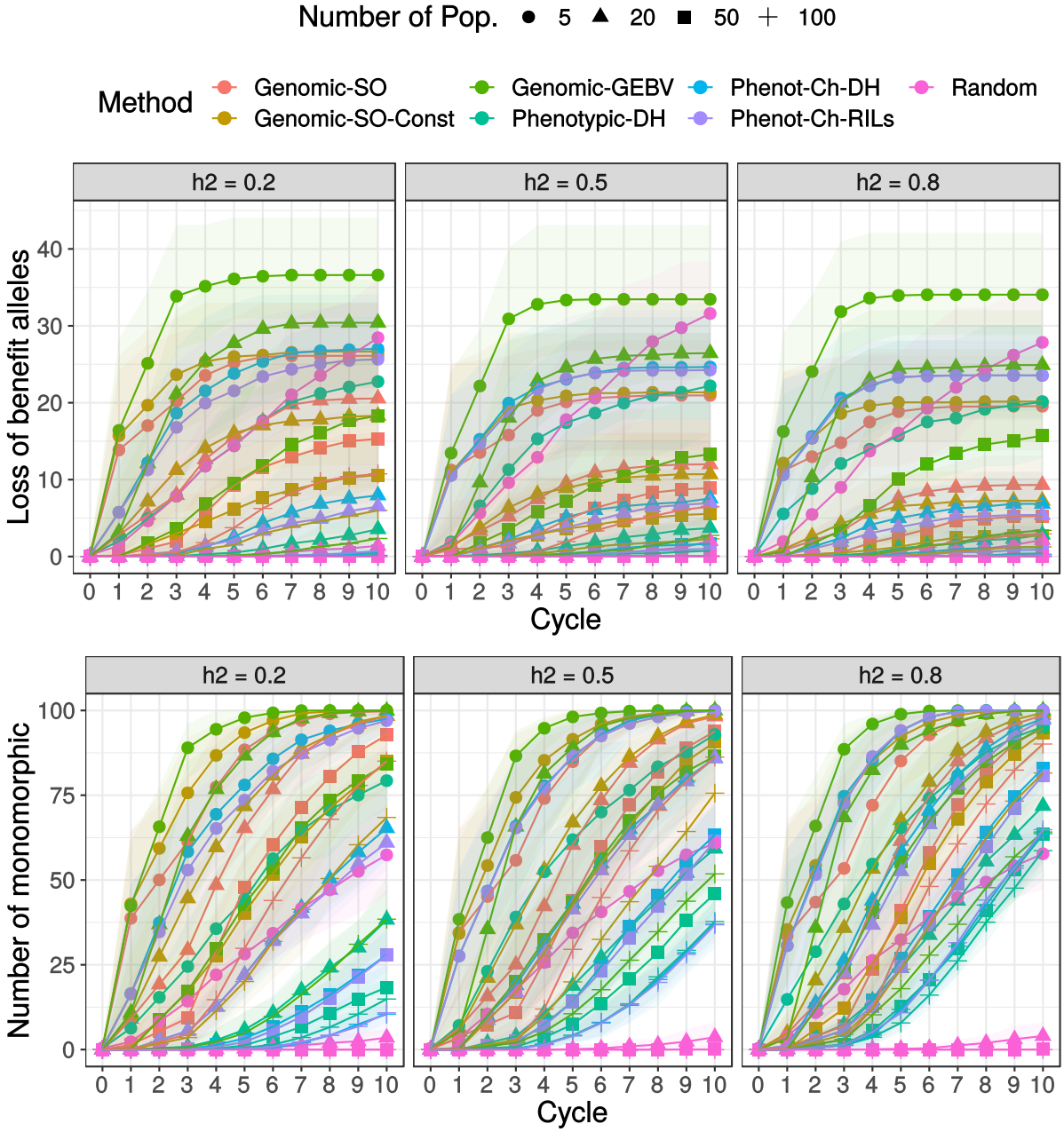

Figure S10. Number of monorphic loci and number of loci where the benefit allele was fixed for the *In silico* simulation under different scenarios of heritability (h^2), number of population, and methods of selection.

Annex C

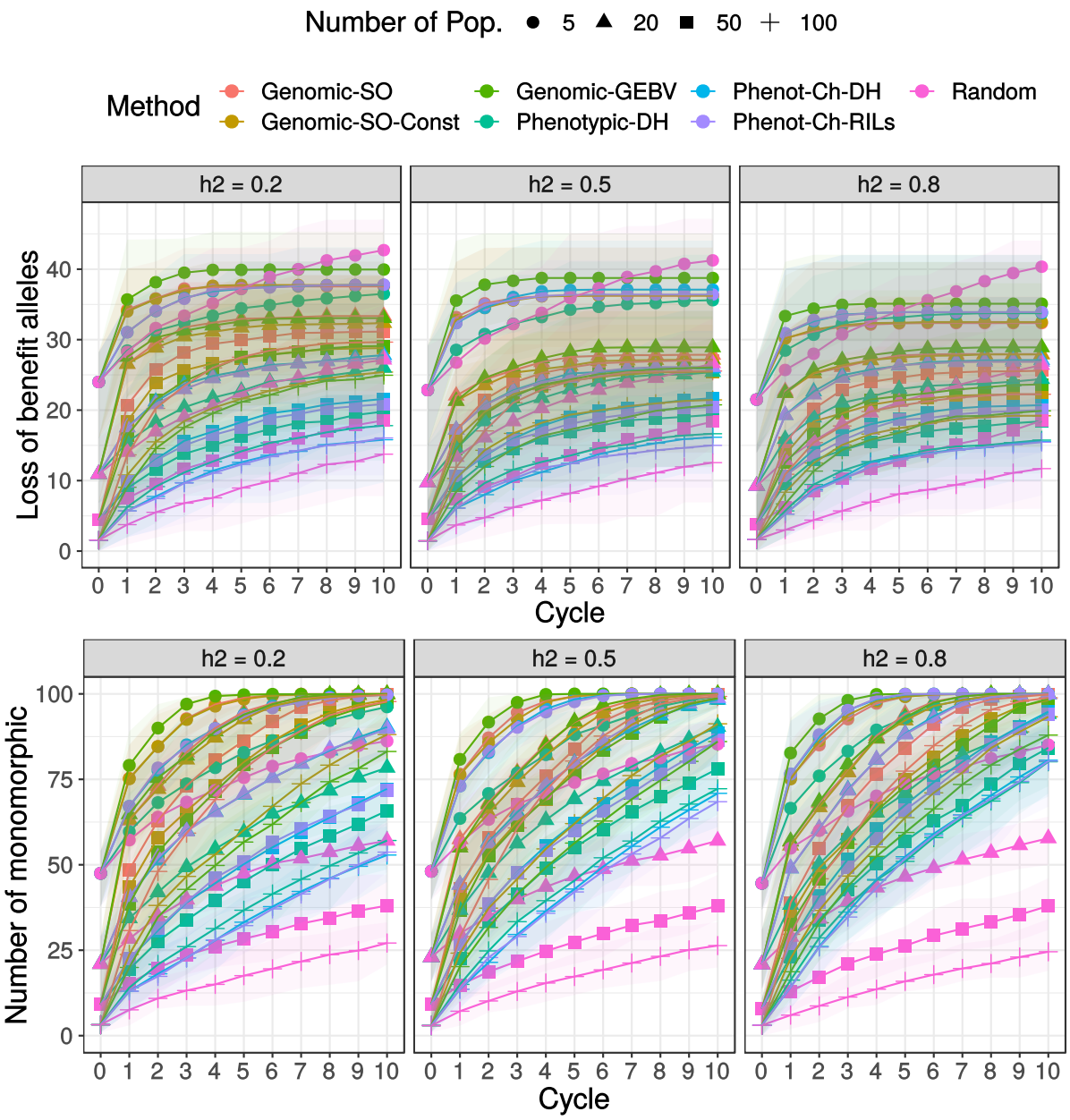

Figure S11. Number of monorphic loci and number of loci where the benefit allele was fixed for the data-based simulation under different scenarios of heritability (h^2), number of population, and methods of selection.

Annex D

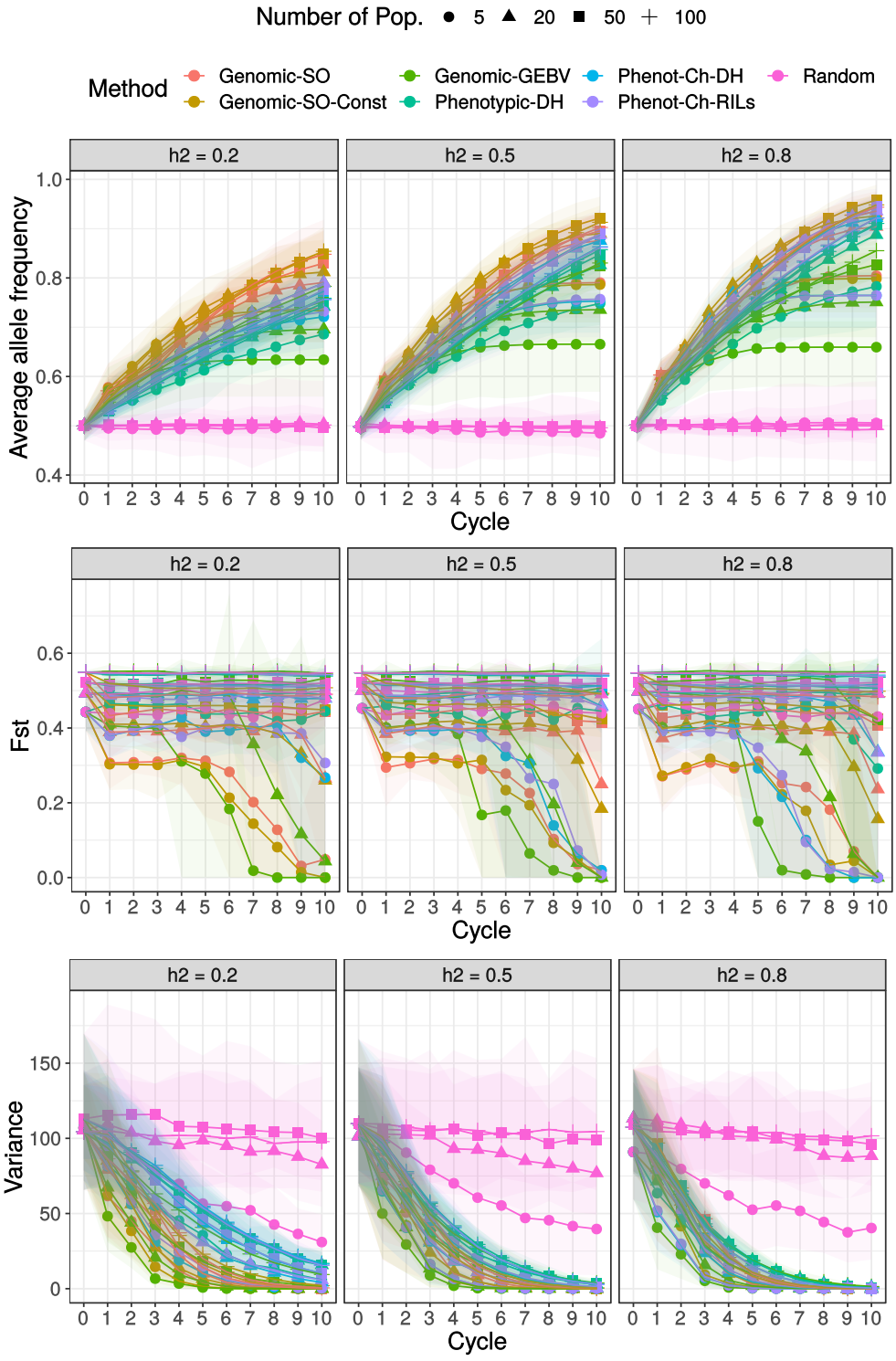

Figure S12. Average allele frequency (p), F_{ST} statistics, and genetic variance (σ^2_G) for the *In silico* simulation under different scenarios of heritability (h^2), number of population, and methods of selection.

Annex E

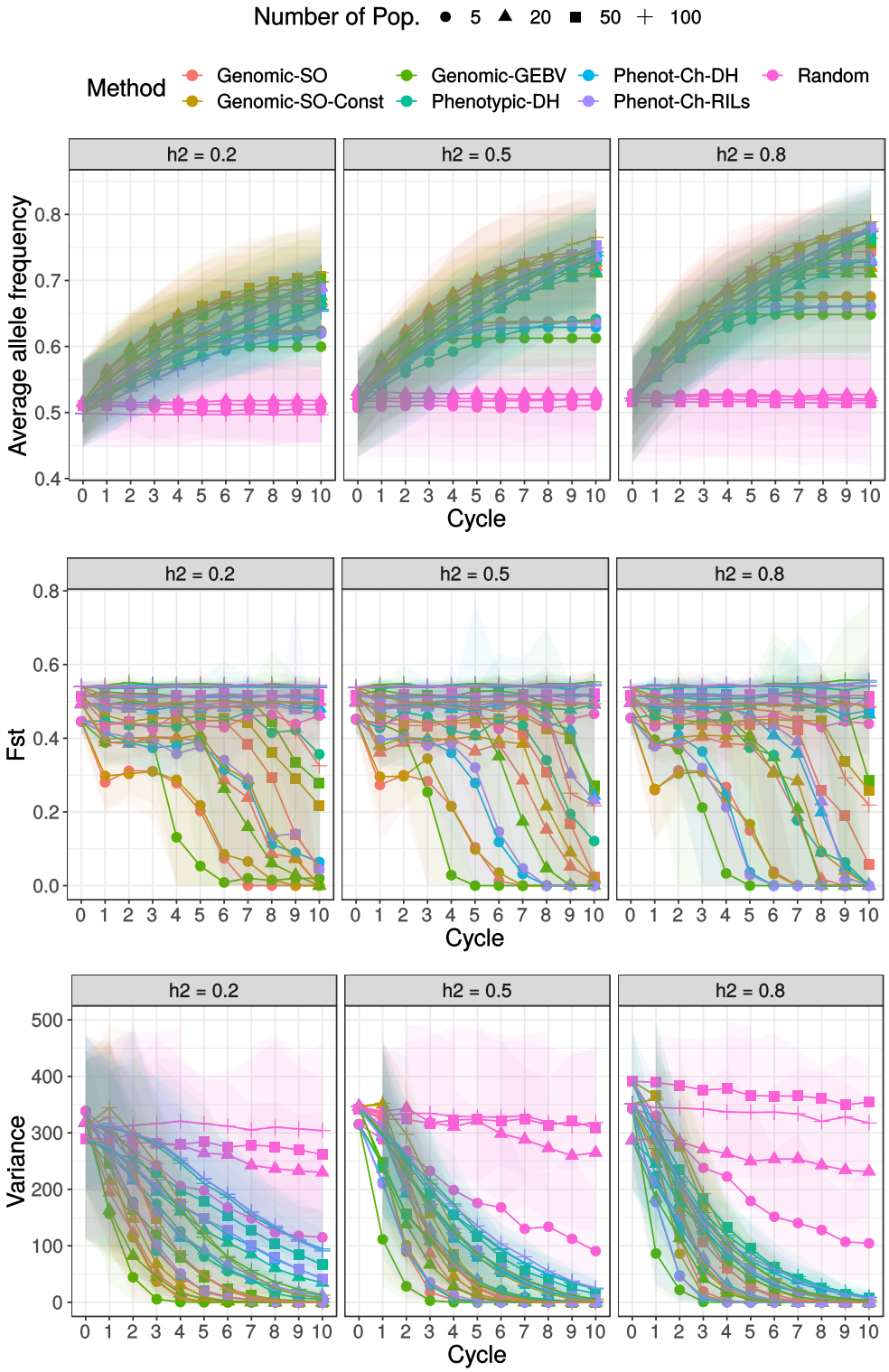

Figure S13. Average allele frequency (p), F_{ST} statistics, and genetic variance (σ^2_G) for the data-based simulation under different scenarios of heritability (h^2), number of population, and methods of selection.

Annex F

Table S1. *In silico* simulation results for the cycles 0, 1, 5, and 10, for the combination of heritability (h^2), number of population (N_{POP}), and method of selection. The variables listed below refers to the overall mean of the breeding values ($\overline{B.V.}$), genetic variance (σ^2_G), , genetic diversity according to the Nei's estimator ($G.D.$), average allele frequency (p), number of lost benefit alleles (Lost +), number of loci with fixed allele (Monom).

Cycle	h^2	N_{POP}	Method	$\overline{B.V.}$	σ^2_G	F_{st}	G.D.	p	Lost +	Monom.
0	0.2	5	Genomic-SO	100.88	107.01	0.44	0.45	0.50	0.20	0.40
1	0.2	5	Genomic-SO	127.29	61.64	0.31	0.24	0.58	13.85	38.70
5	0.2	5	Genomic-SO	161.61	4.08	0.31	0.04	0.70	25.20	88.40
10	0.2	5	Genomic-SO	166.62	0.00	0.05	0.00	0.74	26.10	99.75
0	0.5	5	Genomic-SO	98.06	109.11	0.45	0.44	0.50	0.05	0.30
1	0.5	5	Genomic-SO	129.85	69.58	0.29	0.25	0.59	11.25	34.50
5	0.5	5	Genomic-SO	168.61	4.74	0.29	0.05	0.74	20.10	84.90
10	0.5	5	Genomic-SO	173.73	0.00	0.02	0.00	0.79	20.95	99.95
0	0.8	5	Genomic-SO	95.67	90.95	0.45	0.44	0.50	0.15	0.30
1	0.8	5	Genomic-SO	126.60	73.44	0.27	0.26	0.59	11.15	33.55
5	0.8	5	Genomic-SO	166.59	2.72	0.31	0.05	0.75	18.85	85.10
10	0.8	5	Genomic-SO	170.82	0.00	0.00	0.00	0.80	19.55	100.00
0	0.2	20	Genomic-SO	100.34	105.38	0.49	0.48	0.50	0.00	0.00
1	0.2	20	Genomic-SO	124.57	79.15	0.39	0.31	0.57	2.80	9.05
5	0.2	20	Genomic-SO	164.84	10.11	0.40	0.10	0.71	15.60	65.30
10	0.2	20	Genomic-SO	177.15	0.02	0.27	0.00	0.79	20.55	98.40
0	0.5	20	Genomic-SO	98.70	101.39	0.50	0.48	0.50	0.00	0.00
1	0.5	20	Genomic-SO	129.75	74.88	0.38	0.33	0.59	1.45	6.35
5	0.5	20	Genomic-SO	172.30	9.09	0.40	0.12	0.77	9.55	60.35
10	0.5	20	Genomic-SO	184.41	0.01	0.25	0.00	0.88	12.00	98.25
0	0.8	20	Genomic-SO	104.11	113.27	0.50	0.48	0.50	0.00	0.00
1	0.8	20	Genomic-SO	139.66	82.43	0.37	0.34	0.60	0.80	4.75
5	0.8	20	Genomic-SO	187.60	7.91	0.41	0.11	0.79	7.55	61.50
10	0.8	20	Genomic-SO	199.51	0.01	0.24	0.00	0.90	9.30	98.60
0	0.2	50	Genomic-SO	102.03	113.03	0.52	0.49	0.50	0.00	0.00
1	0.2	50	Genomic-SO	127.78	87.45	0.43	0.37	0.56	0.15	1.40
5	0.2	50	Genomic-SO	170.25	14.06	0.44	0.15	0.71	9.30	47.80
10	0.2	50	Genomic-SO	188.02	0.14	0.44	0.02	0.83	15.30	92.75
0	0.5	50	Genomic-SO	98.75	110.02	0.52	0.49	0.50	0.00	0.00
1	0.5	50	Genomic-SO	129.03	80.20	0.44	0.36	0.59	0.90	3.75
5	0.5	50	Genomic-SO	176.67	12.04	0.45	0.15	0.77	4.75	43.95
10	0.5	50	Genomic-SO	192.87	0.11	0.42	0.01	0.90	8.90	93.85
0	0.8	50	Genomic-SO	100.77	109.45	0.52	0.49	0.50	0.00	0.00
1	0.8	50	Genomic-SO	135.51	81.14	0.43	0.37	0.60	0.05	0.60
5	0.8	50	Genomic-SO	182.95	10.18	0.45	0.15	0.79	2.60	41.00
10	0.8	50	Genomic-SO	198.24	0.03	0.41	0.01	0.94	5.25	94.80
0	0.2	100	Genomic-SO	99.30	104.35	0.55	0.49	0.50	0.00	0.00
1	0.2	100	Genomic-SO	123.72	86.43	0.48	0.40	0.57	0.00	0.05
5	0.2	100	Genomic-SO	166.26	20.11	0.49	0.18	0.71	3.75	29.85

10	0.2	100	Genomic-SO	188.70	0.56	0.49	0.03	0.85	10.75	85.05
0	0.5	100	Genomic-SO	100.11	109.83	0.55	0.49	0.50	0.00	0.00
1	0.5	100	Genomic-SO	129.00	88.12	0.48	0.40	0.58	0.00	0.05
5	0.5	100	Genomic-SO	175.45	15.06	0.49	0.18	0.75	1.95	29.55
10	0.5	100	Genomic-SO	196.06	0.34	0.49	0.03	0.90	6.50	86.25
0	0.8	100	Genomic-SO	100.07	107.45	0.55	0.49	0.50	0.00	0.00
1	0.8	100	Genomic-SO	134.66	78.01	0.48	0.40	0.60	0.00	0.10
5	0.8	100	Genomic-SO	181.45	10.01	0.49	0.17	0.79	0.70	31.35
10	0.8	100	Genomic-SO	197.34	0.06	0.48	0.02	0.94	2.95	90.10
0	0.2	5	Genomic-SO-Const	100.88	107.01	0.44	0.45	0.50	0.20	0.40
1	0.2	5	Genomic-SO-Const	126.99	62.39	0.30	0.23	0.58	15.70	42.95
5	0.2	5	Genomic-SO-Const	163.22	1.37	0.30	0.02	0.72	26.00	93.45
10	0.2	5	Genomic-SO-Const	164.97	0.00	0.00	0.00	0.73	26.65	100.00
0	0.5	5	Genomic-SO-Const	98.06	109.11	0.45	0.44	0.50	0.05	0.30
1	0.5	5	Genomic-SO-Const	129.40	70.34	0.32	0.26	0.59	10.95	34.20
5	0.5	5	Genomic-SO-Const	171.57	1.53	0.31	0.03	0.76	20.90	91.50
10	0.5	5	Genomic-SO-Const	173.82	0.00	0.00	0.00	0.79	21.35	100.00
0	0.8	5	Genomic-SO-Const	95.67	90.95	0.45	0.44	0.50	0.15	0.30
1	0.8	5	Genomic-SO-Const	126.51	73.91	0.27	0.26	0.59	12.15	35.40
5	0.8	5	Genomic-SO-Const	167.26	1.15	0.30	0.03	0.77	20.05	92.30
10	0.8	5	Genomic-SO-Const	168.66	0.00	0.00	0.00	0.80	20.15	100.00
0	0.2	20	Genomic-SO-Const	100.34	105.38	0.49	0.48	0.50	0.00	0.00
1	0.2	20	Genomic-SO-Const	123.16	85.22	0.41	0.34	0.57	3.05	10.20
5	0.2	20	Genomic-SO-Const	170.84	6.60	0.40	0.09	0.74	16.05	71.75
10	0.2	20	Genomic-SO-Const	179.02	0.02	0.26	0.01	0.81	18.30	98.40
0	0.5	20	Genomic-SO-Const	98.70	101.39	0.50	0.48	0.50	0.00	0.00
1	0.5	20	Genomic-SO-Const	126.81	82.16	0.40	0.37	0.59	1.20	5.40
5	0.5	20	Genomic-SO-Const	177.85	4.68	0.41	0.10	0.80	8.90	66.95
10	0.5	20	Genomic-SO-Const	185.41	0.00	0.18	0.00	0.89	10.70	98.75
0	0.8	20	Genomic-SO-Const	104.11	113.27	0.50	0.48	0.50	0.00	0.00
1	0.8	20	Genomic-SO-Const	136.43	91.56	0.40	0.37	0.59	0.95	4.90
5	0.8	20	Genomic-SO-Const	193.16	3.76	0.41	0.10	0.83	6.55	67.65
10	0.8	20	Genomic-SO-Const	200.26	0.00	0.16	0.00	0.93	7.25	99.30
0	0.2	50	Genomic-SO-Const	102.03	113.03	0.52	0.49	0.50	0.00	0.00
1	0.2	50	Genomic-SO-Const	123.89	99.42	0.46	0.42	0.55	0.10	0.65
5	0.2	50	Genomic-SO-Const	174.06	16.22	0.46	0.18	0.72	6.10	40.35
10	0.2	50	Genomic-SO-Const	192.57	0.84	0.45	0.04	0.85	10.60	84.90
0	0.5	50	Genomic-SO-Const	98.75	110.02	0.52	0.49	0.50	0.00	0.00
1	0.5	50	Genomic-SO-Const	124.79	94.30	0.47	0.42	0.57	0.05	0.25
5	0.5	50	Genomic-SO-Const	180.15	11.38	0.46	0.16	0.79	3.60	42.65
10	0.5	50	Genomic-SO-Const	195.14	0.09	0.42	0.02	0.92	5.55	90.65
0	0.8	50	Genomic-SO-Const	100.77	109.45	0.52	0.49	0.50	0.00	0.00
1	0.8	50	Genomic-SO-Const	129.99	96.38	0.46	0.43	0.59	0.00	0.00
5	0.8	50	Genomic-SO-Const	187.13	8.39	0.47	0.17	0.82	1.25	38.95
10	0.8	50	Genomic-SO-Const	200.06	0.04	0.42	0.01	0.96	2.70	93.45
0	0.2	100	Genomic-SO-Const	99.30	104.35	0.55	0.49	0.50	0.00	0.00

1	0.2	100	Genomic-SO-Const	117.76	102.44	0.52	0.46	0.55	0.00	0.00
5	0.2	100	Genomic-SO-Const	167.30	22.95	0.51	0.23	0.72	1.50	20.05
10	0.2	100	Genomic-SO-Const	190.47	2.07	0.51	0.08	0.85	6.00	68.45
0	0.5	100	Genomic-SO-Const	100.11	109.83	0.55	0.49	0.50	0.00	0.00
1	0.5	100	Genomic-SO-Const	122.50	100.98	0.52	0.46	0.56	0.00	0.05
5	0.5	100	Genomic-SO-Const	176.66	16.46	0.51	0.23	0.76	0.80	19.95
10	0.5	100	Genomic-SO-Const	196.96	0.48	0.49	0.06	0.91	2.30	75.60
0	0.8	100	Genomic-SO-Const	100.07	107.45	0.55	0.49	0.50	0.00	0.00
1	0.8	100	Genomic-SO-Const	125.84	96.93	0.52	0.46	0.57	0.00	0.00
5	0.8	100	Genomic-SO-Const	181.65	10.94	0.51	0.21	0.79	0.05	23.70
10	0.8	100	Genomic-SO-Const	197.70	0.15	0.50	0.04	0.95	1.25	81.55
0	0.2	5	Genomic-GEBV	100.88	107.01	0.44	0.45	0.50	0.20	0.40
1	0.2	5	Genomic-GEBV	125.91	48.21	0.41	0.23	0.57	16.40	42.40
5	0.2	5	Genomic-GEBV	142.67	0.78	0.28	0.01	0.63	36.10	97.85
10	0.2	5	Genomic-GEBV	143.38	0.00	0.00	0.00	0.63	36.60	100.00
0	0.5	5	Genomic-GEBV	98.06	109.11	0.45	0.44	0.50	0.05	0.30
1	0.5	5	Genomic-GEBV	129.08	49.99	0.40	0.24	0.59	13.45	38.40
5	0.5	5	Genomic-GEBV	149.57	0.36	0.17	0.01	0.66	33.35	98.10
10	0.5	5	Genomic-GEBV	150.08	0.00	0.00	0.00	0.67	33.45	100.00
0	0.8	5	Genomic-GEBV	95.67	90.95	0.45	0.44	0.50	0.15	0.30
1	0.8	5	Genomic-GEBV	126.56	40.58	0.40	0.22	0.59	16.25	43.35
5	0.8	5	Genomic-GEBV	144.91	0.18	0.15	0.00	0.66	33.95	98.90
10	0.8	5	Genomic-GEBV	145.15	0.00	0.00	0.00	0.66	34.05	100.00
0	0.2	20	Genomic-GEBV	100.34	105.38	0.49	0.48	0.50	0.00	0.00
1	0.2	20	Genomic-GEBV	121.96	67.34	0.49	0.33	0.56	3.15	10.85
5	0.2	20	Genomic-GEBV	155.98	4.38	0.49	0.05	0.66	27.75	86.70
10	0.2	20	Genomic-GEBV	160.73	0.00	0.04	0.00	0.70	30.40	99.90
0	0.5	20	Genomic-GEBV	98.70	101.39	0.50	0.48	0.50	0.00	0.00
1	0.5	20	Genomic-GEBV	125.85	68.29	0.48	0.37	0.58	1.20	5.25
5	0.5	20	Genomic-GEBV	160.48	1.44	0.49	0.04	0.71	24.55	89.15
10	0.5	20	Genomic-GEBV	163.46	0.00	0.00	0.00	0.74	26.45	100.00
0	0.8	20	Genomic-GEBV	104.11	113.27	0.50	0.48	0.50	0.00	0.00
1	0.8	20	Genomic-GEBV	135.19	71.53	0.50	0.38	0.59	0.60	3.30
5	0.8	20	Genomic-GEBV	170.21	1.87	0.48	0.04	0.72	24.40	89.85
10	0.8	20	Genomic-GEBV	174.20	0.00	0.00	0.00	0.75	24.90	100.00
0	0.2	50	Genomic-GEBV	102.03	113.03	0.52	0.49	0.50	0.00	0.00
1	0.2	50	Genomic-GEBV	122.51	87.93	0.52	0.43	0.55	0.00	0.15
5	0.2	50	Genomic-GEBV	162.05	17.89	0.52	0.19	0.67	9.50	42.90
10	0.2	50	Genomic-GEBV	179.98	1.55	0.53	0.05	0.75	18.30	84.10
0	0.5	50	Genomic-GEBV	98.75	110.02	0.52	0.49	0.50	0.00	0.00
1	0.5	50	Genomic-GEBV	123.21	80.42	0.53	0.44	0.57	0.00	0.10
5	0.5	50	Genomic-GEBV	166.03	15.80	0.52	0.18	0.71	7.20	43.45
10	0.5	50	Genomic-GEBV	185.26	0.68	0.51	0.04	0.82	13.30	86.55
0	0.8	50	Genomic-GEBV	100.77	109.45	0.52	0.49	0.50	0.00	0.00
1	0.8	50	Genomic-GEBV	128.32	82.07	0.52	0.44	0.58	0.00	0.00
5	0.8	50	Genomic-GEBV	172.62	9.68	0.52	0.14	0.73	10.05	55.50

10	0.8	50	Genomic-GEBV	186.95	0.10	0.42	0.01	0.83	15.70	95.05
0	0.2	100	Genomic-GEBV	99.30	104.35	0.55	0.49	0.50	0.00	0.00
1	0.2	100	Genomic-GEBV	116.21	93.59	0.55	0.47	0.55	0.00	0.00
5	0.2	100	Genomic-GEBV	153.33	40.73	0.55	0.33	0.67	0.15	4.85
10	0.2	100	Genomic-GEBV	180.73	9.42	0.55	0.18	0.78	2.35	38.45
0	0.5	100	Genomic-GEBV	100.11	109.83	0.55	0.49	0.50	0.00	0.00
1	0.5	100	Genomic-GEBV	120.41	92.68	0.55	0.47	0.56	0.00	0.00
5	0.5	100	Genomic-GEBV	163.76	26.78	0.55	0.29	0.70	0.35	10.55
10	0.5	100	Genomic-GEBV	189.07	3.76	0.55	0.13	0.83	2.75	51.80
0	0.8	100	Genomic-GEBV	100.07	107.45	0.55	0.49	0.50	0.00	0.00
1	0.8	100	Genomic-GEBV	123.62	84.50	0.55	0.47	0.57	0.00	0.00
5	0.8	100	Genomic-GEBV	169.81	18.81	0.55	0.25	0.72	0.80	17.95
10	0.8	100	Genomic-GEBV	191.52	1.47	0.55	0.10	0.86	3.40	64.15
0	0.2	5	Phenotypic-DH	100.88	107.01	0.44	0.45	0.50	0.20	0.40
1	0.2	5	Phenotypic-DH	111.43	82.34	0.46	0.38	0.53	2.00	6.30
5	0.2	5	Phenotypic-DH	141.82	35.62	0.44	0.20	0.61	14.40	44.35
10	0.2	5	Phenotypic-DH	161.18	5.85	0.44	0.08	0.69	22.75	79.30
0	0.5	5	Phenotypic-DH	98.06	109.11	0.45	0.44	0.50	0.05	0.30
1	0.5	5	Phenotypic-DH	115.72	83.73	0.46	0.37	0.55	1.90	7.10
5	0.5	5	Phenotypic-DH	154.16	18.67	0.41	0.14	0.67	17.40	61.90
10	0.5	5	Phenotypic-DH	170.06	0.73	0.45	0.03	0.75	22.20	92.90
0	0.8	5	Phenotypic-DH	95.67	90.95	0.45	0.44	0.50	0.15	0.30
1	0.8	5	Phenotypic-DH	115.47	63.57	0.46	0.34	0.55	5.55	14.80
5	0.8	5	Phenotypic-DH	155.73	9.45	0.44	0.13	0.70	15.70	65.35
10	0.8	5	Phenotypic-DH	168.45	0.17	0.29	0.02	0.78	20.10	95.20
0	0.2	20	Phenotypic-DH	100.34	105.38	0.49	0.48	0.50	0.00	0.00
1	0.2	20	Phenotypic-DH	110.63	96.99	0.49	0.46	0.53	0.00	0.00
5	0.2	20	Phenotypic-DH	144.95	52.31	0.49	0.33	0.63	0.50	5.85
10	0.2	20	Phenotypic-DH	172.96	15.76	0.49	0.19	0.74	3.55	38.30
0	0.5	20	Phenotypic-DH	98.70	101.39	0.50	0.48	0.50	0.00	0.00
1	0.5	20	Phenotypic-DH	115.23	91.28	0.48	0.44	0.55	0.00	0.00
5	0.5	20	Phenotypic-DH	160.24	23.01	0.49	0.26	0.70	1.35	18.80
10	0.5	20	Phenotypic-DH	183.94	2.77	0.50	0.12	0.83	3.70	59.25
0	0.8	20	Phenotypic-DH	104.11	113.27	0.50	0.48	0.50	0.00	0.00
1	0.8	20	Phenotypic-DH	127.00	90.73	0.50	0.43	0.56	0.00	0.15
5	0.8	20	Phenotypic-DH	177.70	18.85	0.49	0.24	0.74	1.05	24.05
10	0.8	20	Phenotypic-DH	200.25	0.93	0.49	0.08	0.89	2.85	71.85
0	0.2	50	Phenotypic-DH	102.03	113.03	0.52	0.49	0.50	0.00	0.00
1	0.2	50	Phenotypic-DH	114.10	105.06	0.51	0.47	0.53	0.00	0.00
5	0.2	50	Phenotypic-DH	150.65	51.30	0.51	0.37	0.64	0.00	1.25
10	0.2	50	Phenotypic-DH	178.35	15.42	0.51	0.25	0.75	0.45	18.35
0	0.5	50	Phenotypic-DH	98.75	110.02	0.52	0.49	0.50	0.00	0.00
1	0.5	50	Phenotypic-DH	116.82	93.63	0.51	0.47	0.55	0.00	0.00
5	0.5	50	Phenotypic-DH	163.90	27.68	0.52	0.31	0.70	0.15	7.45
10	0.5	50	Phenotypic-DH	189.88	3.34	0.51	0.15	0.85	0.75	46.00
0	0.8	50	Phenotypic-DH	100.77	109.45	0.52	0.49	0.50	0.00	0.00

1	0.8	50	Phenotypic-DH	123.89	93.25	0.51	0.46	0.57	0.00	0.00
5	0.8	50	Phenotypic-DH	175.43	19.03	0.51	0.27	0.75	0.05	12.70
10	0.8	50	Phenotypic-DH	197.95	0.74	0.51	0.09	0.91	0.40	63.40
0	0.2	100	Phenotypic-DH	99.30	104.35	0.55	0.49	0.50	0.00	0.00
1	0.2	100	Phenotypic-DH	110.40	101.62	0.54	0.48	0.53	0.00	0.00
5	0.2	100	Phenotypic-DH	147.06	53.95	0.54	0.39	0.64	0.00	0.50
10	0.2	100	Phenotypic-DH	176.11	16.83	0.54	0.26	0.76	0.15	14.90
0	0.5	100	Phenotypic-DH	100.11	109.83	0.55	0.49	0.50	0.00	0.00
1	0.5	100	Phenotypic-DH	117.89	97.22	0.54	0.47	0.55	0.00	0.00
5	0.5	100	Phenotypic-DH	165.54	29.44	0.54	0.33	0.71	0.00	4.20
10	0.5	100	Phenotypic-DH	192.20	3.23	0.54	0.16	0.86	0.05	37.75
0	0.8	100	Phenotypic-DH	100.07	107.45	0.55	0.49	0.50	0.00	0.00
1	0.8	100	Phenotypic-DH	123.21	85.83	0.54	0.47	0.57	0.00	0.00
5	0.8	100	Phenotypic-DH	173.88	18.24	0.54	0.29	0.75	0.00	7.85
10	0.8	100	Phenotypic-DH	195.99	0.67	0.54	0.10	0.91	0.00	58.65
0	0.2	5	Phenot-Ch-DH	100.88	107.01	0.44	0.45	0.50	0.20	0.40
1	0.2	5	Phenot-Ch-DH	118.69	81.49	0.38	0.33	0.55	5.75	16.45
5	0.2	5	Phenot-Ch-DH	155.25	10.61	0.39	0.08	0.67	23.80	78.05
10	0.2	5	Phenot-Ch-DH	165.56	0.34	0.27	0.01	0.72	27.00	97.50
0	0.5	5	Phenot-Ch-DH	98.06	109.11	0.45	0.44	0.50	0.05	0.30
1	0.5	5	Phenot-Ch-DH	124.05	64.77	0.39	0.29	0.57	10.55	27.60
5	0.5	5	Phenot-Ch-DH	162.52	2.17	0.39	0.05	0.71	23.05	86.20
10	0.5	5	Phenot-Ch-DH	166.87	0.00	0.02	0.00	0.75	24.65	99.85
0	0.8	5	Phenot-Ch-DH	95.67	90.95	0.45	0.44	0.50	0.15	0.30
1	0.8	5	Phenot-Ch-DH	125.00	50.96	0.39	0.27	0.58	10.65	30.65
5	0.8	5	Phenot-Ch-DH	163.04	0.51	0.29	0.02	0.74	23.30	94.00
10	0.8	5	Phenot-Ch-DH	164.56	0.00	0.00	0.00	0.76	23.55	100.00
0	0.2	20	Phenot-Ch-DH	100.34	105.38	0.49	0.48	0.50	0.00	0.00
1	0.2	20	Phenot-Ch-DH	114.96	95.97	0.49	0.44	0.54	0.00	0.00
5	0.2	20	Phenot-Ch-DH	157.47	31.33	0.49	0.26	0.68	2.55	22.00
10	0.2	20	Phenot-Ch-DH	182.54	4.19	0.48	0.11	0.79	7.95	65.30
0	0.5	20	Phenot-Ch-DH	98.70	101.39	0.50	0.48	0.50	0.00	0.00
1	0.5	20	Phenot-Ch-DH	122.14	81.01	0.49	0.42	0.57	0.00	0.45
5	0.5	20	Phenot-Ch-DH	170.48	11.17	0.49	0.18	0.75	5.10	43.00
10	0.5	20	Phenot-Ch-DH	187.77	0.35	0.46	0.04	0.88	7.55	85.90
0	0.8	20	Phenot-Ch-DH	104.11	113.27	0.50	0.48	0.50	0.00	0.00
1	0.8	20	Phenot-Ch-DH	133.81	79.57	0.48	0.41	0.58	0.10	1.15
5	0.8	20	Phenot-Ch-DH	190.78	4.53	0.48	0.13	0.81	5.00	58.60
10	0.8	20	Phenot-Ch-DH	201.35	0.00	0.34	0.01	0.92	6.85	97.60
0	0.2	50	Phenot-Ch-DH	102.03	113.03	0.52	0.49	0.50	0.00	0.00
1	0.2	50	Phenot-Ch-DH	115.12	105.20	0.52	0.47	0.53	0.00	0.00
5	0.2	50	Phenot-Ch-DH	155.80	44.45	0.52	0.36	0.65	0.00	3.30
10	0.2	50	Phenot-Ch-DH	183.76	10.32	0.51	0.22	0.77	0.45	27.95
0	0.5	50	Phenot-Ch-DH	98.75	110.02	0.52	0.49	0.50	0.00	0.00
1	0.5	50	Phenot-Ch-DH	118.94	96.53	0.52	0.46	0.55	0.00	0.00
5	0.5	50	Phenot-Ch-DH	170.77	19.65	0.51	0.27	0.74	0.40	14.25

10	0.5	50	Phenot-Ch-DH	194.10	1.14	0.51	0.09	0.89	1.65	63.30
0	0.8	50	Phenot-Ch-DH	100.77	109.45	0.52	0.49	0.50	0.00	0.00
1	0.8	50	Phenot-Ch-DH	127.38	85.68	0.52	0.45	0.58	0.00	0.00
5	0.8	50	Phenot-Ch-DH	182.88	10.56	0.51	0.21	0.79	0.35	26.75
10	0.8	50	Phenot-Ch-DH	199.99	0.11	0.52	0.04	0.95	1.20	82.95
0	0.2	100	Phenot-Ch-DH	99.30	104.35	0.55	0.49	0.50	0.00	0.00
1	0.2	100	Phenot-Ch-DH	110.25	101.48	0.55	0.48	0.53	0.00	0.00
5	0.2	100	Phenot-Ch-DH	147.31	54.81	0.55	0.40	0.64	0.00	0.20
10	0.2	100	Phenot-Ch-DH	176.38	16.32	0.54	0.27	0.76	0.05	10.45
0	0.5	100	Phenot-Ch-DH	100.11	109.83	0.55	0.49	0.50	0.00	0.00
1	0.5	100	Phenot-Ch-DH	117.52	98.36	0.55	0.48	0.55	0.00	0.00
5	0.5	100	Phenot-Ch-DH	166.45	27.86	0.54	0.33	0.71	0.05	4.10
10	0.5	100	Phenot-Ch-DH	192.83	2.87	0.54	0.15	0.86	0.05	37.00
0	0.8	100	Phenot-Ch-DH	100.07	107.45	0.55	0.49	0.50	0.00	0.00
1	0.8	100	Phenot-Ch-DH	123.04	88.44	0.55	0.47	0.56	0.00	0.00
5	0.8	100	Phenot-Ch-DH	176.53	15.18	0.55	0.27	0.76	0.00	11.85
10	0.8	100	Phenot-Ch-DH	196.53	0.47	0.54	0.08	0.92	0.30	63.65
0	0.2	5	Phenot-Ch-RILs	100.88	107.01	0.44	0.45	0.50	0.20	0.40
1	0.2	5	Phenot-Ch-RILs	118.79	82.02	0.38	0.33	0.55	5.75	16.45
5	0.2	5	Phenot-Ch-RILs	156.04	14.68	0.40	0.10	0.67	21.55	73.55
10	0.2	5	Phenot-Ch-RILs	168.93	0.30	0.31	0.01	0.73	25.65	97.00
0	0.5	5	Phenot-Ch-RILs	98.06	109.11	0.45	0.44	0.50	0.05	0.30
1	0.5	5	Phenot-Ch-RILs	124.07	65.82	0.39	0.29	0.57	10.55	27.60
5	0.5	5	Phenot-Ch-RILs	164.52	2.27	0.38	0.05	0.71	23.15	86.70
10	0.5	5	Phenot-Ch-RILs	169.05	0.00	0.01	0.00	0.76	24.25	99.85
0	0.8	5	Phenot-Ch-RILs	95.67	90.95	0.45	0.44	0.50	0.15	0.30
1	0.8	5	Phenot-Ch-RILs	124.88	51.95	0.39	0.27	0.58	10.65	30.65
5	0.8	5	Phenot-Ch-RILs	163.23	0.45	0.35	0.02	0.74	23.35	94.10
10	0.8	5	Phenot-Ch-RILs	164.45	0.00	0.00	0.00	0.76	23.55	100.00
0	0.2	20	Phenot-Ch-RILs	100.34	105.38	0.49	0.48	0.50	0.00	0.00
1	0.2	20	Phenot-Ch-RILs	114.92	94.34	0.49	0.44	0.54	0.00	0.00
5	0.2	20	Phenot-Ch-RILs	156.58	30.92	0.48	0.26	0.68	2.15	21.90
10	0.2	20	Phenot-Ch-RILs	182.03	5.74	0.48	0.13	0.79	6.50	61.00
0	0.5	20	Phenot-Ch-RILs	98.70	101.39	0.50	0.48	0.50	0.00	0.00
1	0.5	20	Phenot-Ch-RILs	122.01	80.49	0.49	0.42	0.57	0.00	0.45
5	0.5	20	Phenot-Ch-RILs	171.39	11.04	0.48	0.19	0.76	3.80	41.30
10	0.5	20	Phenot-Ch-RILs	188.36	0.32	0.46	0.04	0.89	6.95	86.00
0	0.8	20	Phenot-Ch-RILs	104.11	113.27	0.50	0.48	0.50	0.00	0.00
1	0.8	20	Phenot-Ch-RILs	133.70	79.07	0.48	0.41	0.58	0.10	1.15
5	0.8	20	Phenot-Ch-RILs	190.73	5.66	0.49	0.15	0.80	3.80	52.60
10	0.8	20	Phenot-Ch-RILs	202.86	0.01	0.33	0.01	0.94	5.40	97.20
0	0.2	50	Phenot-Ch-RILs	102.03	113.03	0.52	0.49	0.50	0.00	0.00
1	0.2	50	Phenot-Ch-RILs	115.20	106.45	0.52	0.47	0.53	0.00	0.00
5	0.2	50	Phenot-Ch-RILs	155.85	45.78	0.51	0.36	0.65	0.00	2.25
10	0.2	50	Phenot-Ch-RILs	184.14	10.18	0.51	0.22	0.78	0.70	27.85
0	0.5	50	Phenot-Ch-RILs	98.75	110.02	0.52	0.49	0.50	0.00	0.00

1	0.5	50	Phenot-Ch-RILs	119.01	94.36	0.52	0.46	0.55	0.00	0.00
5	0.5	50	Phenot-Ch-RILs	170.72	19.79	0.51	0.27	0.73	0.15	14.15
10	0.5	50	Phenot-Ch-RILs	194.09	1.30	0.52	0.10	0.89	1.00	61.45
0	0.8	50	Phenot-Ch-RILs	100.77	109.45	0.52	0.49	0.50	0.00	0.00
1	0.8	50	Phenot-Ch-RILs	127.55	85.44	0.52	0.45	0.58	0.00	0.00
5	0.8	50	Phenot-Ch-RILs	182.12	10.94	0.52	0.23	0.78	0.15	24.45
10	0.8	50	Phenot-Ch-RILs	200.23	0.12	0.51	0.05	0.94	0.90	80.90
0	0.2	100	Phenot-Ch-RILs	99.30	104.35	0.55	0.49	0.50	0.00	0.00
1	0.2	100	Phenot-Ch-RILs	110.24	101.47	0.55	0.48	0.53	0.00	0.00
5	0.2	100	Phenot-Ch-RILs	148.02	53.84	0.55	0.40	0.64	0.00	0.20
10	0.2	100	Phenot-Ch-RILs	176.80	15.43	0.54	0.27	0.76	0.00	10.85
0	0.5	100	Phenot-Ch-RILs	100.11	109.83	0.55	0.49	0.50	0.00	0.00
1	0.5	100	Phenot-Ch-RILs	117.59	98.09	0.55	0.48	0.55	0.00	0.00
5	0.5	100	Phenot-Ch-RILs	166.56	27.81	0.54	0.34	0.71	0.00	4.05
10	0.5	100	Phenot-Ch-RILs	192.96	2.97	0.54	0.16	0.86	0.05	36.80
0	0.8	100	Phenot-Ch-RILs	100.07	107.45	0.55	0.49	0.50	0.00	0.00
1	0.8	100	Phenot-Ch-RILs	122.94	87.30	0.55	0.47	0.56	0.00	0.00
5	0.8	100	Phenot-Ch-RILs	176.87	14.97	0.54	0.27	0.76	0.00	10.85
10	0.8	100	Phenot-Ch-RILs	196.78	0.40	0.54	0.08	0.92	0.20	64.85
0	0.2	5	Random	100.88	107.01	0.44	0.45	0.50	0.20	0.40
1	0.2	5	Random	100.12	101.52	0.44	0.41	0.49	1.40	2.30
5	0.2	5	Random	102.06	56.56	0.45	0.27	0.50	14.40	28.20
10	0.2	5	Random	102.54	31.00	0.48	0.16	0.50	28.45	57.40
0	0.5	5	Random	98.06	109.11	0.45	0.44	0.50	0.05	0.30
1	0.5	5	Random	98.88	101.87	0.43	0.40	0.50	1.70	3.30
5	0.5	5	Random	95.86	60.56	0.44	0.25	0.49	17.80	34.40
10	0.5	5	Random	96.23	39.68	0.44	0.15	0.48	31.60	60.85
0	0.8	5	Random	95.67	90.95	0.45	0.44	0.50	0.15	0.30
1	0.8	5	Random	96.23	89.18	0.47	0.39	0.50	1.95	3.95
5	0.8	5	Random	97.72	52.54	0.46	0.25	0.50	16.05	32.40
10	0.8	5	Random	98.47	40.34	0.43	0.16	0.51	27.85	57.75
0	0.2	20	Random	100.34	105.38	0.49	0.48	0.50	0.00	0.00
1	0.2	20	Random	100.04	109.36	0.49	0.47	0.50	0.00	0.00
5	0.2	20	Random	99.79	98.70	0.49	0.43	0.50	0.05	0.25
10	0.2	20	Random	99.38	82.63	0.50	0.37	0.50	1.40	3.50
0	0.5	20	Random	98.70	101.39	0.50	0.48	0.50	0.00	0.00
1	0.5	20	Random	98.27	102.89	0.50	0.47	0.50	0.00	0.00
5	0.5	20	Random	97.45	92.47	0.49	0.42	0.50	0.05	0.10
10	0.5	20	Random	96.98	76.91	0.49	0.37	0.50	2.00	3.65
0	0.8	20	Random	104.11	113.27	0.50	0.48	0.50	0.00	0.00
1	0.8	20	Random	104.49	111.72	0.50	0.47	0.50	0.00	0.00
5	0.8	20	Random	105.28	100.74	0.50	0.42	0.51	0.05	0.20
10	0.8	20	Random	103.81	88.35	0.49	0.37	0.50	2.15	4.00
0	0.2	50	Random	102.03	113.03	0.52	0.49	0.50	0.00	0.00
1	0.2	50	Random	102.76	115.52	0.52	0.48	0.50	0.00	0.00
5	0.2	50	Random	102.35	107.52	0.52	0.47	0.50	0.00	0.00

10	0.2	50	Random	102.01	100.22	0.52	0.44	0.50	0.00	0.00
0	0.5	50	Random	98.75	110.02	0.52	0.49	0.50	0.00	0.00
1	0.5	50	Random	98.68	105.94	0.52	0.48	0.50	0.00	0.00
5	0.5	50	Random	99.16	102.24	0.52	0.46	0.50	0.00	0.00
10	0.5	50	Random	98.60	99.11	0.52	0.44	0.50	0.00	0.05
0	0.8	50	Random	100.77	109.45	0.52	0.49	0.50	0.00	0.00
1	0.8	50	Random	100.50	107.93	0.52	0.48	0.50	0.00	0.00
5	0.8	50	Random	99.84	104.23	0.52	0.46	0.50	0.00	0.00
10	0.8	50	Random	100.75	96.11	0.52	0.44	0.50	0.05	0.05
0	0.2	100	Random	99.30	104.35	0.55	0.49	0.50	0.00	0.00
1	0.2	100	Random	99.16	104.91	0.55	0.49	0.50	0.00	0.00
5	0.2	100	Random	99.31	101.76	0.55	0.48	0.50	0.00	0.00
10	0.2	100	Random	100.11	97.82	0.55	0.46	0.50	0.00	0.00
0	0.5	100	Random	100.11	109.83	0.55	0.49	0.50	0.00	0.00
1	0.5	100	Random	100.05	109.53	0.55	0.49	0.50	0.00	0.00
5	0.5	100	Random	100.00	105.20	0.55	0.48	0.50	0.00	0.00
10	0.5	100	Random	99.93	104.51	0.55	0.47	0.50	0.00	0.00
0	0.8	100	Random	100.07	107.45	0.55	0.49	0.50	0.00	0.00
1	0.8	100	Random	99.64	104.56	0.55	0.49	0.50	0.00	0.00
5	0.8	100	Random	98.87	103.67	0.55	0.48	0.50	0.00	0.00
10	0.8	100	Random	98.38	101.66	0.55	0.47	0.49	0.00	0.00

Annex G

Table S2. Data-based simulation results for the cycles 0, 1, 5, and 10, for the combination of heritability (h^2), number of population (N_{POP}), and method of selection. The variables listed below refers to the overall mean of the breeding values ($\overline{B.V.}$), genetic variance (σ^2_G), , genetic diversity according to the Nei's estimator ($G.D.$), average allele frequency (p), number of lost benefit alleles (Lost +), number of loci with fixed allele (Monom).

Cycle	h^2	N_{POP}	Method	$\overline{B.V.}$	σ^2_G	F_{st}	G.D.	p	Lost +	Monom.
0	0.2	5	Genomic-SO	39.08	338.65	0.45	0.16	0.51	24.00	47.50
1	0.2	5	Genomic-SO	79.85	245.59	0.28	0.09	0.55	34.25	75.30
5	0.2	5	Genomic-SO	135.26	0.90	0.20	0.00	0.62	37.50	98.85
10	0.2	5	Genomic-SO	135.77	0.00	0.00	0.00	0.62	37.55	99.95
0	0.5	5	Genomic-SO	-0.33	315.15	0.45	0.15	0.51	22.85	48.10
1	0.5	5	Genomic-SO	54.02	246.43	0.27	0.09	0.56	33.20	76.50
5	0.5	5	Genomic-SO	107.68	0.07	0.10	0.00	0.64	36.20	99.30
10	0.5	5	Genomic-SO	107.82	0.00	0.00	0.00	0.64	36.25	100.00
0	0.8	5	Genomic-SO	31.32	342.72	0.45	0.17	0.53	21.50	44.60
1	0.8	5	Genomic-SO	92.75	252.30	0.26	0.09	0.59	30.25	75.50
5	0.8	5	Genomic-SO	147.03	0.12	0.15	0.00	0.67	32.30	99.20
10	0.8	5	Genomic-SO	147.21	0.00	0.00	0.00	0.68	32.35	99.95
0	0.2	20	Genomic-SO	24.18	319.83	0.49	0.17	0.52	10.95	21.00
1	0.2	20	Genomic-SO	72.38	194.60	0.39	0.11	0.57	26.55	62.95
5	0.2	20	Genomic-SO	137.20	4.74	0.39	0.01	0.65	32.95	94.75
10	0.2	20	Genomic-SO	140.12	0.00	0.00	0.00	0.67	33.35	99.90
0	0.5	20	Genomic-SO	29.57	346.41	0.50	0.17	0.53	9.75	23.00
1	0.5	20	Genomic-SO	80.75	334.30	0.36	0.13	0.58	22.05	57.50
5	0.5	20	Genomic-SO	171.91	9.98	0.36	0.02	0.70	27.55	91.75
10	0.5	20	Genomic-SO	176.79	0.00	0.02	0.00	0.72	27.85	99.85
0	0.8	20	Genomic-SO	28.18	286.78	0.50	0.17	0.52	9.25	20.85
1	0.8	20	Genomic-SO	83.97	290.96	0.38	0.13	0.58	22.70	56.75
5	0.8	20	Genomic-SO	168.05	5.28	0.38	0.01	0.71	27.75	94.15
10	0.8	20	Genomic-SO	170.85	0.00	0.00	0.00	0.72	27.95	99.90
0	0.2	50	Genomic-SO	33.80	289.01	0.51	0.17	0.51	4.40	9.10
1	0.2	50	Genomic-SO	73.55	268.05	0.44	0.14	0.55	20.70	48.50
5	0.2	50	Genomic-SO	151.63	20.87	0.45	0.03	0.66	29.90	86.30
10	0.2	50	Genomic-SO	162.25	0.04	0.05	0.00	0.69	31.15	99.75
0	0.5	50	Genomic-SO	28.52	341.80	0.52	0.18	0.52	4.60	9.15
1	0.5	50	Genomic-SO	83.83	306.13	0.44	0.14	0.57	17.15	43.00
5	0.5	50	Genomic-SO	181.00	24.51	0.44	0.04	0.70	26.05	83.80
10	0.5	50	Genomic-SO	191.46	0.00	0.02	0.00	0.73	27.05	99.85
0	0.8	50	Genomic-SO	29.77	391.15	0.52	0.18	0.52	3.85	7.90
1	0.8	50	Genomic-SO	93.48	326.63	0.44	0.14	0.57	15.35	38.90
5	0.8	50	Genomic-SO	199.71	26.35	0.44	0.03	0.71	24.60	84.10
10	0.8	50	Genomic-SO	210.75	0.00	0.06	0.00	0.74	25.60	99.85
0	0.2	100	Genomic-SO	4.14	311.57	0.54	0.17	0.50	1.55	3.20
1	0.2	100	Genomic-SO	43.26	327.64	0.49	0.16	0.54	13.65	30.75
5	0.2	100	Genomic-SO	145.34	45.91	0.48	0.05	0.65	27.50	78.05

10	0.2	100	Genomic-SO	166.13	0.19	0.33	0.00	0.70	29.65	97.75
0	0.5	100	Genomic-SO	20.02	346.31	0.54	0.17	0.52	1.45	2.95
1	0.5	100	Genomic-SO	71.51	324.71	0.49	0.16	0.57	11.90	29.05
5	0.5	100	Genomic-SO	177.78	26.39	0.48	0.04	0.70	24.10	80.35
10	0.5	100	Genomic-SO	192.46	0.01	0.22	0.00	0.74	25.55	99.25
0	0.8	100	Genomic-SO	30.60	351.63	0.54	0.18	0.52	1.65	3.05
1	0.8	100	Genomic-SO	91.51	303.72	0.49	0.15	0.58	10.90	29.65
5	0.8	100	Genomic-SO	198.40	25.96	0.48	0.05	0.72	20.60	77.40
10	0.8	100	Genomic-SO	215.60	0.02	0.22	0.00	0.77	22.25	98.85
0	0.2	5	Genomic-SO-Const	39.08	338.65	0.45	0.16	0.51	24.00	47.50
1	0.2	5	Genomic-SO-Const	79.52	248.27	0.30	0.10	0.55	34.00	75.15
5	0.2	5	Genomic-SO-Const	136.22	1.05	0.22	0.00	0.62	37.75	98.45
10	0.2	5	Genomic-SO-Const	136.93	0.00	0.00	0.00	0.62	37.75	99.95
0	0.5	5	Genomic-SO-Const	-0.33	315.15	0.45	0.15	0.51	22.85	48.10
1	0.5	5	Genomic-SO-Const	52.91	243.90	0.30	0.09	0.56	32.80	76.30
5	0.5	5	Genomic-SO-Const	109.96	0.14	0.10	0.00	0.64	36.20	99.30
10	0.5	5	Genomic-SO-Const	110.12	0.00	0.00	0.00	0.64	36.25	100.00
0	0.8	5	Genomic-SO-Const	31.32	342.72	0.45	0.17	0.53	21.50	44.60
1	0.8	5	Genomic-SO-Const	91.77	264.12	0.26	0.10	0.59	30.15	75.10
5	0.8	5	Genomic-SO-Const	147.47	0.08	0.17	0.00	0.67	32.40	99.20
10	0.8	5	Genomic-SO-Const	147.58	0.00	0.00	0.00	0.67	32.50	99.95
0	0.2	20	Genomic-SO-Const	24.18	319.83	0.49	0.17	0.52	10.95	21.00
1	0.2	20	Genomic-SO-Const	70.42	214.56	0.40	0.12	0.56	26.55	63.25
5	0.2	20	Genomic-SO-Const	141.05	9.55	0.41	0.02	0.66	31.70	92.60
10	0.2	20	Genomic-SO-Const	145.51	0.00	0.00	0.00	0.68	32.35	99.95
0	0.5	20	Genomic-SO-Const	29.57	346.41	0.50	0.17	0.53	9.75	23.00
1	0.5	20	Genomic-SO-Const	78.72	350.02	0.38	0.14	0.58	21.30	55.75
5	0.5	20	Genomic-SO-Const	174.76	14.77	0.41	0.03	0.70	26.30	88.90
10	0.5	20	Genomic-SO-Const	182.26	0.00	0.02	0.00	0.73	27.15	99.85
0	0.8	20	Genomic-SO-Const	28.18	286.78	0.50	0.17	0.52	9.25	20.85
1	0.8	20	Genomic-SO-Const	81.89	291.58	0.39	0.14	0.58	22.40	55.65
5	0.8	20	Genomic-SO-Const	165.68	5.50	0.41	0.02	0.70	27.45	92.25
10	0.8	20	Genomic-SO-Const	169.30	0.00	0.00	0.00	0.72	27.90	99.90
0	0.2	50	Genomic-SO-Const	33.80	289.01	0.51	0.17	0.51	4.40	9.10
1	0.2	50	Genomic-SO-Const	69.51	279.32	0.46	0.16	0.55	18.40	43.45
5	0.2	50	Genomic-SO-Const	154.84	47.52	0.46	0.05	0.66	27.35	80.00
10	0.2	50	Genomic-SO-Const	173.45	0.20	0.22	0.01	0.71	28.80	98.30
0	0.5	50	Genomic-SO-Const	28.52	341.80	0.52	0.18	0.52	4.60	9.15
1	0.5	50	Genomic-SO-Const	75.60	324.00	0.47	0.16	0.56	15.90	39.95
5	0.5	50	Genomic-SO-Const	175.86	34.23	0.45	0.06	0.69	24.15	78.00
10	0.5	50	Genomic-SO-Const	193.34	0.07	0.26	0.00	0.74	25.45	98.70
0	0.8	50	Genomic-SO-Const	29.77	391.15	0.52	0.18	0.52	3.85	7.90
1	0.8	50	Genomic-SO-Const	83.38	365.29	0.46	0.16	0.56	13.70	35.00
5	0.8	50	Genomic-SO-Const	199.68	44.09	0.45	0.07	0.71	21.40	74.85
10	0.8	50	Genomic-SO-Const	223.04	0.21	0.26	0.00	0.77	22.25	98.20
0	0.2	100	Genomic-SO-Const	4.14	311.57	0.54	0.17	0.50	1.55	3.20

1	0.2	100	Genomic-SO-Const	34.67	344.54	0.52	0.18	0.53	10.80	23.75
5	0.2	100	Genomic-SO-Const	131.74	123.94	0.50	0.11	0.64	21.60	60.15
10	0.2	100	Genomic-SO-Const	177.68	6.91	0.49	0.03	0.71	25.40	88.10
0	0.5	100	Genomic-SO-Const	20.02	346.31	0.54	0.17	0.52	1.45	2.95
1	0.5	100	Genomic-SO-Const	59.80	352.58	0.52	0.17	0.56	10.30	24.15
5	0.5	100	Genomic-SO-Const	163.83	80.62	0.50	0.10	0.68	19.35	62.05
10	0.5	100	Genomic-SO-Const	205.12	1.99	0.49	0.02	0.77	21.45	91.25
0	0.8	100	Genomic-SO-Const	30.60	351.63	0.54	0.18	0.52	1.65	3.05
1	0.8	100	Genomic-SO-Const	76.29	359.71	0.51	0.17	0.56	8.35	22.60
5	0.8	100	Genomic-SO-Const	185.91	68.23	0.49	0.09	0.70	17.45	63.80
10	0.8	100	Genomic-SO-Const	224.16	1.08	0.48	0.01	0.79	19.20	93.30
0	0.2	5	Genomic-GEBV	39.08	338.65	0.45	0.16	0.51	24.00	47.50
1	0.2	5	Genomic-GEBV	81.31	156.62	0.39	0.08	0.55	35.70	79.10
5	0.2	5	Genomic-GEBV	121.89	0.01	0.05	0.00	0.60	39.90	99.75
10	0.2	5	Genomic-GEBV	121.90	0.00	0.02	0.00	0.60	39.95	99.90
0	0.5	5	Genomic-GEBV	-0.33	315.15	0.45	0.15	0.51	22.85	48.10
1	0.5	5	Genomic-GEBV	52.74	111.40	0.44	0.07	0.56	35.55	80.85
5	0.5	5	Genomic-GEBV	90.43	0.00	0.00	0.00	0.61	38.75	100.00
10	0.5	5	Genomic-GEBV	90.43	0.00	0.00	0.00	0.61	38.75	100.00
0	0.8	5	Genomic-GEBV	31.32	342.72	0.45	0.17	0.53	21.50	44.60
1	0.8	5	Genomic-GEBV	93.80	86.38	0.40	0.07	0.59	33.35	82.70
5	0.8	5	Genomic-GEBV	127.14	0.00	0.00	0.00	0.65	35.10	99.95
10	0.8	5	Genomic-GEBV	127.14	0.00	0.00	0.00	0.65	35.10	99.95
0	0.2	20	Genomic-GEBV	24.18	319.83	0.49	0.17	0.52	10.95	21.00
1	0.2	20	Genomic-GEBV	68.04	163.03	0.49	0.12	0.56	27.55	64.80
5	0.2	20	Genomic-GEBV	134.27	6.38	0.40	0.02	0.65	32.60	93.90
10	0.2	20	Genomic-GEBV	139.69	0.00	0.00	0.00	0.67	33.05	99.95
0	0.5	20	Genomic-GEBV	29.57	346.41	0.50	0.17	0.53	9.75	23.00
1	0.5	20	Genomic-GEBV	78.87	236.32	0.49	0.14	0.58	21.30	55.05
5	0.5	20	Genomic-GEBV	166.45	7.82	0.48	0.02	0.68	28.25	91.95
10	0.5	20	Genomic-GEBV	172.72	0.00	0.00	0.00	0.71	28.90	99.95
0	0.8	20	Genomic-GEBV	28.18	286.78	0.50	0.17	0.52	9.25	20.85
1	0.8	20	Genomic-GEBV	81.17	183.93	0.49	0.13	0.58	22.45	56.90
5	0.8	20	Genomic-GEBV	161.99	3.79	0.40	0.02	0.69	28.30	93.15
10	0.8	20	Genomic-GEBV	166.91	0.00	0.00	0.00	0.71	28.90	99.95
0	0.2	50	Genomic-GEBV	33.80	289.01	0.51	0.17	0.51	4.40	9.10
1	0.2	50	Genomic-GEBV	66.94	245.71	0.53	0.16	0.55	16.40	38.50
5	0.2	50	Genomic-GEBV	147.45	44.84	0.52	0.06	0.65	27.55	79.40
10	0.2	50	Genomic-GEBV	171.83	1.06	0.28	0.01	0.70	29.10	97.60
0	0.5	50	Genomic-GEBV	28.52	341.80	0.52	0.18	0.52	4.60	9.15
1	0.5	50	Genomic-GEBV	72.72	247.14	0.53	0.16	0.56	14.45	36.65
5	0.5	50	Genomic-GEBV	166.66	30.98	0.53	0.06	0.67	24.35	76.75
10	0.5	50	Genomic-GEBV	192.99	0.08	0.27	0.00	0.74	26.05	98.75
0	0.8	50	Genomic-GEBV	29.77	391.15	0.52	0.18	0.52	3.85	7.90
1	0.8	50	Genomic-GEBV	80.50	289.41	0.51	0.16	0.56	12.55	32.40
5	0.8	50	Genomic-GEBV	186.48	42.48	0.51	0.07	0.68	22.00	73.05

10	0.8	50	Genomic-GEBV	218.80	0.06	0.29	0.00	0.76	23.65	98.55
0	0.2	100	Genomic-GEBV	4.14	311.57	0.54	0.17	0.50	1.55	3.20
1	0.2	100	Genomic-GEBV	32.01	308.52	0.54	0.17	0.52	9.05	19.95
5	0.2	100	Genomic-GEBV	120.99	115.09	0.55	0.11	0.62	21.15	56.65
10	0.2	100	Genomic-GEBV	173.27	12.57	0.54	0.04	0.70	24.95	83.15
0	0.5	100	Genomic-GEBV	20.02	346.31	0.54	0.17	0.52	1.45	2.95
1	0.5	100	Genomic-GEBV	57.50	297.85	0.55	0.17	0.55	8.45	20.20
5	0.5	100	Genomic-GEBV	151.04	86.71	0.55	0.11	0.66	17.25	56.35
10	0.5	100	Genomic-GEBV	204.23	5.77	0.55	0.03	0.75	20.75	86.05
0	0.8	100	Genomic-GEBV	30.60	351.63	0.54	0.18	0.52	1.65	3.05
1	0.8	100	Genomic-GEBV	73.01	272.33	0.54	0.16	0.56	6.70	18.70
5	0.8	100	Genomic-GEBV	171.78	68.43	0.54	0.10	0.67	16.80	59.45
10	0.8	100	Genomic-GEBV	216.96	3.87	0.56	0.03	0.76	19.90	87.95
0	0.2	5	Phenotypic-DH	39.08	338.65	0.45	0.16	0.51	24.00	47.50
1	0.2	5	Phenotypic-DH	56.24	260.59	0.45	0.13	0.53	28.45	59.70
5	0.2	5	Phenotypic-DH	109.84	80.82	0.45	0.06	0.58	34.45	82.85
10	0.2	5	Phenotypic-DH	134.25	5.03	0.36	0.01	0.62	36.50	96.20
0	0.5	5	Phenotypic-DH	-0.33	315.15	0.45	0.15	0.51	22.85	48.10
1	0.5	5	Phenotypic-DH	31.88	242.11	0.43	0.12	0.54	28.55	63.55
5	0.5	5	Phenotypic-DH	95.07	35.46	0.41	0.04	0.61	34.25	88.10
10	0.5	5	Phenotypic-DH	113.83	0.60	0.12	0.00	0.64	35.60	99.25
0	0.8	5	Phenotypic-DH	31.32	342.72	0.45	0.17	0.53	21.50	44.60
1	0.8	5	Phenotypic-DH	71.43	223.80	0.46	0.12	0.57	28.45	66.55
5	0.8	5	Phenotypic-DH	132.93	9.50	0.37	0.02	0.64	33.15	93.00
10	0.8	5	Phenotypic-DH	139.40	0.00	0.00	0.00	0.66	33.75	99.95
0	0.2	20	Phenotypic-DH	24.18	319.83	0.49	0.17	0.52	10.95	21.00
1	0.2	20	Phenotypic-DH	43.77	300.83	0.49	0.16	0.53	15.45	34.35
5	0.2	20	Phenotypic-DH	105.97	127.41	0.49	0.12	0.60	22.75	59.55
10	0.2	20	Phenotypic-DH	149.26	34.75	0.48	0.06	0.66	26.00	78.40
0	0.5	20	Phenotypic-DH	29.57	346.41	0.50	0.17	0.53	9.75	23.00
1	0.5	20	Phenotypic-DH	60.91	292.50	0.49	0.16	0.56	14.55	36.80
5	0.5	20	Phenotypic-DH	143.90	89.23	0.48	0.09	0.64	22.95	69.20
10	0.5	20	Phenotypic-DH	184.50	5.40	0.49	0.03	0.71	25.40	90.35
0	0.8	20	Phenotypic-DH	28.18	286.78	0.50	0.17	0.52	9.25	20.85
1	0.8	20	Phenotypic-DH	65.89	245.29	0.48	0.16	0.56	15.90	38.30
5	0.8	20	Phenotypic-DH	151.14	51.36	0.49	0.08	0.66	23.00	70.95
10	0.8	20	Phenotypic-DH	185.96	1.35	0.46	0.02	0.73	24.40	94.10
0	0.2	50	Phenotypic-DH	33.80	289.01	0.51	0.17	0.51	4.40	9.10
1	0.2	50	Phenotypic-DH	51.81	267.83	0.51	0.17	0.53	8.90	19.45
5	0.2	50	Phenotypic-DH	114.76	179.01	0.51	0.14	0.60	16.25	45.15
10	0.2	50	Phenotypic-DH	168.40	66.33	0.51	0.09	0.67	19.80	65.80
0	0.5	50	Phenotypic-DH	28.52	341.80	0.52	0.18	0.52	4.60	9.15
1	0.5	50	Phenotypic-DH	60.40	297.76	0.51	0.17	0.55	9.40	22.55
5	0.5	50	Phenotypic-DH	149.58	115.92	0.51	0.12	0.64	16.90	55.05
10	0.5	50	Phenotypic-DH	204.17	15.30	0.51	0.06	0.73	19.35	78.10
0	0.8	50	Phenotypic-DH	29.77	391.15	0.52	0.18	0.52	3.85	7.90

1	0.8	50	Phenotypic-DH	73.37	310.98	0.51	0.17	0.55	9.50	23.20
5	0.8	50	Phenotypic-DH	174.46	95.23	0.51	0.11	0.66	16.30	56.90
10	0.8	50	Phenotypic-DH	229.81	7.51	0.51	0.04	0.76	18.40	84.20
0	0.2	100	Phenotypic-DH	4.14	311.57	0.54	0.17	0.50	1.55	3.20
1	0.2	100	Phenotypic-DH	23.03	309.16	0.54	0.17	0.52	6.25	14.05
5	0.2	100	Phenotypic-DH	91.49	210.21	0.54	0.15	0.58	14.15	36.70
10	0.2	100	Phenotypic-DH	151.46	90.93	0.54	0.11	0.65	17.80	57.10
0	0.5	100	Phenotypic-DH	20.02	346.31	0.54	0.17	0.52	1.45	2.95
1	0.5	100	Phenotypic-DH	53.02	308.93	0.54	0.17	0.55	6.70	15.00
5	0.5	100	Phenotypic-DH	143.78	126.39	0.54	0.13	0.64	13.55	45.40
10	0.5	100	Phenotypic-DH	204.04	23.67	0.54	0.07	0.74	16.65	72.25
0	0.8	100	Phenotypic-DH	30.60	351.63	0.54	0.18	0.52	1.65	3.05
1	0.8	100	Phenotypic-DH	72.11	293.30	0.53	0.17	0.56	5.70	16.25
5	0.8	100	Phenotypic-DH	170.76	87.51	0.54	0.11	0.67	13.50	51.00
10	0.8	100	Phenotypic-DH	227.26	8.95	0.54	0.05	0.78	15.75	80.15
0	0.2	5	Phenot-Ch-DH	39.08	338.65	0.45	0.16	0.51	24.00	47.50
1	0.2	5	Phenot-Ch-DH	66.80	261.42	0.41	0.12	0.54	31.05	67.10
5	0.2	5	Phenot-Ch-DH	127.27	19.93	0.39	0.02	0.60	37.20	93.95
10	0.2	5	Phenot-Ch-DH	135.97	0.04	0.06	0.00	0.62	37.75	99.70
0	0.5	5	Phenot-Ch-DH	-0.33	315.15	0.45	0.15	0.51	22.85	48.10
1	0.5	5	Phenot-Ch-DH	44.21	212.91	0.39	0.10	0.55	32.35	73.05
5	0.5	5	Phenot-Ch-DH	103.17	1.04	0.28	0.01	0.62	37.05	98.40
10	0.5	5	Phenot-Ch-DH	104.18	0.00	0.00	0.00	0.63	37.10	100.00
0	0.8	5	Phenot-Ch-DH	31.32	342.72	0.45	0.17	0.53	21.50	44.60
1	0.8	5	Phenot-Ch-DH	89.25	177.66	0.38	0.09	0.59	30.90	76.55
5	0.8	5	Phenot-Ch-DH	138.83	0.00	0.04	0.00	0.66	33.90	99.80
10	0.8	5	Phenot-Ch-DH	138.85	0.00	0.00	0.00	0.66	33.90	99.95
0	0.2	20	Phenot-Ch-DH	24.18	319.83	0.49	0.17	0.52	10.95	21.00
1	0.2	20	Phenot-Ch-DH	52.15	272.54	0.48	0.16	0.54	17.55	40.60
5	0.2	20	Phenot-Ch-DH	123.62	91.56	0.48	0.09	0.63	25.60	70.50
10	0.2	20	Phenot-Ch-DH	164.14	8.05	0.48	0.03	0.69	27.80	90.25
0	0.5	20	Phenot-Ch-DH	29.57	346.41	0.50	0.17	0.53	9.75	23.00
1	0.5	20	Phenot-Ch-DH	71.49	295.53	0.48	0.16	0.57	17.20	44.20
5	0.5	20	Phenot-Ch-DH	164.06	38.89	0.47	0.06	0.67	25.20	79.85
10	0.5	20	Phenot-Ch-DH	188.46	0.10	0.24	0.00	0.74	25.90	98.30
0	0.8	20	Phenot-Ch-DH	28.18	286.78	0.50	0.17	0.52	9.25	20.85
1	0.8	20	Phenot-Ch-DH	80.09	208.68	0.49	0.14	0.58	19.30	48.95
5	0.8	20	Phenot-Ch-DH	169.16	13.20	0.47	0.04	0.69	26.30	88.35
10	0.8	20	Phenot-Ch-DH	179.48	0.00	0.00	0.00	0.73	27.10	99.90
0	0.2	50	Phenot-Ch-DH	33.80	289.01	0.51	0.17	0.51	4.40	9.10
1	0.2	50	Phenot-Ch-DH	54.23	269.99	0.51	0.17	0.53	9.85	22.50
5	0.2	50	Phenot-Ch-DH	121.00	148.04	0.51	0.13	0.61	18.30	50.20
10	0.2	50	Phenot-Ch-DH	175.65	41.88	0.51	0.07	0.69	21.60	72.15
0	0.5	50	Phenot-Ch-DH	28.52	341.80	0.52	0.18	0.52	4.60	9.15
1	0.5	50	Phenot-Ch-DH	66.39	290.94	0.51	0.17	0.55	10.65	26.75
5	0.5	50	Phenot-Ch-DH	162.96	80.57	0.51	0.10	0.66	19.10	62.20

10	0.5	50	Phenot-Ch-DH	209.22	4.15	0.51	0.03	0.75	21.70	88.40
0	0.8	50	Phenot-Ch-DH	29.77	391.15	0.52	0.18	0.52	3.85	7.90
1	0.8	50	Phenot-Ch-DH	79.58	292.83	0.51	0.16	0.56	10.70	27.05
5	0.8	50	Phenot-Ch-DH	190.62	51.33	0.51	0.09	0.68	18.85	67.50
10	0.8	50	Phenot-Ch-DH	231.78	0.83	0.48	0.01	0.78	20.70	94.80
0	0.2	100	Phenot-Ch-DH	4.14	311.57	0.54	0.17	0.50	1.55	3.20
1	0.2	100	Phenot-Ch-DH	23.08	311.32	0.54	0.17	0.52	5.70	13.00
5	0.2	100	Phenot-Ch-DH	91.67	218.41	0.54	0.16	0.58	12.60	33.20
10	0.2	100	Phenot-Ch-DH	154.04	94.57	0.54	0.12	0.66	15.80	52.85
0	0.5	100	Phenot-Ch-DH	20.02	346.31	0.54	0.17	0.52	1.45	2.95
1	0.5	100	Phenot-Ch-DH	52.51	309.90	0.54	0.17	0.55	6.10	13.60
5	0.5	100	Phenot-Ch-DH	143.57	123.97	0.54	0.13	0.64	12.45	42.90
10	0.5	100	Phenot-Ch-DH	206.03	25.14	0.55	0.07	0.74	16.15	70.90
0	0.8	100	Phenot-Ch-DH	30.60	351.63	0.54	0.18	0.52	1.65	3.05
1	0.8	100	Phenot-Ch-DH	72.00	287.28	0.55	0.17	0.56	5.25	14.20
5	0.8	100	Phenot-Ch-DH	172.16	73.68	0.54	0.11	0.67	13.30	52.35
10	0.8	100	Phenot-Ch-DH	225.32	8.81	0.55	0.05	0.78	15.55	80.60
0	0.2	5	Phenot-Ch-RILs	39.08	338.65	0.45	0.16	0.51	24.00	47.50
1	0.2	5	Phenot-Ch-RILs	66.71	256.90	0.42	0.12	0.53	31.05	67.10
5	0.2	5	Phenot-Ch-RILs	125.64	17.88	0.38	0.02	0.60	37.30	93.65
10	0.2	5	Phenot-Ch-RILs	134.47	0.01	0.04	0.00	0.62	37.80	99.75
0	0.5	5	Phenot-Ch-RILs	-0.33	315.15	0.45	0.15	0.51	22.85	48.10
1	0.5	5	Phenot-Ch-RILs	44.03	210.02	0.40	0.10	0.55	32.35	73.05
5	0.5	5	Phenot-Ch-RILs	104.53	1.53	0.32	0.01	0.63	36.35	97.70
10	0.5	5	Phenot-Ch-RILs	105.86	0.00	0.00	0.00	0.64	36.40	100.00
0	0.8	5	Phenot-Ch-RILs	31.32	342.72	0.45	0.17	0.53	21.50	44.60
1	0.8	5	Phenot-Ch-RILs	89.24	178.36	0.38	0.09	0.59	30.90	76.55
5	0.8	5	Phenot-Ch-RILs	139.52	0.00	0.03	0.00	0.66	33.90	99.90
10	0.8	5	Phenot-Ch-RILs	139.52	0.00	0.00	0.00	0.66	33.90	99.95
0	0.2	20	Phenot-Ch-RILs	24.18	319.83	0.49	0.17	0.52	10.95	21.00
1	0.2	20	Phenot-Ch-RILs	52.23	270.96	0.48	0.16	0.54	17.55	40.60
5	0.2	20	Phenot-Ch-RILs	124.24	91.93	0.49	0.09	0.62	25.35	70.70
10	0.2	20	Phenot-Ch-RILs	166.01	11.79	0.47	0.03	0.69	27.45	89.75
0	0.5	20	Phenot-Ch-RILs	29.57	346.41	0.50	0.17	0.53	9.75	23.00
1	0.5	20	Phenot-Ch-RILs	71.52	302.75	0.48	0.16	0.57	17.20	44.20
5	0.5	20	Phenot-Ch-RILs	164.53	35.66	0.49	0.06	0.68	24.55	79.50
10	0.5	20	Phenot-Ch-RILs	186.91	0.05	0.23	0.00	0.73	26.10	98.80
0	0.8	20	Phenot-Ch-RILs	28.18	286.78	0.50	0.17	0.52	9.25	20.85
1	0.8	20	Phenot-Ch-RILs	79.86	213.23	0.49	0.14	0.58	19.30	48.95
5	0.8	20	Phenot-Ch-RILs	170.95	11.95	0.45	0.04	0.69	26.35	88.15
10	0.8	20	Phenot-Ch-RILs	181.18	0.00	0.00	0.00	0.73	26.85	99.90
0	0.2	50	Phenot-Ch-RILs	33.80	289.01	0.51	0.17	0.51	4.40	9.10
1	0.2	50	Phenot-Ch-RILs	54.59	276.42	0.51	0.17	0.53	9.85	22.50
5	0.2	50	Phenot-Ch-RILs	123.94	148.49	0.51	0.13	0.61	17.55	50.85
10	0.2	50	Phenot-Ch-RILs	178.49	41.58	0.51	0.08	0.69	20.85	71.65
0	0.5	50	Phenot-Ch-RILs	28.52	341.80	0.52	0.18	0.52	4.60	9.15

1	0.5	50	Phenot-Ch-RILs	66.37	289.76	0.51	0.17	0.55	10.65	26.75
5	0.5	50	Phenot-Ch-RILs	161.71	90.25	0.51	0.11	0.66	18.05	59.95
10	0.5	50	Phenot-Ch-RILs	211.80	4.55	0.52	0.03	0.75	20.45	86.50
0	0.8	50	Phenot-Ch-RILs	29.77	391.15	0.52	0.18	0.52	3.85	7.90
1	0.8	50	Phenot-Ch-RILs	79.45	307.69	0.52	0.16	0.56	10.70	27.05
5	0.8	50	Phenot-Ch-RILs	191.23	57.31	0.52	0.09	0.69	18.60	66.00
10	0.8	50	Phenot-Ch-RILs	232.73	1.05	0.51	0.01	0.78	20.00	93.90
0	0.2	100	Phenot-Ch-RILs	4.14	311.57	0.54	0.17	0.50	1.55	3.20
1	0.2	100	Phenot-Ch-RILs	23.10	316.78	0.54	0.17	0.52	5.70	13.00
5	0.2	100	Phenot-Ch-RILs	92.31	218.90	0.54	0.15	0.58	12.30	32.20
10	0.2	100	Phenot-Ch-RILs	156.26	92.10	0.54	0.12	0.65	16.05	53.70
0	0.5	100	Phenot-Ch-RILs	20.02	346.31	0.54	0.17	0.52	1.45	2.95
1	0.5	100	Phenot-Ch-RILs	52.77	308.63	0.54	0.17	0.55	6.10	13.60
5	0.5	100	Phenot-Ch-RILs	144.12	135.49	0.54	0.14	0.65	12.40	42.05
10	0.5	100	Phenot-Ch-RILs	209.30	24.36	0.54	0.07	0.74	15.00	68.45
0	0.8	100	Phenot-Ch-RILs	30.60	351.63	0.54	0.18	0.52	1.65	3.05
1	0.8	100	Phenot-Ch-RILs	71.86	288.25	0.54	0.17	0.56	5.25	14.20
5	0.8	100	Phenot-Ch-RILs	172.89	78.71	0.54	0.11	0.67	12.95	51.80
10	0.8	100	Phenot-Ch-RILs	228.25	7.76	0.56	0.05	0.78	15.45	80.35
0	0.2	5	Random	39.08	338.65	0.45	0.16	0.51	24.00	47.50
1	0.2	5	Random	40.85	300.51	0.44	0.14	0.51	28.25	57.25
5	0.2	5	Random	36.03	197.81	0.43	0.09	0.50	37.20	75.50
10	0.2	5	Random	39.61	115.05	0.46	0.05	0.50	42.70	86.20
0	0.5	5	Random	-0.33	315.15	0.45	0.15	0.51	22.85	48.10
1	0.5	5	Random	1.11	288.43	0.44	0.14	0.51	26.75	56.45
5	0.5	5	Random	2.19	175.76	0.47	0.10	0.51	35.90	74.10
10	0.5	5	Random	2.70	90.72	0.47	0.05	0.51	41.25	85.30
0	0.8	5	Random	31.32	342.72	0.45	0.17	0.53	21.50	44.60
1	0.8	5	Random	26.59	330.32	0.43	0.15	0.53	25.70	54.85
5	0.8	5	Random	29.19	179.43	0.46	0.09	0.53	34.10	73.75
10	0.8	5	Random	26.01	104.44	0.44	0.06	0.52	40.35	85.10
0	0.2	20	Random	24.18	319.83	0.49	0.17	0.52	10.95	21.00
1	0.2	20	Random	23.39	289.97	0.50	0.17	0.52	14.05	28.35
5	0.2	20	Random	21.98	264.10	0.49	0.15	0.52	22.45	47.40
10	0.2	20	Random	23.21	229.87	0.50	0.14	0.52	27.15	57.05
0	0.5	20	Random	29.57	346.41	0.50	0.17	0.53	9.75	23.00
1	0.5	20	Random	29.03	337.58	0.49	0.16	0.53	13.35	29.90
5	0.5	20	Random	27.76	321.41	0.49	0.15	0.53	21.75	46.65
10	0.5	20	Random	29.46	264.83	0.49	0.13	0.53	26.65	57.05
0	0.8	20	Random	28.18	286.78	0.50	0.17	0.52	9.25	20.85
1	0.8	20	Random	27.75	289.34	0.49	0.17	0.52	12.90	28.95
5	0.8	20	Random	30.12	249.57	0.50	0.15	0.52	21.15	46.50
10	0.8	20	Random	29.62	231.43	0.49	0.13	0.53	26.40	57.75
0	0.2	50	Random	33.80	289.01	0.51	0.17	0.51	4.40	9.10
1	0.2	50	Random	34.25	289.74	0.52	0.17	0.51	7.50	15.30
5	0.2	50	Random	35.40	284.55	0.52	0.17	0.51	13.95	28.35

10	0.2	50	Random	35.57	261.53	0.52	0.16	0.51	18.55	38.05
0	0.5	50	Random	28.52	341.80	0.52	0.18	0.52	4.60	9.15
1	0.5	50	Random	28.87	324.31	0.52	0.18	0.52	7.35	14.80
5	0.5	50	Random	29.21	324.04	0.51	0.17	0.52	13.60	27.40
10	0.5	50	Random	32.40	308.09	0.52	0.16	0.52	18.35	37.95
0	0.8	50	Random	29.77	391.15	0.52	0.18	0.52	3.85	7.90
1	0.8	50	Random	29.93	389.19	0.52	0.18	0.52	6.25	12.95
5	0.8	50	Random	29.62	366.12	0.52	0.17	0.52	12.80	26.30
10	0.8	50	Random	27.92	354.61	0.51	0.16	0.51	18.50	37.95
0	0.2	100	Random	4.14	311.57	0.54	0.17	0.50	1.55	3.20
1	0.2	100	Random	3.70	308.02	0.54	0.17	0.50	3.75	7.55
5	0.2	100	Random	3.48	316.33	0.54	0.17	0.50	8.95	17.55
10	0.2	100	Random	3.57	303.99	0.54	0.16	0.50	13.75	27.10
0	0.5	100	Random	20.02	346.31	0.54	0.17	0.52	1.45	2.95
1	0.5	100	Random	20.88	330.51	0.54	0.17	0.52	3.70	7.15
5	0.5	100	Random	20.93	328.16	0.54	0.17	0.52	8.20	17.30
10	0.5	100	Random	18.94	318.28	0.54	0.17	0.52	12.55	26.35
0	0.8	100	Random	30.60	351.63	0.54	0.18	0.52	1.65	3.05
1	0.8	100	Random	30.47	344.59	0.54	0.18	0.52	3.00	6.00
5	0.8	100	Random	28.81	336.17	0.54	0.17	0.52	8.10	15.90
10	0.8	100	Random	28.62	317.33	0.54	0.17	0.52	11.70	24.55

REFERENCES

Abreu, G. B., D. F. Ferreira, M. A. P. Ramalho, F. H. R. B. Toledo, and J. S. d. S. Bueno Filho, 2013 Computer simulation for the evaluation of recombination strategies in intrapopulation recurrent selection in Eucalyptus. Silvae Genetica **62**: 68–79.

Abreu, G. B., M. A. P. Ramalho, F. H. R. B. Toledo, and J. C. de Souza, 2010 Strategies to improve mass selection in maize. Maydica **55**: 219–225.

Baenziger, P. S. and C. Peterson, 1991 Genetic variation: Its origin and use for breeding self-pollinated species. In *Plant Breeding in the 1990's*, edited by H. Stalker and J. Murphy, pp. 69–100, North Carolina State University, North Carolina.

Beckett, T. J., T. R. Rocheford, and M. Mohammadi, 2019 Re-imagining maize inbred potential: identifying breeding crosses using genetic variance of simulated progeny.

Bernardo, R., 2001 What if we knew all the genes for a quantitative trait in hybrid crops? Crop Science **41**: 1–4.

Bernardo, R., 2003 Parental selection, number of breeding populations, and size of each population in inbred development. Theoretical and Applied Genetics **107**: 1252–1256.

Bernardo, R., 2009 Should maize doubled haploids be induced among F1 or F2 plants? Theoretical and Applied Genetics **119**: 255–262.

Bernardo, R., 2010 *Breeding for quantitative traits in plants*. Stemma Press, Woodbury, second edition.

Bernardo, R., 2014 Genomewide selection of parental inbreds: Classes of loci and virtual biparental populations. Crop Science **54**: 2586–2595.

Bernardo, R., 2016 Bandwagons I, too, have known. Theoretical and Applied Genetics **129**: 2323–2332.

Bernardo, R., 2017 Prospective targeted recombination and genetic gains for quantitative traits in maize. The Plant Genome **10**: 1–9.

Bernardo, R. and J. Yu, 2007 Prospects for genomewide selection for quantitative traits in maize. Crop Science **47**: 1082–1090.

Bloom, J. S., I. Kotenko, M. J. Sadhu, S. Treusch, F. W. Albert, and L. Kruglyak, 2015 Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. Nature Communications **6**: 1–6.

Bohn, M., H. F. Utz, and A. E. Melchinger, 1999 Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. Crop Science **39**: 228–237.

Borges, A., A. González-Reymundez, O. Ernst, M. Cadenazzi, J. Terra, and L. Gutiérrez, 2019 Can spatial modeling substitute for experimental design in agricultural experiments? Crop Science **59**: 44–53.

Breseghello, F. and M. E. Sorrells, 2006 Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics **172**: 1165–1177.

Broman, K. W., 2019 simcross: Simulate Experimental Crosses.

Browning, S., 2000 The relationship between count-location and stationary renewal models for the chiasma process. Genetics **155**: 1955–1960.

Carpentieri-Pípolo, V., A. E. Pípolo, F. A. M. da Silva, and M. R. Petek, 2000 Soybean parent selection based on genetic diversity. Brazilian Archives of Biology and Technology **43**: 295–300.

Casali, V. and E. Tigchelaar, 1975 Computer simulation studies comparing pedigree, bulk and single seed descent in self pollinated populations. Journal of the American Society for Horticultural Science **100**: 364–367.

Chen, G. K., P. Marjoram, and J. D. Wall, 2009 Fast and flexible simulation of DNA sequence data. Genome Research **19**: 136–142.

Cobb, J. N., R. U. Juma, P. S. Biswas, J. D. Arbelaez, J. Rutkoski, G. Atlin, T. Hagen, M. Quinn, and E. H. Ng, 2019 Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. Theoretical and Applied Genetics **132**: 627–645.

Copenhaver, G. P., E. A. Housworth, and F. W. Stahl, 2002 Crossover interference in arabidopsis. Genetics **160**: 1631–1639.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, G. de los Campos, J. Burgueño, J. M. González-Camacho, S. Pérez-Elizalde, Y. Beyene, S. Dreisigacker, R. Singh, X. Zhang, M. Gowda, M. Roorkiwal, J. Rutkoski, and R. K. Varshney, 2017 Genomic selection in plant breeding: Methods, models, and perspectives. Trends in Plant Science **22**: 961–975.

Cruz, C. D., 2001 A informática no melhoramento genético. In *Recursos Genéticos e Melhoramento de Plantas*, edited by L. L. Nass, A. C. C. Valois, I. S. de Melo, and M. C. Valadares-Inglis, chapter 34, pp. 1085–1118, Fundação MT, Rondonópolis.

Cruz, C. D., C. C. Salgado, L. L. Bhering, and R. S. Tomaz, 2013 Uso da simulação em análises genômicas. In *Genômica aplicada*, edited by C. D. Cruz, C. C. Salgado, and L. L. Bhering, p. 424, UFV, Visconde do Rio Branco, MG, first edition.

Dachs, N., 1988 *Estatística computacional: Uma Introdução ao Turbo Pascal*. Livros Técnicos e Científicos, Rio de Janeiro.

Dudley, J., 2007 From means to QTL : The Illinois long-term selection experiment as a case study in quantitative genetics. Crop Science **47**: S20–S31.

Eddelbuettel, D., 2013 *Seamless R and C++ Integration with Rcpp*. Springer, New York.

Edwards, M. D. and J. Page, 1994 Evaluation of marker-assisted selection through computer simulation. Theoretical and Applied Genetics **88**: 376–382.

Endelman, J. B., 2011 Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. The Plant Genome **4**: 250.

Endelman, J. B., G. N. Atlin, Y. Beyene, K. Semagn, X. Zhang, M. E. Sorrells, and J. L. Jannink, 2014 Optimal design of preliminary yield trials with genome-wide markers. Crop Science **54**: 48–59.

Falconer, D. S. and T. F. Mackay, 1996 *Introduction to quantitative genetics*. Longmans Green, Harlow, Essex, UK., fourth edition.

Falque, M., R. Mercier, C. Mézard, D. De Vienne, and O. C. Martin, 2007 Patterns of recombination and MLH1 foci density along mouse chromosomes: Modeling effects of interference and obligate chiasma. Genetics **176**: 1453–1467.

Fehr, W., 1987 *Principles of cultivar development. Theory and technique*. Macmillan, New York.

Fernandes, S. B., 2016 *Estratégias para melhorar a eficiência da seleção em progênies de meios irmãos por meio de simulação*. Ph.D. thesis, Universidade Federal de Lavras.

Ferreira, D. F., 2001 Uso de simulação no melhoramento. In *Recursos Genéticos e Melhoramento de Plantas*, edited by L. L. Nass, A. C. C. Valois, I. S. de Melo, and M. C. Valadares-Inglis, chapter 35, pp. 1119–1141, Fundação MT, Rondonópolis.

Fouilloux, G. and H. Bannerot, 1988 Selection methods in common bean (Phaseolus vulgaris L.). In *Genetic Resources of Phaseolus Beans*, edited by P. Gepts, pp. 503–541, Klüwer Academic Publishers, Dordrecht.

Fraser, A. S., 1957a Simulation of genetic systems by automatic digital computers: I. Introduction. Australian Journal of Biological Sciences **10**: 484–491.

Fraser, A. S., 1957b Simulation of genetic systems by automatic digital computers: II. Effects of linkage on rates of advance under selection. Australian Journal of Biological Sciences **10**: 492–500.

Gaynor, R. C., 2018 AlphaSimR: Breeding Program Simulations.

Ghaderi, A., M. Adams, and A. Nassib, 1984 Relationship between genetic distance and heterosis for yield and morphological traits in dry edible bean and faba bean. Crop Science **24**: 37–42.

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

Goldman, I., 2004 The intellectual legacy of the Illinois long-term selection experiment. In *Plant Breeding Reviews*, edited by J. Janick, volume 24, pp. 61–78, John Wiley & Sons, Inc., New Jersey.

Gorjanc, G., R. C. Gaynor, and J. M. Hickey, 2018 Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. Theoretical and Applied Genetics pp. 1–14.

Haldane, J., 1919 The combination of linkage values, and the calculation of distances between the loci of linked factors. Journal of Genetics **8**: 299–309.

Hallauer, A. R., M. J. Carena, and J. M. Filho, 2010 *Quantitative genetics in maize breeding*. Springer-Verlag New York, New York, third edition.

Heffner, E. L., A. J. Lorenz, J. L. Jannink, and M. E. Sorrells, 2010 Plant breeding with Genomic selection: Gain per unit time and cost. Crop Science **50**: 1681–1690.

Jacobson, A., L. Lian, S. Zhong, and R. Bernardo, 2015 Minimal loss of genetic diversity after genomewide selection within biparental maize populations. Crop Science **55**: 783–789.

Jannink, J. L., 2010 Dynamics of long-term genomic selection. Genetics Selection Evolution **42**: 35.

Kearsey, M. J. and A. G. L. Farquhar, 1998 QTL analysis in plants: where are we now? Heredity **80**: 137–142.

Kemphthorne, O., 1988 An overview of the field of quantitative genetics. In *Proceedings of the 2nd International Conference on Quantitative Genetics*, edited by B. S. Weir, E. Eisen, M. Goodman, and G. Namkoong, pp. 47–56, Sunderland, Massachusetts, Sinauer associates.

Lado, B., P. G. Barrios, M. Quincke, P. Silva, and L. Gutiérrez, 2016 Modeling genotype × Environment interaction for genomic selection with unbalanced data from a wheat breeding program. Crop Science **56**: 2165–2179.

Lado, B., S. Battenfield, C. Guzmán, M. Quincke, R. P. Singh, S. Dreisigacker, R. J. Peña, A. Fritz, P. Silva, J. Poland, and L. Gutierrez, 2017 Strategies for selecting crosses using genomic prediction in two wheat breeding programs. The Plant Genome **10**: 1–12.

Lado, B., I. Matus, A. Rodríguez, L. Inostroza, J. Poland, F. Belzile, A. del Pozo, M. Quincke, M. Castro, and J. von Zitzewitz, 2013 Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. G3: Genes, Genomes, Genetics **3**: 2105–2114.

Laidò, G., D. Marone, M. A. Russo, S. A. Colecchia, A. M. Mastrangelo, P. De Vita, and R. Papa, 2014 Linkage disequilibrium and genome-wide association mapping in tetraploid wheat (Triticum turgidum L.). PLoS ONE **9**: 1–18.

Lande, R. and F. W. Stahl, 1993 Chiasma interference and the distribution of exchanges in Drosophila melanogaster. In *Cold Spring Harbor Symposium of Quantitative Biology*, pp. 543–552.

Lande, R. and R. Thompson, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics **124**: 743–756.

Lehermeier, C., S. Teyssèdre, and C.-C. Schön, 2017 Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. Genetics **Early Onli**: 1–10.

Li, X., C. Zhu, J. Wang, and J. Yu, 2012 Computer simulation in plant breeding. Advances in Agronomy **116**: 219–264.

Lorenz, A. J., 2013 Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. G3: Genes, Genomes, Genetics **3**: 481–491.

Lynch, M. and B. Walsh, 1998 *Genetics and analysis of quantitative traits*. Sinauer associates, Sunderland, Massachusetts, first edition.

Malosetti, M., D. Bustos-Korts, M. P. Boer, and F. A. Van Eeuwijk, 2016 Predicting responses in multiple environments: Issues in relation to genotype × Environment interactions. Crop Science **56**: 2210–2222.

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

Moehring, J., E. R. Williams, and H. P. Piepho, 2014 Efficiency of augmented p-rep designs in multi-environmental trials. Theoretical and Applied Genetics **127**: 1049–1060.

Mohammadi, M., T. Tiede, and K. P. Smith, 2015 Popvar: A genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. Crop Science **55**: 2068–2077.

Monteverde, E., J. E. Rosas, P. Blanco, F. Pérez de Vida, V. Bonnecarrère, G. Quero, L. Gutierrez, and S. McCouch, 2018 Multienvironment models increase prediction accuracy of complex traits in advanced breeding lines of rice. Crop Science **58**: 1519–1530.

Mramba, L., G. Peter, V. Whitaker, and S. Gezan, 2018 Generating improved experimental designs with spatially and genetically correlated observations using mixed models. Agronomy **8**: 1–20.

Mueller, D. and K. W. Broman, 2017 Meiosis: Simulation of meiosis in plant breeding research.

Nei, M., 1987 *Molecular evolutionary genetics*. Columbia University Press.

Neyhart, J. L. and K. P. Smith, 2019 Validating genomewide predictions of genetic variance in a contemporary breeding program.

Osthushenrich, T., M. Frisch, and E. Herzog, 2017 Genomic selection of crossing partners on basis of the expected mean and variance of their derived lines. PLoS ONE **12**: 1–12.

Pinheiro, J., D. Bates, S. DebRoy, and D. Sarkar, 2018 nlme: Linear and Nonlinear Mixed Effect Models.

Poland, J. A. and T. W. Rife, 2012 Genotyping-by-Sequencing for plant breeding and genetics. The Plant Genome **5**: 92–102.

R CORE TEAM, 2015 R: A language and environment for statistical computing.

Ramalho, M. A. P., A. d. F. B. Abreu, J. B. dos Santos, and J. A. R. Nunes, 2012a *Aplicações da genética quantitativa no melhoramento de plantas autógamas*. Editora UFLA, Lavras.

Ramalho, M. A. P., L. A. d. S. Dias, and B. L. Carvalho, 2012b Contributions of plant breeding in Brazil – progress and perspectives. Crop Breeding and Applied Biotechnology **S2**: 111–120.

Robertsen, C. D., R. L. Hjortshøj, and L. L. Janss, 2019 Genomic selection in cereal breeding. Agronomy **9**: 1–16.

Santos, A. H., E. Bearzoti, D. F. Ferreira, and J. L. D. Silva Filho, 2002 Simulation of mixed models in augmented block design. Scientia Agricola **59**: 483–489.

Schwantes, I. A., A. T. Amaral Júnior, J. E. Almeida Filho, M. Vivas, P. D. S. Cabral, A. G. Guimarães, F. H. d. L. e. Silva, P. H. A. D. Santos, M. G. Pereira, A. P. Viana, G. F. Pena, and F. R. A. Ferreira, 2018 First report of recurrent genomic selection with real Data in popcorn and genetic gain increases. bioRxiv **pre-print**: 1–18.

Souza, E. and M. Sorrells, 1991 Prediction of progeny variation in oat from parental genetic relationships. Theoretical and Applied Genetics **82**: 233–241.

Speed, T. and M. S. Waterman, 1996 *Genetic mapping and DNA sequencing*, volume 81. Springer, New York.

Suess, E. A. and B. E. Trumbo, 2008 *Introduction to probability simulation and gibbs sampling with R*. Springer, Hayward.

Sun, X., T. Peng, and R. H. Mumm, 2011 The role and basics of computer simulation in support of critical decisions in plant breeding. Molecular Breeding **28**: 421–436.

Tambasco-Talhari, D., M. M. de Alencar, C. C. P. de Paz, G. M. da Cruz, A. de Andrade Rodrigues, I. U. Packer, L. L. Coutinho, and L. C. de Almeida Regitano, 2005 Molecular marker heterozygosities and genetic distances as correlates of production traits in F1 bovine crosses. Genetics and Molecular Biology **28**: 218–224.

Thavamanikumar, S., R. Dolferus, and B. R. Thumma, 2015 Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. G3: Genes, Genomes, Genetics **5**: 1991–1998.

Tiede, T., L. Kumar, M. Mohammadi, and K. P. Smith, 2015 Predicting genetic variance in bi-parental breeding populations is more accurate when explicitly modeling the segregation of informative genomewide markers. Molecular Breeding **35**: 1–13.

Toledo, F. H. R. B. and CIMMYT, 2018 isqg: In Silico Quantitative Genetics.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. Journal of Dairy Science **91**: 4414–4423.

Veiga, R. D., D. F. Ferreira, and M. A. P. Ramalho, 2000 Eficiência dos dialelos circulantes na escolha de genitores. Pesquisa Agropecuaria Brasileira **35**: 1395–1406.

Vencovsky, R., M. A. P. Ramalho, and F. H. R. B. Toledo, 2012 Contribution and perspectives of quantitative genetics to plant breeding in Brazil. Crop Breeding and Applied Biotechnology **S2**: 7–14.

Visscher, P. M. and M. E. Goddard, 2019 From R.A. Fisher's 1918 Paper to GWAS a century later. Genetics **211**: 1125–1130.

Walsh, B. and M. Lynch, 2018 *Evolution and selection of quantitative traits*. OUP Oxford, Oxford, UK, first edition.

Wang, J., S. C. Chapman, D. G. Bonnett, G. J. Rebetzke, and J. Crouch, 2007 Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. Crop Science **47**: 582–590.

Wang, J., M. van Ginkel, D. Podlich, G. Ye, R. Trethowan, W. Pfeiffer, I. H. DeLacy, M. Cooper, and S. Rajaram, 2003 Comparison of two breeding strategies by computer simulation. Crop Science **43**: 1764–1773.

Wang, J., M. Van Ginkel, R. Trethowan, G. Ye, I. DeLacy, D. Podlich, and M. Cooper, 2004 Simulating the effects of dominance and epistasis on selection response in the CIMMYT wheat breeding program using QuCim. Crop Science **44**: 2006–2018.

Watson, A., S. Ghosh, M. J. Williams, W. S. Cuddy, J. Simmonds, M.-D. Rey, M. Asyraf Md Hatta, A. Hinchliffe, A. Steed, D. Reynolds, N. M. Adamski, A. Breakspear, A. Korolev, T. Rayner, L. E. Dixon, A. Riaz, W. Martin, M. Ryan, D. Edwards, J. Batley, H. Raman, J. Carter, C. Rogers, C. Domoney, G. Moore, W. Harwood, P. Nicholson, M. J. Dieters, I. H. Delacy, J. Zhou, C. Uauy, S. A. Boden, R. F. Park, B. B. Wulff, and L. T. Hickey, 2018 Speed breeding is a powerful tool to accelerate crop research and breeding. Nature Plants **4**: 23–29.

Weir, B. S. and C. C. Cockerham, 1984 Estimating F-Statistics for the analysis of population structure. Evolution **38**: 1358–1370.

Witcombe, J. R., S. Gyawali, M. Subedi, D. S. Virk, and K. D. Joshi, 2013 Plant breeding can be made more efficient by having fewer, better crosses. BMC plant biology **13**: 22.

Wricke, G. and W. Weber, 1986 *Quantitative genetics and selection in plant breeding.* de Gruyter, Berlin.

Wright, S., 1949 The genetical structure of populations. Annals of Eugenics **15**: 323–354.

Xavier, A., W. M. Muir, and K. M. Rainey, 2016 Assessing predictive properties of genome-wide selection in soybeans. G3: Genes, Genomes, Genetics **6**: 2611–2616.

Yabe, S., H. Iwata, and J.-L. Jannink, 2017 A simple package to script and simulate breeding schemes: The Breeding Scheme Language. Crop Science **57**: 1–8.

Zeng, Z.-B. and C. C. Cockerham, 1991 Variance of neutral genetic variances within and between populations for a quantitative character. Genetics **129**: 535–553.

Zhang, W. and C. Smith, 1992 Computer simulation of marker-assisted selection utilizing linkage disequilibrium. Theoretical and Applied Genetics **83**: 813–820.

Zhang, W. and C. Smith, 1993 Simulation of marker-assisted selection utilizing linkage disequilibrium: the effects of several additional factors. Theoretical and Applied Genetics **86**: 492–496.