HELCIO DUARTE PEREIRA

HAPLOTYPE-BASED GENOMIC SELECTION EFFICIENCY OVER GENERATIONS

Thesis submitted to the Breeding and Genetics Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements to obtain the degree on *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRAZIL
2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
P436h
2019

Pereira, Helcio Duarte, 1987-
Haplotype-based genomic selection efficiency over generations / Helcio Duarte Pereira. – Viçosa, MG, 2019.
iv, 22 f. : il. ; 29 cm.

Texto em inglês.
Orientador: José Marcelo Soriano Viana.
Tese (doutorado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 15-17.

1. Plantas - Melhoramento genético. 2. Genômica. 3. Modelos matemáticos. 4. Genética molecular. 5. Marcadores genéticos. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 631.52

HELCIO DUARTE PEREIRA

HAPLOTYPE-BASED GENOMIC SELECTION EFFICIENCY OVER GENERATIONS

Thesis submitted to the Breeding and Genetics Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements to obtain the degree on *Doctor Scientiae*.

APPROVED: March 11, 2019.

Fabyano Fonseca e Silva

Plínio César Soares

Renata Veroneze

Daniele Botelho Diniz Marques

José Marcelo Soriano Viana
(Advisor)

SUMMARY

ABSTRACT

PEREIRA, Helcio Duarte, D.Sc., Universidade Federal de Viçosa, March, 2019. **Haplotype-based genomic selection efficiency over generations.** Advisor: José Marcelo Soriano Viana.

It is expected that genomic selection based on haplotypes makes better use of linkage disequilibrium among markers and QTLs than models based on single markers, and hence, happened a lower decrease in prediction accuracies over time. The objective of this work is to evaluate the efficiency of genomic selection based on haplotypes in a context of advanced generations of prediction and high marker density. We simulated phenotypes and genotypes of 5,000 unrelated individuals (founders) that gave rise to another 5,000 individuals in the next five generations, without selection. A total number of 60,000 SNPs with density of 0.01 cM was adopted in a genome of 10 chromosomes. An additional scenario of lower marker density (6,000 SNPs) was also evaluated from a sample of that whole set. The haplotype length of 10 SNPs was adopted and the haplotypes calling was made within each chromosome after SNP phasing. The trait investigated in predictions, under a GBLUP model, had broad sense heritability of 0.30. The use of lower marker density harmed the prediction accuracies and coincidence of the best individuals predicted. On average the accuracies decreased in 24% for the model based on SNPs and in 65% for the model based on haplotypes. For coincidence those reductions were 18 and 35% for SNPs and haplotypes, respectively. Considering the best scenario, namely, high marker density, the predictions based on single SNP ranged from 0.3891 to 0.4864 while the predictions based on haplotypes ranged from 0.2515 to 0.3886, for the first and the last prediction generations, respectively. For coincidence of the best predicted individuals those values ranged from 0.2466 to 0.3859 for SNP model and from 0.1959 to 0.3349 for haplotype model. In general the accuracies based on haplotypes were 30% lower than based on SNPs. For coincidence this difference was 21% on average. There is reduction, both in prediction accuracy and coincidence, over the generations with the use of SNPs or haplotypes. Our investigation does not support the use of haplotypes for prediction over generations.

RESUMO

PEREIRA, Helcio Duarte, D.Sc., Universidade Federal de Viçosa, março de 2019. **Eficiência da seleção genômica baseada em haplótipos ao longo de gerações.** Orientador: José Marcelo Soriano Viana.

Espera-se que a seleção genômica baseada em haplótipos faça melhor uso do desequilíbrio de ligação entre marcadores e QTLs do que os modelos baseados em marcas únicas, e consequentemente, ocorra um menor decréscimo nas acurácias de predição ao longo do tempo. O objetivo desse trabalho é avaliar a eficiência da seleção genômica baseada em haplótipos em um contexto de gerações avançadas de predição e alta densidade de marcas. Foram simulados fenótipos e genótipos de 5000 indivíduos não relacionados (fundadores) que originaram outros 5000 indivíduos nas próximas cinco gerações, sem seleção. Um total de 60000 SNPs com densidade de 0,01 cM foi adotado tendo o genoma 10 cromossomos. Também foi avaliado um cenário adicional de menor densidade de marcas (6,000 SNPs) a partir de uma amostragem desse conjunto total. O comprimento de haplótipo de 10 SNPs foi adotado e a chamada de haplótipos foi feita para cada cromossomo após a obtenção da fase dos SNPs. O caráter investigado nas predições, sob um modelo GBLUP, possuía herdabilidade em sentido amplo de 0,30. O uso de baixa densidade de marcadores comprometeu as acurácias de predição e coincidência dos melhores indivíduos preditos. Na média as acurácias se reduziram em 24% para o modelo baseado em SNPs e em 65% para o modelo baseado em haplótipos. Para a coincidência essas reduções foram de 18 e 35% para SNPs e haplótipos, respectivamente. Considerando o melhor cenário, ou seja, alta densidade de marcas, as predições baseadas em SNP simples variaram de 0,3891 a 0,4864 enquanto as predições baseadas em haplótipos variaram de 0,2515 a 0,3886, para a primeira e a última geração de predição, respectivamente. Para a coincidência dos melhores indivíduos preditos esses valores variaram de 0,2466 a 0,3859 para o modelo de SNP e de 0,1959 a 0,3349 para o modelo de haplótipos. No geral as acurácias baseadas em haplótipos foram 30% menores que as baseadas em SNPs. Para coincidência esta diferença foi de 21% na média. Há uma redução, tanto na acurácia de predição quanto na coincidência, ao longo das gerações com o uso de SNPs ou haplótipos. Nossa investigação não dá suporte ao uso de haplótipos para a predição ao longo de gerações

1. INTRODUCTION

Genomic selection relies on linkage disequilibrium (LD) between markers and QTLs, but it cannot persist due to recombination over time and this is a concern about the efficiency of genomic selection along generations. The genomic selection was initially proposed with high marker density in order to ensure that each QTL controlling the trait will be in strong LD with at least one marker (Habier et al., 2007). So, it was expected that with the advancements in the platforms of genotyping, and hence the availability of more covered genomes, the genomic selection would provide higher prediction accuracies (Meuwissen & Goddard, 2010). However, this expectation was not confirmed in practice (VanRaden et al., 2013) and a better use of a denser genotyping is still required.

In order to better explore the LD among markers, the use of haplotypes is a recent approach proposed in genomic selection. A haplotype defines a region of the genome that comprises a set of neighboring genetic markers (i.e. SNPs), whereby their phased alleles are likely inherited together (Hess et al., 2017). It is expected stronger and persistent LD of a QTL with a haploblock than with a single SNP, because a QTL in weak LD with any individual SNP may be in strong LD with a multi-marker haplotype (Cuyabano et al., 2015), what can boost the predictions over time. Studies were already conducted investigating the best haplotype length (Villumsen et al., 2009; Ferdosi et al., 2016). Villumsen et al. (2009) found highest reliabilities for 10 SNP haplotypes for traits with heritability of 0.30 and 0.02, whereas Ferdosi et al. (2016) found that the best haplotype length is trait dependent, but 7 or 8 SNP haplotypes was the best for several scenarios investigated by them. The length of haplotypes influenced directly the number of possible alleles and, hence, the computational burden of a haplotype prediction analysis, as well as the results in prediction accuracy (Hess et al., 2017). When haplotypes are based on pairs of SNPs, four alleles are possible and as the number of SNPs in each segment increases, so does the number of possible alleles. Varying the length of haplotypes can assist in the modeling of the LD between SNPs and QTL (Ferdosi et al., 2016).

In a breeding population, there is correlation between the loci due to linkage and various other factors. It is thus important to consider this LD covariance structure in the computation of the genomic relationship matrix (Mathew et al., 2018). The potential of haplotypes to be used for prediction purposes had been investigated in plant populations (Bekele et al., 2018; Ma et al., 2016), animal populations (Cuyabano et al., 2014; Hess

et al., 2017) and simulation studies (Ferdosi et al., 2016; Jiang et al., 2018) with varying results. Bekele et al. (2018), for example, did not find any difference between predictions based on haplotypes or single SNPs. On the other hand, Cuyabano et al. (2014) found accuracies for haplotypes up to 3.1% larger than using single SNPs for milk protein trait in dairy cattle. However, those authors did not find any difference for other two traits investigated by them with low heritability. Hess et al. (2017) found an advantage up to 5.5% using haplotypes, depending of the haplotype length chosen, in genomic prediction compared to single SNPs. A few studies pointed out the properties of the genomic relationship matrix derived from haplotypes instead of the traditional SNP derived matrix (Edwards, 2015; Lawson et al., 2012; Mathew et al., 2008). Lawson et al. (2012) verified by a simulation analysis that when the coancestry matrix was construct by a haplotype based algorithm the individuals were better allocated to groups than the traditional Principal Component Analysis or STRUCTURE bayesian model. Using a real data set of humans the authors identified more coherent cluster results, by previous information from geographic and family data, employing the haplotype based approach to obtain the covariance matrix. Ogawa et al. (2018) performed a GWAS study exploring haplotype alleles, with both, qualitative and quantitative traits in rice, in multi-parental population. The authors reported larger power, for both traits considered, in gene identification with the use of haplotypes. Different genomic models were already investigated with the use of haplotypes or SNP with distinct behavior (Cuyabano et al., 2014 ; Cuyabano et al., 2015). Cuyabano et al. (2014) found superior prediction accuracies for a Bayesian model compared to mixed model, for milk protein and fertility traits studied by them, in dairy cattle data set. Cuyabano et al. (2015) also found superior accuracies, when investigated five traits in Nordic Red cattle, for haplotype model instead of single SNP, mainly when the Bayesian mixture approach was employed compared to Bayesian BLUP.

One important concern about the use of genomic selection is its applicability over generations to save money in phenotyping trials. The traditional models, employing single SNPs, are not successful to predict distant or unrelated individuals from the ones assessed for the traits of interest. The predictive ability is expected to reduce owing to genetic phenomena such as recombination (loss of LD associations), mutation and change of the additive effects of loci due to allele frequency changes at loci with dominance and / or epistasis (Villumsen et al., 2009). However, efforts have been done to estimate breeding value of individuals along several generations of

recurrent genomic selection in order maximize long term genetic gain (Muller et al., 2018).

There is a lack of studies investigating the efficiency of haplotype based prediction along generations. Considering that the genetic relationships are increasing over the time within each generation, but reducing between them with the advancement of the cycles, the traditional SNP based model is not suited to make predictions and a comparison with the haplotype based prediction is still missing. The better use of LD, stated in the literature, for the haplotypes should be compared with high density SNP genotyping over the time to assess the suitability of haplotype prediction for some generations. The objective of this work is to evaluate the efficiency of genomic selection based on haplotypes in a context of advanced generations of prediction and high marker density.

2. MATERIAL AND METHODS

2.1. DATA SETS

The software *REALbreeding* (available by request) was employed to simulate SNP genotypes and phenotypes of 5,000 founder parents (generation 1) and the next five generations, each one with 5,000 individuals too, totaling 30,000 individuals. The parents of every generation came from the previous generation, without selection, originating full sib progenies of different size, but keeping the size of the next generation constant at 5,000 individuals. The software does not assume a distribution for the LD values (nor for gene effects), but computes the true LD values for QTLs based on quantitative genetics theory (Viana, 2004) to determine the parametric additive, dominance, and genotypic values. SNP and QTL frequencies follow a beta distribution.

Based on our input, *REALbreeding* randomly distributed 60,000 SNPs, 20 QTLs, and 380 minor genes (QTLs of lower effect) in 10 chromosomes (6,000 SNPs and 40 genes by chromosome). The average SNP density was 0.01 cM and each QTL had a SNP within it (same frequency). The trait simulated had minimum and maximum genotypic values for homozygotes of 0 and 2, respectively. These values are used to compute the deviations a (the difference between the genotypic value of the homozygote of greater expression and the mean of the homozygotes) and d (dominance deviation). The dominance deviation is computed from the degree of dominance (d/a).

We defined positive dominance ($0 < d/a \leq 1.2$). The true additive and dominance genetic values were computed from the population gene frequencies, LD values, average effects of gene substitution, and dominance deviations. The phenotypic values were computed from the true population mean, additive and dominance values, and from error effects sampled from a Normal distribution. The error variance was computed from the broad sense heritability, assumed as 30%. The narrow sense heritability was 25%. This implies in a maximum theoretical accuracy of phenotypic selection of 0.50. The QTL heritabilities ranged from 0.1 to 3.2%.

Derived from that data set just described, a scenario with lower SNP density was investigated. For do this each chromosome had 600 SNPs selected at random, keeping the individuals, phenotypic values and genome length as before. So, in this scenario the SNP density was reduced ten times to study its effect on the predictions make with haplotypes or single SNP.

2.2. HAPLOTYPES

The haplotypes calling was made by each chromosome grouping 10 SNPs at a time to form the block of haplotypes. So, the different haplotype alleles within each block were scored to form the sample haplotype allele genotypes based on copy number (i.e., 0, 1 or 2 copies) for each diploid individual (only two alleles for each one). Then, the haplotype allele genotypes of each chromosome were gathered together after applying the criteria of MAF greater than 0.01 to perform the analysis. The GHap R package (Utsunomiya and Milanesi, 2017) was used to get the haplotype alleles from phased SNP data. To obtain the phase of SNP data the software AlphaImpute version 1.9.8 (Hickey et al., 2011) was employed providing the genealogical information of all generations to the software.

From the matrix of haplotype allele genotypes for all individuals the genomic relationship matrix was constructed by the expression:

$$\boldsymbol{G = qMDM'}$$

where $\boldsymbol{M}$ is the centered N x H matrix of haplotype alleles genotype, where N is the number of individuals and H is the number of haplotype alleles, $\boldsymbol{D}$ = diag(di), di is the weight of haplotype allele i (default di =1), and $\boldsymbol{q} = tr(\text{MDM'})^{-1}$ N. The genomic relationship matrix derived from single SNP was the matrix proposed by Endelman and Jannink (2012).

A LD analysis was performed in both scenarios, namely, higher and lower SNP density, to study the values of LD within haplotypes in order to better understand the results and support the comparisons between SNP or haplotype based model. From the information of the simulated genome, and the markers selected to compose the lower SNP density scenario, we can know the distances among markers and average length of haplotypes. As all the chromosomes are equivalent only one of them is enough to provide the necessary information. So, the chromosome 1 of generation 2, the first generation used to make predictions, was chosen to perform the LD analysis employing the Haploview software (Barrett et al., 2005).

2.3. GENOMIC PREDICTION

We fitted a genomic BLUP (GBLUP) model, for prediction based on SNP or haplotypes, given by:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z\alpha} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ is the vector of phenotypic values, μ is the population mean, $\boldsymbol{\alpha}$ is the additive genetic value of the individual, $\boldsymbol{Z}$ is the incidence matrix, and $\boldsymbol{\varepsilon}$ is the error vector. The variance of $\boldsymbol{\alpha}$ is $\boldsymbol{G\sigma_A^2}$, where $\boldsymbol{G}$ is the additive genomic relationship matrix and $\boldsymbol{\sigma_A^2}$ is the additive variance.

The analyses were done with the R package *sommer* (Covarrubias-Pazaran 2016). Only SNPs with MAF (minor allele frequency) greater than 0.05 were used in the analysis. We computed additive value prediction accuracies for all the individuals to be predicted and coincidence of the best 500 individuals (10% of the population size). The coincidence was computed after ranking the predicted individuals and their corresponding additive genetic value from the simulated data. So, the first 500 individuals in each of these ranked data were compared. The parametric coincidence, used to make comparisons, was computed from the ranked phenotypic and additive genetic values from the simulated data. The training set size was 2,000 individuals randomly sampled from the first generation after the founder parents' generation (generation 2). The founder parents did not have any genetic relationship among them. This procedure was repeated 50 times for each generation to be predicted. The additive value prediction accuracy was computed as the average correlation between the true additive values computed by *REALbreeding* and the values predicted by GBLUP on the independent validation set over the 50 resampling. As a threshold for the maximum

accuracy achievable we performed the genomic prediction within each generation employing the whole 5,000 individuals as training population.

2.4 COMPARING SNP AND HAPLOTYPE ACCURACIES FOR PREDICTING THE ADDITIVE VALUE OF A QTL

The theoretical investigation on the additive value genomic prediction using haplotypes, in a similar approach adopted for SNPs by Viana et al. (2016) or Grattapaglia and Resende (2010), is a much more complex work. Even defining the number of haplotypes and the minimum number of SNPs per haplotypes (two), there would be four haplotypes alleles per haplotype locus, with different frequencies, and six LD values. Thus, we decided to investigate the SNP and haplotype accuracies for predicting the additive value of a single QTL. The following derivations will show how complex this apparently simple task is.

Initially assume as linked, two SNPs and a QTL in linkage disequilibrium in a population (generation 0) in Hardy-Weinberg equilibrium, in the order SNP1 (alleles A and a) - SNP2 (alleles B and b) - QTL (alleles C and c). Define p_a, p_b, and p_c as the frequencies for alleles A, B, and C, respectively, $f_{222}^{(0)}$, $f_{221}^{(0)}$, $f_{220}^{(0)}$, ..., and $f_{000}^{(0)}$ (36 genotypes, respecting the SNP and QTL alleles in the homologues chromosomes) as the genotype probabilities (for the genotypes ABC/ABC, ABC/ABc, ABc/ABc, ..., abc/abc, respectively), and θ_{ab}, θ_{bc} , and $\theta_{ac} = \theta_{ab} + \theta_{bc}$ (complete interference) as the recombination frequencies for SNP1 and SNP2, SNP2 and QTL, and SNP1 and QTL, respectively. The assumption of Hardy-Weinberg equilibrium implies that, for example, $f_{2..}^{(0)} = p_a^2$, $f_{1..}^{(0)} = 2p_a q_a$, and $f_{0..}^{(0)} = q_a^2$. Define additionally $P_{111}^{(0)}$, $P_{110}^{(0)}$, ..., and $P_{000}^{(0)}$ as the haplotype frequencies (for the haplotypes ABC, ABc, ..., and abc, respectively). Then, the LD values are $\Delta_{ab}^{(0)} = P_{11.}^{(0)} P_{00.}^{(0)} - P_{10.}^{(0)} P_{01.}^{(0)}$, $\Delta_{bc}^{(0)} = P_{.11}^{(0)} P_{.00}^{(0)} - P_{.10}^{(0)} P_{.01}^{(0)}$, and $\Delta_{ac}^{(0)} = P_{1.1}^{(0)} P_{0.0}^{(0)} - P_{1.0}^{(0)} P_{0.1}^{(0)}$ (Kempthorne, 1957).

Using the procedures of Kempthorne (1957), the frequencies of the haplotype ABC in generation 1 and after n generations of random cross are, respectively:

$P_{111}^{(1)} = f_{222}^{(0)} + \frac{f_{221}^{(0)}}{2} + \cdots = \left[P_{111}^{(0)}\right]^2 + P_{111}^{(0)} P_{110}^{(0)} + \cdots = P_{111}^{(0)} - \theta_{ab}\Delta_{ab}^{(0)} - \theta_{bc}\Delta_{bc}^{(0)} - \theta_{ac}\Delta_{ac}^{(0)} - \theta_{ab} f_1^{(0)} - \theta_{bc} f_2^{(0)}$ and

$$P_{111}^{(n)} \cong P_{111}^{(0)} - [1-(1-\theta_{ab})^n]\Delta_{ab}^{(0)} - [1-(1-\theta_{bc})^n]\Delta_{bc}^{(0)} - [1-(1-\theta_{ac})^n]\Delta_{ac}^{(0)} - [1-(1-\theta_{ab})^n]f_1^{(0)} - [1-(1-\theta_{bc})^n]f_2^{(0)} + \theta_{ab}\theta_{bc}\left\{\left[\frac{1-(1-\theta_{ab}-2\theta_{bc})^n}{(\theta_{ab}+2\theta_{bc})}\right]\left(q_a\Delta_{bc}^{(0)} + f_2^{(0)}\right) + \left[\frac{1-(1-2\theta_{ab}-\theta_{bc})^n}{(2\theta_{ab}+\theta_{bc})}\right]\left(q_c\Delta_{ab}^{(0)} + f_1^{(0)}\right)\right\}$$

where $f_1^{(0)} = q_a P_{100}^{(0)} - p_a P_{000}^{(0)}$ and $f_2^{(0)} = q_c P_{001}^{(0)} - p_c P_{000}^{(0)}$.

Because

$$\lim_{n\to\infty} P_{111}^{(n)} \cong$$

$$P_{111}^{(0)} - \Delta_{ab}^{(0)} - \Delta_{bc}^{(0)} - \Delta_{ac}^{(0)} - f_1^{(0)} - f_2^{(0)} + \theta_{ab}\theta_{bc}\left\{\left[\frac{1}{(\theta_{ab}+2\theta_{bc})}\right]\left(q_a\Delta_{bc}^{(0)} + f_2^{(0)}\right) + \left[\frac{1}{(2\theta_{ab}+\theta_{bc})}\right]\left(q_c\Delta_{ab}^{(0)} + f_1^{(0)}\right)\right\} = p_a p_b p_c$$

Then

$$P_{111}^{(0)} \cong p_a p_b p_c + \Delta_{ab}^{(0)} + \Delta_{bc}^{(0)} + \Delta_{ac}^{(0)} + f_1^{(0)} + f_2^{(0)} - \theta_{ab}\theta_{bc}\left\{\left[\frac{1}{(\theta_{ab}+2\theta_{bc})}\right]\left(q_a\Delta_{bc}^{(0)} + f_2^{(0)}\right) + \left[\frac{1}{(2\theta_{ab}+\theta_{bc})}\right]\left(q_c\Delta_{ab}^{(0)} + f_1^{(0)}\right)\right\}$$, where $q_c\Delta_{ab}^{(0)} + f_1^{(0)} = q_a\Delta_{bc}^{(0)} + f_2^{(0)}$.

Because ignoring the last term provides an additional bias very close to zero, we can define that

$$P_{111}^{(0)} \cong p_a p_b p_c + \Delta_{ab}^{(0)} + \Delta_{bc}^{(0)} + \Delta_{ac}^{(0)} + f_1^{(0)} + f_2^{(0)}$$

The approximate probabilities for the other haplotypes in generation 0 are

$$P_{110}^{(0)} \cong p_a p_b q_c - \Delta_{bc}^{(0)} - \Delta_{ac}^{(0)} - f_1^{(0)} - f_2^{(0)}$$

$$P_{101}^{(0)} \cong p_a q_b p_c - \Delta_{ab}^{(0)} - \Delta_{bc}^{(0)} - f_1^{(0)} - f_2^{(0)}$$

$$P_{100}^{(0)} \cong p_a q_b q_c + \Delta_{bc}^{(0)} + f_1^{(0)} + f_2^{(0)}$$

$$P_{011}^{(0)} \cong q_a p_b p_c - \Delta_{ab}^{(0)} - \Delta_{ac}^{(0)} - f_1^{(0)} - f_2^{(0)}$$

$$P_{010}^{(0)} \cong q_a p_b q_c + \Delta_{ac}^{(0)} + f_1^{(0)} + f_2^{(0)}$$

$$P_{001}^{(0)} \cong q_a q_b p_c + \Delta_{ab}^{(0)} + f_1^{(0)} + f_2^{(0)}$$

$$P_{000}^{(0)} \cong q_a q_b q_c - f_1^{(0)} - f_2^{(0)}$$

Note that, as expected,

$$P_{11.}^{(0)} = p_a p_b + \Delta_{ab}^{(0)}$$

$$P_{10.}^{(0)} = p_a q_b - \Delta_{ab}^{(0)}$$

$$P_{01.}^{(0)} = q_a p_b - \Delta_{ab}^{(0)}$$

$$P_{00.}^{(0)} = q_a q_b + \Delta_{ab}^{(0)}$$

And,

$$P_{1.1}^{(0)} = p_a p_c + \Delta_{ac}^{(0)}$$

$$P_{1.0}^{(0)} = p_a q_c - \Delta_{ac}^{(0)}$$

$$P_{0.1}^{(0)} = q_a p_c - \Delta_{ac}^{(0)}$$

$$P_{0.0}^{(0)} = q_a q_c + \Delta_{ac}^{(0)}$$

And,

$$P_{.11}^{(0)} = p_b p_c + \Delta_{bc}^{(0)}$$

$$P_{.10}^{(0)} = p_b q_c - \Delta_{bc}^{(0)}$$

$$P_{.01}^{(0)} = q_b p_c - \Delta_{bc}^{(0)}$$

$$P_{.00}^{(0)} = q_b q_c + \Delta_{bc}^{(0)}$$

Furthermore, $P_{...}^{(0)} = 1$. To compute the accuracies of the breeding value prediction for the QTL, from each SNP, from both SNPs, from each pair of haplotype alleles (four alleles and six accuracies), and the average haplotype prediction accuracy (weighted by the sum of the corresponding haplotype probabilities), we defined random values between 0 and 0.1 for the recombination frequencies θ_{ab} and θ_{bc}, and three scenarios for the frequencies of the SNPs A and B and the QTL C. Defining random values between 0 and 0.05 (0.028 on average) for the 36 genotype probabilities, we generated a scenario of similar frequencies for SNPs and QTL (Table 3; simulations 1 to 25). The other additional scenarios were SNPs with similar frequencies but contrasting with the QTL frequency (Table 3; simulations 26 to 50), and a SNP and the QTL with similar frequencies but contrasting with the other SNP frequency (Table 3; simulations 51 to 100). We computed the prediction accuracies based on the function presented by Viana et al. (2016).

3. RESULTS

3.1 GENOMIC PREDICTION ACCURACY AND COINCIDENCE

The prediction accuracy and coincidence of the best individuals based on single SNP markers are superior than employing haplotype alleles, both in the higher and lower marker density (Table 1). In the scenario of lower marker density the prediction accuracies based on SNP ranged from 0.2360 to 0.4135, and based on haplotypes the values ranged from 0.0737 to 0.1625. For coincidence, those values ranged from 0.1833 to 0.3412 for SNP based model, and from 0.1277 to 0.1788 for haplotype based model. It is worthy to note that this scenario of lower marker density provided accuracies of 24

and 65% lower comparing to the whole set of markers (60,000), for SNP and haplotype based model respectively. For coincidence, the lower marker density provided values 18% lower for SNP and 35% lower for haplotypes, comparing to the whole set of markers. So, thereafter we will consider only the best scenario of higher marker density

Using the higher marker density the model based on single SNP provided accuracies of unevaluated individuals ranging from 0.3891 to 0.4864 for generation 6 and 2, respectively. For the haplotype based model those accuracies ranged from 0.2515 to 0.3886 for generation 6 and 2, respectively. In general the accuracies based on haplotypes (0.3089) are 30% lower than based on SNPs (0.4404). The coincidence of the best individuals (10%) ranged from 0.2466 to 0.3859 for SNP based model and from 0.1959 to 0.3349 for haplotype based model for generation 6 and 2, respectively. For coincidence, the haplotype based model provided results 21% lower on average. Comparing the average parametric coincidence of 0.3156 the SNP based model provided 7.07% lower coincidence and the haplotype based model 26.93%. On average, the coincidences obtained from the SNP based model was 21.38% higher than the haplotype based model. There is a trend of reduction, both in prediction accuracy and coincidence, over the generations. From generation 2 to the generation 6 the decay in prediction accuracy of unevaluated individuals is 20% using single SNP markers in the model and 35% using haplotypes. For coincidence of the best individuals those reductions are 36 and 41% for SNPs and haplotypes, respectively. In any scenario, for both, prediction based on SNP or haplotypes, the maximum accuracy found for unevaluated individuals was lower than the maximum phenotypic accuracy achievable (0.50). So, in our investigation, with the training size, model, heritability and other factors keeping constant the genomic selection does not overcome the best result achievable with phenotypic selection.

As expected, when the training size becomes larger and closer with the validation set, as the case when the whole population was used to prediction, the predictions were higher (Table 2). The prediction accuracies ranged from 0.6626 to 0.7024 for SNP based model and from 0.5744 to 0.6121 for haplotype based model. So, the genomic selection, both based on single SNP or haplotypes, outperformed the phenotypic selection in all generations. On average, the superiority in prediction, compared to the maximum theoretical phenotypic accuracy achievable, was 37.12% for SNP based model and 19.42% for haplotype based model. For coincidence the average value was 0.4524 for SNP based model and 0.3748 for haplotype based model. There is a slightly increase on prediction accuracies and coincidence over the generations.

Again, the analysis based on single SNPs is more suited than based on haplotypes, even in the context of increasing genetic relationship within the generations. On average, the prediction accuracies and coincidences are 15 and 21% larger, respectively, based on single SNP.

The LD analysis showed to us small values of LD in both scenarios. Considering the scenario with higher SNP density the measure r^2, among SNPs within distances encompassing 10 SNPs (haplotype length, approximately 267 Kb), presents an average value of 0.0061. The maximum r^2 found in that situation was only 0.2780. As r^2 represents a correlation it is clear the weak association among SNPs in the population studied. The measure D' presents an average value of 0.1818, in the situation just explained. Considering the scenario with lower SNP density the measure r^2, within a haplotype of 10 SNP (approximately length of 2,669 Kb), presents an average value of 0.0045. The maximum r^2 found in that situation was only 0.1270. The D' measure presents an average value of only 0.1632. Even if the haplotype was composed by 2 SNPs, the least haplotype length possible, the values of LD in that population remain low. In the scenario of higher SNP density the average D' is 0.1825 and the average r^2 is only 0.0058. In the scenario of lower SNP density those values are 0.1595 and 0.0051 for D' and r^2, respectively. So, in that population the use of haplotypes for prediction can not explore high LD among markers.

3.2 COMPARISON BETWEEN SNP AND HAPLOTYPE PREDICTION ACCURACIES

For comparing the SNP and haplotype prediction accuracies of the QTL additive value we computed ratios between the prediction accuracy of a QTL by the two linked SNPs and the average haplotype prediction accuracy of the QTL (Table 3), assuming the described scenarios. Note that, with one exception, the prediction accuracy of the QTL additive value based on both SNPs was greater than the maximum prediction based on a single SNP. The results showed that if SNPs and QTL have similar allelic frequencies (differences in the range 0.003 to 0.111), in 36% of the simulated cases the prediction accuracy based on SNPs was greater than the haplotype prediction accuracy. However, if both SNPs (differences in the range 0.425 to 0.450) or at least one SNP (differences in the range 0.358 to 0.476) have different allele frequencies relative to the QTL, the prediction accuracy based on SNPs was greater than the haplotype prediction accuracy in at least 98% of the cases. The ratios greater than one ranged from 1.01 to

1.86 (1.38 on average). Thus, based on our very limited but complex investigation, we can generally state that: 1) in real situations, is probably that at least one haplotype allele pair will show greater accuracy compared with the prediction accuracy for the SNPs in the haplotype block (we observed this in 91% of the cases simulated) but because this pair of haplotype alleles represents a subsample of the gametic pool of haplotypes, a general inference on the relative efficiency of the haplotype prediction accuracy should be based on the average haplotype prediction accuracy; and 2) in real situations, the SNP prediction accuracy of a quantitative trait should be higher than the haplotype prediction accuracy if there is significant differences between the SNPs and the QTLs frequencies; if SNPs and QTLs frequencies are similar, the haplotype prediction accuracy should be higher than the SNP prediction accuracy.

4. DISCUSSION

Exploring the LD between closed markers, as happens in the case of haplotypes, to perform genomic prediction does not warrant better results in order to make selection than that obtained by single SNPs. Using a real data set of oat lines, Bekele et al. (2018) did not find significant differences in prediction accuracies, for the trait heading date, using GBS-SNPs or haplotypes in a cross-validation scheme. When the authors studied those predictions in an independent validation scheme the GBS-SNP markers gave slightly higher prediction accuracy (lower than 1%), different of our study when the SNP based model outperformed a lot the haplotype based one. However those authors found values of accuracy much larger than obtained here, ranging from 0.724 to 0.743. Cuyabano et al. (2014) did not find difference between the use of haplotypes or single SNP when the trait under investigation had very low heritability (0.04). However, they found significant difference, by Hotelling's test, in prediction accuracy and coincidence of the best individuals in a dairy cattle population for haplotype approach, either using a BLUP or a Bayesian mixture model when the heritability was not so low (0.39). In GWAS studies, Ogawa et al. (2018) detected with higher precision the loci controlling two traits in oat using haplotypes than using single SNP. In turn, Bekele et al. (2018) found more associations in a GWAS analysis using haplotypes alleles than GBS-SNPs.

The LD seems to be more useful for studies that directly estimate the allele effects in association studies than using those alleles to estimate genetic relationships to be employed in a BLUP model, as done in our study. The maintenance of strong LD among tight linked SNPs, as expected in a haplotype block over generations of recurrent

selection, did not ensure good predictions in our investigation. It is worthy to note that in our population there is no much LD to be exploited by the block of haplotypes and this could be a factor limiting the haplotype based prediction. On the other hand, Cuyabano et al. (2014, 2015) found some advantage with the use of haplotypes for prediction in a dairy cattle population when the haplotypes were constructed by the criteria of D' higher than 0,45. Bekele et al. (2018) and Ma et al. (2016) exploited the LD for prediction purposes in plant populations using haplotypes constructed by the criteria of Gabriel et al. (2002), a criteria also based on D' and initially proposed for human studies.

Different ways to obtain the haplotypes result in different efficiency to use in prediction (Jónás et al., 2016). Jónás et al. (2016) investigated methods to construct haplotypes after a previous analysis of the marker effects individually. The authors obtained the haplotypes by merging flanking markers around the ones of greatest effect on the trait at a determined haplotype size (3, 4, 5) or applying a threshold for allele frequency. An average increase prediction of 2% based on haplotypes instead of single SNP was found by them. Edwards (2015) proposed two measures of genetic relationship based on shared segment haplotypes among relatives and compared for prediction purposes. The measures of genetic relationship based on haplotypes resulted in higher prediction accuracies than using the traditional single-SNP derived matrix or pedigree information in a dairy cattle data set. Additionally, the matrices based on haplotype sharing were closer to pedigree matrix in capture genealogical relatedness, even for distantly related individuals, outperforming the SNP derived matrix. Mathew et al. (2018) found only slightly improvement (0.75%) in prediction accuracies using a LD corrected relationship matrix than the traditional SNP genomic matrix with real data set of plant (maize and rice) and animal (mice and cattle) populations. Ferdosi et al. (2016) employing three different methods to construct the genetic relationship matrix among beef cattle individuals also found better prediction accuracies for haplotype based than for single SNP based model. The traditional genomic relationship matrix considers complete LD between SNPs and QTL, but ignores LD between SNPs, especially in short regions (Habier et al., 2013). On the other hand, Uemoto et al. (2017) found slightly larger prediction accuracies for BLUP model employing genomic relationship matrix constructed from SNPs than haplotypes, likewise our results. This superiority was about 6.51% using real data set of closed lines of Duroc pigs. The authors stated about the property of relationship matrix based on SNP capture very old

relationships, whereas haplotype based relationship matrix captures intermediate-aged relationships.

The length and the method to determine the haplotype alleles is another factor influencing the results in genomic prediction studies. The length of 10 SNP markers and the haplotype blocks without overlapping, adopted here, can be considered suited to construct the haplotype alleles based on previous studies of Ferdosi et al. (2016) and Villumsen et al. (2009). Those authors demonstrated that if haplotype similarity is the basis of determining relationships, increasing the number of haplotype alleles, as happens adopting larger haplotype blocks, will generally result in lower relationships within any segment. Cuyabano et al. (2015) grouped SNPs into haploblocks based on LD ($D' \geq 0.45$) and found that more than 90% of the haploblocks contained less than 10 SNPs. Besides that, Ferdosi et al. (2016) concluded that longer haplotypes resulted in a relationship matrix that was similar to an identity matrix. So, an appropriate haplotype allele construction is crucial to capture the genetic relationship among the individuals and, hence, to predict breeding values to further perform selection. A pre-selection of SNPs, to form the haplotypes, closer associated to the QTLs can boost the predictions, as demonstrated by Jónás et al. (2016). Those authors found that haplotypes containing those previously analyzed SNPs (larger effects on the phenotypic expression) provide an increase in prediction accuracies than single SNP model, ranging between 0.8 and 2.9% for five traits of a real dairy cattle data set. Ma et al. (2016) also investigated different strategies of sampling marker to study the prospects of a cost efficient genomic selection for two traits in soybean. The authors found that the approach based on haplotypes provided the best prediction accuracy for yield per plant. In this approach, one SNP was selected within each haplotype block plus SNPs not forming haplotype blocks. Compared to random sampling of SNPs or evenly spaced sampled SNPs this strategy of pre-selection gave accuracies approximately 4% higher.

Our simulation did not consider the presence of epistatic effects in trait expression and this can be pointed as a factor favoring the SNP based model. Jiang et al. (2018) demonstrated by simulation the property of haplotype based models to capture epistatic effects, specially higher order epistasis, even so they are not included explicitly in the model. The authors found a general better performance of haplotype based model than SNP based model when epistasis of higher order was present. The covariance matrix from haplotype alleles considers correlated epistatic effects and this is the main difference between the SNP based model when it includes epistatic effects as a random factor. The role of epistatic effects on the expression of quantitative traits is not well

understood yet, thus, the choice between a model based on single SNP or haplotype should be done cautiously. Their efficiencies differ by this factor, what is seldom considered in genomic prediction models.

As the efficiency of genomic prediction is reduced over the generations, at some time the breeder has to train the model again. However, in our investigation we found that the reduction in accuracy and coincidence of the best individuals is slow, so that the breeder should pay attention on the quality of phenotypic data because it will be utilized for some generations. Villumsen et al. (2009) pointed out the prospects of genomic selection in a dairy cattle context using haplotypes, keeping high values of reliabilities up to seven generations. Those authors simulated seven generations of a dairy cattle population, without selection, as done here, and found a general superiority of haplotype based predictions over the single SNPs. Those authors found similar trend of decrease in accuracy over the generations in a range of heritability from 0.02 to 0.30 and haplotype length from 2 to 40 SNPs. However, it is worthy to note that longer haplotypes are more suited for prediction in the first generations, but the precision may decay quickly in later generations owing to recombination.

The real usefulness of haplotypes in genomic selection is not well understood so far. The studies differ among them in what is the best strategy to be employed considering the use of haplotypes. Calus et al. (2009), for example, did not find significant differences in the efficiency of haplotypes for prediction ranging the haplotype length from 2 to 20 SNPs. Frischknecht et al. (2016) investigated different definitions of haplotypes to predict four traits in Brown Swiss cattle data set varying the physical length or LD measure to construct the haplotypes. The authors stated that only marginal improvements were found comparing the accuracies obtained from haplotypes versus 50K SNP. Also employing relationship matrix derived from single SNP or haplotypes, as done in our study, Karimi et al. (2018) evaluated the prediction accuracy for 57 Holstein traits. Those authors also constructed the block of haplotypes considering a fixed number of adjacent SNPs (5, 10, 15 or 20) and found differences in accuracies from SNP based model and haplotype based model ranging from -4.2% to +3.3%. For the majority of traits with low heritability the haplotype based model performed poor than SNP based one. The haplotype length of 20 SNPs resulted in the largest bias of prediction and poorest accuracies, whereas the length of 5 and 10 SNPs were similar. Mucha et al. (2019) evaluated four models based on haplotypes to predict breeding values of ten traits in Holstein-Friesian dairy cattle. Even exploring high LD in the population ($r^2 \geq 0.8$ or $r^2 \geq 0.9$) and discarding the low frequency haplotype alleles

(MAF = 0.25) all the models underestimated the breeding values and the efficiency of prediction was worse for low heritability traits.

Many studies do not confirm a clear advantage of using haplotypes for prediction, likewise our results pointed out. Our investigation does not support the use of haplotypes for prediction over generations.

5. REFERENCES

Barrett, J. C., Fry, B., Maller, J. D. M. J., Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, 21:263-265.

Bekele, W. A., Wight, C. P., Chao, S., Howarth, C. J., Tinker, N. A. (2018). Haplotype-based genotyping-by-sequencing in oat genome research. Plant biotechnology journal, 16:1452-1463.

Calus, M. P., Meuwissen, T. H., Windig, J. J., Knol, E. F., Schrooten, C., Vereijken, A. L., & Veerkamp, R. F. (2009). Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. Genetics Selection Evolution, 41:11.

Covarrubias-Pazaran G. (2016). Genome assisted prediction of quantitative traits using the R package sommer. PLoS ONE, 11:1-15.

Cuyabano, B. C. D., Su, G., Lund, M. S. (2014). Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. BMC genomics, 15:1171.

Cuyabano, B. C. D., Su, G., Rosa, G. J. M., Lund, M. S., Gianola, D. (2015). Bootstrap study of genome-enabled prediction reliabilities using haplotype blocks across Nordic Red cattle breeds. Journal of dairy science, 98:7351-7363.

Edwards, D. (2015). Two molecular measures of relatedness based on haplotype sharing. BMC bioinformatics, 16:383.

Endelman, J. B., J.-L. Jannink. (2012). Shrinkage estimation of the realized relationship matrix. G3:Genes, Genomes, Genetics, 2:1405-1413.

Ferdosi, M. H., Henshall, J., Tier, B. (2016). Study of the optimum haplotype length to build genomic relationship matrices. Genetics Selection Evolution, 48:75.

Frischknecht, M., Seefried, F. R., Bapst, B., Flury, C., Signer-Hasler, H., Garrick, D., ... & Sölkner, J. (2016). Genomic Prediction using Haplotypes in Brown Swiss. Interbull Bulletin, 50:34-38.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., ... & Liu-Cordero, S. N. (2002). The structure of haplotype blocks in the human genome. Science, 296:2225-2229.

Grattapaglia, D., Resende, M. D. V. (2010). Genomic selection in forest tree breeding Tree Genetics & Genomes, 7:241-255.

Habier, D., Fernando, R. L., Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. Genetics, 177:2389-97.

Habier, D., Fernando, R. L., Garrick, D. J. (2013). Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics, 194:597-607.

Hess, M., Druet, T., Hess, A., Garrick, D. (2017). Fixed-length haplotypes can improve genomic prediction accuracy in an admixed dairy cattle population. Genetics Selection Evolution, 49:54.

Hickey, J. M., Kinghorn, B. P., Tier, B., Wilson, J. F., Dunstan, N., van der Werf, J. H. (2011). A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. Genetics Selection Evolution, 43:12.

Jiang, Y., Schmidt, R. H., Reif, J. C. (2018). Haplotype-Based Genome-Wide Prediction Models Exploit Local Epistatic Interactions Among Markers. G3: Genes, Genomes, Genetics, 8:1687-1699.

Jónás, D., Ducrocq, V., Fouilloux, M. N., Croiseau, P. (2016). Alternative haplotype construction methods for genomic evaluation. Journal of dairy science, 99:4537-4546.

Karimi, Z., Sargolzaei, M., Robinson, J. A. B., & Schenkel, F. S. (2018). Assessing haplotype-based models for genomic evaluation in Holstein cattle. Canadian Journal of Animal Science, 98:750-759.

Kempthorne, O. (1957). An introduction to genetic statistics. Iowa State University Press, Ames, 545 pp.

Lawson, D. J., Hellenthal, G., Myers, S., Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS genetics, 8:e1002453.

Ma, Y., Reif, J. C., Jiang, Y., Wen, Z., Wang, D., Liu, Z., ... & Wang, H. (2016). Potential of marker selection to increase prediction accuracy of genomic selection in soybean (Glycine max L.). Molecular Breeding, 36:113.

Mathew, B., Léon, J., Sillanpää, M. J. (2018). A novel linkage-disequilibrium corrected genomic relationship matrix for SNP-heritability estimation and genomic prediction. Heredity, 120:356-368.

Meuwissen, T. H. E., Goddard, M. K. (2010). Accurate prediction of genetic values for complex traits by whole genome resequencing. Genetics, 185:623-631.

Mucha, A., Wierzbicki, H., Kamiński, S., Oleński, K., & Hering, D. (2019). High-frequency marker haplotypes in the genomic selection of dairy cattle. Journal of applied genetics, 1-8.

Müller, D., Schopp, P., Melchinger, A. E. (2018). Selection on Expected Maximum Haploid Breeding Values Can Increase Genetic Gain in Recurrent Genomic Selection. G3: Genes, Genomes, Genetics, 8:1173-1181.

Ogawa, D., Yamamoto, E., Ohtani, T., Kanno, N., Tsunematsu, H., Nonoue, Y., Yano, M., Yamamoto, T., Yonemaru, J. I. (2018). Haplotype-based allele mining in the Japan-MAGIC rice population. Scientific reports, 8:4379.

Uemoto, Y., Sato, S., Kikuchi, T., Egawa, S., Kohira, K., Sakuma, H., Miyashita, S., Arata, S., Kojima, T., Suzuki, K. (2017). Genomic evaluation using SNP-and haplotype-based genomic relationship matrices in a closed line of Duroc pigs. Animal Science Journal, 88(10), 1465-1474.

Utsunomiya, Y. T., Milanesi, M. (2017). GHap: Genome-Wide Haplotyping. R package version 1.2.2. https://CRAN.R-project.org/package=GHap

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of dairy science, 91:4414-4423.

VanRaden, P. M., Null, D. J., Sargolzaei, M., Wiggans, G. R., Tooker, M. E., Cole, J. B., Sonstegard, T. S., Connor, E. E., Winters, M., Van Kaam, J. B. C. H. M., Valentini, A., Van Doormaal, B. J., Faust, M. A., Doak, G. A. (2013). Genomic imputation and evaluation using high-density Holstein genotypes. Journal of Dairy Science, 96:668-678.

Viana, J. M. S. (2004). Quantitative genetics theory for non-inbred populations in linkage disequilibrium Genetics and Molecular Biology 27:594-601.

Viana, J. M. S., Piepho, H. P., Silva, F. F. (2016). Quantitative genetics theory for genomic selection and efficiency of breeding value prediction in open-pollinated populations Scientia Agricola, 73:243-251.

Villumsen, T. M., Janss, L., Lund, M. S. (2009). The importance of haplotype length and heritability using genomic selection in dairy cattle. Journal of Animal Breeding and Genetics, 126:3-13.

Table 1 Prediction accuracy and coincidence (standard deviation) of the best 500 individuals (10% of population size) in function of the marker density, model used, validation generation and validation size.

Marker density	Model	Val. generation	Val. size	Accuracy *ref=0.5	Coincidence *ref=0.3156
60,000	SNP	2	3000[a]	0.4864 (0.0185)	0.3859 (0.0243)
		2	5000	0.5355 (0.0122)	0.3296 (0.0171)
		3	5000	0.4765 (0.0136)	0.3086 (0.0191)
		4	5000	0.4357 (0.0155)	0.2649 (0.0180)
		5	5000	0.4143 (0.0148)	0.2606 (0.0159)
		6	5000	0.3891 (0.0158)	0.2466 (0.0180)
	Haplotype	2	3000[a]	0.3886 (0.0180)	0.3349 (0.0214)
		2	5000	0.4388 (0.0119)	0.2922 (0.0201)
		3	5000	0.3480 (0.0200)	0.2364 (0.0201)
		4	5000	0.3038 (0.0251)	0.2004 (0.0175)
		5	5000	0.2528 (0.0276)	0.1853 (0.0175)
		6	5000	0.2515 (0.0192)	0.1959 (0.0171)
6,000	SNP	2	3000[a]	0.4135 (0.0178)	0.3412 (0.0173)
		2	5000	0.4861 (0.0157)	0.3079 (0.0200)
		3	5000	0.3967 (0.0186)	0.2532 (0.0172)
		4	5000	0.3374 (0.0186)	0.2186 (0.0184)
		5	5000	0.2799 (0.0203)	0.2010 (0.0164)
		6	5000	0.2360 (0.0170)	0.1833 (0.0177)
	Haplotype	2	3000[a]	0.0590 (0.0597)	0.1788 (0.0348)
		2	5000	0.1586 (0.0871)	0.1444 (0.0470)
		3	5000	0.1625 (0.0151)	0.1611 (0.0180)
		4	5000	0.1398 (0.0180)	0.1508 (0.0129)
		5	5000	0.1037 (0.0189)	0.1295 (0.0119)
		6	5000	0.0737 (0.0155)	0.1277 (0.0133)

[*]: Parametric values

[a]: Independent validation within the generation 2

Table 2 Maximum prediction accuracy and coincidence of the best 500 individuals (10% of population size), using the total marker set, in function of the model used and validation generation.

Model	Val. Generation	Accuracy	Coincidence
SNP	2	0.6626	0.418
	3	0.6911	0.486
	4	0.7024	0.478
	5	0.6983	0.452
	6	0.6737	0.428
Haplotypes	2	0.5744	0.374
	3	0.5974	0.394
	4	0.6121	0.362
	5	0.6076	0.362
	6	0.5942	0.382

Table 3 Ratios between the prediction accuracy of a QTL by the two linked SNPs (Acc3) and the average haplotype prediction accuracy (Acc10) of the QTL, assuming similar frequencies for SNPs (p_a and p_b) and QTL (p_c) (simulations 1 to 25), SNPs with similar frequencies but contrasting with the QTL frequency (simulations 26 to 50), and a SNP and the QTL with similar frequencies but contrasting with the other SNP frequency (simulations 51 to 100), assuming recombination frequencies θ_{ab} and θ_{bc} between 0 and 0.1[1]

Sim.	θ_{ab}	θ_{bc}	p_a	p_b	p_c	$\Delta_{ab}^{(0)}$	$\Delta_{bc}^{(0)}$	$\Delta_{ac}^{(0)}$	Acc1	Acc2	Acc3	$\Delta_{12}^{(0)}$	$\Delta_{13}^{(0)}$	$\Delta_{14}^{(0)}$	$\Delta_{23}^{(0)}$	$\Delta_{24}^{(0)}$	$\Delta_{34}^{(0)}$	Acc4	Acc5	Acc6	Acc7	Acc8	Acc9	Acc10	Ratio
1	0.005	0.078	0.4109	0.4736	0.4007	0.0638	0.0571	0.0266	0.1103	0.2335	0.2356	-0.0770	-0.0973	0.0498	-0.0152	0.1102	0.1387	0.3121	0.3688	0.2297	0.0509	0.5294	0.6045	0.3672	0.64
2	0.048	0.071	0.4274	0.4710	0.3601	0.0358	0.0151	0.0099	0.0419	0.0629	0.0710	-0.0814	-0.0783	0.0089	0.0041	0.0844	0.0845	0.3315	0.3133	0.0444	0.0130	0.3791	0.3659	0.2435	0.29
3	0.049	0.005	0.5218	0.4765	0.4408	0.0076	-0.0018	0.0345	0.1392	0.0073	0.1396	0.0028	0.0407	0.0318	0.0378	0.0290	-0.0091	0.0112	0.1648	0.1278	0.1534	0.1166	0.0376	0.1012	1.38
4	0.029	0.015	0.5170	0.5023	0.4794	0.0019	0.0413	0.0032	0.0127	0.1653	0.1657	-0.0703	-0.1119	0.0484	-0.0418	0.1187	0.1606	0.2786	0.3915	0.2618	0.1081	0.5362	0.6495	0.3690	0.45
5	0.005	0.096	0.4882	0.4596	0.4737	0.0033	0.0184	0.0133	0.0531	0.0740	0.0905	0.0272	0.0225	0.0319	-0.0048	0.0050	0.0097	0.1062	0.0865	0.1268	0.0195	0.0206	0.0400	0.0657	1.38
6	0.012	0.009	0.5282	0.4795	0.4965	0.0088	0.0033	0.0344	0.1378	0.0134	0.1380	-0.0545	-0.0283	0.0374	0.0257	0.0919	0.0656	0.1921	0.1122	0.1823	0.0794	0.3721	0.2983	0.2074	0.67
7	0.059	0.055	0.4746	0.4833	0.5453	-0.0096	0.0171	0.0327	0.1316	0.0688	0.1509	0.0086	0.0237	0.0507	0.0152	0.0424	0.0272	0.0295	0.0883	0.2149	0.0583	0.1827	0.1239	0.1167	1.29
8	0.067	0.043	0.3889	0.4310	0.4410	0.0261	0.0017	0.0267	0.1102	0.0071	0.1096	-0.0940	-0.0558	0.0041	0.0372	0.0888	0.0578	0.3334	0.2243	0.0212	0.1054	0.3786	0.2571	0.2188	0.50
9	0.02	0.033	0.4890	0.5130	0.4441	0.0100	-0.0168	0.0135	0.0544	0.0674	0.0884	-0.0759	-0.0417	-0.0029	0.0344	0.0731	0.0388	0.3018	0.1709	0.0124	0.1301	0.2893	0.1584	0.1753	0.50
10	0.028	0.061	0.3263	0.3408	0.3461	0.0369	0.0506	0.0471	0.2113	0.2245	0.2855	-0.0396	-0.0397	0.0437	-0.0005	0.0794	0.0825	0.1623	0.1633	0.2220	0.0018	0.3812	0.3892	0.2542	1.12
11	0.009	0.074	0.4695	0.4909	0.5347	-0.0062	0.0178	-0.0017	0.0069	0.0713	0.0715	0.0711	0.0498	0.0205	-0.0210	-0.0505	-0.0294	0.2865	0.1913	0.0724	0.0945	0.2112	0.1176	0.1607	0.44
12	0.024	0.037	0.3863	0.3881	0.3513	0.0134	0.0493	0.0175	0.0753	0.2118	0.2209	-0.1322	-0.1659	0.0034	-0.0348	0.1293	0.1619	0.5559	0.6880	0.0166	0.1368	0.5709	0.7074	0.4452	0.50
13	0.003	0.026	0.4260	0.4932	0.4874	0.0289	0.0243	-0.0049	0.0199	0.0973	0.1017	0.0207	-0.0116	0.0164	-0.0319	-0.0040	0.0280	0.0853	0.0424	0.0679	0.1254	0.0179	0.1110	0.0744	1.37
14	0.05	0.012	0.4616	0.4650	0.4064	0.0146	0.0114	0.0197	0.0803	0.0466	0.0905	0.0071	0.0160	0.0285	0.0088	0.0216	0.0129	0.0280	0.0637	0.1183	0.0357	0.0903	0.0547	0.0666	1.36
15	0.097	0.034	0.4471	0.5016	0.4583	0.0159	-0.0035	0.0002	0.0010	0.0141	0.0142	-0.0647	-0.0541	-0.0080	0.0107	0.0554	0.0460	0.2603	0.2198	0.0376	0.0354	0.2230	0.1850	0.1589	0.09
16	0.029	0.036	0.5279	0.4920	0.5291	-0.0040	-0.0055	-0.0014	0.0056	0.0222	0.0230	0.0239	0.0302	-0.0083	0.0062	-0.0321	-0.0385	0.0962	0.1227	0.0306	0.0270	0.1271	0.1538	0.0924	0.25
17	0.049	0.017	0.4153	0.4408	0.3876	0.0246	0.0140	0.0147	0.0614	0.0578	0.0804	-0.1413	-0.1313	-0.0045	0.0095	0.1274	0.1217	0.5750	0.5359	0.0204	0.0362	0.5443	0.5131	0.3689	0.22
18	0.093	0.018	0.4888	0.3928	0.4490	0.0246	0.0169	0.0226	0.0908	0.0695	0.1092	-0.0715	-0.0873	0.0184	-0.0150	0.0908	0.0971	0.2717	0.3325	0.0920	0.0453	0.3805	0.4350	0.2626	0.42
19	0.061	0.069	0.3532	0.3487	0.3513	0.0458	0.0329	0.0570	0.2497	0.1445	0.2662	-0.1489	-0.1184	0.0188	0.0307	0.1406	0.1144	0.6217	0.4948	0.0899	0.1273	0.6549	0.5365	0.4231	0.63
20	0.089	0.049	0.3834	0.4104	0.3935	0.0423	0.0480	0.0442	0.1861	0.1998	0.2517	-0.0212	-0.0207	0.0586	0.0006	0.0751	0.0782	0.0713	0.0692	0.2961	0.0018	0.3740	0.3754	0.2292	1.10
21	0.021	0.044	0.5152	0.4832	0.4160	0.0208	-0.0242	0.0097	0.0393	0.0983	0.1090	-0.0567	-0.0220	-0.0135	0.0344	0.0432	0.0087	0.2240	0.0965	0.0599	0.1247	0.1635	0.0365	0.1179	0.92
22	0.091	0.065	0.3911	0.3746	0.3468	0.0417	0.0347	0.0347	0.1492	0.1507	0.1955	-0.0949	-0.1011	0.0222	-0.0061	0.1061	0.1080	0.3678	0.3900	0.1170	0.0191	0.4996	0.5172	0.3321	0.59
23	0.017	0.094	0.5026	0.4775	0.4507	-0.0132	0.0253	0.0071	0.0285	0.1018	0.1072	0.0102	-0.0077	0.0335	-0.0180	0.0232	0.0413	0.0403	0.0285	0.1417	0.0691	0.1003	0.1697	0.0914	1.17
24	0.011	0.073	0.4136	0.4572	0.4373	0.0258	-0.0168	0.0135	0.0551	0.0680	0.0924	-0.0851	-0.0444	-0.0140	0.0403	0.0654	0.0289	0.3291	0.1878	0.0650	0.1380	0.2629	0.1218	0.1778	0.52
25	0.036	0.08	0.4304	0.4619	0.4121	0.0396	0.0616	-0.0083	0.0341	0.2512	0.2591	0.1046	0.0187	0.0500	-0.0867	-0.0497	0.0315	0.4290	0.0743	0.2050	0.3570	0.2147	0.1307	0.2257	1.15
26	0.069	0.096	0.2685	0.2986	0.7128	-0.0284	0.0360	-0.0292	0.1454	0.1741	0.2127	0.0386	-0.0051	0.0083	-0.0705	-0.0324	0.0294	0.2208	0.0290	0.0624	0.3127	0.1568	0.1361	0.1520	1.40
27	0.042	0.052	0.2815	0.2969	0.7275	-0.0247	0.0232	-0.0268	0.1339	0.1139	0.1662	0.0014	-0.0331	-0.0083	-0.0540	-0.0203	0.0272	0.0077	0.1822	0.0600	0.2417	0.0988	0.1283	0.1205	1.38
28	0.037	0.008	0.2826	0.2803	0.7050	-0.0341	0.0412	-0.0324	0.1579	0.2009	0.2369	0.0310	-0.0108	0.0071	-0.0779	-0.0312	0.0375	0.2030	0.0568	0.0555	0.3405	0.1521	0.1689	0.1648	1.44
29	0.032	0.082	0.2549	0.3216	0.7175	-0.0393	0.0457	-0.0250	0.1276	0.2172	0.2330	0.0435	0.0026	0.0147	-0.0710	-0.0260	0.0395	0.2782	0.0151	0.1188	0.3151	0.1309	0.1765	0.1704	1.37
30	0.025	0.072	0.2720	0.3309	0.7459	-0.0420	0.0458	-0.0260	0.1340	0.2236	0.2403	0.0126	-0.0236	0.0020	-0.0683	-0.0142	0.0506	0.0767	0.1519	0.0159	0.3146	0.0694	0.2379	0.1527	1.57
31	0.089	0.062	0.2796	0.3257	0.7208	-0.0311	0.0357	-0.0358	0.1780	0.1697	0.2295	0.0327	-0.0157	0.0021	-0.0740	-0.0381	0.0291	0.2053	0.0781	0.0147	0.3338	0.1928	0.1264	0.1578	1.45
32	0.065	0.059	0.2447	0.3469	0.7436	-0.0430	0.0377	-0.0292	0.1555	0.1816	0.2175	0.0389	-0.0001	0.0091	-0.0660	-0.0337	0.0284	0.2420	0.0006	0.0747	0.3087	0.1685	0.1314	0.1541	1.41

33	0.072	0.005	0.2628	0.3448	0.7336	-0.0341	0.0485	-0.0320	0.1643	0.2306	0.2636	0.0471	-0.0071	0.0121	-0.0804	-0.0339	0.0413	0.2640	0.0423	0.0870	0.3683	0.1669	0.1903	0.1866	1.41
34	0.059	0.040	0.2730	0.3224	0.7165	-0.0352	0.0346	-0.0260	0.1297	0.1641	0.1938	0.0106	-0.0242	-0.0016	-0.0605	-0.0187	0.0372	0.0681	0.1212	0.0122	0.2591	0.0954	0.1616	0.1243	1.56
35	0.087	0.060	0.2361	0.3878	0.7452	-0.0426	0.0455	-0.0249	0.1345	0.2143	0.2324	0.0157	-0.0186	0.0021	-0.0605	-0.0156	0.0470	0.1034	0.1050	0.0167	0.2698	0.0865	0.2062	0.1418	1.64
36	0.047	0.010	0.2998	0.2929	0.7353	-0.0332	0.0311	-0.0211	0.1044	0.1549	0.1744	-0.0117	-0.0449	-0.0092	-0.0544	-0.0042	0.0447	0.0711	0.2485	0.0692	0.2406	0.0203	0.2107	0.1360	1.28
37	0.022	0.007	0.3084	0.2467	0.6872	-0.0120	0.0187	-0.0296	0.1381	0.0938	0.1625	0.0256	-0.0126	0.0003	-0.0542	-0.0335	0.0133	0.1515	0.0566	0.0017	0.2455	0.1589	0.0617	0.1096	1.48
38	0.027	0.099	0.2759	0.3243	0.7255	-0.0331	0.0360	-0.0268	0.1344	0.1725	0.2036	0.0221	-0.0174	0.0029	-0.0637	-0.0233	0.0352	0.1509	0.0762	0.0217	0.2758	0.1253	0.1492	0.1349	1.51
39	0.059	0.042	0.2470	0.3299	0.7390	-0.0404	0.0429	-0.0217	0.1148	0.2078	0.2196	-0.0086	-0.0336	-0.0061	-0.0598	-0.0042	0.0536	0.0541	0.2234	0.0519	0.2711	0.0210	0.2487	0.1496	1.47
40	0.089	0.001	0.2680	0.3479	0.7516	-0.0435	0.0418	-0.0348	0.1817	0.2032	0.2484	0.0130	-0.0257	-0.0034	-0.0722	-0.0273	0.0418	0.1013	0.1218	0.0273	0.3184	0.1543	0.1750	0.1586	1.57
41	0.048	0.088	0.2936	0.2697	0.7040	-0.0268	0.0297	-0.0204	0.0982	0.1467	0.1666	0.0027	-0.0320	-0.0026	-0.0555	-0.0106	0.0372	0.0161	0.1601	0.0197	0.2366	0.0511	0.1723	0.1087	1.53
42	0.045	0.058	0.2679	0.3007	0.7212	-0.0383	0.0316	-0.0411	0.2068	0.1537	0.2371	0.0668	0.0232	0.0170	-0.0773	-0.0605	0.0078	0.5975	0.1016	0.1121	0.4152	0.3277	0.0301	0.2336	1.02
43	0.011	0.042	0.2763	0.3043	0.7174	-0.0355	0.0492	-0.0239	0.1187	0.2376	0.2490	0.0110	-0.0296	0.0028	-0.0740	-0.0095	0.0580	0.0825	0.0833	0.0249	0.2667	0.0557	0.2398	0.1341	1.86
44	0.079	0.028	0.2887	0.2928	0.7173	-0.0282	0.0287	-0.0265	0.1301	0.1402	0.1794	0.0221	-0.0148	0.0022	-0.0589	-0.0259	0.0263	0.1525	0.0626	0.0159	0.2621	0.1349	0.1122	0.1230	1.46
45	0.001	0.022	0.2779	0.2824	0.7067	-0.0436	0.0340	-0.0248	0.1218	0.1658	0.1874	0.0120	-0.0141	0.0016	-0.0599	-0.0189	0.0350	0.1035	0.0559	0.0146	0.2473	0.1017	0.1431	0.1165	1.61
46	0.095	0.022	0.2937	0.2615	0.7143	-0.0331	0.0295	-0.0278	0.1349	0.1487	0.1860	0.0110	-0.0223	-0.0011	-0.0615	-0.0228	0.0309	0.0906	0.0811	0.0094	0.2657	0.1215	0.1283	0.1192	1.56
47	0.091	0.049	0.2539	0.3336	0.7496	-0.0449	0.0519	-0.0258	0.1369	0.2540	0.2654	0.0473	0.0053	0.0178	-0.0750	-0.0257	0.0452	0.3331	0.0315	0.1544	0.3429	0.1376	0.2058	0.1968	1.35
48	0.024	0.078	0.2418	0.3463	0.7556	-0.0408	0.0409	-0.0293	0.1593	0.2003	0.2341	0.0382	-0.0030	0.0087	-0.0692	-0.0320	0.0332	0.2436	0.0201	0.0726	0.3284	0.1655	0.1554	0.1643	1.42
49	0.082	0.003	0.2563	0.3027	0.7524	-0.0349	0.0441	-0.0273	0.1446	0.2224	0.2464	0.0387	-0.0028	0.0112	-0.0747	-0.0274	0.0392	0.2769	0.0146	0.0962	0.3415	0.1502	0.1808	0.1756	1.40
50	0.064	0.079	0.2830	0.3314	0.7468	-0.0515	0.0425	-0.0354	0.1808	0.2077	0.2472	0.0294	-0.0072	0.0066	-0.0735	-0.0346	0.0351	0.2579	0.0314	0.0537	0.3374	0.1961	0.1414	0.1733	1.43
51	0.024	0.065	0.2587	0.7540	0.3747	-0.0338	0.0373	-0.0328	0.1546	0.1787	0.2178	-0.0513	-0.0553	0.0120	-0.0133	0.0638	0.0604	0.2212	0.2774	0.0506	0.0785	0.2729	0.3106	0.2112	1.03
52	0.002	0.074	0.2881	0.7029	0.2642	-0.0230	0.0198	-0.0263	0.1318	0.0981	0.1562	-0.0208	-0.0390	-0.0102	-0.0166	0.0111	0.0320	0.1083	0.1961	0.0522	0.0925	0.0565	0.1572	0.1251	1.25
53	0.092	0.082	0.2743	0.7479	0.3158	-0.0300	0.0401	-0.0415	0.2001	0.1986	0.2624	0.0190	-0.0410	0.0021	-0.0391	-0.0176	0.0395	0.0975	0.1975	0.0102	0.2231	0.0896	0.1965	0.1643	1.60
54	0.026	0.084	0.2264	0.7736	0.3638	-0.0331	0.0448	-0.0434	0.2157	0.2225	0.2842	-0.0013	-0.0462	0.0024	-0.0315	0.0036	0.0477	0.0069	0.2331	0.0121	0.1860	0.0186	0.2408	0.1696	1.68
55	0.086	0.053	0.2783	0.7423	0.3025	-0.0292	0.0255	-0.0347	0.1686	0.1271	0.1975	-0.0150	-0.0437	-0.0097	-0.0225	0.0062	0.0342	0.0784	0.2136	0.0491	0.1258	0.0310	0.1700	0.1370	1.44
56	0.097	0.062	0.2461	0.7624	0.3800	-0.0367	0.0503	-0.0463	0.2213	0.2435	0.3004	0.0243	-0.0424	0.0093	-0.0431	-0.0156	0.0466	0.1197	0.2066	0.0439	0.2440	0.0781	0.2321	0.1886	1.59
57	0.096	0.03	0.2897	0.7121	0.2874	-0.0386	0.0267	-0.0216	0.1053	0.1304	0.1541	-0.0045	-0.0282	0.0079	-0.0207	0.0122	0.0337	0.0211	0.1427	0.0376	0.1139	0.0584	0.1720	0.1117	1.38
58	0.038	0.037	0.2528	0.7253	0.2614	-0.0252	0.0420	-0.0339	0.1776	0.2140	0.2619	0.0064	-0.0397	0.0074	-0.0318	0.0005	0.0479	0.0312	0.2148	0.0350	0.2026	0.0026	0.2517	0.1661	1.58
59	0.017	0.085	0.2578	0.7283	0.3039	-0.0299	0.0377	-0.0293	0.1459	0.1842	0.2191	0.0022	-0.0351	0.0099	-0.0274	0.0070	0.0438	0.0101	0.1839	0.0438	0.1659	0.0323	0.2247	0.1459	1.50
60	0.086	0.022	0.2305	0.7357	0.3444	-0.0271	0.0451	-0.0361	0.1804	0.2150	0.2624	-0.0048	-0.0437	0.0048	-0.0292	0.0087	0.0528	0.0249	0.2135	0.0248	0.1632	0.0480	0.2478	0.1618	1.62
61	0.072	0.029	0.2759	0.7220	0.2824	-0.0262	0.0296	-0.0259	0.1286	0.1466	0.1835	-0.0208	-0.0402	0.0049	-0.0175	0.0253	0.0441	0.1069	0.1985	0.0254	0.0970	0.1320	0.2183	0.1466	1.25
62	0.015	0.029	0.2556	0.7499	0.3357	-0.0319	0.0416	-0.0415	0.2014	0.2032	0.2646	-0.0095	-0.0496	0.0017	-0.0307	0.0112	0.0498	0.0470	0.2452	0.0082	0.1750	0.0552	0.2483	0.1724	1.54
63	0.002	0.087	0.2622	0.7530	0.3393	-0.0299	0.0471	-0.0489	0.2349	0.2308	0.3060	0.0038	-0.0541	0.0016	-0.0395	-0.0025	0.0522	0.0213	0.2502	0.0083	0.2112	0.0135	0.2469	0.1753	1.75
64	0.044	0.041	0.2476	0.7523	0.3220	-0.0244	0.0334	-0.0370	0.1837	0.1658	0.2328	-0.0510	-0.0586	-0.0056	-0.0134	0.0459	0.0548	0.2605	0.2927	0.0286	0.0783	0.2319	0.2726	0.2032	1.15
65	0.066	0.004	0.2698	0.7291	0.2949	-0.0308	0.0361	-0.0310	0.1529	0.1784	0.2187	-0.0076	-0.0402	0.0076	-0.0258	0.0147	0.0457	0.0393	0.1968	0.0393	0.1420	0.0779	0.2255	0.1492	1.47
66	0.062	0.075	0.2973	0.7074	0.2753	-0.0262	0.0256	-0.0292	0.1428	0.1259	0.1795	-0.0184	-0.0427	-0.0042	-0.0210	0.0146	0.0390	0.0877	0.2153	0.0194	0.1201	0.0690	0.1975	0.1403	1.28
67	0.071	0.074	0.3047	0.7065	0.2300	-0.0221	0.0130	-0.0086	0.0447	0.0677	0.0774	-0.0651	-0.0409	0.0079	0.0110	0.0736	0.0459	0.3231	0.2171	0.0389	0.0669	0.3636	0.2475	0.1957	0.40
68	0.094	0.044	0.2381	0.7584	0.3950	-0.0305	0.0369	-0.0396	0.1902	0.1763	0.2401	-0.0185	-0.0483	-0.0055	-0.0249	0.0132	0.0455	0.0893	0.2381	0.0262	0.1428	0.0635	0.2217	0.1645	1.46
69	0.055	0.029	0.2697	0.7447	0.3377	-0.0447	0.0426	-0.0482	0.2297	0.2065	0.2785	0.0115	-0.0453	-0.0059	-0.0435	-0.0175	0.0391	0.0530	0.2294	0.0265	0.2427	0.0797	0.2029	0.1773	1.57
70	0.072	0.016	0.2720	0.7444	0.3645	-0.0302	0.0350	-0.0458	0.2139	0.1666	0.2527	-0.0449	-0.0663	-0.0111	-0.0227	0.0354	0.0552	0.2259	0.3113	0.0561	0.1184	0.1714	0.2622	0.2086	1.21
71	0.033	0.033	0.2648	0.7282	0.3116	-0.0301	0.0373	-0.0436	0.2134	0.1809	0.2607	-0.0369	-0.0617	-0.0116	-0.0258	0.0256	0.0548	0.1846	0.3070	0.0581	0.1431	0.1264	0.2681	0.2062	1.26
72	0.023	0.056	0.2680	0.7437	0.3173	-0.0318	0.0422	-0.0326	0.1579	0.2077	0.2424	0.0038	-0.0388	0.0176	-0.0301	0.0129	0.0490	0.0183	0.1921	0.0829	0.1738	0.0628	0.2501	0.1638	1.48
73	0.093	0.051	0.2793	0.7157	0.2595	-0.0335	0.0222	-0.0234	0.1191	0.1121	0.1515	-0.0190	-0.0341	-0.0027	-0.0159	0.0163	0.0327	0.0915	0.1817	0.0128	0.0940	0.0786	0.1728	0.1230	1.23
74	0.098	0.052	0.2870	0.7319	0.2692	-0.0258	0.0240	-0.0288	0.1437	0.1219	0.1775	-0.0107	-0.0377	-0.0040	-0.0204	0.0072	0.0328	0.0532	0.1926	0.0192	0.1223	0.0351	0.1715	0.1247	1.42
75	0.017	0.042	0.2837	0.7158	0.2650	-0.0241	0.0347	-0.0298	0.1497	0.1743	0.2174	-0.0305	-0.0504	0.0067	-0.0189	0.0367	0.0556	0.1446	0.2627	0.0310	0.1147	0.1745	0.2894	0.1902	1.14

76	0.015	0.026	0.7227	0.3720	0.3117	-0.0087	0.0392	0.0291	0.1402	0.1750	0.2287	0.0308	0.0144	0.0809	-0.0103	0.0401	0.0655	0.1363	0.0646	0.3592	0.0559	0.2016	0.3007	0.1725	1.33
77	0.043	0.098	0.7255	0.3784	0.2900	-0.0058	0.0341	0.0287	0.1419	0.1548	0.2129	0.0255	0.0130	0.0737	-0.0072	0.0397	0.0599	0.1137	0.0601	0.3364	0.0399	0.2038	0.2845	0.1595	1.33
78	0.061	0.046	0.7186	0.3820	0.3306	0.0075	0.0329	0.0400	0.1889	0.1441	0.2338	0.0079	-0.0078	0.0857	-0.0111	0.0678	0.0917	0.0341	0.0373	0.3759	0.0603	0.3181	0.4129	0.1860	1.26
79	0.033	0.029	0.7521	0.3824	0.3305	-0.0117	0.0329	0.0423	0.2081	0.1438	0.2598	0.0193	0.0165	0.0958	-0.0002	0.0609	0.0847	0.0850	0.0779	0.4308	0.0009	0.3083	0.3653	0.1850	1.40
80	0.057	0.011	0.7571	0.3938	0.3922	0.0016	0.0427	0.0462	0.2206	0.1791	0.2831	0.0226	0.0072	0.1059	-0.0077	0.0710	0.1039	0.0881	0.0306	0.4813	0.0405	0.3643	0.5286	0.2221	1.27
81	0.017	0.055	0.7605	0.3841	0.3325	-0.0042	0.0365	0.0369	0.1837	0.1591	0.2455	0.0269	0.0185	0.0869	-0.0022	0.0495	0.0705	0.1108	0.0802	0.3964	0.0125	0.2663	0.3611	0.1812	1.35
82	0.100	0.022	0.7413	0.3832	0.3018	-0.0017	0.0402	0.0292	0.1455	0.1800	0.2323	0.0253	0.0008	0.0819	-0.0153	0.0474	0.0838	0.1086	0.0033	0.3810	0.0849	0.2535	0.4205	0.1874	1.24
83	0.012	0.086	0.7555	0.3945	0.3737	-0.0101	0.0361	0.0396	0.1903	0.1525	0.2498	0.0173	0.0046	0.0968	-0.0075	0.0648	0.1029	0.0725	0.0209	0.4333	0.0397	0.3253	0.4497	0.1924	1.30
84	0.006	0.02	0.6950	0.3725	0.2697	-0.0066	0.0356	0.0308	0.1509	0.1661	0.2278	0.0259	0.0136	0.0775	-0.0084	0.0427	0.0610	0.1209	0.0649	0.3539	0.0460	0.2120	0.2835	0.1694	1.34
85	0.086	0.084	0.6975	0.3940	0.3020	-0.0059	0.0260	0.0338	0.1601	0.1160	0.2002	0.0089	0.0038	0.0719	-0.0036	0.0543	0.0686	0.0386	0.0168	0.3310	0.0180	0.2682	0.3279	0.1510	1.33
86	0.075	0.095	0.7026	0.3724	0.2640	-0.0076	0.0376	0.0193	0.0959	0.1764	0.2038	0.0256	-0.0018	0.0671	-0.0197	0.0338	0.0683	0.1167	0.0075	0.3155	0.1081	0.1829	0.3429	0.1665	1.22
87	0.034	0.026	0.7490	0.3583	0.3015	-0.0080	0.0350	0.0413	0.2076	0.1589	0.2664	0.0290	0.0297	0.0935	0.0030	0.0496	0.0598	0.1268	0.1340	0.4254	0.0177	0.2649	0.3025	0.1917	1.39
88	0.066	0.098	0.7194	0.3682	0.2899	-0.0084	0.0395	0.0322	0.1579	0.1806	0.2446	0.0239	0.0030	0.0878	-0.0141	0.0514	0.0844	0.1029	0.0123	0.4142	0.0755	0.2786	0.4413	0.1993	1.23
89	0.085	0.012	0.7517	0.3775	0.3436	-0.0001	0.0360	0.0327	0.1595	0.1563	0.2234	0.0244	0.0098	0.0808	-0.0072	0.0467	0.0709	0.1047	0.0461	0.3582	0.0404	0.2317	0.3211	0.1666	1.34
90	0.093	0.077	0.7165	0.3858	0.3217	0.0003	0.0307	0.0318	0.1510	0.1348	0.2023	0.0219	0.0155	0.0708	-0.0025	0.0419	0.0525	0.0902	0.0653	0.3178	0.0133	0.2135	0.2690	0.1493	1.35
91	0.002	0.08	0.7334	0.3710	0.2696	-0.0045	0.0328	0.0356	0.1816	0.1529	0.2399	0.0166	0.0041	0.0845	-0.0076	0.0552	0.0804	0.0768	0.0202	0.3959	0.0442	0.2871	0.3821	0.1807	1.33
92	0.049	0.013	0.6995	0.3974	0.2923	-0.0120	0.0331	0.0341	0.1637	0.1486	0.2273	0.0084	-0.0084	0.0848	-0.0131	0.0649	0.0971	0.0355	0.0350	0.4153	0.0635	0.3395	0.4919	0.2026	1.12
93	0.036	0.049	0.7133	0.3777	0.2448	-0.0032	0.0236	0.0274	0.1408	0.1131	0.1819	0.0111	0.0040	0.0611	-0.0045	0.0421	0.0561	0.0512	0.0191	0.3014	0.0254	0.2267	0.2961	0.1390	1.31
94	0.046	0.012	0.7469	0.3649	0.3297	0.0014	0.0462	0.0443	0.2166	0.2040	0.2965	0.0324	0.0158	0.1065	-0.0074	0.0602	0.0856	0.1370	0.0708	0.4760	0.0425	0.3085	0.4253	0.2211	1.34
95	0.091	0.079	0.7259	0.3887	0.3304	0.0024	0.0376	0.0374	0.1784	0.1639	0.2409	0.0188	0.0022	0.0876	-0.0105	0.0592	0.0855	0.0789	0.0096	0.3966	0.0554	0.2941	0.4135	0.1882	1.28
96	0.031	0.029	0.7235	0.3706	0.2843	-0.0142	0.0348	0.0338	0.1677	0.1597	0.2396	0.0234	0.0135	0.0860	-0.0064	0.0492	0.0732	0.1058	0.0607	0.3976	0.0353	0.2589	0.3528	0.1816	1.32
97	0.022	0.090	0.7204	0.3642	0.3119	-0.0137	0.0395	0.0321	0.1544	0.1770	0.2427	0.0326	0.0209	0.0873	-0.0075	0.0422	0.0650	0.1402	0.0846	0.3879	0.0407	0.2234	0.3233	0.1827	1.33
98	0.014	0.009	0.7358	0.3534	0.3139	-0.0083	0.0422	0.0365	0.1785	0.1901	0.2661	0.0275	0.0076	0.0984	-0.0122	0.0547	0.0889	0.1172	0.0308	0.4514	0.0675	0.2936	0.4611	0.2118	1.26
99	0.057	0.066	0.7587	0.4066	0.4005	-0.0050	0.0407	0.0432	0.2059	0.1692	0.2697	0.0204	0.0058	0.1028	-0.0082	0.0702	0.1094	0.0794	0.0241	0.4743	0.0414	0.3661	0.5460	0.2172	1.24
100	0.077	0.028	0.7118	0.3983	0.2936	-0.0008	0.0262	0.0296	0.1437	0.1175	0.1859	0.0132	0.0064	0.0649	-0.0040	0.0452	0.0588	0.0573	0.0294	0.3000	0.0211	0.2261	0.2829	0.1394	1.33

[1] $\Delta_{ab}^{(0)}, \Delta_{bc}^{(0)}, and\ \Delta_{ac}^{(0)}$ are the LD values for SNPs and QTL;
accuracies 1 and 2 are the prediction accuracies of the QTL additive value based on SNPs 1 and 2, respectively;
$\Delta_{12}^{(0)}, \Delta_{13}^{(0)}, \Delta_{14}^{(0)}, \Delta_{23}^{(0)}, \Delta_{24}^{(0)}, and\ \Delta_{34}^{(0)}$ are the LD values for the haplotypes alleles pairs and the QTL;
accuracies 4 to 9 are the prediction accuracies of the QTL additive value based on the haplotypes alleles 1 and 2 to 3 and 4, respectively.