Research, Society and Development, v. 15, n. 2, e7215250700, 2026
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v15i2.50700

Integration of dominance effects into genomic models for enhancing the understanding of heterosis in dairy cattle

Integração de efeitos de dominância em modelos genômicos para aprimorar a compreensão da heterose em bovinos leiteiros

Integración de efectos de dominancia en modelos genómicos para mejorar la comprensión de la eterosis en bovinos lecheros

Received: 02/20/2026 | Revised: 02/24/2026 | Accepted: 02/24/2026 | Published: 02/25/2026

Fabrício Pilonetto
ORCID: https://orcid.org/0000-0002-2161-0831
Brazilian Association of Girolando Breeders, Brazil
E-mail: fpilonetto@alumni.usp.br
Giovanni Coelho Ladeira
ORCID: https://orcid.org/0000-0003-2925-3266
University of Florida, Gainesville, USA
E-mail: fpilonetto@alumni.usp.br
Juliana Petrini
ORCID: https://orcid.org/0000-0003-3458-1619
Embrapa Dairy Cattle, Juíz de Fora, Brazil
E-mail: julianapetrini@gmail.com
Mayara Salvian
ORCID: https://orcid.org/0000-0002-8336-4082
Luiz de Queiroz College of Agriculture, University of São Paulo, Brazil
E-mail: fpilonetto@alumni.usp.br
Izally Carvalho Gervásio
ORCID: https://orcid.org/0000-0003-3334-6721
Luiz de Queiroz College of Agriculture, University of São Paulo, Brazil
E-mail: iza_cgervasio@alumni.usp.br
Flavia de Oliveira Scarpino van Cleef
ORCID: https://orcid.org/0000-0002-2153-8059
College of Agriculture and Bioresources, University of Saskatchewan, Canada
E-mail: flavia.vancleef@usask.ca
Darlene dos Santos Daltro
ORCID: https://orcid.org/0000-0003-2278-8129
Brazilian Association of Girolando Breeders, Brazil
E-mail: darlenedaltro@gmail.com
Aline Zampar
ORCID: https://orcid.org/0000-0002-2269-7932
Santa Catarina State University, Brazil
E-mail: aline.zampar@udesc.br
Paulo Fernando Machado
ORCID: https://orcid.org/0000-0002-4441-1851
Luiz de Queiroz College of Agriculture, University of São Paulo, Brazil
E-mail: paulo@clinicadoleite.com.br
Luiz Lehmann Coutinho
ORCID: https://orcid.org/0000-0002-7266-8881
Luiz de Queiroz College of Agriculture, University of São Paulo, Brazil
E-mail: llcoutinho@usp.br
Gerson Barreto Mourão
ORCID: https://orcid.org/0000-0002-0990-4108
Luiz de Queiroz College of Agriculture, University of São Paulo, Brazil
E-mail: gbmourao@usp.br

Abstract
Molecular heterozygosity and heterosis are gaining importance in the evaluation of various species. However, the exploration of estimated heterosis based on the genome in purebred populations remains limited. We aimed at investigating the use of heterozygosity as a potential indicator of genomic heterosis in purebred populations. Using a simulated dataset for milk production and quality traits, we considered three scenarios (h^2 = 0.10, 0.30, and 0.50). We adapted the Genome-Wide Association Study (GWAS) to capture non-additive variations for milk yield (MY), fat

percentage (FP), protein percentage (PP), casein percentage (CP), and polyunsaturated fatty acid percentage (PUFA). Heterozygosity ranged from 35.4% to 35.6% in the simulated scenario, with genomic heterosis ranging from 3.2% to 17.0%. Regression coefficients emphasized the significance of heterozygosity for genomic heterosis, varying from 4.8 to 6.07. In real data, most identified genomic regions showed consistency between additive and non-additive models. An increase of 1.96 kg/day in MY was associated with a one-unit increase in heterozygosity, along with a 0.0059% increase in PUFA. Eight genes, including the SOX5 gene associated with PUFAs, have been identified in the literature as important for human health. Selection based on heterozygosity is proposed to favor genomic heterosis, and considering dominance effects in GWAS contributes to marker and QTL identification with potential heterozygous advantages. Our study significantly contributes to understanding how heterosis can be utilized for animal selection in pure breeds using genomic information. Our findings suggest predictive approaches tailored to include genetic dominance effects in genetic evaluation and genome-wide association studies.
Keywords: Dairy cattle; Fatty acid; GWAS; Holstein; Non-additive effect; Simulation; SOX5 gene.

Resumo
A heterozigosidade molecular e a heterose estão ganhando importância na avaliação de várias espécies. No entanto, a exploração da heterose estimada com base no genoma em populações de raça pura permanece limitada. O objetivo deste estudo foi investigar o uso da heterozigose como um potencial indicador de heterose genômica em populações de raças puras. Usando um conjunto de dados simulado para características de produção e qualidade do leite, consideramos três cenários (h^2 = 0,10, 0,30 e 0,50). Adaptamos o Estudo de Associação Genômica Ampla (GWAS) para capturar variações não aditivas para a produção de leite (MY), porcentagem de gordura (FP), porcentagem de proteína (PP), porcentagem de caseína (CP) e porcentagem de ácidos graxos poli-insaturados (PUFA). A heterozigosidade variou de 35,4% a 35,6% no cenário simulado, com a heterose genômica variando de 3,2% a 17,0%. Os coeficientes de regressão enfatizaram a significância da heterozigosidade para a heterose genômica, variando de 4,8 a 6,07. Em dados reais, a maioria das regiões genômicas identificadas mostrou consistência entre os modelos aditivos e não-aditivos. Um aumento de 1,96 kg/dia na MY foi associado a um aumento de uma unidade na heterozigosidade, juntamente com um aumento de 0,0059% nos PUFA. Oito genes, incluindo o gene SOX5 associado aos PUFA, foram identificados na literatura como importantes para a saúde humana. A seleção baseada na heterozigosidade é proposta para favorecer a heterose genômica, e considerar os efeitos de dominância no GWAS contribui para a identificação de marcadores e QTLs com potenciais vantagens heterozigóticas. Nosso estudo contribui significativamente para a compreensão de como a heterose pode ser utilizada para a seleção animal em raças puras utilizando informações genômicas. Nossos achados sugerem abordagens preditivas adaptadas para incluir os efeitos de dominância genética nas avaliações genéticas e nos estudos de associação genômica ampla.
Palavras-chave: Bovinos leiteiros; Ácido graxo; GWAS; Holandesa; Efeito não-aditivo; Simulação; Gene SOX5.

Resumen
La heterocigosidad molecular y la heterosis están ganando importancia en la evaluación de diversas especies. Sin embargo, la exploración de la heterosis estimada basada en el genoma en poblaciones de raza pura sigue siendo limitada. El objetivo de este estudio fue investigar el uso de la heterocigosidad como un indicador potencial de heterosis genómica en poblaciones de razas puras. Utilizando un conjunto de datos simulado para características de producción y calidad de la leche, consideramos tres escenarios (h^2 = 0,10, 0,30 y 0,50). Adaptamos el Estudio de Asociación de Genoma Completo (GWAS) para capturar variaciones no aditivas para la producción de leche (MY), porcentaje de grasa (FP), porcentaje de proteína (PP), porcentaje decaseína (CP) y porcentaje de ácidos grasos poliinsaturados (PUFA). La heterocigosidad varió de 35,4% a 35,6% en el escenario simulado, con una heterosis genómica que osciló entre 3,2% y 17,0%. Los coeficientes de regresión enfatizaron la importancia de la heterocigosidad para la heterosis genómica, variando de 4,8 a 6,07. En datos reales, la mayoría de las regiones genómicas identificadas mostraron consistencia entre los modelos aditivos y no aditivos. Un aumento de 1,96 kg/día en MY se asoció con un aumento de una unidad en la heterocigosidad, junto con un aumento de 0,0059% en PUFA. Ocho genes, incluido el gen SOX5 asociado con PUFA, fueron identificados en la literatura como importantes para la salud humana. Se propone la selección basada en la heterocigosidad para favorecer la heterosis genómica, y la consideración de los efectos de dominancia en el GWAS contribuye a la identificación de marcadores y QTL con posibles ventajas heterocigóticas. Nuestro estudio contribuye significativamente a la comprensión de cómo se puede utilizar la heterosis para la selección animal en razas puras utilizando información genómica. Nuestros hallazgos sugieren enfoques predictivos adaptados para incluir efectos de dominancia genética en la evaluación genética y en los estudios de asociación del genoma completo.
Palabras clave: Bovinos lecheros; Ácido graso; GWAS; Holstein; Efecto no-aditivo; Simulación; Gen SOX5.

1. Introduction

Non-additive genetic models have been extensively explored for selecting crossbred animals in various livestock production systems, such as beef cattle, poultry, and swine (Gengler et al., 1997; Norris et al., 2010; Vitezica et al., 2013; Amuzu-Aweh et al., 2013; Tsairidou et al., 2018; Akanno et al., 2018). This approach assumes that dominance deviations play a pivotal

role in generating heterosis. Furthermore, these effects might contribute to genetic and phenotypic diversity, emphasizing their significance in comprehending the genetic architecture of complex traits, including quantitative traits and common diseases (Nicolini et al., 2018; Hill et al., 2008).

In purebred populations, selecting superior individuals can expedite trait improvement, even in crossbred populations such as dairy cattle, where crossing Jersey and Holstein breeds enhances milk yield and fat content (Sorensen et al., 2008; Kargo et al., 2021). Genetic contributions from each breed provide insights into the genetic progress of crossbred populations, while selecting parents based on their individual genetic potential shows great promise (Clasen et al., 2017; Saborío-Montero et al., 2018). This prompts the hypothesis that genetic and genomic factors linked to heterozygosity could provide informative insights into the evaluated phenotype.

We aimed at investigating the use of heterozygosity as a potential indicator of genomic heterosis in purebred populations. Additionally, we proposed an adapted genome-wide association analysis to capture non-additive variations in regions with potential heterozygous advantages, evaluated through both additive and additive-dominant models. At the protein level, heterosis may result from enhanced protein-protein interactions where heterozygous alleles produce complementary or more stable protein complexes, leading to superior phenotypic expression compared to additive genetic effects. Additionally, dominance effects may arise when the functional product of one allele compensates for a less functional or deleterious allele, while epistasis can reflect complex regulatory interactions between different gene loci that influence the trait expression (Birchler et al., 2010; Veitia, 2009).

2. Methodology

All procedures using animals followed the Institution's ethical research standards, protocol nº 2017.5.1197.11.3 approved by the Institutional Animal Care and Use Committee - CEUA, ESALQ / USP. We conducted experimental, field research of a quantitative nature (Pereira et al., 2018; Risemberg et al., 2026) using descriptive statistics with data classes, mean values, standard deviation, maximums, minimums, absolute frequency in quantity (Vieira, 2021; Shitsuka et al., 2014) and regression and statistical analysis (Chein, 2019; Perez Lopez, 2022).

2.1 Simulation dataset

We employed a stochastic simulation to estimate the role of heterozygosity factors and non-additive effects in purebred population performance. The population structure and genomic information were simulated using QMSim software (Sargolzaei and Schenkel 2009). Three genetic scenarios were simulated to print phenotypes with heritability (h^2) 0.10, 0.30, and 0.50. These values were chosen to mirror the empirical genomic heritabilities observed in our real dataset, which ranged from 0.09 for milk yield to 0.25 for protein and casein percentages (Table 1). A relatively small dairy cattle population was simulated like previous studies, with 20 replicates in each scenario (Perez et al., 2018; Compton et al., 2017; Plieshke et al., 2016).

Table 1 - Records number (N), mean, standard deviation (SD), minimum (MIN), maximum (MAX), and additive genomic heritability (h^2) for milk yield (MY), protein percentage (PP), fat percentage (FP), casein percentage (CP) and polyunsaturated fatty acid content in milk (PUFA).

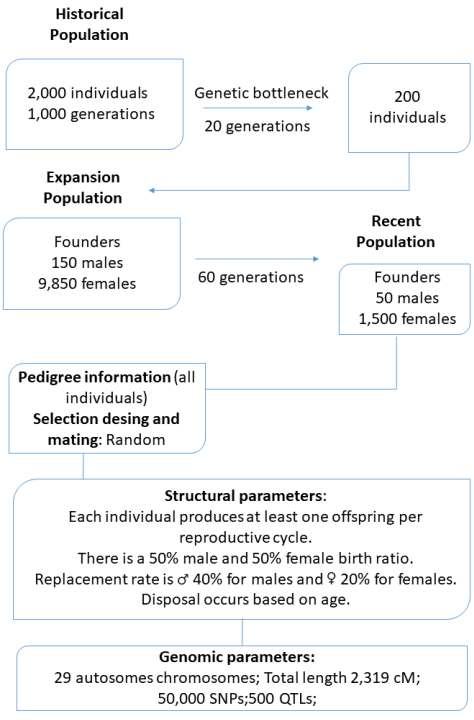
Trait	N	Mean	SD	MIN	MAX	h^2
MY (kg/day)	15,780	32.19	11.53	3.02	80.10	0.09
PP (%)	17,894	3.13	0.36	1.40	5.83	0.25
FP (%)	17,809	3.49	0.79	1.50	11.92	0.12
CP (%)	13,623	2.38	0.32	0.92	4.68	0.25
PUFA (%)	12,438	0.16	0.05	0.01	1.09	0.11

Source: Authors.

A population of 2,000 unrelated animals with a balanced sex ratio was generated to create the historical population (Figure 1). These animals were randomly mated for 1,000 generations, and a genetic bottleneck was introduced in generation 1,001, reducing the historical population to 200 animals after 20 generations. The genetic bottleneck was used to create a scenario where there is consistent linkage disequilibrium (LD) in the simulated population (Bradford et al., 2019).

An expanding population with 10,000 offspring (150 males and 9,850 females) was generated through random mating over 60 generations (1,020 to 1,080 generation). After 10 generations, 16,500 animals were obtained, which were the offspring of 50 males and 1,500 females. The bovine genome was used as a reference for the simulation which comprised 29 pairs of autosomal chromosomes and 2,319 cM total length. The traits were controlled by 500 biallelic QTLs randomly spaced. Additionally, there were 50,000 SNP (Single Nucleotide Polymorphism) molecular markers distributed equidistantly throughout the genome. For SNP, the allelic frequency was 0.50 to start the historical population and a 2.5 x 10-5 mutation rate, randomly distributed throughout the genome. Additive and non-additive genetic actions were obtained through simulation using R software 4.3.3 (R Code Team, 2024), according to Abdollahi-Arpanahi et al. (2020).

Figure 1 - Depiction of the simulation of the population structure.



Source: Authors.

2.2 Additive genetic effect or purely additive action

The additive genetic action effects (α) were obtained using a standard normal distribution α ~N (μ = 0, σ = 1). To obtain the purely additive effect, the dominance effect (*d*) was set to 0, and incidence elements of the dominance matrix ***A*** were recoded as 0, 1 and 2, for the *A1A1*, *A1A2*/*A2A1* and *A2A2* genotypes. Each effect was formulated as: $\alpha = a + d\,(q - p)$; where a and d are additive and dominance effects, respectively, and p is the allelic frequency with q = 1- p. Additionally, additive breeding values were estimated and defined as the true breeding value (TBV), performing the multiplication of genotypic codes by the substitution effects of quantitative trait nucleotides (QTN). The additive and non-additive effects were distributed so that approximately 50% were positive and 50% negative. Phenotypic values were obtained for each individual *i* (y_i). Each *i* received a phenotype value and a normally distributed residual value was adding $e_i \sim N(0, \sigma_e^2)$, to the sum over QTN (genetic values) as:

$$y_i = \sum_{k=1}^{m} X_{ik}\alpha_k + e_i$$

where X_{ik} (i = 1,...n; k = 1,...m) is an incidence component of the matrix of genetic markers for additive effects (α_k), and e_i is residual effect (σ_e^2).

2.3 Non-additive genetic effect or non-additive action

To model the genetic effects of dominance, the incidence elements of the dominance matrix ***D*** were recoded as 0, 1 and 0, for the A1A1, A1A2/A2A1 and A2A2 genotypes, respectively. According to Abdollahi-Arpanahi et al. (2020), genetic events such as incomplete dominance can be partially captured in an additive model, thus, the simulation structure was for overdominance.

Statistically, the dominance deviations represent the interaction between alleles, in which the mean value is obtained through the mean deviation between heterozygous and homozygous genotypes (Falconer and Mackay, 1996). Thus, to model the dominance action, the distribution follows the study by Abdollahi-Arpanahi et al. (2020), where $d \sim N(\mu = 0, \sigma = 0.50)$.

Each dominance effect was formulated as $d = \alpha * (\sum d_k)$, where d is the dominance deviation; α the allelic substitution effects (additive effect); and d_k is the dominance for k QTN, distributed along the genome. The TBV for the dominance deviations was obtained by multiplying the genotypic codes and dominance effects of each QTN. The dominance effects were distributed so that approximately 50% were positive and 50% negative.

The phenotypes for dominance effect were generated by adding the additive effects, dominance effects and the residual effect being $e_i \sim N(0, \sigma_e^2)$, as demonstrated below:

$$y_i = \sum_{k=1}^{m} X_{ik}\alpha_k + \sum_{k=1}^{m} Z_{ik}d_k + e_i$$

where y_i is the phenotype for each individual *i*; X_{ik} and Z_{ik} (i = 1,...n; k = 1,...m) are the incidence components of the matrix of genetic markers for additive and dominance effects (α_k e d_k, respectively) e e_i is the random residual value, σ_e^2.

2.4 Real dataset

Monthly records of milk yield (MY, kg), fat percentage (FP, %), protein percentage (PP, %), casein percentage (CP, %) and polyunsaturated fatty acid percentage in milk (PUFA, %) were collected from Holstein cows from four Brazilian farms between 2012 and 2016. Cows were raised in free-stall barns and milked three times a day with an automatic system. In addition, genealogical records were available for all herds.

2.5 Phenotypes

From the initial 7,664 Holstein cows, animals with invalid measurements, extreme values, or missing genotypic records were excluded (Table 1). Milk components were determined by mid-infrared spectroscopy (*Delta Instruments CombiScope Filter, Advanced Instruments Inc., Norwood*) carried out in partnership with "Milk Clinic", from the Department of Animal Science, Superior School of Agriculture "Luiz de Queiroz" (University of São Paulo, Piracicaba – SP, Brazil). The details of this analysis can be found in Rodriguez et al. (2014) and Petrini et al. (2016).

Data from animals without valid measurements or with measurements outside the acceptable range (mean ±3 standard deviations) were excluded. Animals over 10 years old with days in milk (DIM) lower than 5 or greater than 305, or without genotypic record and birth, test and calving date were removed; records with milk yield below 3.0 and above 99.0 kg of milk/day, with below 1.0 and above 9.0% protein and below 1.5 and above 12.0% fat were considered outside the range (ICAR, 2017) (Table 1).

Contemporary groups (CG) were formed by the combination of farm, calving season (month and year), and milk sample. An analysis of variance (ANOVA) using R software tested the significance level of the variables included in the CG. The three factors tested were significantly different from zero according to the F test (P-value < 0.05). The covariance components were estimated using the restricted maximum likelihood method with an average information algorithm under genomic-polygenic model (1), using the AIREMLF90 software (Misztal et al., 2018).

2.6 Genotypes

Genomic information was obtained from 1,157 animals genotyped for 50,000 SNPs. Genotypic data were further enriched through imputation analysis, as detailed by Petrini et al. (2016) and Iung et al. (2019). The reference population was formed by bulls, sire of these cows, genotyped for 79,294 SNPs on the Illumina BovineSNP50 (Illumina, San Diego, UEA) or GeneSeek Genomic Profiler HD (Neogen Agrigenomics, Lexington, USA) platform. This imputation analysis was developed by the Animal Genomics and Improvement Laboratory (Agriculture Research Service, USDA) under the responsibility of Dr. George Wiggans, who used the findhap.f90 program (VanRaden, 2015).

After imputation analysis, eight animals with duplicate genotypes or incompatible pedigree were excluded, resulting in 1,149 animals with information of 79,294 markers. Quality control (QC) was applied to reduce genotyping errors and the computational effort of the analyses. In this control, 7,899 markers with call-rate lower than 90%; 32 monomorphic SNPs; and 1,241 with minor allele frequency (MAF) lower than 0.02 were excluded. No marker was excluded by the frequency of heterozygotes criterion (value greater than 0.15 for Hardy-Weinberg equilibrium). Markers on sex chromosomes and 56 animals that presented a conflict between parents and progenies were excluded. The QC was applied individually for each trait using the PREGSF90 program (Aguilar et al., 2014). Thus, approximately 1,096 animals and 70,193 SNPs remained.

2.7 GWAS adapted for heterozygosity

We conducted a Genome-wide Association Study (GWAS) with a unique focus on exploring how non-additive genetic variations in molecular markers relate to the observed traits (as introduced by Tsairidou et al., 2018). This phase of our study complements the simulation phase and aims to uncover potential genomic regions and specific molecular markers that exhibit a heterozygous advantage.

In the traditional additive model, we assume a linear relationship between the number of copies of a specific allele and the phenotype. In this model, we code the SNPs as 0 (for individuals with two identical alleles, A1A1 or A2A2), 1 (for individuals with one of each allele, A1A2 or A2A1), and 2 (for individuals with two different alleles, A1A2).

In contrast, our study also considers a 'heterozygote (dis)advantage' model, where we code homozygous genotypes (A1A1 or A2A2) as having the same effect, while heterozygotes (A1A2 or A2A1) are assigned to different effects. This unique coding approach helps us explore the potential impact of heterozygosity on the phenotype, as suggested by Tsairidou et al. (2018) and Vidotti et al. (2019). In this model, we code SNPs as 0 for homozygotes and 1 for heterozygotes.

The genomic model additive (M1) and additive-dominant (M2) was adjusted for GWAS, as shown below:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{Sc} + \boldsymbol{\varepsilon} \qquad \text{(M1)}$$

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Za} + \boldsymbol{Wd} + \boldsymbol{Sc} + \boldsymbol{\varepsilon} \qquad \text{(M2)}$$

where $\boldsymbol{y}$ is the phenotypic record; $\boldsymbol{X}$ is the incidence matrix for the fixed effects of contemporary group (CG); $\boldsymbol{\beta}$ is the vector of fixed effects; $\boldsymbol{Z}$ is the incidence matrix for the random additive genetic effect; $\boldsymbol{a}$ is the vector of solutions for animal random effect; $\boldsymbol{W}$ is the incidence matrix for the random dominance genetic effect; $\boldsymbol{d}$ is the vector of solutions for dominance random effect; $\boldsymbol{S}$ is the incidence matrix for the random effect of the permanent environment; $\boldsymbol{c}$ is the vector of solutions for the random effect of the permanent environment; $\boldsymbol{\varepsilon}$ is the vector for the residual effects.

Specifically for the additive-dominant model (M2), the incidence matrix *W* was constructed to directly reflect this 'heterozygote (dis)advantage' parameterization. The elements of the genotype matrix used to build the genomic dominance relationship matrix (or to compute the SNP dominance effects) were populated with 1 for heterozygous genotypes (A1A2 or A2A1) and 0 for homozygous genotypes (A1A1 or A2A2). This incidence structure isolates the non-additive deviation specifically associated with the heterozygous state in the vector *d*.

The programs BLUPF90 and POSTGSF90 were used to obtain the solution matrices of the GEBVs, the *P*-value of the SNPs, and the proportion of genetic variance explained by the SNPs (Misztal et al., 2009; Aguilar et al., 2014). The proportion of variance explained by *n* SNP was calculated based on window segments of 40 SNPs adjacent to -log10 (*P*-value), rather than being individual (Tsairidou et al., 2018). From the estimates of the effects for the SNP, the explained variance for the SNPs was estimated according to Wang et al. (2014):

$$\frac{Var\ (a_i)}{\sigma_a^2}\ x\ 100 = \frac{Var\ (\sum_{j=1}^{40} Z_j \hat{u}_j)}{\sigma_a^2}\ x\ 100$$

where a_i is the genetic value of each genomic region that consisted of 40 adjacent SNPs; σ_a^2 the total additive genetic variance; Z_j the vector that contains the SNP genotype for all individuals; $\hat{u}_j$ the effect of each marker within the stipulated genomic region.

A weighted single step GWAS (WssGWAS) was used iteratively three times, and a vector of ones (w) was used as the initial weight SNP. During the three iterations of the WssGWAS, the total additive genetic variance (σ_a^2) used in the denominator to calculate the proportion of variance explained was updated dynamically in each iteration based on the sum of the weighted SNP variances.

This methodology allows unequal variables between markers for better accuracy in estimating the effects of SNPs (Wang et al., 2012). Therefore, if the population structure is uninformative or when the number of individuals genotyped is small for a polygenic trait, single-step GWAS can outperform traditional approaches (Marques et al., 2018). Furthermore, the use of weight can minimize the identification of false positives associated with the phenotype (Habier et al., 2011). This same method has been adopted in species such as dairy cattle (Tiezzi et al., 2015), pigs (Marques et al., 2018), chickens (Fragomeni et al., 2014), and sheep (Rovadoscki et al., 2018). The *p* values were obtained based on the (co)variance matrix prediction error for the SNPs:

$$pval_i = 2\left(1 - \phi\left(\left|\frac{\widehat{a_i}}{sd(\widehat{a_i})}\right|\right)\right)$$

where ϕ is a standard cumulative normal function; $sd(\widehat{a_i})$ is the square root of the prediction error variance (PEV) for the *j-th* SNP effect obtained from the inverse of the coefficient matrix of the mixed model equations.

Once the variance for the SNPs was calculated, the regions that explain at least 1% of the additive genetic variance were recorded. The chromosome and genome-wide significance thresholds were set to -log10 ($1e^{-5}$) and -log10 ($5e^{-8}$), respectively. These thresholds were adopted as established empirical standards to stringently control false-positive associations in complex traits, following the methodology applied in dairy cattle by Tsairidou et al. (2018).

Additionally, the annotation of the genes and the obtaining of the Gene Ontology terms (GO Terms) were performed using the BioMart platform (https://www.ensembl.org/biomart) for SNPs significantly associated with the phenotype. The reference genome version was ARS-UCD1.2 for *Bos Taurus*, last updated in 2018 (https://www.ncbi.nlm.nih.gov/assembly/GCA_002263795.2). Genes with functional relevance for milk yield and solid constituents were considered important.

2.8 Estimation of heterozygosity factors

In our analysis, we calculated heterozygosity factors based on the information from 50,000 molecular markers, drawing on the research of Tsairidou et al. (2018) and Liu et al. (2018). To do this, we selected 500 Quantitative Trait Nucleotides (QTNs) and computed the associated dominance deviations for individuals with heterozygous genotypes. For instance, if a particular genomic locus had a dominance effect of 0.50 or -0.50, the heterozygous genotype for that locus was assigned the same value for each individual (i) across all k QTNs considered.

We referred to the sum of k QTN dominance deviations for each individual (i) as 'genomic heterosis.' Genomic loci exhibiting negative dominance deviations were regarded as non-advantages of heterozygosity. This led us to believe that to enhance and maximize heterosis, it is preferable to select individuals displaying positive genomic heterosis alongside heterozygosity.

Heterozygosity was calculated by dividing the total count of heterozygous genotypes for each individual by the total number of QTNs assessed, in line with the findings of Cao et al. (2019). Additionally, the degree of heterozygosity was achieved by summing the dominance effects of genomic loci and dividing it by the total number of QTNs considered, based on the studies of Iversen et al. (2019) and Tambasco-Talhari et al. (2005).

To gain a better understanding of the effects of heterozygosity and the occurrence of genomic heterosis, we conducted univariate linear regression analyses on heterozygosity factors for each replicate within the three simulated genetic scenarios. For this analysis, each replicate was treated as a distinct population. We further assessed the estimated regression coefficients by dividing them by the standard deviation of each parameter to ascertain the significance of heterozygosity in genomic heterosis and the phenotype, as suggested by Iversen et al. (2019). In the case of real data, we estimated heterozygosity for the regions identified through our adapted Genome-wide Association Study (GWAS) and performed a linear regression for the phenotype after adjusting for fixed effects (M1 and M2), considering the influence of molecular heterozygosity.

3. Results

3.1 Simulated data

In our analysis, we assessed heterozygosity as the proportion of heterozygotes for each individual, which is calculated by summing the number of heterozygous genotypes and dividing it by the total number of marker loci under consideration.

In the simulated populations with a heritability (h^2) value of 0.10, we observed an average heterozygosity rate of 35.6% ± 1.00. For the other scenarios, the heterozygosity rates ranged from 35.4% ± 0.5 to 35.7% ± 0.7. Notably, genomic heterosis was more pronounced in the scenario with high heritability. In this case, the mean genomic heterosis was 17.0% ± 13.9, with a maximum value of 60.1%. For the scenario with $h^2 = 0.30$, the mean genomic heterosis was 11.0% ± 10.8, with a maximum of 49.5%. The scenario with the lowest heritability magnitude exhibited lower rates, where the maximum heterosis was 25.8%, and the population mean was 3.2% ± 6.0.

Table 2 displays the regression coefficients for heterozygosity as it relates to genomic heterosis. Each simulation cycle, or repetition, was treated as a distinct population. Notably, the regression estimates varied among populations within each scenario. However, across these scenarios, the coefficients exhibited a consistent trend in terms of the direction of the regression estimate. The scenario with higher heritability values showed higher estimates compared to scenarios with lower to moderate heritability. In this table, Hetr denotes genomic heterosis; Hetz, heterozygosity; h^2, heritability scenarios; β, estimated coefficient; SD, standard deviation; α, intercept; a, standard deviation of genomic heterosis; b, regression coefficient divided by the standard deviation of genomic heterosis; and * indicates that the regression coefficient was significantly different from zero ($P < 0.05$).

Table 2 - Estimates of regression coefficients for genomic heterosis (Hetr) on heterozygosity (Hetz) for three simulation scenarios.

E (Hetr/Hetz) = $\alpha + \beta \text{Hetz}_i$															
Repetition	$h^2 = 0.10$					$h^2 = 0.30$					$h^2 = 0.50$				
	β	SD	α	$\sigma_{Hetr}a$	$\beta/\sigma_{Hetr}\,b$	β	SD	α	$\sigma_{Hetr}a$	$\beta/\sigma_{Hetr}\,b$	β	SD	α	$\sigma_{Hetr}a$	$\beta/\sigma_{Hetr}\,b$
1	0.20*	0.095	-0.06	0.27	0.71	0.85*	0.178	0.08	0.52	1.64	0.75*	0.223	-0.13	0.64	1.18
2	-0.69*	0.095	0.08	0.27	-2.56	-0.91*	0.165	0.05	0.48	-1.91	-1.58*	0.219	0.25	0.63	-2.53
3	0.06	0.093	0.10	0.27	0.22	-0.26	0.148	0.00	0.43	-0.61	-0.07	0.206	-0.04	0.59	-0.12
4	-0.13	0.106	0.00	0.30	-0.42	-1.22*	0.185	-0.05	0.53	-2.31	-0.43	0.238	0.22	0.68	-0.63
5	-1.03*	0.091	0.03	0.26	-3.97	-2.38*	0.158	0.07	0.46	-5.17	-3.23*	0.196	0.13	0.57	-5.70
6	0.17	0.099	0.03	0.29	0.57	-0.68*	0.180	0.30	0.50	-1.35	-0.14	0.220	0.05	0.63	-0.23
7	1.34*	0.097	-0.02	0.28	4.80	1.69*	0.162	0.05	0.48	3.53	2.20*	0.217	0.03	0.62	3.54
8	0.59*	0.097	0.15	0.28	2.09	0.43*	0.168	-0.07	0.48	0.89	1.32*	0.218	0.03	0.63	2.09
9	0.97*	0.096	0.12	0.28	3.47	2.05*	0.175	-0.03	0.50	4.13	2.22*	0.222	0.10	0.65	3.43
10	0.26*	0.086	0.08	0.25	1.03	0.48*	0.165	0.03	0.46	1.04	0.81*	0.198	0.09	0.58	1.41
11	0.52*	0.091	-0.11	0.26	1.99	1.45*	0.173	-0.27	0.50	2.91	0.56*	0.202	-0.07	0.60	0.95
12	0.01	0.087	0.10	0.25	0.02	-0.80*	0.167	0.15	0.48	-1.67	0.60*	0.194	-0.18	0.54	1.10
13	-0.04	0.096	-0.12	0.28	-0.15	0.75*	0.162	-0.11	0.47	1.61	-1.59*	0.060	0.09	0.72	-2.21
14	0.55*	0.091	0.04	0.26	2.10	-0.57*	0.181	0.18	0.51	-1.11	1.94*	0.224	-0.04	0.65	3.01
15	0.05	0.091	0.00	0.26	0.18	-2.85*	0.159	0.08	0.46	-6.19	-1.24*	0.220	-0.11	0.62	-1.99
16	-0.22*	0.091	0.06	0.27	-0.80	0.69*	0.166	-0.12	0.48	1.43	1.00*	0.217	-0.17	0.62	1.63
17	-0.98*	0.084	0.07	0.24	-4.09	1.63*	0.172	0.02	0.50	3.25	-0.43	0.227	0.21	0.65	-0.66
18	0.10	0.089	-0.12	0.25	0.38	0.84*	0.162	0.03	0.46	1.82	-2.51*	0.204	-0.14	0.59	-4.28
19	-0.82*	0.097	-0.04	0.29	-2.99	3.04*	0.173	-0.08	0.50	6.07	0.33	0.220	-0.07	0.64	0.52
20	0.14	0.088	-0.10	0.25	0.55	1.05*	0.151	-0.03	0.43	2.43	2.92*	0.219	0.14	0.62	4.69

Source: Authors.

For the scenario I ($h^2 = 0.10$), the regression estimates were in a favorable direction for most populations, that is, it is expected that the change in heterozygosity unit can positively change genomic heterosis. In scenario I, some substantial estimates were found. For example, in population 5 is expected proportional reduction of 1% in heterosis, if heterozygosity increases 1%. However, for population 7 the increased heterozygosity is estimated to increase by 1.34% genomic heterosis. The estimates were also positive for 8, 9, 11 and 14 populations (0.59, 0.97, 0.52 and 0.55, respectively), and population 1, but with a lower value estimate (0.20). In the other populations, the estimates were negative or not different from zero ($P > 0.05$).

In scenarios II and III (moderate and high heritability), more favorable and high estimates were found, compared to scenario I. In scenario II, the highest estimated regression coefficient was 3.04 (population 19), followed by 2.05 for population 9, 1.68 in 7, 1.63 in 17, 1.45 and 1.05 for 11 and 20 inhabitants, respectively. For scenario III, the highest estimate was 2.92 in population 20, followed by 2.22, 2.20, and 1.00 for populations 9, 7 and 16, respectively. For the three scenarios, a reversal in the direction of some estimates was noted. For example, for population 12, the estimates were -0.80 for scenario II and 0.60 for scenario III, both with a significant regression coefficient.

When dividing the regression coefficient estimates by the standard deviation of heterozygosity, it is possible to get an indication of the importance of heterozygosity (Table 2). We found greater effect of heterozygosity under genomic heterosis for many populations: ***a)*** 7 (4.80), 9 (3.47), 14 (2.10), 8 (2.09), 11 (1.99) and 10 (1.03) belonging to scenario I; ***b)*** 19 (6.07), 9 (4.13), 7 (3.53), 17 (3.25), 11 (2.91), 20 (2.43), 1 (1.64), 13 (1.61), 16 (1.43), 10 (1.04) for scenario II; ***c)*** 20 (4.69), 7 (3.54), 3 (3.43), 14 (3.01), 8 (2.09), 16 (1.63), 10 (1.41), 12 (1.10) and 1 (1.18) for scenario III. For populations not mentioned, there was no advantage of heterozygosity and for those populations with a positive effect, the estimate of the regression coefficient was not significant.

3.2 Adapted GWAS analysis

There is no specific reason and no consistent report regarding the size of the SNPs window to be considered in GWAS. Therefore, the additive and additive-dominant models were fitted considering a moving window of 40 adjacent SNPs, to explain the additive genetic variance of the phenotype. At this stage, only real milk yield and quality traits were evaluated. Many genomic regions were similar between models, but with some minor differences in total percentage.

3.2.1 Milk yield (MY)

A total of 15 regions were above the 1% threshold in the additive model, located on chromosomes 1, 3, 7, 9, 11, 12, 16, 20, and 29 (Table 3), having the greatest impact on the phenotype (22.77% of the additive genetic variance). When accounting for the dominance effects, 10 windows explained 13.92%. The greatest number of regions above the variation limit and with higher percentages of variance were recorded when the additive model was used. However, several of these markers were also identified in the additive-dominant model. Interestingly, on chromosome 11 the window 47378387_48847011 Mbp had a greater proportion of explained variance in the additive model (2.36%), compared to the additive-dominant model (1.68%).

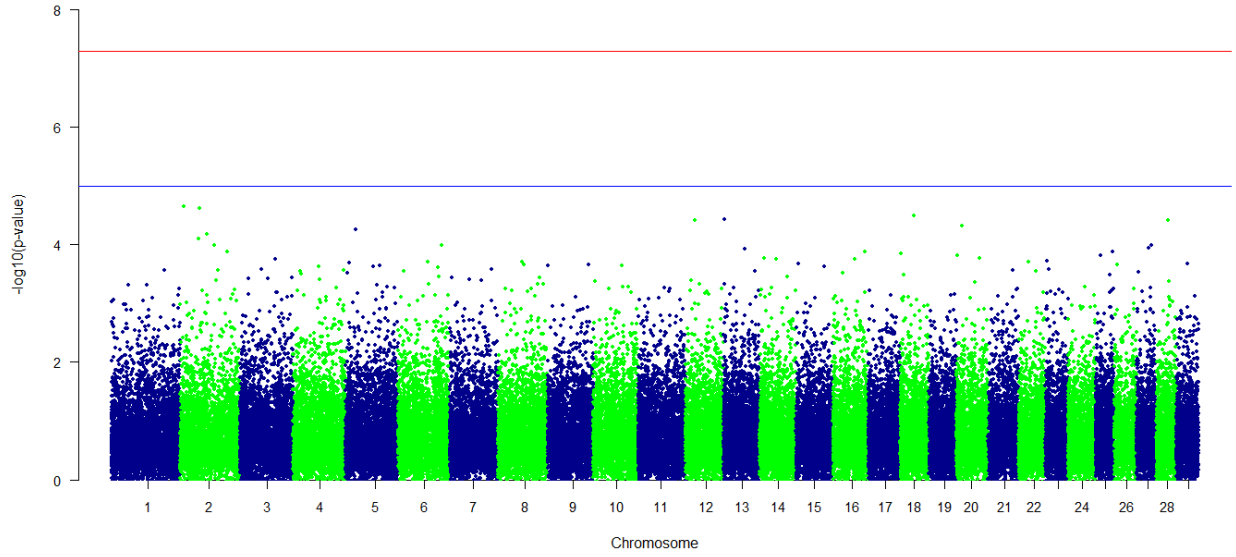
No SNP reached the significance limit by the additive model (Figure. 2). However, including the dominance effects SNPs BTA-30857-no-rs at position 76081859 Mbp on chromosome 12, SNP BovineHD1300023864 on chromosome 13 (position 81532044 Mbp), and SNP BovineHD1900002699 on chromosome 19 (9334455 Mpb) were significant at the chromosome-wide level with -log10 $P = 5.49$, $P = 5.08$ e $P = 5.52$, respectively.

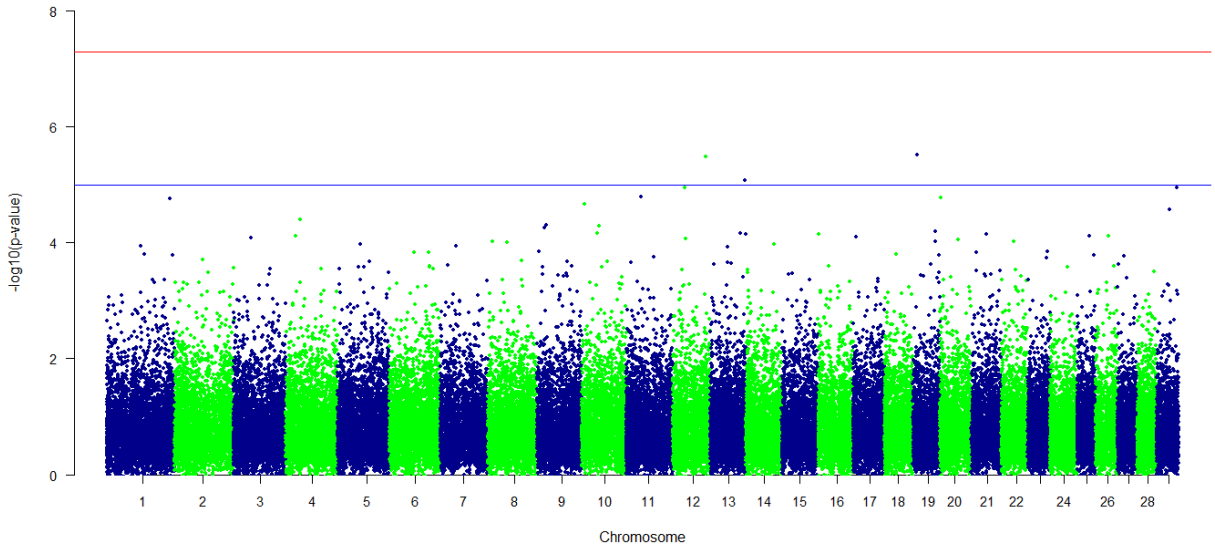
Table 3 - Chromosome, start and final position of the window (CHR:POSstart:POSend, Mpb), and sum variance explained by 40 adjacent SNP ($\%\sigma_a^2$) for additive and additive-dominant genomic model for milk yield (kg/day).

Additive model		Additive-dominant model	
$CHR:POS_{start}:POS_{end}$	$\%\sigma_a^2$	$CHR:POS_{start}:POS_{end}$	$\%\sigma_a^2$
11:47378387:48847011	2.36	11:47378387:48847011	1.89
29:40988761:42330369	2.17	16:25152110:26870109	1.62
3:30484376:31575300	1.76	29:40988761:42330369	1.61
9:69124337:70477400	1.70	9:69124337:70477400	1.33
11:80160161:81840635	1.66	12:15767379:16896392	1.32
7:24494025:25725796	1.59	3:30484376:31575300	1.30
12:50850657:52114901	1.53	7:24494025:25725796	1.25
12:15767379:16896392	1.50	16:19740175:21187584	1.24
16:25152110:26870109	1.46	12:50850657:52114901	1.23
16:19740175:21187584	1.35	11:80182206:81943182	1.13
1:10724714:13177821	1.33	-	-
20:24156101:26083947	1.11	-	-
11:22485728:23678911	1.10	-	-
25:36560015:37411428	1.09	-	-
9:64464857:65955264	1.06	-	-
Total	**22.77**		**13.92**

Source: Authors.

Figure 2 - Manhattan plot of the association of SNPs with milk yield (kg/day). Significance threshold for chromosome-wide (blue line) and genome-wide (red line) for additive (above) and additive-dominant (below) genomic model.





Source: Authors.

3.2.2 Fat percentage (FP)

Small differences between models were found for the fat percentage in milk (Table 4). The main regions explained approximately 37.87% of the additive genetic variance and were all are located on chromosomes 3, 5, 6, 7, 9, 13, 14, 15, and 17, with the position window 146715 to 1681783 Mbp (chr 14) having the greatest impact on genetic variance (26.23% and 26.28% for the additive and additive-dominant model, respectively).

Table 4 - Chromosome, start and final position of the window (CHR:POSstart:POSend, Mpb), and sum variance explained by 40 adjacent SNP ($\%\sigma_a^2$) for additive and additive-dominant genomic model for fat percentage (%).

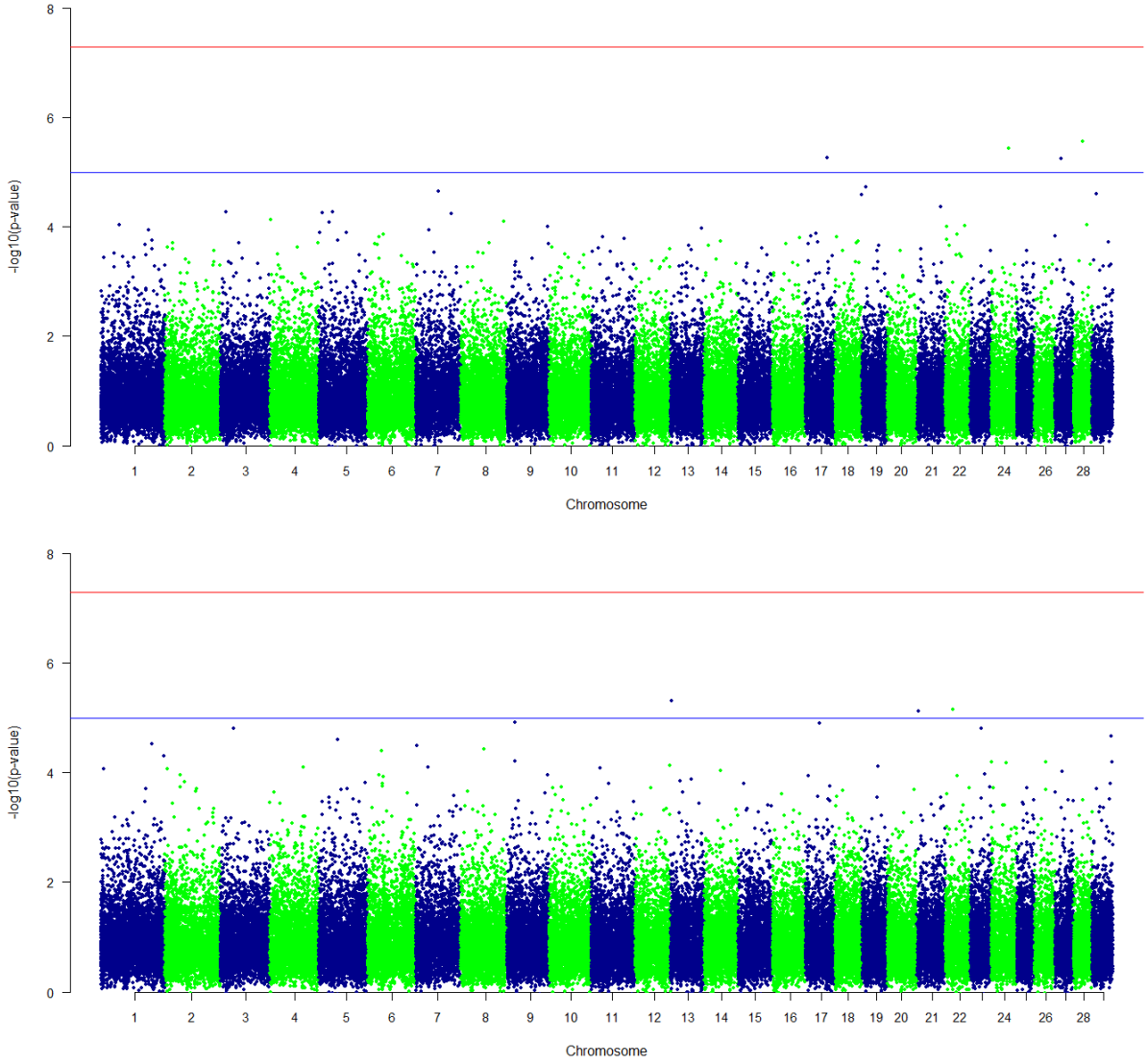
Additive model		Additive-dominant model	
$CHR:POS_{start}:POS_{end}$	$\%\sigma_a^2$	$CHR:POS_{start}:POS_{end}$	$\%\sigma_a^2$
14:146715:1681783	26.23	14:146715:1681783	26.28
3:37209678:39904161	2.25	3:37209678:39904161	2.26
5:92860844:93719732	1.44	5:92860844:93719732	1.46
17:49224250:50522879	1.42	17:49224250:50522879	1.40
7:16632028:17996900	1.20	7:16632028:17996900	1.20
15:72976813:74332306	1.12	15:72976813:74332306	1.11
14:2840340:3819126	1.09	14:2840340:3819126	1.08
13:66163790:67177128	1.07	13:66163790:67177128	1.06
9:68260591:69659376	1.02	9:68260591:69659376	1.02
6:84000396:85422786	1.01	6:84000396:85422786	1.01
Total	**37.85**		**37.87**

Source: Authors.

Significantly different SNPs associated at the chromosome-wide level were recorded among the models and no marker reached the genome-wide significance threshold (Figure 3). In the additive model, the associated markers are SNP ARS-BFGL-NGS-57149, Hapmap30536-BTA-137818, BovineHD2700003888, and BTB-00981253; located on chromosomes 17 (position

54062683 Mbp), 24 (position 42090544 bp), 27 (position 14315992 Mbp) and 28 (position 22219573 Mbp), respectively. For the additive-dominant model, the markers are located on chromosomes 13 position 2545062 Mbp (SNP BovineHD4100009853), 21 position 2498610 Mbp (SNP BovineHD2100000332), and 22 position 19093982 Mbp (SNP BTB-00839323).

Figure 3 - Manhattan plot of the association of SNPs with fat percentage (%). Significance threshold for chromosome-wide (blue line) and genome-wide (red line) for additive (above) and additive-dominant (below) genomic model.



Source: Authors.

3.2.3 Protein percentage (PP)

In both models, eight regions had the greatest impact on additive genetic variance for milk protein concentration, located on chromosomes 1, 7, 9, 15, 19, 22, 27, and 28 (Table 5). For the additive model, the proportion was 13.92%, about 2.03% higher compared to the additive-dominant model (total of 11.89%). Eight markers were significantly associated with milk protein concentration, including dominance effects in the model (Figure 4). The markers are SNP ARS-BFGL-NGS-85942 (chr 3, position 32529922 Mbp); BovineHD0500011755 (chr 5, position 40936260 Mbp); ARS-BFGL-NGS-117043 (chr 6, position 13415776 Mbp); BTB-00359112 (chr 8, position 75072384 Mbp); BTB-00402806 (chr 9, position 85181556 Mbp); RS - USMARC-Parent-DQ837646-rs29012894 (chr 11, position 1714947 Mbp); ARS-BFGL-NGS-25084 (chr 13, position 81126899

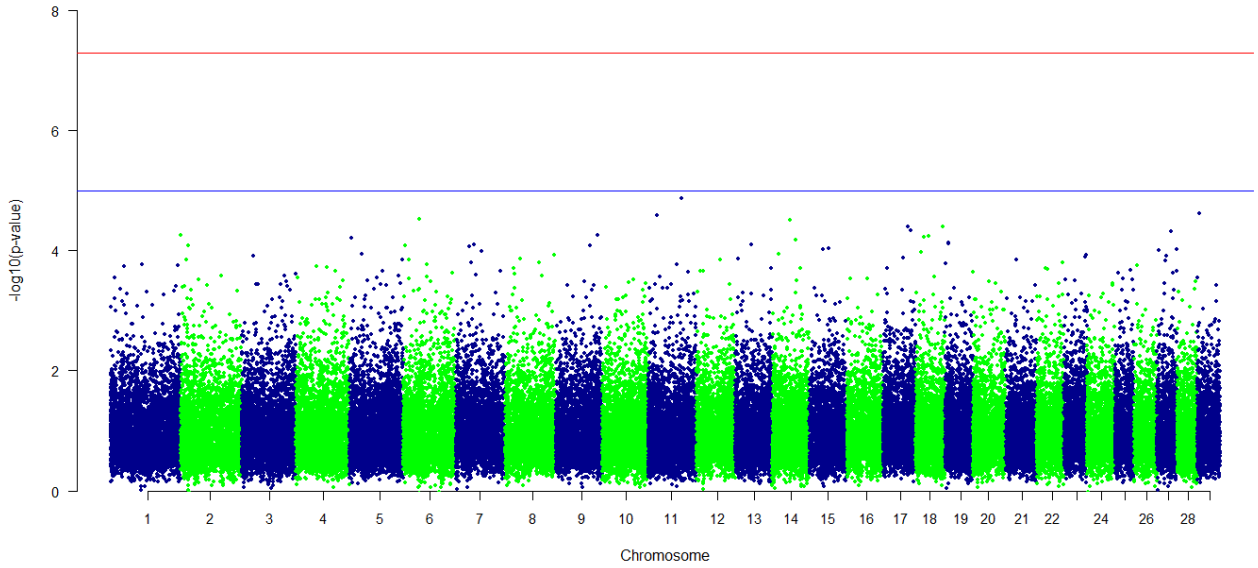
Mbp); and SNP BovineHD1400008879 (chr 14, heading 29039917 Mbp). None of the markers were associated with the phenotype by the additive model.

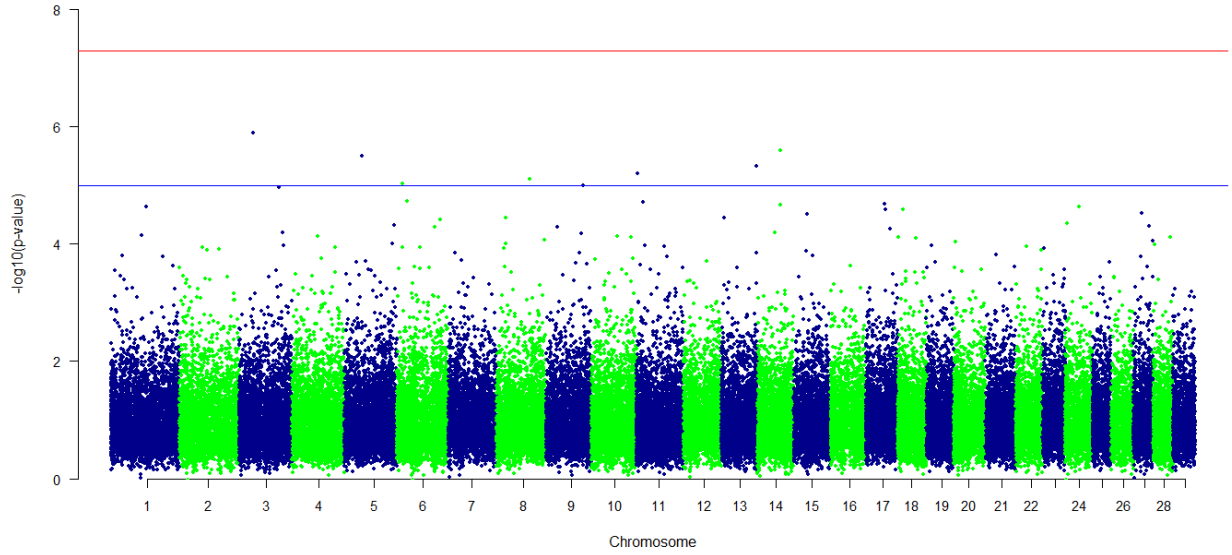
Table 5 - Chromosome, start and final position of the window (CHR:POSstart:POSend, Mpb), and sum variance explained by 40 adjacent SNP ($\%\sigma_a^2$) for additive and additive-dominant genomic model for protein percentage (%).

Additive model		Additive-dominant model	
CHR:POS_{start}:POS_{end}	$\%\sigma_a^2$	CHR:POS_{start}:POS_{end}	$\%\sigma_a^2$
22:40916767:42054278	3.12	22:40916767:42054278	3.12
1:71190486:73143183	2.60	1:71190486:73143183	2.60
15:35856607:37078463	2.01	15:35856607:37078463	2.02
9:6450799:8405821	1.65	9:6450799:8405821	1.64
19:57724102:59090292	1.30	19:57724102:59090292	1.30
28:42346190:43243315	1.17	28:42346190:43243315	1.17
7:92574720:94403482	1.06	7:92574720:94403482	1.05
27:11736207:13523197	1.01	27:11736207:13523197	1.01
Total	**13.92**		**13.91**

Source: Authors.

Figure 4 - Manhattan plot of the association of SNPs with protein percentage (%). Significance threshold for chromosome-wide (blue line) and genome-wide (red line) for additive (above) and additive-dominant (below) genomic model.





Source: Authors.

3.2.4 Casein percentage (CP)

For the concentration of casein in the milk, 17 windows explained 28.3% of the additive genetic variance and are located on chromosomes 1, 3, 4, 6, 8, 9, 11, 15, 19, 22, 25, and 27. The regions are identical between the models (Table 6).

Three significant SNPs at the chromosome-wide level were identified by the additive model and two by the additive-dominant model (Figure 5). For the first model, the markers were SNP 18714 (Hapmap50621-BTA-21320) on chromosome 6 (position 62773878 Mbp), SNP 35910 (Hapmap42878-BTA-23796) on chromosome 12 (position 49425612 Mbp) and SNP 65117 (BovineHD2700009481) located on chromosome 27 (position 33945088 Mbp). By the additive-dominant model, the markers associated with the phenotype are different: SNP Hapmap47398-BTA-74932 located on chromosome 5 (position 105138348 Mbp) and SNP Hapmap48564-BTA-103192 on chromosome 8 (position 24139857 Mbp). No markers were significant at the genome level.

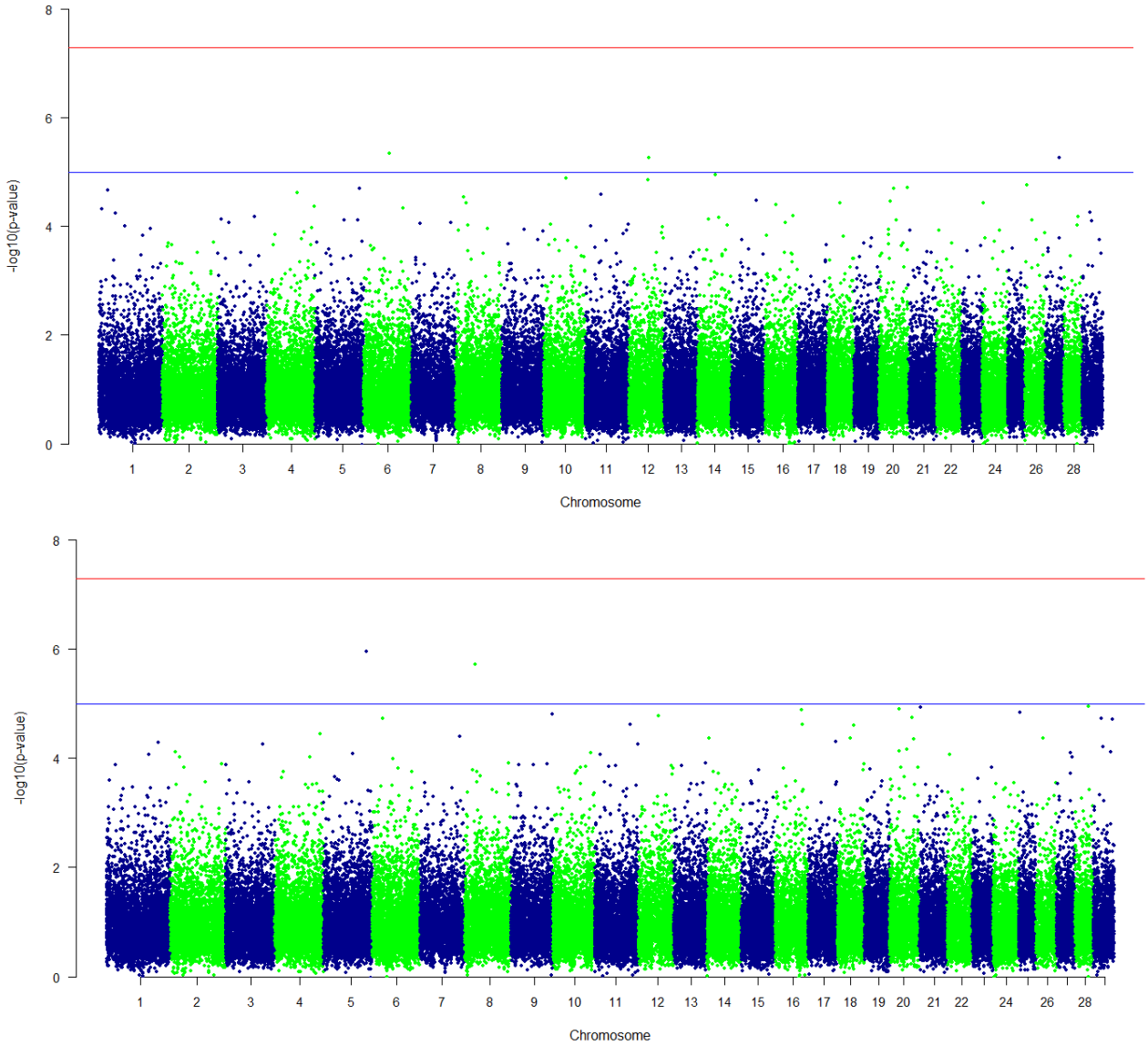
Table 6 - Chromosome, start and final position of the window (CHR:POSstart:POSend, Mpb), and sum variance explained by 40 adjacent SNP ($\%\sigma_a^2$) for additive and additive-dominant genomic model for casein percentage (%).

Additive model		Additive-dominant model	
$CHR:POS_{start}:POS_{end}$	$\%\sigma_a^2$	$CHR:POS_{start}:POS_{end}$	$\%\sigma_a^2$
9:6450799:8405821	3.10	9:6450799:8405821	3.10
1:71190486:73143183	2.94	1:71190486:73143183	2.93
15:35856607:37078463	2.21	15:35856607:37078463	2.20
3:99320849:100987554	1.85	3:99320849:100987554	1.85
8:32526176:34413649	1.72	8:32526176:34413649	1.72
25:22329517:23907681	1.63	25:22329517:23907681	1.63
15:74763309:75726555	1.60	15:74763309:75726555	1.61
8:80869483:81631309	1.58	8:80869483:81631309	1.58
9:60328344:61996687	1.55	9:60328344:61996687	1.55
16:58664360:59498662	1.50	16:58664360:59498662	1.50
22:40936789:42071416	1.49	22:40936789:42071416	1.48

27:11669357:13486135	1.38	27:11669357:13486135	1.38
1:29520198:31171747	1.28	1:29520198:31171747	1.27
6:47353083:49140436	1.25	6:47353083:49140436	1.25
11:16528054:17815368	1.15	11:16528054:17815368	1.14
4:96201713:98497972	1.07	4:96201713:98497972	1.07
19:57724102:59090292	1.04	19:57724102:59090292	1.04
Total	**28.34**		**28.32**

Source: Authors.

Figure 5 - Manhattan plot of the association of SNPs with casein percentage (%). Significance threshold for chromosome-wide (blue line) and genome-wide (red line) for additive (above) and additive-dominant (below) genomic model.



Source: Authors.

3.2.5 Polyunsaturated fatty acid percentage in milk (PUFA)

Absolute values varied considerably between windows and 13 regions in both models (Table 7). When accounting for dominance effects, 18 regions explained about 33.46% of the additive genetic variance, superior to the additive model, in which

17 regions explained 31.29% of the additive genetic variance. The regions with the greatest impact on genetic variance are on chromosome 14 (position 146715 to 1681783 Mbp) and on chromosome 9 (position 68260591 to 69659376 Mbp). These regions also had an important impact on milk fat. For PUFA, these regions explained about 2.23% and 1.56% of the genetic variance in the additive model, and 2.41 and 1.63% when dominance was included in the model.

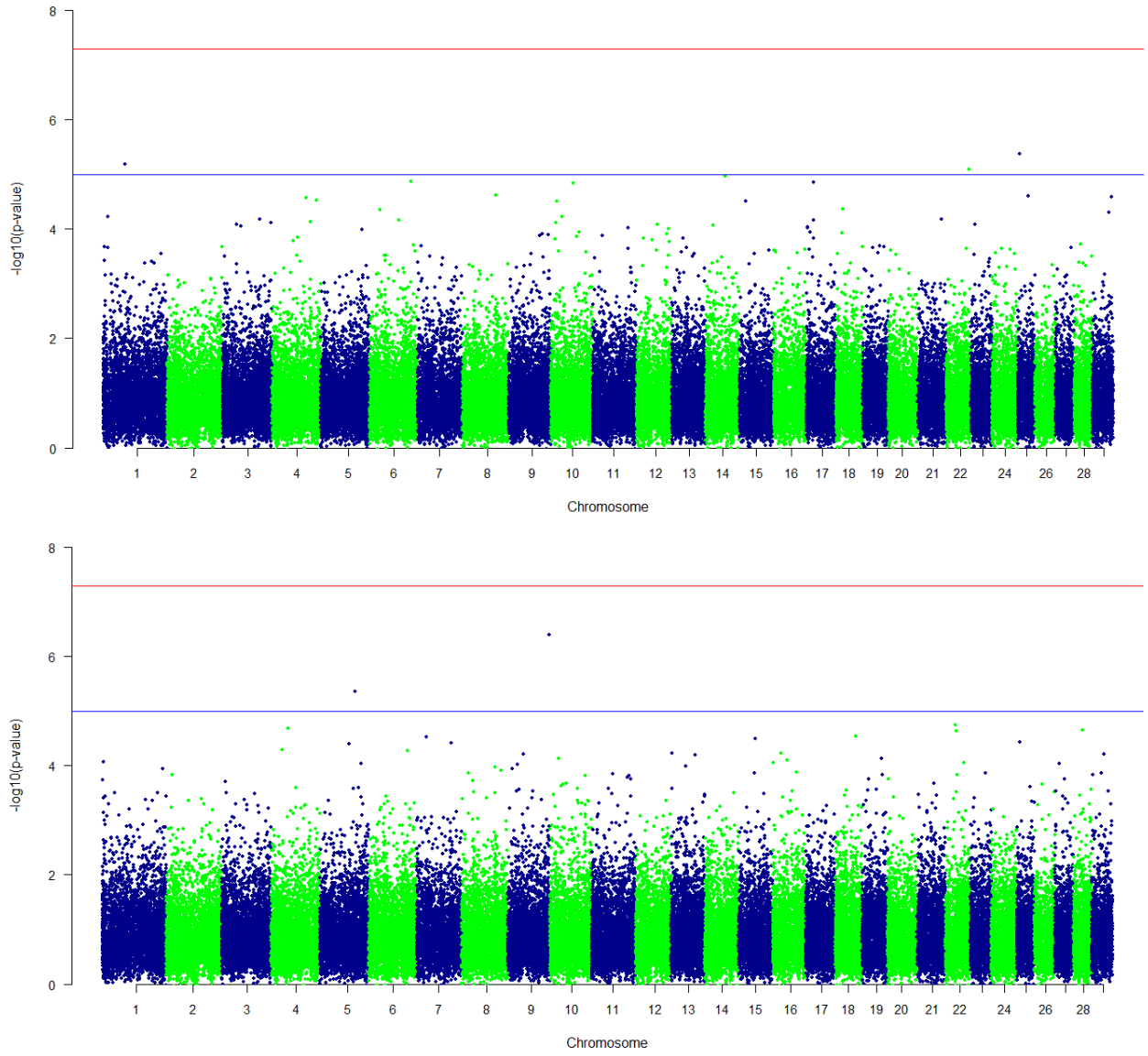
Table 7 - Chromosome, start and final position of the window (CHR:POSstart:POSend, Mpb), and sum variance explained by 40 adjacent SNP ($\%\sigma_a^2$) for additive and additive-dominant genomic model for polyunsaturated fatty acid content in milk (%).

Additive model		Additive-dominant model	
$CHR:POS_{start}:POS_{end}$	$\%\sigma_a^2$	$CHR:POS_{start}:POS_{end}$	$\%\sigma_a^2$
26:22847961:24988516	3.89	26:22467951:24855737	4.07
3:118896461:120546834	3.14	3:118896461:120546834	3.21
28:34912784:36299478	2.53	28:34912784:36299478	2.69
15:80810699:81959353	2.52	15:81025239:82292544	2.46
29:3439848:4457515	2.27	29:3439848:4457515	2.44
14:146715:1681783	2.23	14:146715:1681783	2.41
13:9220838:10403183	1.81	13:9220838:10403183	2.01
19:26735925:28206737	1.57	9:68260591:69659376	1.63
9:68260591:69659376	1.56	19:26735925:28206737	1.61
22:40159906:41633675	1.48	22:40159906:41633675	1.54
20:33569467:34603769	1.41	18:5010603:5936520	1.49
11:39194600:41327004	1.35	20:33569467:34603769	1.45
18:5010603:5936520	1.29	11:39194600:41327004	1.24
17:53928009:55324818	1.09	18:9749447:10827317	1.11
18:9749447:10827317	1.09	17:53928009:55324818	1.05
3:48432235:49898786	1.05	3:48950219:50608425	1.04
11:42618575:43853640	1.01	13:48540388:50502256	1.01
-	-	5:65378999:66734512	1.00
Total	**31.29**		**33.46**

Source: Authors.

Five SNPs were found associated with the phenotype across the chromosome-wide, three by the additive model and two by the additive-dominant model (Figure 6). The SNPs ARS-BFGL-NGS-43783 being located on chromosome 1 (position 54347232 Mbp); BovineHD2200015974 on chromosome 22 (position 55192606 Mpb); and SNP BovineHD2500001118 located on chromosome 25 (position 4800735 Mbp), by additive model. When accounting for dominance effects only the SNPs Hapmap36729-SCAFFOLD106566_3245 on chromosome 5 (position 86395589 Mbp) and ARS-BFGL-NGS-19753 on chromosome 9 (position 102980032 Mbp), were associated with the phenotype.

Figure 6 - Manhattan plot of the association of SNPs with polyunsaturated fatty acid content in milk (%). Significance threshold for chromosome-wide (blue line) and genome-wide (red line) for additive (above) and additive-dominant (below) genomic model.



Source: Authors.

3.3 Associated genes

Eight genes were annotated in regions adjacent to the markers associated with milk yield, fat, protein, casein, and polyunsaturated fatty acid. The identified genes play a role in cell structure and coding for different proteins, as *RNF43*, *GRM7*, *CD53*, *SLC2A13*, *NFX1*, *NPHP*1, and *TACC1* gene. None of them have a direct biological function with the phenotype. However, for polyunsaturated fatty acid, the *SOX5* located on chromosome 5, position 86395589 Mpb (SNP: Hapmap36729-SCAFFOLD106566_3245), responds to the stimulation of transforming growth factor-beta, which is important for human health. Also, the candidate gene Diacylglycerol acyl-CoA acyltransferase (*DGAT1*) activity was found in the 14: 603,813-612,791 regions (Ensembl: BTAG00000026356). This gene plays a fundamental role in the metabolism of cellular triacylglycerol during physiological processes, such as intestinal fat absorption, lipoprotein assembly, fat tissue formation and lactation.

3.4 Exploring genomic regions for heterozygosity

We calculated the heterozygosity level of the genomic locus as the proportion of heterozygous genotypes of each selected QTL, divided by the total of selected QTLs (Table 8). For fat, approximately 50.5% of mean heterozygosity was found in 10 QTLs with the greatest impact on genetic variance. Casein concentration presented 50.4% heterozygosity in 17 QTLs, followed by PUFA with 50.3% for 22 QTLS, protein with 50.1% for eight QTLs. The lesser heterozygosity was observed for milk production, with 49.4% for 16 QTLs.

Table 8 - Number of observations, number of genomic regions that explained more than 1% of the additive genetic variance, minimum (MIN), maximum (MAX) and mean values for the heterozygosity factors calculated at the genomic locus and individually, for milk yield (MY, kg/day), fat percentage (FP, %), protein percentage (PP, %), casein percentage (CP, %) and polyunsaturated fatty acid content in milk (PUFA, %).

Trait	Number of observations	Number of genomic regions	MIN	MAX	Mean
Heterozygosity in locus					
MY (kg/day)	1092	16	0.467	0.550	0.494
FP (%)	1092	10	0.439	0.585	0.505
PP (%)	1092	8	0.486	0.541	0.501
CP (%)	951	17	0.445	0.583	0.504
PUFA (%)	1089	22	0.470	0.568	0.503
Individual heterozygosity					
MY (kg/d)	1092	16	0	0.875	0.377
FP (%)	1092	10	0	1	0.369
PP (%)	1092	8	0	1	0.379
CP (%)	951	17	0	0.882	0.353
PUFA (%)	1089	22	0	0.864	0.373

Source: Authors.

Furthermore, the mean individual heterozygosity was 0 to 1, for fat and protein percentages (up to 100% heterozygosity). For the other traits, the variation was from 0 to 0.875 (0 to 87.5%). The highest proportion of heterozygous genotypes was for protein (37.9%), milk production (37.7%), fatty acid (37.3%), fat (36.9%), and casein (35.3%). The mean levels of heterozygosity at the genomic locus and individual were uniform among traits and relatively high. This may indicate the possible heterozygous advantage for phenotypes.

Estimates of the regression coefficients between the adjusted phenotype and the level of individual heterozygosity (Table 9) were favorable for MY (1.9642 ± 0.3376), protein (0.0026 ± 0.0092), and PUFA (0.0059 ± 0.0021). However, only MY and PUFA were significant (P-value < 0.05). For FP and CP, the coefficients were unfavorable and not significant.

There was a predicted increase of 1.96 kg/day of milk yield when the heterozygosity increased by one unit. This should also happen with PUFA, but to a lesser degree (0.0059% more PUFA in milk). When we divide the estimated regression coefficient by the standard deviation of the phenotype, it is observed that the level of heterozygosity had a greater impact on the volume of milk production and the content of polyunsaturated fatty acids. For the percentage of protein, heterozygosity was also important, but not significant (Table 9).

Table 9 - Regression coefficients for heterozygosity (Hetz) as a function of phenotype adjusted (y*) for milk yield (MY, kg/day), fat percentage (FP, %), protein percentage (PP, %), casein percentage (CP, %), and polyunsaturated fatty acid content in milk (PUFA, %).

Trait	$E\,(y^*/\,\text{Hetz}) = \alpha + \beta\text{Hetz}_i$				
	β	SE	α	σ_{y^*} [1]	β/σ_{y^*} [2]
MY (kg/day)	1,9642*	0,3376	31,3916	7,7663	0,2530
FP (%)	-0,0048	0,0265	3,4958	0,6922	-0,0007
PP (%)	0,0026	0,0092	3,1269	0,2521	0,0103
CP (%)	-0,0094	0,0115	2,3875	0,2280	-0,0412
PUFA (%)	0,0059*	0,0021	0,1559	0,0404	0,1460

Hetz: heterozygosity; β: estimated coefficient; SE: Standard error; α: intercept; [1]Standard deviation of adjusted phenotype (σ_{y^*}[1]); [2] Regression coefficient divided by the standard deviation of adjusted phenotype; * Regression coefficient significantly different from zero ($P < 0.05$).

Source: Authors.

4. Discussion

We proposed to explore molecular heterozygosity as a possible criterion in the selection of animals in purebred populations to generate genomic heterosis. Initially, data simulation was used to assess the relationships among heterozygosity factors in different genetic scenarios. Soresen et al. (2008) defines the degree of heterosis of a breed by the chance that a specific genomic locus will have genes from different breeds used for crossing. In addition, breed heterozygosity would be calculated as the proportion of the genetic or heterozygous contribution of the parental lines. Therefore, the heterosis of the F1 population would be due to dominance deviations. Thus, we calculated genomic heterosis due to simulated dominance deviations and found favorable and significant coefficient estimates regardless of the scenario.

In some populations, it became clear that increased heterozygosity could lead to an increase in genomic heterosis. In these cases, it is possible that a group of individuals has regions of the genome with greater heterozygote effect and, therefore, prone to greater genomic heterosis. Furthermore, we observed higher coefficients of regression for the lowest heritability scenarios, and this may indicate that heterozygosity and heterosis should be factors with a greater impact on complex traits (Kargo et al., 2021; Saborío-Montero et al., 2018; Clasen et al., 2017).

While our simulation provided robust theoretical insights into the behavior of heterozygosity factors, it is pertinent to contextualize the magnitude of the simulated results. The dominance deviations were simulated using a standard deviation of 0.50 ($d \sim N(0, 0.50)$), which deliberately creates a genetic architecture leaning towards overdominance to ensure the non-additive effects were fully captured by our models. Consequently, the maximum simulated genomic heterosis observed (e.g., 60.1% in the high heritability scenario) represents an upper bound and might be an overestimation compared to typical biological variations in dairy cattle, where dominance variance generally accounts for a smaller fraction of the phenotypic variance. Nonetheless, this parameterization effectively serves as a proof of concept for selecting towards a heterozygous advantage.

Transitioning from the simulated environment — where genomic heterosis is explicitly known — to the empirical data required by estimating these effects indirectly. Therefore, the GWAS analysis was adapted for heterozygous effects to identify regions of the genome with non-additive variations. Through a linear regression analysis of the phenotype and estimated heterozygosity in some genomic regions, favorable and significant regression coefficients were found for milk production and polyunsaturated fatty acid content, where it is expected that the increase in heterozygosity can increase the production levels. Previous studies have shown advantages of including dominance and heterozygosity effects in mating programs in different

animal species (Alillo et al., 2017; Gutiérrez-Gil et al., 2014; Bagheri et al., 2013; Tambasco-Talhari et al., 2005). We found average heterozygosity between 0.353 and 0.377, as reported by Keller et al (2017) in Holstein cows.

The molecular basis of heterosis would be strongly correlated with genomic locus heterozygosity and with dominance and overdominance effects (Tambasco-Talhari et al., 2005; Xiao et al., 1995). When we included the genomic relationship matrix for dominance effects in the GWAS model, some extra regions were identified, rather than just an additive model. It is possible that the set of heterozygous QTLs with greater dominance effect contributed to the favorable estimates. In the study by Deng & Fu (1998) it is pointed out that the associations between heterozygosity and phenotype are not due to the dominance of a single locus since homozygous genotypes for the dominant allele could share the same phenotype for heterozygotes.

Thus, if multiple loci were involved in the increased heterozygous expression, there would be a correlation between heterozygosity and phenotype (Tambasco-Talhari et al., 2005). Our results suggest considering a group of QTLs as a basis for understanding genomic heterosis, therefore, selection based on genomic regions with superior heterozygosity should be appropriate for mattings between purebred individuals. There are no studies directly comparable to the current one, but our results indicate the use of molecular heterozygosity as a parameter associated with genomic heterosis to improve the productive performance of purebred populations.

In addition, molecular markers associated with milk production and quality traits were identified. A higher number of SNPs was associated with the phenotype with a non-additive model compared to an additive model. Gene annotation analysis was performed only for the genomic regions of these SNPs and most of them are associated with protein-coding, solute transport, hormone receptors, and cell membrane constituents.

Moreover, for polyunsaturated fatty acids, the *SOX5* gene is important for cartilage development and extension, resulting from the stimulation of transforming growth factor-beta. In mammals, especially in dairy cattle, when this growth factor is present in milk, it stimulates the growth of cells, especially in the connective tissue, in addition to participating in the formation of bones and cartilage, being the focus of studies aimed at human health (Gauthier et al., 2006; Silva et al., 2016). Factors closely linked to the health of the consumer are the main motivations in the assessment of fatty acid content in milk. Additionally, this is the first *SOX5* record for polyunsaturated fatty acids in dairy cattle, in a region with a possible heterozygous advantage.

5. Conclusion

The heterozygosity evaluated empirically and through real data revealed optimistic results regarding the occurrence of genomic heterosis in the purebred population. Considering the selection of individuals through information about genomic loci for heterozygosity should favor genomic heterosis. Furthermore, the additive genetic variance explained by the marker windows was similar or superior between milk yield and quality traits, using the additive-dominant model in the adapted GWAS. Therefore, including dominance effects in the adapted GWAS model can contribute to the identification of markers associated with the phenotype, in regions with possible heterozygous advantage. Our results suggest the continuation of studies on molecular heterozygosity and genomic heterosis as an alternative to purebred selection for genetic improvement.

Acknowledgements

The authors thank the “Milk Clinic” and the Department of Animal Science, from Superior School of Agriculture “Luiz de Queiroz” (University of São Paulo, Piracicaba – SP, Brazil) for their support and for providing the database, and the Cooperative Dairy DNA Repository (CDDR) for providing the genotypes used as reference population in imputation analysis.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Declaration of funding

This research received the financial support from The São Paulo Research Foundation (FAPESP, 2019/03373-9), National Council for Scientific and Technological Development (CNPq), and Coordination for the Improvement of Higher Education Personnel (CAPES).

References

Abdollahi-Arpanahi, R., Gianola, D., & Peñagaricano, F. (2020). Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genetics Selection Evolution*, 52, 1-15. https://doi.org/10.1186/s12711-020-00531-z

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, 93, 743-752. https://doi.org/10.3168/jds.2009-2730

Aguilar, I. (2014). *Short introduction to BLUPF90 family programs*. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=slides_2014_lva.pdf

Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., & Misztal, I. (2019). Frequentist p-values for large-scale-single step Genomic-wide association, with an application to birth weight in American Angus cattle. *Genetics Selection Evolution*, 51, 1-8. https://doi.org/10.1186/s12711-019-0469-3

Alillo, H., Pryce, J. E., González-Recio, O., Cocks, B. G., Goddard, M. E., & Hayes, B. J. (2017). Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. *Journal of Dairy Science*, 100, 1203-1222. https://doi.org/10.3168/jds.2016-11261

Amadeu, R. R., Cellon, C., Lara, L., Resende, M., Oliveira, I., Ferrao, L., Muñoz, P., & Garcia, A. (2021). *AGHmatrix: Relationship Matrices for Diploid and Autopolyploid Species*. https://github.com/prmunoz/AGHmatrix

Bagheri, M., Miraie-Ashtiani, R., Moradi-Shahrbabak, M., Nejati-Javaremi, A., Pakdel, A., von Borstel, U. U., Pimentel, E. C. G., & Konig, S. (2013). Selective genotyping and logistic regression analyses to identify favorable SNP-genotypes for clinical mastitis and production traits in Holstein dairy cattle. *Livestock Science*, 151, 140-151. https://doi.org/10.1016/j.livsci.2012.11.018

Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., & Veitia, R. A. (2010). Heterosis. *The Plant Cell*, 22(7), 2105-2112. https://doi.org/10.1105/tpc.110.076133

Cao, Z., Gu, Y., Yang, Q., He, Y., Fetouh, M. I., Warner, R. M., & Deng, Z. (2019). Genome-wide identification of quantitative trait loci for important plant and flower traits in petunia using a high-density linkage map and an interspecific recombinant inbred population derived from Petunia integrifolia and P. axillaris. *Horticulture Research*, 6, 1-15. https://doi.org/10.1038/s41438-018-0091-5

Chein, F. (2019). Introdução aos modelos de regressão linear. Brasilia. ENAP. ISBN: 978-85-256-0115-5. https://bibliotecadigital.enap.gov.br/bitstream/1/4788/1/Livro_Regress%C3%A3o%20Linear.pdf

Christensen, O. F., & Lund, M. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, 42, 1-8. https://doi.org/10.1186/1297-9686-42-2

Clasen, J. B., Norberg, E., Madsen, P., Pederson, J., & Kargo, M. (2017). Estimation of genetic parameters and heterosis for longevity in crossbred Danish dairy cattle. *Journal of Dairy Science*, 100, 6337-6342. https://doi.org/10.3168/jds.2017-12627

Deng, H. W., & Fu, Y. X. (1998). Conditions for positive and negative correlations between fitness and heterozygosity in equilibrium populations. *Genetics*, 148, 1333-1340. https://doi.org/10.1093/genetics/148.3.1333

Fragomeni, B. O., Misztal, I., Lourenco, D. L., Aguilar, I., Okimoto, R., & Muir, W. M. (2014). Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. *Frontiers in Genetics*, 1, 1-7. https://doi.org/10.3389/fgene.2014.00332

Gauthier, S. F., Pouliot, S. F., & Maoubois, J. L. (2006). Growth factors from bovine milk and colostrum: composition, extraction and biological activities. *Le Lait*, 86, 99-125. https://doi.org/10.1051/lait:2005048

Gutiérrez-Gil, B., Arranz, J. J., Pong-Wong, R., Garcia-Gámez, E., Kijas, J., & Wiener, P. (2014). Application of selection mapping to identify genomic associated with dairy production in sheep. *PLoS ONE*, 9, 1-13. https://doi.org/10.1371/journal.pone.0094623

Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension for the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12, 1-12. https://doi.org/10.1186/1471-2105-12-186

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS ONE*, 4, 1-10. https://doi.org/10.1371/journal.pgen.1000008

ICAR – International Committee for Animal Recording. (2017). *Section 2: Cattle milk recording*. https://www.icar.org/Guidelines/02-Overview-Cattle-Milk-Recording.pdf

Iung, L. H. S., Petrini, J., Ramírez-Díaz, J., Salvian, M., Rovadoscki, G. A., Pilonetto, F., Dauria, B. D., Machado, P. F., Coutinho, L. L., Wiggans, G. R., & Mourão, G. B. (2019). Genome-wide association study for milk production traits in a Brazilian Holstein population. *Journal of Dairy Science*, 102, 5305-5314. https://doi.org/10.3168/jds.2018-14811

Kargo, M., Clasen, J. B., Nielsen, H. M., Byskov, K., & Norberg, E. (2021). Short communication: Heterosis and breed effects for milk production and udder health traits in crosses between Danish Holstein, Danish Red, and Danish Jersey. *Journal of Dairy Science*, 104, 678-682. https://doi.org/10.3168/jds.2019-17866

Kelleher, M. M., Berry, D. R., Keraney, J. F., McParland, S., Buckley, F., & Purifield, D. C. (2017). Inference of population structure of purebred dairy and beef cattle using high-density genotype data. *Animal*, 11, 15-23. https://doi.org/10.1017/S1751731116001099

Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, 92, 4656-4663.

Marques, D. B. D., Bastiaansen, J. W. M., Broekhuijse, M. L. W. J., Lopes, M. S., Knol, E. F., Harlizius, B., Guimarães, S. E. F., Silva, F. F., & Lopes, P. S. (2018). Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. *Genetics Selection Evolution*, 50, 1-14. https://doi.org/10.1186/s12711-018-0412-z

Masuda, Y. (2021). *Introduction to BLUPF90 suit programs Standard edition*. University of Georgia. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=tutorial_blupf90.pdf

Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, 92, 4648-4655. https://doi.org/10.3168/jds.2009-2064

Misztal, I., Tsuruta, S., Lourenco, D., & Masuda, Y. (2018). *Manual for BLUPF90 family of programs*. University of Georgia, Athens, USA. http://nce.ads.uga.edu/wiki/doku.php?id=start

Nicolini, P., Amorín, R., Han, Y., & Peñagarciano, F. (2018). Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. *BMC Genetics*, 19, 1-18. https://doi.org/10.1186/s12863-018-0600-4

Petrini, J., Iung, L. H. S., Rodriguez, M. A. P., Salvian, M., Pértille, F., Rovadoscki, G. A., Cassoli, L. D., Coutinho, L. L., Machado, P. F., Wiggans, G. R., & Mourão, G. B. (2016). Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. *Animal Breeding and Genetics*, 133, 384-395. https://doi.org/10.1111/jbg.12205

Pereira, A. S. et al. (2018). Metodologia da pesquisa científica. [free ebook]. Santa Maria: Editora da UFSM.

Perez Lopez, C. (2022). Técnicas de análise preditiva de dados. Modelos de regressão linear múltipla. Editora Scientific Books. ISBN-13: 978-1446702123

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/

Risemberg, R. I. C. et al. (2026). A importância da metodologia científica no desenvolvimento de artigos científicos. E-Acadêmica Journal, 7(1), e0171675. https://eacademica.org/eacademica/article/view/675

Rodriguez, M. A. P., Petrini, J., Ferreira, E. M., Mourão, L. R. M. B., Salvian, M., Cassoli, L. D., Pires, A. V., Machado, P. F., & Mourão, G. B. (2014). Concordance analysis between estimation methods of milk fatty acids content. *Food Chemistry*, 156, 170–175. https://doi.org/10.1016/j.foodchem.2014.01.092

Rovadoscki, G. A., Pertile, S. F. N., Alvarenga, A. B., Cesar, A. S. M., Pértille, F., Petrini, J., Franzo, V., Soares, W. V. B., Morota, G., Spangler, M. L., Pinto, L. F. B., Carvalho, G. G. P., Lanna, D. P. D., Coutinho, L. L., & Mourão, G. B. (2018). Estimates of Genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics*, 19, 1-14. https://doi.org/10.1186/s12864-018-4777-8

Saborío-Montero, A., Vargas-Leitón, B., Romero-Zúñiga, J. J., & Camacho-Sandoval, J. (2018). Additive genetic and heterosis effects for milk fever in a population of Jersey, Holstein x Jersey, and Holstein cattle under grazing conditions. *Journal of Dairy Science*, 101, 9128-9134. https://doi.org/10.3168/jds.2017-142234

Silva, F. L., Pinto, M. S., Carvalho, A. F., & Perrone, I. T. (2016). Fator de crescimento transformador beta (TGF-β). *Revista do Instituto de Laticínios Cândido Tostes*, 70, 226-238. https://doi.org/10.14295/2238-6416.v70i4.384

Shitsuka, R. et al. (2014). Matemática fundamental para tecnologia. (2nd Ed.). Editora Érica.

Sorensen, M. K., Norberg, E., Pedersen, J., & Christensen, L. G. (2008). Invited Review: Crossbreeding in Dairy Cattle: A Danish Perspective. *Journal of Dairy Science*, 91, 4116-4128. https://doi.org/10.3168/jds.2008-1273

Tambasco-Talhari, D., Alencar, M. M., Paz, C. C. P., Cruz, G. M., Rodrigues, A. A., Packer, I. U., Coutinho, L. L., & Regitano, L. C. A. (2005). Molecular marker heterozygosities and genetic distances as correlates of production traits in F1 bovine crosses. *Genetics and Molecular Biology*, 28, 218-224. https://doi.org/10.1590/S1415-47572005000200007

Tiezzi, F., Parker-Gaddis, K. L., Cole, J. B., Clay, J. S., & Maltecca, C. (2015). A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS ONE*, 10, 1-15. https://doi.org/10.1371/journal.pone.0114919

Tsairidou, S., Allen, A. R., Pong-Wong, R., McBride, S. H., Wright, D. M., Matika, O., Pooley, C. M., McDowell, S. W. J., Glass, E. J., Skuce, R. A., Bishop, S. C., & Woolliams, J. A. (2018). An analysis of effects of heterozygosity in dairy cattle for bovine tuberculosis resistance. *Animal Genetics*, 49, 103-109. https://doi.org/10.1111/age.12637

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91, 4414-4423. https://doi.org/10.3168/jds.2007-0980

VanRaden, P. M. (2015). *findhap.f90 – Find haplotypes and impute genotypes using multiple chip sets and sequence data*. http://aipl.arsusda.gov/software/findhap

Veitia, R. A. (2010). A generalized model of gene dosage and dominant negative effects in macromolecular complexes. *FASEB Journal*, 24, 994–1002. https://doi.org/10.1096/fj.09-146969

Vidotti, M. S., Lyra, D. H., Morosini, J. S., Granato, I. S. C., Quecine, M. C., Azevedo, J. L., & Fritsche-Neto, R. (2019). Additive and heterozygous (dis)advantages GWAS models reveal candidate genes involved in the genotypic variation of maize hybrids to Azospirillum brasilense. *PLoS ONE*, 14, 1-21. https://doi.org/10.1371/journal.pone.0222788

Vieira, S. (2021). Introdução à bioestatística. Editora GEN/Guanabara Koogan.

Vitezica, Z., Varona, L., & Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics*, 195, 1223-1230. https://doi.org/10.1534/genetics.113.155176

Wang, H., Misztal, I., Aguilar, I., & Legarra, A. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research*, 94, 73-83. https://doi.org/10.1017/S0016672312000274

Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., Okimoto, R., Wing, T., Hawken, R., & Muir, W. M. (2014). Genome-wide association mapping including phenotypes from relatives without genotypes in single-step (ssGWAS) for 6-week body weight in broiler chickens. *Frontiers in Genetics*, 5, 1-10. https://doi.org/10.3389/fgene.2014.00134

Xiao, J., Li, J., Yuan, L., & Tanksley, S. D. (1995). Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics*, 140, 745-754. https://doi.org/10.1093/genetics/140.2.745