UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

USE OF HAPLOTYPES AND INDIVIDUAL SNPS IN GENOMIC STUDIES FOR REPRODUCTIVE TRAITS IN NELORE CATTLE

Andres Chaparro Pinzon

Zootecnista

2019

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

USE OF HAPLOTYPES AND INDIVIDUAL SNPS IN GENOMIC STUDIES FOR REPRODUCTIVE TRAITS IN NELORE CATTLE

Andres Chaparro Pinzon

Orientadora: Profa. Dra. Lucia Galvão de Albuquerque

Corientadora: Dra. Ana Fabrícia Braga Magalhães

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Campus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Genética e Melhoramento Animal

2019

P661u

Pinzon, Andres Chaparro
Use of haplotypes and SNPs in genomic selection and genomic-wide association studies for reproductive traits in Nelore cattle / Andres Chaparro Pinzon. -- Jaboticabal, 2019
58 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientadora: Lucia Galvão de Albuquerque
Coorientadora: Ana Fabricia Braga Magalhães

1. age at first calving. 2. Beef cattle. 3. Bos indicus. 4. genomic breeding values. 5. scrotal circumference. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: USE OF HAPLOTYPES AND SNPS IN GENOMIC SELECTION AND GENOMIC-WIDE ASSOCIATION STUDIES FOR REPRODUCTIVE TRAITS IN NELORE CATTLE

AUTOR: ANDRES CHAPARRO PINZON
ORIENTADORA: LUCIA GALVÃO DE ALBUQUERQUE
COORIENTADORA: ANA FABRÍCIA BRAGA MAGALHÃES

Aprovado como parte das exigências para obtenção do Título de Doutor em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:

Pesquisadora Dra. ANA FABRÍCIA BRAGA MAGALHÃES
Instituto de Zootecnia / Sertãozinho/SP

Prof. Dr. HENRIQUE NUNES DE OLIVEIRA
Departamento de Zootecnia / FCAV / Unesp - Jaboticabal

Pesquisador Dr. ROBERTO CARVALHEIRO
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Pesquisadora Dra. LENIRA EL FARO ZADRA
Instituto de Zootecnia / Sertãozinho/SP

Pós-doutorando RAFAEL ESPIGOLAN
FZEA/USP / Pirassununga/SP

Jaboticabal, 30 de abril de 2019

DADOS CURRICULARES DO AUTOR

Andrés Chaparro Pinzón, nasceu no dia 3 de janeiro de 1986, em Bucaramanga, Estado de Santander – Colômbia, filho de Leonor Pinzón Martínez e Carlos Chaparro Becerra. Iniciou a graduação em junho de 2005 e obteve o título de Zootecnista pela "Universidad Industrial de Santander" em 2011. Em agosto de 2013, ingressou no Programa de Pós-graduação em Zootecnia (área de concentração: produção e melhoramento animal) da Universidade Federal do Ceará (UFC), sob orientação do Prof. Dr. Raimundo Nonato Braga Lôbo, obtendo o título de Mestre em Zootecnia em julho de 2015. Em agosto de 2015 iniciou o doutorado no Programa de Pós-graduação em Genética e Melhoramento Animal da Universidade Estadual "Júlio de Mesquita Filho", campus de Jaboticabal (UNESP – FCAV), sob orientação da Prof. Dra. Lucia Galvão de Albuquerque.

“If you can't fly, run. If you can't run, walk. If you can't walk, crawl, but by all means, keep moving”

Martin Luther King Jr.

Aos meus pais **LEONOR PINZÓN MARTÍNEZ e CARLOS CHAPARRO BECERRA**,

in memoriam, pessoas que amo e admiro.

(Dedico)

À minha família. Obrigado por sempre acreditarem e confiarem em mim. **OFEREÇO**

AGRADECIMENTOS

A Deus, pela vida e por ter me concedido uma família maravilhosa, e pelos anjos que guiam meu caminho.

Aos meus pais, que caminham comigo e me iluminam.

À minha família, por ter me dado amor, força e apoio em todo momento, e principalmente, por acreditar em mim.

À minha orientadora, Profa. Dra. Lucia Galvão de Albuquerque, pela oportunidade, e ensinamentos que contribuíram para o desenvolvimento deste trabalho.

À minha coorientadora Dra. Ana Fabrícia Braga Magalhães pela paciência, apoio, confiança, ensinamentos e pela disposição para desenvolver este trabalho e me ajudar a crescer intelectualmente.

Aos membros da banca do Exame Geral de Qualificação, Prof. Dr. Henrique Olivera Nunes, Dr. Roberto Carvalheiro, Dra. Larissa Simielli e Dra. Sirlene Lazaro, pelas sugestões que contribuíram e acrescentaram a este trabalho.

À Mariele pela paciência, amor e força. Por me ajudar a enxergar que sim era possível.

A todos os docentes do Programa de Genética e Melhoramento Animal pelos ensinamentos.

Ao Dr. Raimundo Nonato Braga Lôbo, pela amizade e apoio durante todos estes anos no Brasil.

Aos meus companheiros e colegas da “salinha” Ana Fabrícia, Ana Cristina, Anderson, André M, André V, Andreia, Bianca, Camila, Daiane, Dani Beraldo, Diego, Diércles, Diogo, Elisa, Fabielli, Gabriela, Gerardo, Giovana, Gordo, Henrique, Juan Diego, Laiza, Larissa, Lucas, Lúcio, Marisa, Malane, Medeiros, Natalia, Patricia,

Salvatore, Samuel, Thaise, Tiago, Tonussi e William pela amizade durante estes anos do Doutorado.

Aos amigos de Jaboticabal, pessoal da UNATI, colombianos, mexicanos peruanos e todos aqueles que me acompanharam nesta caminhada.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos.

Muito obrigado!

SUMÁRIO

USO DE HAPLOTIPOS E SNPS INDIVIDUAIS EM ESTUDOS GENÔMICOS PARA CARACTERÍSTICAS REPRODUTIVAS EM BOVINOS DA RAÇA NELORE

Resumo – As características reprodutivas são fundamentais para a rentabilidade do sistema produtivo de gado de corte. Contudo, características como idade ao primeiro parto (IPP) e perímetro escrotal (PE) possuem desvantagens para serem utilizadas em programas de melhoramento genético tradicional, uma vez que são mensuradas em só um sexo e podem apresentar baixa herdabilidade. Nas últimas décadas os avanços nas tecnologias de análise de DNA têm permitido o desenvolvimento de procedimentos estatísticos como estudos de associação (GWAS) e seleção genômica (GS). Comumente têm-se utilizado marcadores de tipo SNP para desenvolver este tipo de estudos. Entretanto, outro tipo de marcador molecular os haplótipos, que são grupos de SNPs que estão em alto desequilíbrio de ligação (DL) podem ser utilizados neste tipo de estudo, uma vez que estes podem estar em maior desequilíbrio de ligação com os QTL quando comparados com os SNPs. O objetivo deste estudo foi verificar o uso de haplótipos em estudos de GWAS e GS, para características reprodutivas, em bovinos da raça Nelore. Para o estudo de GWAS, 2.390 observações de IPP e 4.832 informações para CE, provenientes de três programas de melhoramento da raça Nelore, foram utilizadas em um estudo de associação genômica ampla. 1900 fêmeas e 1500 machos jovens foram genotipados com o painel HD da Ilumina® (777K). 490 fêmeas e 3.332 machos jovens foram genotipados utilizando o painel da GeneSeek 75K. Os animais genotipados com painel de menor densidade foram imputados para HD utilizando o programa FImpute. O fenótipo foi ajustado para os efeitos fixos (Y*). Os haplótipos foram construídos por médio do software HaploView utilizando desequilíbrio de ligação (D') e a metodologia de intervalo de confiança. Para o GWAS foi utilizado um modelo misto univariado no software GEMMA. Vinte e um blocos significativos de haplótipos ao longo de 12 diferentes cromossomos bovinos foram observados. Dez blocos localizados nos cromossomos 5, 9, 11, 13, 19, 21, 22 e 25 foram associados a genes que têm efeitos putativos nos sistemas reprodutivos em mamíferos. Os genes *FAF1, HDAC4, ARHGEF13, CCDC62, SLC25A39, FKBP6* e *NSUN5* foram associados com CE e estão envolvidos em funções reprodutivas como espermatogênese, maturação e mobilidade dos espermatozoides. Os resultados encontrados no presente estudo podem ser uteis para identificar mutações causais e genes candidatos que influenciam as características IPP e CE. No estudo de seleção genômica, fenótipos e genótipos de 2.390 e 4.832 para IPP (h^2 =0,08) e circunferência escrotal (CE, h^2 =0,42), respectivamente, foram analisados. A metodologia para construir os haplótipos foi a mesma descrita no capítulo 2. Foi realizada validação cruzada com 5 grupos e o método GBLUP foi usado para predição do valor genômico usando haplótipos e SNPs como preditores. A acurácia e o viés foram empregados para comparar a habilidade de predição dos haplótipos e SNPs. A acurácia de predição para IPP foi de 0,16 para SNPs e haplótipos e para CE foi de 0,21 para SNPs e 0,22 para haplótipos. O coeficiente de regressão para IPP foi < 1 e para CE foi > 1, indicando que a predição genômica pode estar subestimada para IPP e superestimada para CE. Valores similares de coeficientes de regressão foram encontrados para ambas

as características quando comparados haplótipos e SNPs. O coeficiente de herdabilidade afetou as acurácias de predição, causando um pequeno aumento usando haplótipos em CE (0.42) e igual acurácia entre os preditores para IPP (0.08)

Palavras chave: *Bos indicus*, Circunferência escrotal, Gado de corte, Idade ao primeiro parto, Valores genéticos genômicos.

USE OF HAPLOTYPES AND INDIVIDUAL SNPS IN GENOMIC STUDIES FOR REPRODUCTIVE TRAITS IN NELORE CATTLE

Sumary – The reproductive traits are fundamental for the profitability of the beef cattle production system profitability. However, traits such as age at firs calving (AFC) and scrotal circumference (SC) have disadvantages to be used in traditional breeding programs, since they are measured in only one sex and may have low heritability. In the last few decades, the advances in technology for DNA analysis allowed the development of statistical techniques as genomic-wide (GWAS) and genomic selection studies (GS). Commonly, SNPs markers have been used to perform these studies. Nonetheless, another molecular marker, the haplotype, that are SNPs in high linkage disequilibrium (LD), could be used in these studies, since these could be in higher LD with the QTLs when compared to the individual SNPs. The aim of this study was to verify the use of haplotypes in GWAS and GS, for reproductive traits, in Nelore cattle. In GWAS study, 2,390 and 4,832 animals with information of age at first calving (AFC) and scrotal circumference (SC) belonging to three Nelore breeding programs were used to perform the GWAS. The genotypes of 1900 heifers and 1500 young bulls were obtained with HD panel from Ilumina® (777K) and 490 heifers and 3332 young bulls were genotyped with the GeneSeek Genomic Profiler HDi 75K. Animals genotyped with lower density panel were imputed to HD using the FImpute program. Phenotype was adjusted for the contemporary groups fixed effects (Y*). Missing genotypes and linkage phase were inferred with the Fimpute software. Haplotypes were constructed in the software HaploView using the linkage disequilibrium measurement D' and the confidence interval methodology. The GWAS was performed using univariate mixed model in the software GEMMA. Twenty-one significant haplotype blocks along 12 different chromosomes were associated with AFC. Ten blocks located in the chromosomes 5, 9, 11, 13, 19, 21, 22, and 25 were associated with genes that have putative effects on reproductive systems in mammals. The genes *FAF1, HDAC4, ARHGEF13, CCDC62, SLC25A39, FKBP6* and *NSUN5* were associated with SC and are involved in reproductive functions as spermatogenesis and spermatozoa maturation and motility. The results found in this study can be used to identify causal mutations and candidate genes that underline AFC and SC. In chapter 3, phenotypes and genotypes of 2,390 and 4,832 animals were available for AFC (h^2=0.08) and scrotal circumference (SC, h^2=0.42) respectively. The methodology used to impute the missing genotypes, the haplotype phase and construction was descripting in chapter 2. A cross-validation with 5 folds was performed and GBLUP method was used for genomic prediction using haplotypes and SNPs as predictors. The accuracy and bias were used to compare the prediction models. The accuracy of prediction for AFC was 0.16 for SNPs and haplotypes and for SC was 0.21 for SNPs and 0.22 for haplotypes. The regression coefficient for AFC was >1 and for SC was <1, showing that the prediction accuracy was overestimating for AFC and underestimating for SC. Similar regression coefficient values were found for both traits when compared haplotypes and SNPs. The heritability coefficient affected the prediction accuracies, causing slight increase using haplotypes for trait of high heritability and equal accuracy between predictors for trait of low heritability.

Keywords: Age at first calving, Beef cattle, *Bos indicus*, genomic breeding values, scrotal circumference.

CHAPTER 1 – General considerations

1. Introduction

In beef cattle breeding programs, phenotypic and pedigree information are used to estimate the breeding values (EBV) of the animals that are candidates for selection. Commonly, the phenotypes used as selection criteria correspond to growth traits measured at different ages. Nonetheless, beef cattle herds most attempt equilibrium between biotype, production, and reproduction. According to Brumatti et al. (2011), in a bio-economical model study for beef cattle raised in pasture, variables as cow discard, replacements of heifers and fertility are the most important traits for the index calculation. Reproductive traits could be 13 times more relevant from a genetic-economic point of view than growth and carcass traits. Therefore, selection for reproductive traits brings economic return both for the producer and for the production system.

Traits as scrotal circumference (SC) have been included in genetic evaluation programs as an indirect indicator of fertility, since it is an easy measurement trait, with moderate to high heritability. Authors as Forni and Albuquerque (2005), Yokoo et al. (2007), Boligon et al. (2010), Laureano et al. (2011) and Marques et al. (2013) found heritability varying between 0.42 and 0.61, and genetic correlation with important economical traits as final weight (0.57) and age at first calving (-0.23), as reported by Yokoo et al. (2007) and Boligon et al. (2007), respectively. Age at first calving (AFC) has also been included in breeding programs, aiming for the improvement the reproductive index and consequently the profitability of the herds.

Many reproductive traits can only be measured lately, are expressed just in one sex and may have low heritability that limit the genetic progress when using traditional methods of selection. One way of overcoming this problem is using molecular markers in genomic selection (GS) and genome-wide association studies (GWAS).

Meuwissen et al. (2001) proposed the GS that is based on the evaluation of a large number of markers widely distributed throughout the genome, in order to explain part of the genetic variation of economic interest traits. Two steps are required to perform GS. First, estimate the markers effects in a training population

with phenotypic and genotypic data. Second, use the markers effects of training population to predict the genomic estimated breeding value (GEBVs) of animal's candidates to selection that have genotypic data. With the implementation of GS in breeding programs, it is possible to predict the genetic merit of the animals, without the need to collect the phenotype, thus allowing anticipation of the selection process (Resende et al., 2008; Van Eenennaam et al., 2014), reducing the generation interval (Schaeffer, 2006) and providing greater accuracy of prediction of genomic values (GEBVs). The GS can benefit traits that are expressed late in animal life, difficult to measure, that presents low heritability and/or expressed in a single-sex (Muir, 2007; Meuwissen, 2007).

GWAS are used to identify regions of deoxyribonucleic acid (DNA) associated with traits of economic interest. For this, a large number of single nucleotide polymorphisms (SNPs) are used to find possible associations between the markers and the quantitative trait loci (QTL), in order to identify and study the function of possible candidate genes associated with them (Wu et al., 2014).

The use of haplotypes instead of single-nucleotide polymorphism (SNP) markers have been employed for the estimation of GEBVs (Cuyabano et al., 2014; Cuyabano et al., 2015) and for GWAS (Pryce et al., 2010). Haplotypes are groups of SNPs that are in high linkage disequilibrium (LD) (Tempelman, 2015) and present higher LD with QTLs when compared to individual SNPs (Cuyabano et al., 2014). Moreover, genomic data studies considering haplotypes can be used to identify genetic variants and underlying genetic mechanisms of quantitative traits (Wu et al., 2014), since the use of haplotype blocks, at population level, improves the understanding of the nature of the non-linear association between phenotypes and genes. Consequently, this elucidation of the haplotype structure reduces the information of several SNPs consistently in a single marker (Mokry et al., 2014).

The most researches with GS and GWAS have been used SNPs as molecular markers for reproductive traits in Nelore beef cattle (Costa et al., 2015; Irano, 2015; Fortes et al., 2016; Melo et al., 2017; Regatieri et al., 2017; Teixeira et al., 2017). Nevertheless, GS and GWAS using haplotypes for reproductive traits in Nelore cattle are still scarce. Hence, the proposal of this work is to evaluate whether the use of haplotypes is more efficient than SNP for GS and GWAS for reproductive traits in Nelore cattle.

2. General objective

The objective of this study was to verify the feasibility of perform association and genomic selection studies using haplotypes, seeking to identify those that allow to explain a greater part of the genetic variance and to increase the accuracy in the prediction of genomic values for reproductive traits in Nelore breed, generating information that may be used in breeding programs.

2.1. Specific objectives

- Construct haplotypes based on linkage disequilibrium.
- Identify chromosomal regions of large effect associated with reproductive traits, using haplotypes instead of SNPs.
- Predict the animal GEBV using haplotypes and SNPs information for reproductive traits.
- Compare the prediction accuracy of the animals' GEBVs using SNPs and haplotypes.

3. Literature review

3.1. haplotype

Single nucleotide polymorphisms (SNPs) are molecular markers that are found as variations in DNA and occur in a population with a frequency bigger than 1% (Brookes, 1999). SNPs are the most abundant form of molecular markers, reaching at least 5.8 million distributed throughout the bovine genome (Daetwyler et al., 2014). Due to the linkage disequilibrium (LD) between the SNPs and the QTLs, these markers have been widely used in livestock researches to identify QTLs and causative regions that influence the expression of different traits and, estimate the genetic merit of animals that are candidate for selection. However, the LD is not uniform across the entire genome (Ardlie et al., 2002) and it is possible finding regions in strong LD, known as haplotypes, that are separated by recombination hotspots (The International Hapmap Consortium, 2005).

Haplotypes are groups of SNPs presenting high LD (Villumsen et al., 2009; Garrick e Fernando, 2014; Tempelman, 2015), low recombination rate and that are

segregated together, making possible their conservation over the generations (Meuwissen et al., 2001).

Different studies in GS (Meuwissen et al., 2001; Cuyabano et al., 2014; Cuyabano et al., 2015) and GWAS (Pryce et al., 2010; Braz et al., 2019) have used haplotypes, instead of SNPs, as explanatory variables, because the LD between the QTLs and haplotypes are stronger than with SNPs (Hayes, 2013; Cuyabano et al., 2014). Different methodologies have been proposed to construct haplotypes. Authors as Boichard et al. (2012), Hess et al. (2016) and Jonas et al. (2016) constructed haplotypes with a fixed number of SNPs. Mathias et al. (2006) and Braz et al. (2019) used overlapping slide windows, while Gabriel et al. (2002) and Cuyabano et al. (2015) constructed haplotypes based in the LD between SNPs, determining the location of begining and the end of the haplotype.

In the LD methodology, three pairwise LD measures are commonly used to construct the haplotypes: D, r2 (Hill and Robertson, 1968) and D' (Hill, 1981). However, the most common measures are r^2 and D', because both parameters are less dependent of the individual alleles frequencies and, additionally, are standardized between 0 and 1.

According to Cuyabano et al. (2014), loci with low allele frequencies have higher LD with D' than r2, because D' is estimated dividing a pair of alleles by the minimum allele frequency. So, fewer haplotypes will be constructed using D', which means that the number of variables to be estimated is lower and consequently there is a reduction in the computing time of the genomic prediction model (Cuyabano et al., 2014).

3.2. Genomic selection

Traditional animal breeding programs have been based on the best linear unbiased prediction (BLUP) proposed by Henderson (1975), where pedigree and phenotype data are used to predict the EBV of individuals. Nowadays, advancements of high-throughput genotyping tools have allowed genotyping of samples in commercial scale and the use of molecular markers to predict genetic values. The selection using molecular markers information, also known as genomic selection, was introduced by Meuwissen et al. (2001), and it is based on the evaluation of a large number of markers widely distributed throughout the genome,

in order to explain part of the genetic variation of economic interest traits (Dekkers, 2004).

Implementing genomic selection requires first the use of a training population, where the markers effect are estimated using regression procedure in animals with known genotypes and phenotypes. Second, use the estimated effects to predict the GEBV of animals with genotypes in a validation population.

The prediction ability, measured by the accuracy, is defined as the correlation between the predicted and the true breeding values (Legarra et al., 2008). Several elements may influence in the prediction accuracy: population size (Goddard, 2009), LD between the markers and the QTL (Meuwissen et al., 2001), trait architecture (Hayes et al., 2010), molecular marker i.e. SNPs or haplotypes (Cuyabano et al., 2014; Cuyabano et al 2015), haplotype size (Cuyabano et al., 2014), number of the genotyped animals (Vanraden et al., 2009; Daetwyler et al., 2010; Calus, 2011), heritability of the trait (Goddard e Hayes, 2009) and GEBV prediction method (Moser et al., 2009).

The accuracy of GEBV estimated by genomic selection is higher than the estimated by traditional BLUP (Dekkers, 2002; Meuwissen, 2007; Muir, 2007). It occurs especially for traits that are costly and difficult to measure, such as the traits expressed only in one sex and those presenting low heritability. GS has been widely used to anticipate the selection process in dairy cattle, decreasing the generation interval and reducing the selection cost of young candidates (Schaeffer, 2006).

The type of molecular marker used in GS is important, since it influences the accuracy of GEBV. Haplotypes constructed from the combination of two molecular markers flanked a QTL, were used by Meuwissen et al. (2001) to estimate breeding values in a simulated population. Since this first study of GS, different authors have used haplotypes as an option instead SNPs to perform studies of GS either in simulated (Calus et al., 2008; Villumsen et al., 2009; Cuyabano et al., 2014) and in real data (Hayes et al., 2007; Cuyabano et al., 2015).

Studies using haplotypes lead to an improvement in the accuracy (Villumsen et al., 2008; Boichard et al., 2012; Cuyabano et al., 2014 and Cuyabano et al., 2015) because they are in strong LD with QTL (Hayes 2013; Cuyabano et al., 2014) and could detect an epistasis effect between markers located in the same locus (Clark, 2004; Bardel et al., 2005).

Cuyabano et al. (2015) constructed LD haplotypes, considering D' ≥ 0.45 between SNPs and using Bayesian BLUP or a Bayesian mixture model they estimed the SNPs effects, after they ranked the SNPs by largest effect. A set of *n* SNPs with great effect were defined to finally select the haplotypes containing those SNPs. The authors concluded that in GS, the use of 20,000 to 50,000 SNPs with large effect to select the haplotypes allowed to obtain reliabilities equal or higher than individual SNPs.

3.3. Genome-wide association studies

The theory of quantitative genetic is based on the infinitesimal genetic model proposed by Fisher (1918), where is assumed that most of the important traits (quantitative traits) are determined by infinite genes each one with infinitesimal small effects (Goddard e Hayes, 2009). In the last few decades, the advances in technology for DNA analysis and the development of statistical techniques have allowed a better understanding of the genetic architecture of important traits.

Genome-wide association studies (GWAS) search a marker-trait association, resulted from the LD between a molecular marker and the causative mutation that produce variation in a trait (Hayes e Goddard, 2010; Visscher et al., 2012). The first GWAS (Klein et al., 2005) was performed in humans and becoming the starting point to identify the genetic mechanism of complex traits as cancer. Afterward, GWAS studies were extended for livestock areas as animal breeding and genetics (Chan et al., 2009).

The reproductive traits are important in the dairy and beef cattle production system. However, these traits have a complex nature, because are influenced by many genes, expressed late in the production life, are sex-limited and could be difficult and costly to measure. Nevertheless, it is possible to have a better understanding of the biological mechanism that underlies these traits using genomic information by looking for candidate genes or/and QTLs that influence the genetic variability of these traits. According to Taylor et al. (2014), it is interesting to identify regions where the molecular markers explain more than 1% of the additive genetic variance of a trait. Additionally, the identified markers could be used in customized low-density assays, which are cheap and assist in the genetic evaluation process (Snelling et al., 2012).

Several GWAS studies have been published in *Bos taurus taurus* (Minozzi et al., 2013; Hyeong et al., 2014; Nayeri et al., 2016) and *Bos taurus indicus* cattle (Fortes et al., 2012; Camargo et al., 2014; Costa et al., 2015; Irano et al., 2016; Melo et al., 2017). These studies allowed identifying different genes and regions that affect the genetic variability of the reproductive traits. However, different authors showed that when analyzed the same traits in different populations or breeds, the chromosomal regions associated with such traits change (Fortes et al., 2012; Saatchi et al., 2014). Although genes and chromosomal regions have been identified, the genetic variance explained by these has been small (Bolormaa et al., 2011), a fact that can be explained by the complex nature of the traits. Thus, for a successful GWAS, other molecular markers as SNPs windows (Wang et al., 2012) or haplotypes (Pryce et al., 2010), instead individual SNPs, could be used. Studies of GWAS employing haplotypes were performed to identify markers that affect the calf survival in Holstein population (Kipp et al., 2015), prenatal death (Fritz et al., 2013) and milk production and fertility traits (Pryce et al., 2010).

Braz et al. (2019), in a GWAS, have tested overlapping slide windows methodology, proposed by Mathias et al. (2006), to construct the haplotypes and using Genomic Best Linear Unbiased Predictor (GBLUP), to estimate the haplotypes and SNPs effects. They used a population of 1,616 Nelore animals with haplotypes and SNPs as prediction variables and three size windows of 5, 7 and 9 SNPs for detect QTLs and causal mutations for meat tenderness. These authors verified that haplotypes with the largest size can capture a great proportion of the additive genetic variance and that using haplotypes was possible to identify regions that influence the tenderness that could not be identified just with SNP. Oliveira Junior (2017), in Nelore heifers, found 15 haplotypes for heifer pregnancy and 16 to antral folicle harbored important genes that contributed to explain the genetic control of these reproductive traits, demonstrating that the use of haplotypes could improve the GWAS studies. Nevertheless, studies for reproductive traits using haplotypes as molecular markers in Nelore cattle are still few.

4. References

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. **Nature Reviews Genetics** 3: 299

Bardel C, Danjean V, Hugot JP, Darlu P, Génin E (2005) On the use of haplotype phylogeny to detect disease susceptibility loci. **BMC genetics** 6:24.

Boichard D, Guillaume F, Baur A, et al. (2012) Genomic selection in French dairy cattle. **Animal Production Science** 52:115-120.

Boligon AA, Rorato PRN, Albuquerque LG (2007) Correlações genéticas entre medidas de perímetro escrotal e características produtivas e reprodutivas de fêmeas da raça Nelore. **Revista Brasileira de Zootecnia** 36:565-571.

Boligon AA, Silva JAV, Sesana RC, Sesana JC, Junqueira, JB, Albuquerque LG (2010) Estimation of genetic parameters for body weights, scrotal circumference, and testicular volume measured at different ages in Nelore cattle. **Journal of animal science** 88:1215-1219.

Bolormaa S, Porto Neto LR, Zhang YDA (2011) genome-wide association study of meat and carcass traits in Australian cattle **Journal of animal science** 89, 8, 2297–2309.

Braz CU et al. (2019) Sliding window haplotype approaches overcome single SNP analysis limitations in identifying genes for meat tenderness in Nelore cattle. **BMC Genetics** 20:8.

Brookes AJ (1999) The essence of SNPs. **Gene** 234:177-186.

Brumatti RC, Ferraz JBS, Eler JP, Formigonni IB (2011) Desenvolvimento de índice de seleção em gado corte sob o enfoque de um modelo bioeconômico. **Revista archivos de zootecnia** 60:205-213

Calus MPL, De Roos APW, Veerkamp RF (2008). Accuracy of genomic selection using different methods to define haplotypes. **Genetics** 178:553-561

Calus MPL, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. **Genetic Selection Evolution** 43:26

Camargo GMF, Costa RB, Albuquerque LG, Regitano LCD, Baldi F, Tonhati H (2014) Association between JY-1 gene polymorphisms and reproductive traits in beef cattle. **Gene** 533:477-80.

Chan EKF, Hawken R, Reverter A (2009) The combined effect of SNP-marker and phenotype attributes in genome-wide association studies. **Animal Genetics** 40:149–56.

Clark AG (2004). The role of haplotypes in candidate gene studies. **Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society** 27:321-333.

Costa RB, Camargo GM, Diaz ID, Irano N, Dias MM, Carvalheiro R, Boligon AA, Baldi F, Oliveira HN, Tonhati H, Albuquerque LG (2015) Genome-wide association study of reproductive traits in Nelore heifers using Bayesian inference. **Genetics Selection Evolution** 47:67.

Cuyabano BC, Su G, Lund MS (2014) Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. **BMC Genomics** 15: 1171.

Cuyabano BC, Su G, Lund MS. (2015) Selection of haplotype variables from a high-density marker map for genomic prediction. **Genetics Selection Evolution** 47:61.

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. **Genetics** 185:1021-1031.

Daetwyler HD, Capitan A, Pausch H, et al. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. **Nature genetics** 46:858.

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics** 3:22.

Dekkers JCM (2004) Commercial application of marker- and gene-assisted selection in livestock: Strategies. **Journal of animal science** 82:E313-328.

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. **American Journal of Human Genetics** 52: 399–433.

Forni S, Albuquerque LG (2005) Estimates of genetic correlations between days to calving and reproductive and weight traits in Nelore cattle. **Journal of animal science** 83:511-1515.

Fortes MRS, Lehnert SA, Wolcott M, Zhang Y, Bolormaa S, Holroyd R, Hawken RJ, Reverter-Gomez T (2012) Finding genes for economically important traits: Brahman cattle puberty. **Animal Production Science** 52:143-150.

Fortes, MR, Nguyen, LT, Porto Neto, LR, Reverter, A, Moore, SS, Lehnert, SA, Thomas, MG (2016) Polymorphisms and genes associated with puberty in heifers. **Theriogenology** 86:333-339.

Fritz S, Capitan A, Djari A, et al. (2013) Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2 **PLoS One** 8.

Gabriel SB, Schaffner SF, et al. (2002) The structure of haplotype blocks in the human genome. **Science** 296:2225-2229.

Garrick DJ, Fernando, R (2014) Genomic prediction and genome-wide association studies in beef and dairy cattle Pages 474–501 in The Genetics of Cattle 2nd ed D J Garrick and A Ruvinsky, ed CABI, Wallingford, UK

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews: Genetics** 10: 381-391.

Hayes BJ, Chamberlain AJ, McPartlan H, Macleod I, Sethuraman L, Goddard M (2007) Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. **Genetics Research** 89:215-220.

Hayes BJ (2009) Whole Genome Association and Genomic Selection Jaboticabal, SP: FUNEP, 120p (Course Notes).

Hayes B, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. **Genome** 53:876-883.

Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. **PLoS Genetics** 6:e1001139.

Hayes, B (2013) Overview of statistical methods for genome-wide association studies (GWAS) In Genome-wide association studies and genomic prediction (pp 149-169) Humana Press, Totowa, NJ

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.

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. **Biometrics** 31:423–447.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. Genetics Research 38:209-216.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38:226-231.

Hyeong KE, Iqbal A, Kim JJ (2014) A genome wide association study on age at first calving using high density single nucleotide polymorphism chips in Hanwoo (Bos taurus coreanae). **Asian-Australasian journal of animal sciences** 27: 1406.

Irano N, Camargo GMF, et al (2016) Genome-Wide Association Study for Indicator Traits of Sexual Precocity in Nelore Cattle. **PLOS ONE** v. 11, n. 8, p. e0159502.

Jonas D, Ducrocq V, Fouilloux MN, Croiseau P (2016) Alternative haplotype construction methods for genomic evaluation. Journal of Dairy Science 99: 4537–4546.

Kipp S, Segelke D, Reinhardt F, Reents R, Schierenbeck S, Wurmser C, Pausch H, Fries R, Thaller G, Tetens J, Pott J, Piechotta M, Grünberg W (2015) A new Holstein haplotype affecting calf survival. **Interbull Bulletin** 49.

Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, Sangiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement Factor H Polymorphism in Age-Related Macular Degeneration. **Science** 308: 385– 389.

Laureano MMM, Boligon AA, Costa RB, Forni S, Severo JLP, Albuquerque LGD (2011) Estimativas de herdabilidade e tendências genéticas para características de crescimento e reprodutivas em bovinos da raça Nelore: Estimates of heritability and genetic trends for growth and reproduction traits in Nelore cattle. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia** 63:143-152.

Legarra A, Robert-Granie C Manfredi E, Elsen J (2008) Performance of Genomic Selection in Mice. **Genetics** 180: 611–618 2008

Marques EG, Magnabosco CU, Lopes FB, Silva MC (2013) Estimativas de crescimento, carcaça e perímetro escrotal de animais confinamento. **Bioscience Journal** 29:159–167.

Mathias E, Gao P, Goldstein J, Wilson A, Pugh E, Furbertharris P, Dunston G, Malveaux F, Togias A, Barnes K, Beaty T, Huang S (2006) A graphical assessment of p-values from sliding window haplotype tests of association to identify asthma susceptibility loci on chromosome 11q. BMC Genetics 7:38.

Melo TP, Camargo GMF, Albuquerque LG, Carvalheiro R (2017) Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nelore beef cows. **PloS one** 12: e0178551.

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

Meuwissen THE (2007) Genomic selection: marker assisted selection on genome-wide scale **Journal of Animal Breeding and Genetics**124:321-322.

Minozzi G, Nicolazzi EL, Stella A, Biffani S, Negrini R, Lazzari B, Ajmone-Marsan P, Williams JL (2013) Genome wide analysis of fertility and production traits in Italian Holstein cattle. **PLoS One** 8:e80219.

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.

Mokry FB, Buzanskas ME, Mudadu MA, Grossi DA, Higa R H, Ventura RV, Lima AO, Sargolzaei M, Meirelles SLC, Schenkel FS, Silva MVGB, Niciura SCM, Alencar MM, MunarI DP, Regitano LCA (2014) Linkage disequilibrium and haplotype block structure in a composite beef cattle breed. **BMC Genomics** 15: 1-9.

Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HWA (2009) comparison of five methods to predict genomic breeding values of dairy bulls from genome wide SNP markers Genetics Selection. **Evolution** 41: 56.

Muir WM (2007) Comparison of genomic and traditional BLUP – estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal of Animal Breeding and Genetics** 124:342-355.

Nayeri S, Sargolzaei M, Abo-Ismail MK, May N, Miller SP, Schenkel F, Moore SS, Stothard, P (2016) Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. **BMC genetics** 17:75.

Oliveira Júnior, GA (2017) **Associação genômica ampla (GWAS) aplicada a características reprodutivas de novilhas da raça Nelore** 82 f Tese (Doutorado em Biociência Animal) – USP, Pirassununga

Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME, Hayes BJA (2010) validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes **Journal of Dairy Science** 93:3331-3345.

Regatieri IC, Boligon AA, Costa RB, Souza FRP, Baldi F, Takada L, Venturini, GC, Camargo GMF, Fernandes Junior GA, Tonhati H, Oliveira HN, Albuquerque, LG (2017) Association between single nucleotide polymorphisms and sexual precocity in Nelore heifers. **Animal reproduction science** 177:88-96.

Resende MDV, Lopes PS, Silva RL, Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira** 56:63-77.

Saatchi M, Schnabel RD, Taylor JF, Garrick DJ (2014) Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. **BMC Genomics** 15.

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics** 123:218–223.

Snelling WM, Cushman R, Fortes MRS, Reverter A, Bennett GL, Keele JW, Kuehn LA, Mcdaneld TG, Thallman M, Thomas MG (2012) How single nucleotide polymorphism chips will advance our knowledge of factors controlling puberty and aid in selecting replacement beef females. **Journal of Animal Science** 90:, p1152-1165, 2012

Taylor JF, Decker JE, Schnabel RD (2014) Animal Breeding in the Next-Generation Sequencing Era In: 10th World Congress of Genetics Applied to Livestock Production Proceedings…Vancouver

Teixeira DBA, Fernandes Júnior GA, Silva DBS, Costa RB, Takada L, Gordo DG M, Bresolin T, Cavalheiro R, Baldi F, Albuquerque LG (2017) Genomic analysis of stayability in Nelore cattle. **PloS one** 12:e0179076

Tempelman RJ (2015) Statistical and Computational Challenges in Whole Genome Prediction and Genome-Wide Association Analyses for Plant and Animal Breeding. **Journal of Agricultural, Biological, and Environmental Statistics** 20:442-466.

The International Hapmap Consortium (2015) A haplotype map of the human genome. **Nature** 437:1299–1320.

Van Eenennaam AL, Weigel KA, Young AE, Cleveland MA, Dekkers JCM (2014) Applied Animal Genomics: Results from the Field Annual Review of Animal. **Biosciences** 2:105–139.

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

Villumsen TM, Janss L, Lund MS (2009) The importance of haplotype length and heritability using genomic selection in dairy cattle. **Journal of Animal Breeding and Genetics** 126:3–13

Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery The American Journal of Human. **Genetics** 90:7–24.

Yokoo MJI, Albuquerque LG, Lôbo RB, Sainz RD, Carneiro Júnior JM, Bezerra LAF, Araujo FRC (2007) Estimativas de parâmetros genéticos para altura do posterior, peso e circunferência escrotal em bovinos da raça Nelore**. Revista Brasileira de Zootecnia** 36:1761-1768.

Wu Y, Fan H, Wang Y, Zhang L, Gao X, Chen Y, Li J, Ren H, Gao, H (2014) Genome-Wide Association Studies Using Haplotypes and Individual SNPs in Simmental Cattle. **PLoS ONE** 9

Wang H, Misztal I, Aguilar I, Legarra A, Muir WM (2012) Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research** 94:73-83.

CHAPTER 2 - GENOME WIDE ASSOCIATION STUDY FOR REPRODUCTIVE TRAITS USING HAPLOTYPES IN NELORE CATTLE

Abstract - The aim of this study was to perform genome-wide association using haplotypes to detect chromosomal regions associated with age at first calving (AFC) and scrotal circumference (SC) in Nelore cattle. Genotypes of 1900 heifers and 1500 young bulls were obtained with the HD panel from Ilumina® (777K) and 490 heifers and 3332 young bulls were genotyped with the GeneSeek Genomic Profiler HDi 75K. The animals genotyped with lower density panel were imputed to HD using the FImpute program. Phenotype was adjusted for fixed effects (Y*) to be used as response variable in genomic analysis. The Fimpute software was used to inferring the missing genotypes and the haplotype phase. The linkage disequilibrium measurement D' and the confidence interval methodology were used to construct the haplotypes through software HaploView. Haplotypes effects were estimated performing a univariate mixed model in the software GEMMA. Results indicated that 21 haplotypes were associated with AFC and 16 were associated with SC. Besides that, we identified candidate genes located within haplotype blocks for both traits studied. In general, the annotated genes found for AFC and SC, play a role in reproductive functions, as ovulation, myometrial contraction, spermatozoon motility, spermatogenesis and spermatozoa maturation. The results from haplotype-based association study have supportive evidence, such as colocalization with known QTLs controlling reproductive functions, including AFC and SC. The results found in this study can be used to identify causal mutations and candidate genes that underline AFC and SC in Nelore cattle.

Keywords: Beef cattle, *Bos indicus*, GWAS**,** Genes, Reproductive functions

1. Introduction

Productive and reproductive traits are important to the profitability of the livestock production system. The reproductive traits have a high economic value and could be 13 times economically more important than growth and carcass traits in a selection index for beef cattle (Brumatti et al., 2011). Due their importance, the breeding programs has included reproductive traits in the breeding programs as selection criteria (Chud et al., 2014; Grossi et al., 2016).

The age at first calving (AFC) is a reproductive trait widely used in research and cattle production system for being an indicator of sexual precocity in the herd. According Boligon et al. (2010), the AFC is an easily and early measured phenotype, since it is part of the routine data collection and can be obtained in most of the females considered in genetic evaluations. However, it is a sex-limited trait and there is evidence that is controlled by several genes of small effect (Costa et al., 2015; Melo et al., 2016), with heritability estimates varying between 0.10 and 0.21 (Boligon et al., 2011; Boligon et al., 2012; Caetano et al., 2013; Regatieri et al., 2013).

The scrotal circumference (SC) is commonly used in breading programs due its easy measurement, repeatability and medium to high heritability. Additionally, SC is associated with productive and reproductive male traits as final weight (Yokoo et al., 2007), visual scores at yearling (Boligon et al., 2007), testis development (Coulter et al., 1979) and in female with stayability (Van Melis et al, 2010) and heifer pregnancy (Santana Jr et al., 2012).

The use of molecular markers in genome-wide association studies (GWAS), have allowed a better understanding of the genetic bases of complex traits (Kemper e Goddard, 2012). The GWAS, based on the linkage disequilibrium (LD) between the molecular marker and the quantitative trait loci (QTL), is used to identify genomic regions associated with traits of economic interest (Hayes e Goddard, 2010; Visscher et al., 2012).

Different GWAS has been published for *Bos taurus indicus* and many genes and regions that explain the genetic variability of traits as AFC (Costa et al., 2015; Irano et al., 2016; Melo et al., 2017) and SC (Fortes et al., 2012; Irano et al., 2016) have been described. Although many candidate genes have been identified, the genetic variance explained by them is small (Bolormaa et al., 2011), fact that may

be attributed to the complex nature of the trait or the molecular marker used in the studies (traditionally SNPs).

Groups of SNPs which are in high LD are known as haplotypes (Villumsen et al., 2009; Garrick e Fernando, 2014; Tempelman, 2015). They have a low recombination rate and can be conserved across generations (Meuwissen et al., 2001). LD between QTL and haplotypes could be stronger than between QTL and SNPs (Hayes, 2013; Cuyabano et al., 2014). GWAS has been used with haplotypes instead SNPs as explanatory variables to identify markers that affect milk production and fertility (Pryce et al., 2010), prenatal death (Fritz et al., 2013), calf survival in Holstein population (Kipp et al., 2015), tenderness (Braz et al., 2018) and fertility in heifers (Oliveira Junior, 2018). Nevertheless, there are few works GWAS using haplotypes for AFC and SC in Nelore are still few. Hence, the proposal of this study was to perform GWAS using haplotypes for age at first calving and scrotal circumference in Nelore cattle.

2. Material and methods

2.1 Ethical approval

Procedures were approved by the Animal Care of the School of Agricultural and Veterinary Science, São Paulo State University (UNESP) Ethical Committee (protocol No. 18.340/16).

2.2 Phenotype and genotype data

The phenotypic and pedigree data were obtained from three Nelore breeding programs (DeltaGen®, CRV Lagoa - PAINT® and CIA de Melhoramento®). Animals, females and males, were born in different herds between 1984 and 2015 and were raised under tropical pasture in the central, midwest and southeast regions of Brazil.

According to the breeding program criteria for AFC, three mating systems were used: artificial insemination, controlled breeding and multiple-sire mating system. Regardless the body condition score and the body weight, all the heifers were exposed to reproduction with an age of 14 to 16 months in a breeding season of approximately 60 days (from February to March). If the pregnancy not confirmed by palpation after the early breeding season, heifers have a second change to

conceive in the regular season with the cows between November and February. AFC was obtained as the difference in days between the female birth and her first calving data and SC was taken at yearling and measured with a metric tape.

The fixed effects of birth, weaning and yearling farm, management group at weaning and yearling and year of birth were combined to form the contemporary groups (CG) for both traits. The observations outside the intervals given by the mean ±3 standard deviations and with less than three animals by CG were excluded from the analysis. A total of 258,306 phenotypes from AFC and 463,919 from SC were used. Descriptive statistics for both traits are shown in Table 1.

For AFC, the phenotype was adjusted for fixed effects (y*) of CG and for SC were adjusted using the CG and age of animal at recording. Posteriorly, the y* were used as response variable in genomic analysis. The fixed effects and heritability were estimated using AIREMLF90 software (Misztal et al., 2002) considering a single-trait animal model (Table 1).

Table 1. Descriptive statistics and heritability estimates (h^2) for age at first calving (AFC) and scrotal circumference (SC) in Nelore cattle.

Trait	N_{phe}	Mean ± SD	Min	Max	h^2
AFC (days)	258,306	1,042 ± 111.58	634	1250	0.08
SC (cm)	463,919	26.25 ± 3 .90	15.12	38.50	0.42

N_{phe} = Number of phenotypes; Mean = mean of phenotype; SD = standard deviation, Min= minimum; Max= Maximum.

The Illumina BovineHD BeadChip assay (770k, Illumina Inc., San Diego, CA, USA) was used to genotype 1,900 heifers and 1,500 young bulls, while 490 heifers and 3,332 young bulls were genotyped with the GeneSeek Genomic Profiler HDi 75K (GeneSeek Inc., Lincoln, NE). The FImpute software v.2.2 (Sargolzaei et al., 2014) was used to impute the animals genotyped with lower density panel (75K) to the HD panel (770k). For quality control of genotypes were considered only autosomal SNPs, minor allele frequency for SNPs less than 0.02, Hardy-Weinberg equilibrium (p-value > 10^{-5}), call rate for the SNPs great than 0.92 and call rate for samples great than 0.90. After quality control remained in dataset 2,390 females and 4,832 young bulls with 407,158 and 412,124 SNPs for AFC and SC, respectively.

2.3 Haplotypes construction

The software Fimpute (Sargolzaei et al., 2014) was used to inferring the missing genotypes and haplotype phase. Software HaploView (Barrett et al., 2005) was used to construct the haplotypes, according to Gabriel et al. (2002). A pair of SNPs was considered in “Strong LD” when confidence bound ranged from 0.70 to 0.98. Linkage disequilibrium (LD) measure used in this study was D’, proposed by Hill (1981):

$$D' = \begin{cases} \dfrac{D}{\min\{freq.A^{*}freq.b, freq.a^{*}freq.B\}} \; if \quad D > 0, \\ \dfrac{D}{\min\{freq.A^{*}freq.B, freq.a^{*}freq.b\}} \; if \quad D < 0 \end{cases}$$

Where,

$$D = freq.AB - (freq.A \times freq.B).$$

The freq. AB, are the frequencies of haplotypes AB in the population; the freq. A, freq. a, freq. B and freq. b are the frequencies of alleles A, a, B and b, respectively.

The haplotype incidence matrix was set up with the haplotype variations (alleles) of each haplotype, and the number of copies that every subject possessed of the allele *i* in the haplotype *j* were determined. The alleles were encoded as 0, 1 or 2, depending on the number of copies of the allele (paternal and maternal). The haplotype matrix has n × p dimension, being n the number of haplotype alleles and p the total of animals in the analysis.

2.4 Genome-wide association study (GWAS)

An univariate linear mixed model was performed in GEMMA (Zhou e Stephens, 2012) using the adjusted phenotype and the haplotypes following the model:

$$y^{*} = 1\mu + Xb + Zu + \epsilon.$$

Where, y* is a vector of observed phenotypes adjusted for fixed effects; 1 is a vector of ones; “μ" is the overall mean; X is the haplotype incidence matrix; b is

a vector of haplotypes effects; Z is an n x n identity matrix; u is a vector of the additive genetic effects and "ϵ" is a residual vector. It was assumed that "u" ~ MVN (0, Gσ_g^2), where G is the genomic relationship matrix (GRM) and σ_g^2 is the additive genetic variance component, and "ϵ" ~ MVN(0,Iσ_e^2), where σ_e^2 is a component of residual variance and I is an identity matrix. The GRM was set up using the haplotype incidence matrix using the filtering of the software GEMMA.

The methodology proposed by Lynch e Walsh (1998) was used to calculate the haplotype additive genetic variance(σ_a^2):

$$\sigma_j^2 = \sum_{i=1}^{k_{j-1}} \sum_{l>i}^{k_j} (a_{ij} - a_{lj})^2 \, q_{ij} q_{lj} \, .$$

Where: σ_j^2 is the additive genetic variance of the *j*th haplotype, *k* is the total of alleles in the *j*th haplotype, a_{ij} is the additive effect of the *i*th allele at the *j*th haplotype, a_{lj} is the additive effect of the *l*th allele at the *j*th haplotype, q_{ij} is the *i*th allele in the *j*th haplotype frequency and q_{lj} is the *l*th allele in the *j*th haplotype frequency.

It was assumed that the σ_j^2 followed a gamma probability distribution with parameters shape (θ) and rate (β). Shape and rate parameters were adjusted using σ_j^2, allowing set a quantile value in the gamma distribution corrected for multiple testing by Bonferroni (α ≤ 0.05), which let identify the haplotypes with the greatest amount of σ_j^2.

For modeling θ parameter a cubic regression was used. For estimation of β, it was performed a not linear regression (Brody, 1924) as follows:

$$\hat{\beta} = \mathrm{A}(1 - Be^{-pt}) + \epsilon,$$

where $\hat{\beta}$ are the predicted β; A is the β asymptotic limit; B is the integration constant; p represent the ratio of maximum growth rate to asymptotic limit of $\hat{\beta}$; t is the haplotype alleles number and ϵ is the residual. Using these parameters, it was possible establish the threshold in the gamma distribution that let identify the haplotypes which explained the greatest σ_a^2.

Genome data viewer tool from the NCBI (*National Center for Biotechnology* Information) was used for identification of the genes within haplotypes using the UMD 3.1 assembly as the reference map (https://www.ncbi.nlm.nih.gov/genome/gdv/?org=bos-taurus). The metabolic

function of each gene was studied with the aim of understand their action on the studied trait. Additionally, it was verified if the haplotype block regions for AFC and SC belong to genomic regions identified by quantitative trait loci (QTL) in the QTL animal database (QTLdb) for cattle (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index).

3. Results

The SNPs were clustering in 75,069 and 77,283 haplotypes and, 483,933 and 570,064 alleles using the D' thresholds (Table 2). In both traits (AFC and SC), the largest haplotype was constructed by 104 SNPs, the smallest by only 2 SNPs and the average of SNPs/Haplotype was 4.56 for AFC and 4.46 for SC. Figures 1, 2 and 3 represents the SNPs into haplotypes, haplotypes alleles and haplotypes distribution across the 29 chromosomes. The haplotypes distribution was not uniform and was distributed almost proportionally with the size of the chromosome. The chromosome 1 harbored the highest number of haplotypes (4,621 and 4,782) and the chromosome 25 the smallest (1,407 and 1,434), for AFC and SC respectively.

Table 2. Descriptive statistics for genomic information used to perform GWAS for age at first calving (AFC) and scrotal circumference (SC) in a Nelore cattle population.

	Variable	Total records	Descriptive statistics**		
			Min	Max	Mean
AFC	Haplotypes	75,069	1,407	4,621	2,589
	SNPs*	342,626	2	104	4.56
	Alleles	483,933	2	141	6.45
SC	Haplotypes	77,283	1,434	4,782	2,665
	SNPs*	344,933	2	104	4.46
	Alleles	570,064	2	209	7.38

*****SNPs used to construct haplotypes; ** number of haplotypes (per chromosome), SNPs (per haplotype) and alleles (per haplotype)

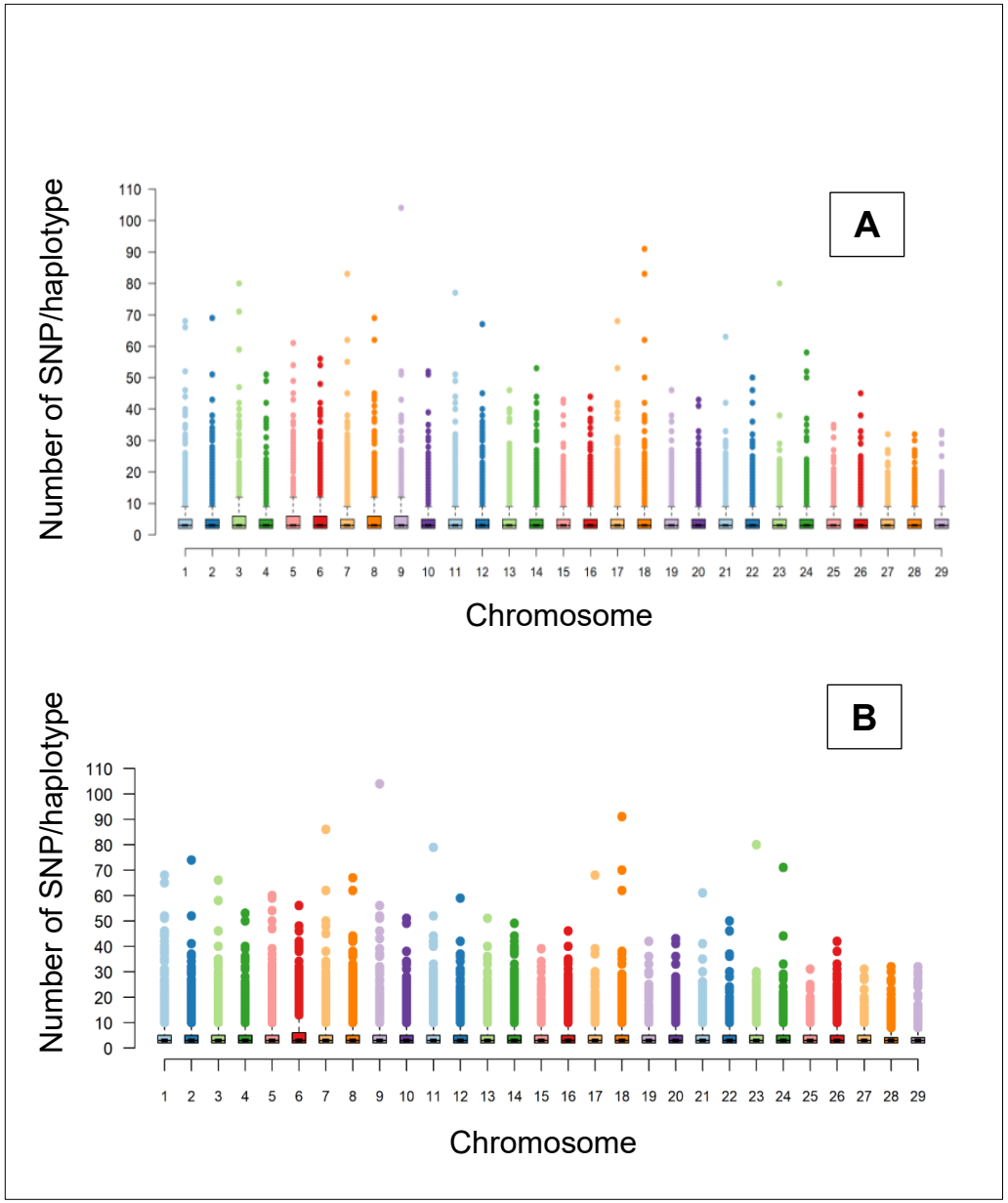

Figure 1 Boxplots showing the number of SNPs per haplotype of a female (A) and male (B) Nelore cattle population used to perform GWAS

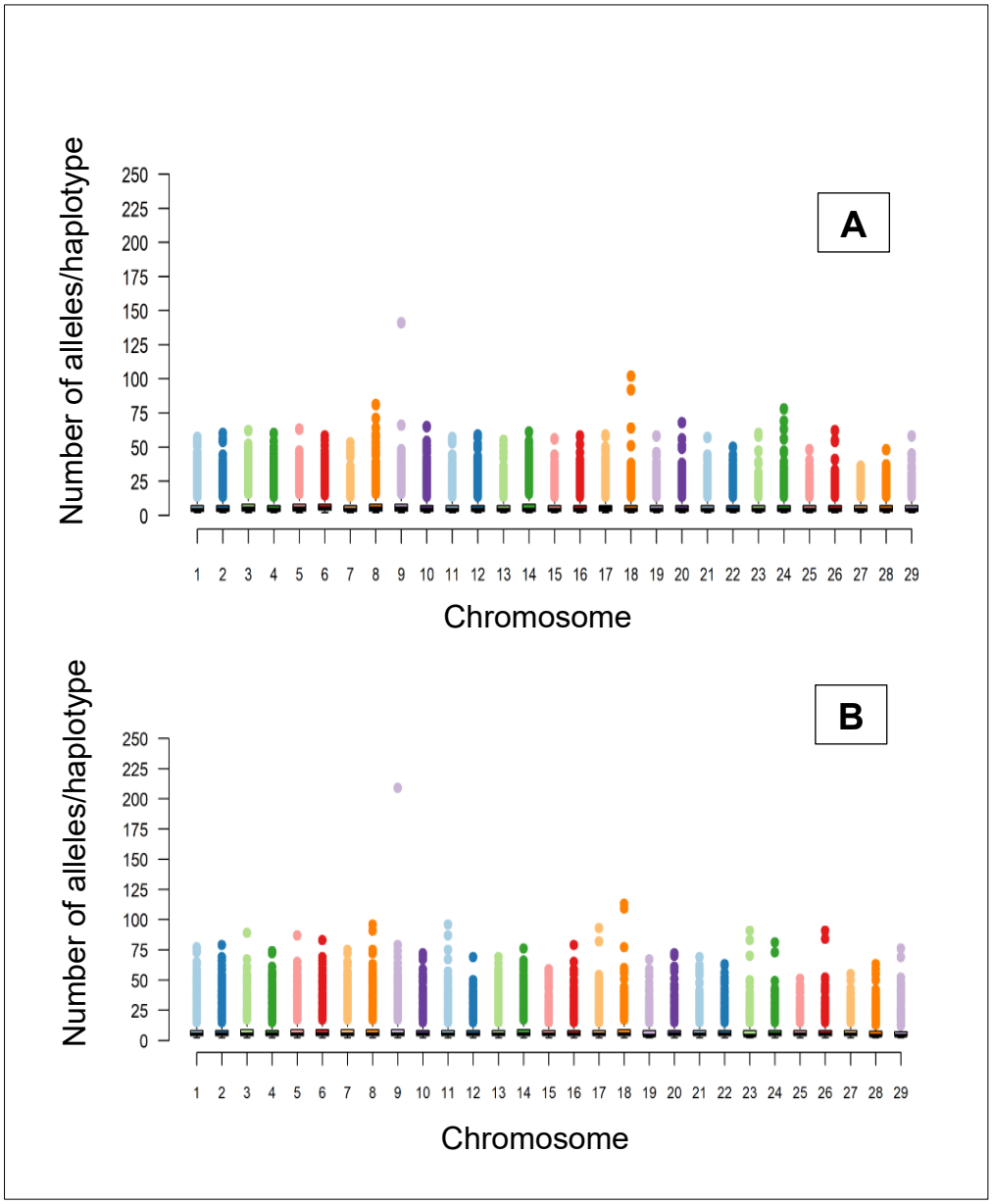

Figure 2 Boxplots showing the number of alleles per haplotype of a female (A) and male (B) Nelore cattle population used to perform GWAS

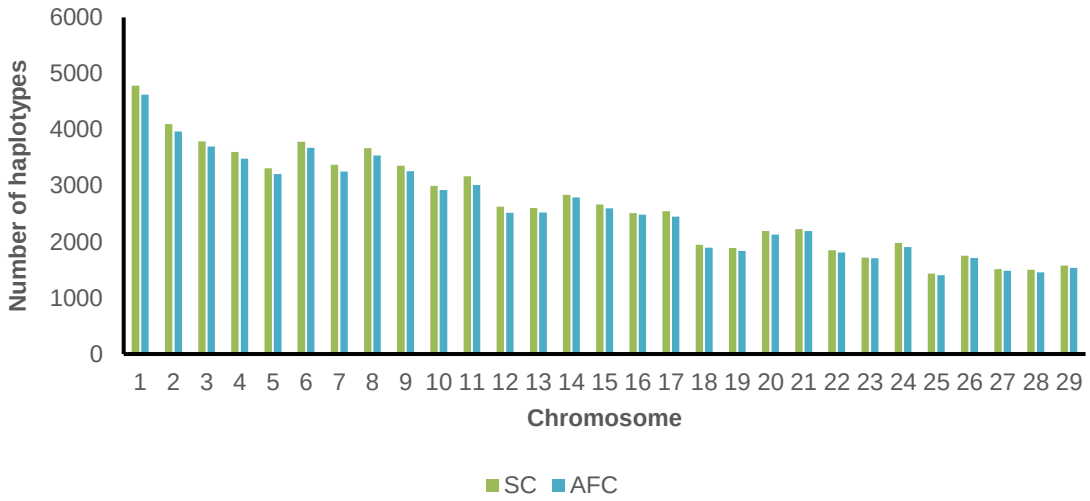

Figure 3. Number of haplotypes per chromosome of a Nelore cattle population used to perform GWAS using haplotypes for age at first calving and scrotal circumference

The results showed 21 haplotypes with the highest additive genetic variance (Bonferroni-corrected p-values ≤ 0.05) along 12 different bovine chromosomes for AFC (Table 3). We investigated if the haplotype regions for AFC belonged to genomic regions identified by QTL. As shown in Supplementary File: Table S1, for AFC, reproductive-related 32 QTLs were located within 14 haplotypes found this study. Associated genes were found within the interval of the 10 haplotypes (Table 3). Some of the coding protein genes that were identified play a role on female reproductive system in mammals, specifically in ovulation, myometrial contraction (Weiner et al., 2010) and regulate of sexual hormones (Shimizu et al., 2010; Parada-Bustamante et al.,2010).

In total, 26 haplotypes with great additive genetic variance (Bonferroni-corrected p-values ≤ 0.05) were identified for SC (Table 4). We found reproductive-related 62 QTL (Supplementary File: Table S2) and 26 candidate genes (Table 4) were located present in haplotypes. Some of the genes that were identified were reported in the literature as involved in male reproductive processes (Adham et al., 2008; Luconi et al., 2011; Makinistoglu and Karsenty, 2015; Turner, 2006). The gene names and its symbols for both traits are reported in Supplementary File table S1.

The results from haplotype-based association showed genes and have supportive evidence, such as colocalization with known QTLs controlling AFC and

SC (see Supplementary File: Tables S2 and S3). Still results showed candidate regions that affect AFC and SC of the Nelore cattle.

Table 3. Genes located within the interval of the haplotype blocks associated with age at first calving of a Nelore cattle population

Chr	Hap	N_{all}	position (bp)	Length (bp)	Gene
5	16003	4	64,58,772 - 6,462,823	4,051	*E2F7*
5	16005	5	6,497,108 - 6,514,900	17,792	Intergenic
5	16312	3	14,326,657 - 14,331,430	4,773	Intergenic
5	16462	9	19,510,761 - 19,526,737	15,976	Intergenic
5	17590	10	67,854,189 - 67,897,838	43,649	*LOC505479 (TTC41P)*
8	28861	2	95,451,584 - 95,452,399	815	Intergenic
9	32415	14	97,836,867 - 97,861,919	25,052	*ORCT2 (SLC22A2)*
10	33030	4	8,243,173 - 8,245,542	2,369	Intergenic
11	35722	3	2,397,460 - 23,997,42	2,282	*ITPRIPL1 (KIAA1754L)*
13	42145	3	30,217,873 - 30,221,912	4,039	*LOC100848562*
13	42405	4	38,389,345 - 38,392,153	2,808	Intergenic
16	50426	2	48,507,095 - 48,510,030	2,935	Intergenic
18	54090	6	3,306,525 - 3,327,732	21,207	Intergenic
18	55311	9	44,675,555 - 44,700,852	25,297	Intergenic
19	57246	4	50,329,783 - 50,336,021	6,238	*B3GNTL1*
21	59864	4	845,880 - 983,837	137,957	Intergenic
21	59977	10	6,712,534 - 6,750,410	37,876	*ADAMTS17*
21	60641	17	25,062,833 - 25,121,384	58,551	*SH3GL3*
22	62055	12	350,129 - 395,272	45,143	*DBNL*
25	68538	3	31,408,989 - 31,409,857	868	Intergenic
25	68879	11	42,719,548 - 42,755,300	35,752	*FAM20C*

Hap = Haplotype; N_{all} = Number of alleles that each haplotype block contains

Table 4. Genes located within the interval of the haplotype blocks associated for scrotal circumference of a Nelore cattle population

Chr	Hap	N_{all}	position (bp)	Length (bp)	Gene
1	1450	4	44844961-44850305	5344	Intergenic
1	2040	4	65530942-65534765	3823	Intergenic
1	2042	10	65562631-65613386	50755	*GPR156*
2	5835	4	29321440-29322967	1527	Intergenic
2	5895	8	30883656-30925831	42175	*LOC101904870*
3	10356	7	43940802-43944009	3207	Intergenic
3	11837	8	96305979-96314088	8109	*FAF1*
3	12605	8	119032361-119078411	46050	*HDAC4*, *LOC104971894*
7	24569	10	40572826-40598898	26072	*N4BP3*, *RMND5B*, *LOC101905866*
8	28559	12	51553965-51593010	39045	Intergenic
8	28813	16	61602700-61703986	101286	*ZCCHC7*
9	32016	7	51304721-51323919	19198	Intergenic
9	32030	5	51921230-51926215	4985	Intergenic
12	40844	4	29079800-29083909	4109	*FRY*
14	46072	8	27064286-27095501	31215	Intergenic
15	48271	4	9437348-9441625	4277	Intergenic
15	48953	8	31446062-31501499	55437	*ARHGEF12*,*TMEM136*
15	48954	11	31558843-31590706	31863	*ARHGEF12*
15	48956	4	31595404-31599942	4538	*ARHGEF12*
15	48957	4	31619191-31621332	2141	*ARHGEF12*
15	48958	4	31647441-31665547	18106	*ARHGEF13*
15	49324	10	41764711-41772616	7905	*GALNT18*
17	55009	14	54901135-54970181	69046	*CCDC62, HIP1R, VPS37B, LOC104974637*
19	58954	4	44743980-44754980	11000	*RUNDC3A, SLC25A39*
25	70716	20	34306509-34373088	66579	*FKBP6, NSUN5, POM121C, TRIM50, LOC104975906*
27	73665	4	29576824-29592610	15786	Intergenic

Hap = Haplotype; N_{all} = Number of alleles that each haplotype block contains

4. Discussion

In this study, we used a haplotype-based approach to identify genomic variants associated with AFC and SC. Haplotype-based information in GWAS approach reduces the number of genomic variants to few haplotypes (Johnson et al., 2001). Several methods to construct haplotypes have considered a fixed number of SNPs to perform GWAS (Nascimento et al., 2018; Braz et al., 2019). Although, in a real population, the haplotypes do not have the same length, because they depend of LD between SNPs, and LD is not uniform across the entire genome (Ardlie et al., 2002). Thus, the use of haplotypes constructed from D' demonstrated to be efficient to perform GWAS, allowing identify 21 haplotypes associated with AFC and 16 with SC.

4.1. Age at first calving

The results showed 21 haplotypes with highest additive genetic variance along 12 different bovine chromosomes for AFC (Table 3). Ten associated genes were found within the interval of the ten haplotypes (Table 3). Among these genes, *ADAMTS17, B3GNTL1, ITPRIPL1 (KIAA1754L)* and *SH3GL3* annotated genes were previously reported to play a role on female reproductive system in mammals, specifically in ovulation, myometrial contraction (Weiner et al., 2010) and regulate of sexual hormones (Shimizu et al., 2010; Parada-Bustamante et al.,2010).

The ADAMTS17 gene is a member of zinc-dependent proteinase family. These enzymes are located in the extracellular matrix (ECM) and play a role in matrix assembly and degradation in morphogenesis, angiogenesis, ovulation and coagulation in many organisms, including bovine (Demircan et al., 2014; Dubail and Apte, 2015; Russell et al., 2015). The *B3GNTL1* gene is a putative glycosyltransferase and was describe by Weiner et al. (2010), as a gene that participates in initiating term labor in human and, consequently, should relates to myometrial contractility and the cellular activities necessary to sustain it.

The ITPRIPL1 (KIAA1754L) and *SH3GL3* genes are regulated by estradiol. Shimizu et al. (2010) showed that when the estradiol was increased in cows, the *ITPRIPL1 (KIAA1754L)* was overexpressed in the bovine endometrium. The same case was observed for *SH3GL3* gene in rats (Parada-Bustamante et al., 2010).

Estradiol is related with the ovulation control and female development. This hormone accelerates oviduct egg transport (Parada-Bustamante et al., 2010). The oviduct provides an optimal microenvironment for fertilization and early embryo development (Jansen, 1984). Estradiol is one of the main regulators of these phenomena modifying expression and secretion of molecules, which assure fertilization and embryo viability (Buhi, 2002; Bhatt et al., 2004). Silva (2018) associated the gene *ITPRIPL1 (KIAA1754L)* with stayability in Nelore heifers.

In this study, the pseudogene LOC505479 (TTC41P) was observed in association with AFC. Pseudogenes are usually characterized by a combination of homology to a known gene and loss of some functionality, which means that they are unable to produce functional final protein products (Mighell et al., 2000). However, these sequences, although they do not produce transcripts that can be translated into proteins, can still produce transcripts that, in some cases, have important regulatory roles and are therefore functional (Pei et al., 2012).

DBNL, E2F7, FAM20C and *ORCT2* (*SLC22A2*) genes were observed associated with AFC in this study. These genes play roles as adapter protein (Fish et al., 2016), transcription factor (Chu et al., 2015), calcium-binding (Nalbant et al., 2005) and solute carrier (Zwart et al., 2001), respectively.

The *DBNL* and *E2F7* genes also have possible effects in the breast cancer metastasis in mouse and human, respectively (Zwart et al., 2001; Fish et al., 2016). *FAM20C* gene was involved with bone mineralization and was related to carcass trait in bovine (Mokry et al., 2013). *ORCT2* (*SLC22A2*) gene was associated with phenotypes of net tubular creatinine secretion and end-stage renal disease in human (Reznichenko et al., 2013). The main effects that these genes have on the female reproductive system are still unclear. Therefore, more studies should be performed to elucidate how these genes affect the AFC and consequently, the female sexual precocity.

The annotated genes and known QTL involvement with female reproductive traits (e.g., reproductive efficiency, age at first calving and ovulation rate, calving ease - maternal) support the results obtained for the haplotypes regions found. Besides that, the results endorse the putative effect of the identified regions in AFC.

4.2. Scrotal circumference

In total, 26 candidate genes were annotated within haplotype blocks associated with SC (Table 4). In the present study six haplotypes were related with the genes *FAF1, HDAC4, ARHGEF13, CCDC62, SLC25A39, FKBP6* and *NSUN5,* reported in the literature as involved in male reproductive processes.

The *FAF1* gene has been associated with cellular apoptosis and plays a key effect during embryonic development, oogenesis, and spermatogenesis in mouse (Adham et al., 2008). Commonly, apoptosis acts as a balance mechanism between cell death and cell production. Meggiolaro et al, (2006), suggested that a *FAF1* antigen can has an anti-apoptotic effect protecting the spermatozoa in cattle since they found the antigen in ejaculated spermatozoa of bulls fertile.

The *HDAC4* gene is a member of the HDAC family with a key role in the male reproductive system in mice. Makinistoglu and Karsenty (2015) demonstrated that the *HDAC4* gene, expressed in the osteoblast affect the osteocalcin production. In this context, the osteocalcin absence lead to reductions in mice male reproductive traits, producing a decrease in sperm counts, seminal vesicle, epididymis, and testis weight as well as the amount of circulating testosterone levels.

Acording to Turner (2006) and Luconi et al. (2011), ARHGEF13 is involved in mammalian spermatozoa maturation and motility. In Nelore cattle, this gene was previously associated with omega-3 and omega-6 fatty acids (Feitosa et al. 2018). The *CCDC62* gene, encodes a nuclear receptor that increase estrogen receptors (Chen et al., 2009). A nonsense mutation in this gene caused spermatogenesis defects and consequently male infertility (Li et al., 2017). Recently Nyman et al. (2019) associated *CCDC62* with progesterone profiles in a GWAS study with Holstein-Friesian cattle. *CCDC62* is expressed in testis and has been found as antigen in prostate cells (Faramarzi e Ghafouri-Fard, 2017).

The *SLC25A39* gene, belongs to the SLC25 family of mitochondrial carrier proteins (Haitina et al., 2006). This gene was found in human and mice testis (Yu et al., 2001). The FK506-binding protein 6 (*FKBP6*) gene, encodes a protein that plays a role in mammalian fertility. Different studies have related that this gene is expressed in testicular tissues of human, horse mouse and yak (Crackower et al., 2003; Miyamato et al., 2006; Raudsepp et al., 2012; Li et al., 2016) and has an

important role in spermatogenesis. Li et al. (2016), studied the *FKBP6* methylation in yak and cattle-yak hybrid and founding a suppression in the gene transcriptional activity, leading to infertility, especially in cattle-yak hybrid. The *NSUN5* gene, encodes a member of the RNA Methyltransferase family. In mice, Harrys et al. (2007) suggested that this gene could be involved in mice male fertility, once this gene is paralog of *NSUN7* that produce motility defects and infertility in mice.

The genes *GPR156, N4BP3, RNMD5B, ZCCHC7, FRY, TMEM136, ARHGEF12, GALANT18, HIP1R, VPS37B, RUNDC3A, TRIM50 and POM121C* were associated with SC in our study. These genes have functions as couple receptors (Vassilatis et al., 2003), binding protein (Dawnson et al., 2007; Hyun et al., 2004; Jia et al., 2012; Nagai e Mizuno, 2014; Zhou et al., 2018), transporting protein (Martinez-Royo et al., 2017) and regulatory protein (Siripurapu et al., 2005).

The *GPR156* gene belongs to the G protein-coupled receptors (GPCRs) group, which encodes for proteins that act as cell surface receptors. Diverse members of the GPCR are major targets of pharmaceutical drugs as they participate in a variety of physiological functions (Vassilatis et al., 2003). The *N4BP3* gene, encodes a binding protein gene and has been associated with cervical cancer in humans (Zhou et al., 2018). In anterior pituitary tissue of Japanese heifers, Pandey et al. (2017) found that *N4BP3* coded the NEDD4 binding protein 3 and their level increase during the pre-ovulation compared to post-ovulation.

Dawson et al. (2007), associated *RMND5B* gene with the protein NKX3.1 in human, which acting in the suppression of prostatic tumors and prostate epithelial cells proliferation. The haplotype 28813 (61602700-61703986) on chromosome 8 harboring the gene *ZCCHC7*, which is considered a part of the RNA damaged degradation complex, *and* was associated with coat color in Holstein Cattle (Hayes et al., 2010), placental functions and development in humans (Jia et al., 2012) and female reproductive traits in pigs (Metodiev et al., 2018).

The Furry protein produced by *FRY* gene has a key role in the organization and stability of the microtubules in mammal mitosis (Nagai e Mizuno, 2014). In sheep, the gene *TMEM136* has an important effect in reproductive traits as progesterone cycling months and total days of anestrus (Martinez-Royo et al., 2017). The *ARHGEF12* gene has been associated with cellular apoptosis (Siripurapu et al., 2005) and intra ocular pressure in humans (Springelkamp et al.,

2015). In humans *TRIM50* gene, is a member of the tripartite motif family. Nishi et al. (2012) demonstrated that *TRIM50* gene is specifically expressed in gastric parietal cells being associated with the dynamic movement of intracellular vesicles.

The POM121 gene is expressed in many reproductive tissues as placenta, uterus and cervix, besides prostate, testis and seminal vesicle. Noguchi et al. (2004), associated this gene with Williams syndrome in humans, a disorder inherited in an autosomal dominant manner, which causes distinctive facial features, mild intellectual disability and an overly sociable personality. This gene coluld be involved with infertility in humans, since most people with Williams syndrome do not reproduce.

5. Conclusion

The use of haplotypes constructed by linkage disequilibrium between the SNPs using D' allowed identifies ten regions in the chromosomes 5, 9, 11, 13, 19, 21, 22 and 25 associated with the AFC. Among the genes found in this study, *ADAMTS17, B3GNTL1, ITPRIPL1 (KIAA1754L)* and *SH3GL3* are implicated in biologic process as in ovulation, myometrial contraction and regulate of sexual hormones.

Seven genes associated with SC (FAF1, HDAC4, ARHGEF13, CCDC62, SLC25A39, FKBP6 and NSUN5) were found in the chromosomes 3, 15, 17, 19 and 25. These genes are involved in reproductive functions as spermatogenesis and spermatozoa maturation and motility.

The results found in this study can be used to identify causal mutations and candidate genes that underline AFC and SC.

6. References

Adham IM, Khulan J, Held T, Schmidt B, Meyer BI, Meinhardt A, Engel W (2008) Fas-associated factor (FAF1) is required for the early cleavage-stages of mouse embryo. **Molecular human reproduction** 14:207-213.

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nature Reviews Genetics 3: 299

Aung T, Ozaki M, Lee MC, et al. (2017). Genetic association study of exfoliation syndrome identifies a protective rare variant at LOXL1 and five new susceptibility loci. **Nature genetics** 49:993.

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

Bhatt P, Kadam K, Saxena A, Natraj U (2004) Fertilization, embryonic development and oviductal environment: role of estrogen induced oviductal glycoprotein. **Indian Journal of Experimental Biology** 42:1043-1055.

Berry DP, Wall E, Pryce JE (2014) Genetics and genomics of reproductive performance in dairy and breed cattle. **Animal** 8:105-121.

Boligon AA, Silva JAV, Sesana RC, Sesana JC, Junqueira JB, Albuquerque LG (2010) Estimation of genetic parameters for body weights, scrotal circumference, and testicular volume measured at different ages in Nelore cattle. **Journal of Animal Science** 88:1215-1219.

Boligon AA, Albuquerque LG (2011) Genetic parameters and relationships of heifer pregnancy and age at first calving with weight gain, yearling and mature weight in Nelore cattle. **Livestock Science** 141:12-16.

Boligon AA, Baldi F, Albuquerque LG (2012) Genetic correlations between heifer subsequent rebreeding and age at first calving and growth traits in Nelore cattle by Bayesian inference. **Genetics and Molecular Research** 11:4516-4524.

Bolormaa S, Porto Neto LR, Zhang YD (2011) A genome-wide association study of meat and carcass traits in Australian cattle. **Journal of Animal Science** 89:2297-2309.

Braz CU et al. (2019) Sliding window haplotype approaches overcome single SNP analysis limitations in identifying genes for meat tenderness in Nelore cattle. **BMC Genetics** 20:8.

Brody S. (1945) Bioenergetics and growth with special reference to the efficiency complex in domestic animals. New York: Reinhold Publishing Corporation. p.1023.

Brumatti RC, Ferraz JBS, Eler JP, Formigonni IB (2011) Desenvolvimento de índice de seleção em gado corte sob o enfoque de um modelo bioeconômico. **Revista Archivos de Zootecnia** 60:205-213.

Buhi WC (2002) Characterization and biological roles of oviduct-specific, oestrogen-dependent glycoprotein. **Reproduction** 123:355-362.

Caetano SL, Savegnago RP, Boligon AA, Ramos SB, Chud TCS, Lôbo RB, Munari, DP (2013) Estimates of genetic parameters for carcass, growth and reproductive traits in Nelore cattle. **Livestock Science** 155:1-7.

Camargo GMF, Costa RB, Albuquerque LG, Regitano LCD, Baldi F, Tonhati H (2014) Association between JY-1 gene polymorphisms and reproductive traits in beef cattle. **Gene** 533:477-80.

Chen M, Ni J, Chang HC, Lin CY, Muyan M, Yeh S (2009) CCDC62/ERAP75 functions as a coactivator to enhance estrogen receptor beta-mediated transactivation and target gene expression in prostate cancer cells. **Carcinogenesis** 30:841-850.

Chu J et al. (2015) E2F7 overexpression leads to tamoxifen resistance in breast cancer cells by competing with E2F1 at miR-15a/16 promoter. **Oncotarget** 6: 31944.

Chud TCS, Caetano SL, Buzanskas ME, Grossi DA, Guidolin DGF, Nascimento GB, Rosa JO, Lôbo RB, Munari DP (2014) Genetic analysis for gestation length, birth weight, weaning weight, and accumulated productivity in Nelore beef cattle. **Livestock Science** 170:16-21.

Costa RB, Camargo GM, Diaz ID, Irano N, Dias MM, Carvalheiro R, Boligon AA, Baldi F, Oliveira HN, Tonhati H, Albuquerque LG (2015) Genome-wide association study of reproductive traits in Nelore heifers using Bayesian inference. **Genetics Selection Evolution** 170:16-21.

Crackower MA, Kolas NK, Noguchi J, et al. (2003). Essential role of Fkbp6 in malefertility and homologous chromosome pairing in meiosis. **Science** 300:1291–1295.

Cuyabano BC, Su G, Lund MS (2014) Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. **BMC Genomics** 15:1171.

Dawson L, Thomas M, Bentel, J (2007). FLJ22318-a novel binding partner of NKX3.1 in prostate cancer cells. In: American association for cancer research annual meeting. Los Angeles, CA. disponivel em: < http://cancerres.aacrjournals.org/content/67/9_Supplement/1013.short >

Demircan K, Cömertoğlu İ, Akyol S, Yiğitoğlu BN, Sarıkaya E (2014) A new biological marker candidate in female reproductive system diseases: Matrix metalloproteinase with thrombospondin motifs (ADAMTS). **Journal of the Turkish German Gynecological Association** 15:250-255.

Dubail J, Apte SS (2015) Insights on ADAMTS proteases and ADAMTS-like proteins from mammalian genetics. **Matrix Biology** 44:24-37.

Faramarzi S, Ghafouri-Fard S (2017). Expression analysis of cancer-testis genes in prostate cancer reveals candidates for immunotherapy. **Immunotherapy** 9:1019-1034.

Feitosa FLB, Braz CU, Lemos MVA , Berton MP, Silva RMOP, Tonussi RL, Peripolli E, Olivieri BF, Pereira ASC, Albuquerque LG, Schenkel FS, BaldI F (2018) Genome-wide association study for omega-3 and omega-6 fatty acids in intramuscular fat of Nelore cattle using haplotypes. In: World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand, WCGALP 2018, 645.

Fish L, Pencheva N, Goodarzi H, Tran H, Yoshida M, Tavazoie SF (2016) Muscleblind-like 1 suppresses breast cancer metastatic colonization and stabilizes metastasis suppressor transcripts. **Genes & Developmentl** 30: 386-398

Fortes MRS, Lehnert SA, Wolcott M, Zhang Y, Bolormaa S, Holroyd R, Hawken RJ, Reverter-Gomez T (2012) Finding genes for economically important traits: Brahman cattle puberty. **Animal Production Science** 52:143-150.

Fortes MR, Nguyen LT, Porto Neto LR, Reverter A, Moore SS, Lehnert, SA, Thomas MG (2016) Polymorphisms and genes associated with puberty in heifers. **Theriogenology** 86:333-339.

Fritz S, Capitan A, Djari A, Rodriguez SC, Barbat A, Baur A, Grohs C, Weiss B, Boussaha M, Esquerré D, Klopp C, Rocha D, Boichard D (2013) Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. **PLoS One** 8:2013.

Gabriel SB, Schaffner SF, et al. (2002) The structure of haplotype blocks in the human genome. **Science** 296:2225-2229.

Garrick DJ, Fernando R (2014) Genomic prediction and genome-wide association studies in beef and dairy cattle. In.: Garrick J, Ruvinsky A (2nd ed) The Genetics of Cattle CABI, Wallingford, UK, p.474-501.

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Reviews: Genetics** 10:381-391.

Grossi DA, Berton MP, Buzanskas ME, Chud TCS, Grupioni NV, Paz CCP, Lôbo RB, Munari DP (2016) Genetic analysis on accumulated productivity and calving intervals in Nelore cattle. **Tropical Animal Health and Production** 48:207-210.

Haitina T, Lindblom J, Renstrom T, Fredriksson R (2006) Fourteen novel human members of mitochondrial solute carrier family 25 (SLC25) widely expressed in the central nervous system. **Genomics** 88:779-790.

Harris T, Marquez B, Suarez S, Schimenti J (2007) Sperm motility defects and infertility in male mice with a mutation in Nsun7, a member of the Sun domain-containing family of putative RNA methyltransferases. **Biology of reproduction** 77:376-382.

Hayes B, Goddard M. (2001) The distribution of the effects of genes affecting quantitative traits in livestock. **Genetics Selection Evolution** 33:209–29.

Hayes B, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. **Genome** 53:876-883.

Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. **PLoS Genetics** 6:e1001139

Hayes B (2013) Overview of statistical methods for genome-wide association studies (GWAS) In.: Gondro C, Van Der Werf JHJ, Hayes B Genome-wide association studies and genomic prediction Humana Press, Totowa, NJ, p.149-169.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. **Genetics Research** 38:209-216.

Hudson GFS, Van Vleck LD (1981) Relationship between production and stayability in Holstein cattle. **Journal of Dairy Science** 64:2246-2250.

Hyun TS, Rao DS, Saint-Dic D, Michael LE, Kumar PD, Bradley SV, Muzukami IF, Oravecz-Wilson KI, Ross TS (2004) HIP1 and HIP1r Stabilize Receptor Tyrosine Kinases and Bind 3-Phosphoinositides via Epsin N-terminal Homology Domains. **The Journal Of Biological Chemistry** 279: 14294-14306.

Irano N, Camargo GMF et al. (2016) Genome-wide association study for indicator traits of sexual precocity in Nelore cattle. **PloS one** 11:e0159502.

Jansen RPS (1984) Endocrine response in the fallopian tube. **Endocrine Reviews** 5:525-551.

Jia RZ, Zhang X, Hu P, Liu XM, Hua XD, Wang X, Ding HJ. (2012) Screening for differential methylation status in human placenta in preeclampsia using a CpG island plus promoter microarray. **International journal of molecular medicine** 30:133-141.

Johnson GC, Esposito L, Barratt BJ, et al. (2001). Haplotype tagging for the identification of common disease genes. **Nature genetics** 29:233.

Kemper KE, Goddard ME (2012) Understanding and predicting complex traits: knowledge from cattle. **Human molecular genetics** 21:R45-R51.

Kipp S, Segelke D et al. (2015) A new Holstein haplotype affecting calf survival **Interbull Bulletin** 49.

Koufariotis L, Chen YPP, Bolormaa S, Hayes BJ (2014) Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. **BMC genomics** 15:436

Li X, Wang J, Li W, Xu Y, Shao D, Xie Y, Xie W, Kubota T, Narimatsu H, Zhang Y (2011). Characterization of ppGalNAc-T18, a member of the vertebrate-specific Y subfamily of UDP-N-acetyl-α-D-galactosamine: polypeptide N-acetylgalactosaminyltransferases. **Glycobiology** 22: 602-615.

Li B, Luo H, Weng Q, Wang S, Pan Z, Xie Z, Wu W, Liu H, Li Q. (2016). Differential DNA methylation of the meiosis-specific gene FKBP 6 in testes of yak and cattle–yak hybrids. **Reproduction in Domestic Animals** 51:1030-1038.

Li Y, Li C, Lin S, et al. (2017) A nonsense mutation in Ccdc62 gene is responsible for spermiogenesis defects and male infertility in repro29/repro29 mice. **Biology of reproduction** 96:587-597.

Levine M, Tjian R (2003) Transcription regulation and animal diversity. **Nature** 424:147.

Lynch M, Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sunderland: Sinauer, p 980.

Luconi M, Cantini G, Baldi E, Forti G (2011) Role of a-kinase anchoring proteins (AKAPs) in reproduction. **Frontiers in Bioscience** 16:1315-1330.

Makinistoglu MP, Karsenty G (2015) The class II histone deacetylase HDAC4 regulates cognitive, metabolic and endocrine functions through its expression in osteoblasts. **Molecular metabolism** 4: 64-69.

Martinez-Royo A, Alabart JL, Sarto P, Serrano M, Lahoz B, Folch J, Calvo JH (2017) Genome-wide association studies for reproductive seasonality traits in Rasa Aragonesa sheep breed. **Theriogenology** 99:21-29.

Mattick JS (2001) Non-coding RNAs: the architects of eukaryotic complexity. **EMBO reports** 2:986-991.

Meggiolaro D, Porcelli F, Carnevali A, Crepaldi P, Savarese E, Ferrandi B (2006) A possible role of Fas antigen in ejaculated spermatozoa of fertile bulls: an immunocytochemical quantitative approach. **Acta histochemica** 107:463-468.

Melo TP, Takada L, Baldi F, Oliveira HN, Dias MM, Neves HHR, SchenkeL FS, Albuquerque LG, Carvalheiro R (2016) Assessing the value of phenotypic information from non-genotyped animals for QTL mapping of complex traits in real and simulated populations. **BMC Genetics** 17:89.

Melo TP, Camargo GMF, Albuquerque LG, Carvalheiro R (2017) Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nelore beef cows. **PloS one** 12:e0178551.

Metodiev S, Thekkoot DM, Young JM, Onteru S, Rothschild MF, Dekkers JCM (2018) A whole-genome association study for litter size and litter weight traits in pigs. **Livestock Science** 211:87-97.

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

Mighell AJ, Smith NR, Robinson PA, Markham AF (2000) Vertebrate pseudogenes. **FEBS letters** 468:109-114.

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH (2002) BLUPF90 and related programs (BGF90). In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier.

Miyamato T, Sato, H, Yogev L, Kleiman S, Namiki M, Koh E, Sakugawa N, Hayashi H, Ishikawa M, Lamb DJ, Sengoku K. (2006) Is a genetic defect in Fkbp6 a common cause of azoospermia in humans? **Cellular and Molecular Biology Letters**, 11:557–569.

Mokry FB, Higa RH, Mudadu MA, Lima AO, Meirelles SLC, Silva MVGB, Tullio RR (2013) Genome-wide association study for backfat thickness in Canchim beef cattle using Random Forest approach. **BMC Genetics** 14:47.

Nagai T, Mizuno K (2014) Multifaceted roles of Furry proteins in invertebrates and vertebrates. **The Journal of Biochemistry** 155:137-146.

Nalbant D, Youn H, Nalbant SI, Sharma S, Cobos E, Beale EG, Williams SC (2005) FAM20: an evolutionarily conserved family of secreted proteins expressed in hematopoietic cells. **BMC Genomics** 6:11.

Nascimento AV, Silva Romero ÂR, Utsunomiya YT, et al. (2018). Genome-wide association study using haplotype alleles for the evaluation of reproductive traits in Nelore cattle. **PloS one** 13:e0201876.

Nishi M, Aoyama F, Kisa F, et al. (2012) TRIM50 Protein Regulates Vesicular Trafficking for Acid Secretion in Gastric Parietal Cells. **Journal of Biological Chemistry** 287:33523-33532.

Noguchi J, Kobayashi E, Akiyama K, Kawai Y, Ozawa M, Ohnuma K, Kikuchi K, Kaneko H, Kunieda T (2004). Fine mapping of a region of rat chromosome 12 close to the aspermia (as) locus and comparison with the human orthologous regions. **Experimental animals** 53:429-435.

Nyman S, Duchemin SI, Koning DJ, Berglund B (2019) Genome-wide association study of normal and atypical progesterone profiles in Holstein-Friesian dairy cows. **Journal of dairy science** 102:3204-3215.

Oliveira PS, Cesar AS, Nascimento ML, Chaves AS, Tizioto PC, Tullio R R, Reecy JM (2014) Identification of genomic regions associated with feed efficiency in Nelore cattle. **BMC genetics** 15:100.

Oliveira Júnior GA (2017) **Associaçao genômica ampla (GWAS) aplicada a características reprodutivas de novilhas da raça Nelore**. 82 f. Tese (Doutorado em Biociência Animal) – USP, Pirassununga.

Parada-Bustamante A, Orihuela PA, Ríos M, Cuevas CA, Oróstica ML, Velásquez LA, Croxatto HBA (2010) Non-genomic signaling pathway shut down by mating changes the estradiol-induced gene expression profile in the rat oviduct. **Reproduction** 139: 631-644.

Pandey K, Mizukami Y, Watanabe K, Sakaguti S, Kadokawa, H (2017) Deep sequencing of the transcriptome in the anterior pituitary of heifers before and after ovulation. **Journal of Veterinary Medical Science** 16:0531.

Pei B, Sisu C, Frankish A, et al. (2012). The GENCODE pseudogene resource. **Genome biology** 13:R51.

Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME, Hayes BJA (2010) Validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. **Journal of Dairy Science** 93: 3331-3345.

Raudsepp T, Mccue ME, Das PJ, Dobson L, Vishnoi M, Fritz KL, Schaefer R, Rendahl AK, Derr JN, Love CC, Varner DD, Chowdhary BP (2012) Genome-wide association study implicates testis-sperm specific FKBP6 as a susceptibility locus for impaired acrosome reaction in stallions. **Plos Genetics** 8:e1003139.

Regatieri IC, Boligon AA, Baldi F, Albuquerque LG (2012) Genetic correlations between mature cow weight and productive and reproductive traits in Nelore cattle. **Genetics and Molecular Research: GMR** 2979-2986.

Reznichenko A, Sinkeler SJ, Snieder H, Van Den Born J, Borst MH, Damman J, Leuvenink HG (2013) SLC22A2 is associated with tubular creatinine secretion and bias of estimated GFR in renal transplantation **Physiological Genomics** 45:201-209.

Rubtsova MP, Vasilkova DP, Naraykina YV, Dontsova OA 2016 Peculiarities of yeasts and human telomerase RNAs processing. **Acta Naturae** 8:31.

Russell DL, Brown HM, Dunning KR (2015) ADAMTS proteases in fertility. **Matrix Biology** 44:54-63.

Saatchi M, Schnabel RD, Taylor JF, Garrick DJ (2014) Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. **BMC Genomics** 15:442.

Santana ML, Eler JP, Ferraz JBS, Mattos EC (2012) Genetic relationship between growth and reproductive traits in Nellore cattle. **Animal** 6:565-570.

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. **BMC Genomics** 15:478.

Schneider A, Hommel G, Blettner M (2010) Linear regression analysis—part 14 of a series on evaluation of scientific publications. **Pesquisa Agropecuária Brasileira** 107:76 - 82.

Shimizu T, Krebs S, Bauersachs S, Blum H, Wolf E, Miyamoto A (2010) Actions and interactions of progesterone and estrogen on transcriptome profiles of the bovine endometrium. **Physiological Genomics** 42:290-300.

Silva DO (2018) **Estudo de associação genômica para habilidade de permanência no rebanho na raça Nelore, considerando diferentes idades.** 43 f. Dissertação (Mestrado em Zootecnia) – Unesp, Jaboticabal.

Silva MR, Pedrosa VB, Silva JCB, Eler JP, Guimarães JD, Albuquerque LG (2011) Testicular traits as selection criteria for young Nelore bulls. **Journal of Animal Science** 89:2061-2067

Silva JAIIV, Eler JP, Ferraz JBS, Golden BL, Oliveira HN (2003) Heritability estimate for stayability in Nelore cows. **Livestock Production Science** 79:97-101.

Siripurapu V, Meth J, Kobayashi N, Hamaguchi M (2005) DBC2 significantly influences cell-cycle, apoptosis, cytoskeleton and membrane-trafficking pathways. **Journal of molecular biology** 346:83-89.

Springelkamp H, Iglesias AI, Cuellar-Partida G, et al. (2015) ARHGEF12 influences the risk of glaucoma by increasing intraocular pressure. **Human molecular genetics** 24:2689-2699.

Tempelman RJ (2015) Statistical and Computational Challenges in Whole Genome Prediction and Genome-Wide Association Analyses for Plant and Animal Breeding. J**ournal of Agricultural, Biological, and Environmental Statistics** 20:442-466.

Turner RM (2006) Moving to the beat: a review of mammalian sperm motility regulation**. Reproduction, Fertility, and Development**. 18:25–38.

Van Melis, MH, Eler JP, Rosa GJM, Ferraz JBS, Figueiredo LGG, Mattos EC, Oliveira HND (2010). Additive genetic relationships between scrotal circumference, heifer pregnancy, and stayability in Nelore cattle. Journal of Animal Science 88:3809-3813.

Vassilatis DK, Hohmann JG, Zeng H, et al. (2003). The G protein-coupled receptor repertoires of human and mouse. **Proceedings of the National Academy of Sciences** 100:4903-4908.

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

Visscher PM, Brown MA, Mccarthy MI, Yang J (2012) Five years of GWAS discovery. **The American Journal of Human Genetics** 90:7-24.

Wang W, Dong L, Saville B, Safe S (1999) Transcriptional activation of E2F1 gene expression by 17β-estradiol in MCF-7 cells is regulated by NF-Y-Sp1/estrogen receptor interactions. **Molecular Endocrinology** 13:1373-1387.

Weiner CP, Mason CW, Dong Y, Buhimschi IA, Swaan PW, Buhimschi CS (2010) Human effector/initiator gene sets that regulate myometrial contractility during term and preterm labor. **American journal of obstetrics and gynecology** 202: 474-e1.

Wu X, Fang M et al. (2013) Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. **BMC Genomics** 14:897.

Wu Y, Fan H, Wang Y, Zhang L, Gao X, Chen Y, Li J, Ren H, Gao H (2014) Genome-wide association studies using haplotypes and individual SNPs in Simmental cattle. **PLoS ONE** 9:e109330

Yokoo MJI, Albuquerque LG, Lôbo RB, Sainz RD, Carneiro Júnior JM, Bezerra LAF, Araujo FRC (2007) Estimativas de parâmetros genéticos para altura do posterior, peso e circunferência escrotal em bovinos da raça Nelore. Revista Brasileira de Zootecnia 36:1761-1768.

Yu, XX, Lewin DA, Zhong A, et al. (2001). Overexpression of the human 2-oxoglutarate carrier lowers mitochondrial membrane potential in HEK-293 cells:

contrast with the unique cold-induced mitochondrial carrier CGI-69. **Biochemical Journal** 353:369-375.

Zhou J, Liu X, Wang C, Li C (2018). The correlation analysis of miRNAs and target genes in metastasis of cervical squamous cell carcinoma. **Epigenomics**, 10:259-275.

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. **Nature Genetics** 44:821-824.

Zwart R, Verhaagh S, de Jong J, Lyon M, Barlow DP (2001) Genetic analysis of the organic cation transporter genes Orct2/Slc22a2 and Orct3/Slc22a3 reduces the critical region for the t haplotype mutant t w73 to 200 kb. **Mammalian Genome** 12:734–740.

7. Supplementary file

Table S1 Gene and gene name associated with age at first calving in a Nelore cattle population.

Gene	Gene name
ADAMTS17	ADAM Metallopeptidase With Thrombospondin Type 1 Motif 17
ARHGEF12	Rho Guanine Nucleotide Exchange Factor 12
ARHGEF13	Protein Kinase A-Anchoring Protein 13
B3GNTL1	UDP-GlcNAc:BetaGal Beta-1,3-N-Acetylglucosaminyltransferase Like 1
CCDC62	Coiled-Coil Domain Containing 62
DBNL	Drebrin Like
E2F7	E2F Transcription Factor 7
FAF1	Fas Associated Factor 1
FAM20C	FAM20C, Golgi Associated Secretory Pathway Kinase
FKBP6	FK506-binding protein 6
FRY	FRY Microtubule Binding Protein
GALNT18	Polypeptide N-Acetylgalactosaminyltransferase 18
GPR156	G Protein-Coupled Receptor 156
HDAC4	Histone Deacetylase 4
HIP1R	Huntingtin Interacting Protein 1 Related
ITPRIPL1 (KIAA1754L)	ITPRIP Like 1
LOC505479 (TTC41P)	Tetratricopeptide Repeat Domain 41
N4BP3	NEDD4 Binding Protein 3
NSUN5	NOP2/Sun RNA Methyltransferase Family Member 5
ORCT2 (SLC22A2)	Solute Carrier Family 22 Member 15
POM121C	POM121 Transmembrane Nucleoporin C
RMND5B	Required For Meiotic Nuclear Division 5 Homolog B
RUNDC3A	RUN Domain Containing 3A
SH3GL3	SH3 Domain Containing GRB2 Like 3, Endophilin A3
SLC25A39	Solute Carrier Family 22 Member 15
TMEM136	Transmembrane Protein 136
TRIM50	Tripartite Motif Containing 50
VPS37B	VPS37B, ESCRT-I Subunit
ZCCHC7	Zinc Finger CCHC-Type Containing

Table S2. Reproductive-related QTL (Quantitative Trait Loci) located within haplotype blocks associated with age at first calving of Nelore cattle.

Haplotype ID	Haplotype interval	QTL ID	QTL interval	Description
16312	14,326,657 - 14,331,430	QTL:10281	5:9511742-15032793	Non-return rate
17590	67,854,189 - 67,897,838	QTL:10570	.5:53310920-73099899	Ovulation rate
28861	95,451,584 - 95,452,399	QTL:106491	8:92459537-96625250	Calving ease (maternal)
32415	97,836,867 - 97,861,919	QTL:15182 QTL:15183 QTL:15184	9:88955721-98245564 9:88955721-98245564 9:88955721-98245564	Calving index Calving ease Stillbirth
42405	38,389,345 - 38,392,153	QTL:11385	13:17709118-39711868	Dystocia
42145	30,217,873 - 30,221,912	QTL:3569	13:771884-56839719	Heat intensity
50426	48,507,095 - 48,510,030	QTL:106504 QTL:11030	16:48342599-55862677 16:42505723-49512207	Calving ease (maternal) Scrotal circumference
54090	3,306,525 - 3,327,732	QTL:30538 QTL:11362	18:2605069-4271633 18:2437515-33011652	Calving index Stillbirth
55311	44,675,555 - 44,700,852	QTL:30537 QTL:4667	18:13346768-57589121 18:37815806-48003912	Birth index Stillbirth
57246	50,329,783 - 50,336,021	QTL:11368 QTL:11369	19:46184287-53032282 19:46184287-53032282	Dystocia Dystocia
59977	6,712,534 - 6,750,410	QTL:11389 QTL:29797	21:3044239-19763101 21:6735720-6735760	Stillbirth Age at puberty
59864	845,880 - 983,837	QTL:143907 QTL:143909 QTL:143919 QTL:143934 QTL:119782	21:903935-903975 21:917230-917270 21:943188-943228 21:961269-961309 21:8725-3028689	Calving ease Calving ease Calving ease Calving ease Heifer pregnancy
60641	25,062,833 - 25,121,384	QTL:9921	21:19915231-32302601	Percentage abnormal sperm
68538	31,408,989 - 31,409,857	QTL:63744 QTL:63758 QTL:4674 QTL:4675 QTL:9926 QTL:11214 QTL:63718 QTL:63754	25:31321194-32320535 25:31321194-32320535 25:27460729-35322340 25:27460729-35322340 25:19324607-36175299 25:27460729-35322340 25:31321194-32320535 25:31321194-32320535	Reproductive efficiency Age at first calving Calving ease Calf size Sperm average path velocity Scrotal circumference Reproductive efficiency Age at first calving

Table S3. Reproductive-related QTL (Quantitative Trait Loci) located within haplotype blocks associated with scrotal circumference of Nelore cattle.

Haplotype ID	Haplotype interval	QTL ID	QTL interval	Description
1450	44844961-44850305	QTL:11319 QTL:3439	1:17595955-46785676 1:25344370-70216318	Fertility treatments Conception rate
2040	65530942-65534765		1:25344370-70216318	
2042	65562631-65613386	QTL:3439 QTL:5659 QTL:5658 QTL:10644	1:62450179-72997940 1:48441931-87234102 1:58404741-69268709	Conception rate Non-return rate Non-return rate Scrotal circumference

5895	30883656-30925831	QTL:30597 QTL:30598 QTL:30599 QTL:30600 QTL:30601 QTL:30602 QTL:30603 QTL:30604 QTL:37608	2:30452407-32030774 2:30452407-32030774 2:30452407-32030774 2:30452407-32030774 2:30452407-32030774 2:30452407-32030774 2:30452407-32030774 2:30452407-32030774 2:30080851-32170510	Inhibin level Inhibin level Inhibin level Inhibin level Inhibin level Inhibin level Inhibin level Inhibin level Luteal activity
10356	43940802-43944009	QTL:11380 QTL:10695	3:34566854-53636529 3:40633135-56210430	Dystocia Calving ease
11837	96305979-96314088	QTL:30798 QTL:30484 QTL:15171 QTL:15172 QTL:15173 QTL:15174 QTL:30487	3:86454284-97532603 3:94364377-97462890 3:84669155-102513835 3:84669155-102513835 3:84669155-102513835 3:84669155-102513835 3:94148947-96323465	Dystocia Calving index Calf size Calving index Calving ease Stillbirth Calving ease
24569	40572826-40598898	QTL:4656 QTL:10795	7:26916745-42831622 7:33416156-42831622	Stillbirth Calving ease
28559	51553965-51593010	QTL:1718 QTL:10839 QTL:10838 QTL:4657 QTL:4658	8:48094361-60937432 8:48094361-58434111 8:48094361-58434111 8:36373569-58434111 8:36373569-58434111	Twinning Calving ease Calving ease Calving ease Stillbirth
32016	51304721-51323919	QTL:10293	9:30894099-74904058	Non-return rate
32030	51921230-51926215	QTL:1574	9:49563648-59664163	Calving ease
40844	29079800-29083909	QTL:15186 QTL:15187 QTL:15188 QTL:15189 QTL:11354	12:23991213-33037434 12:23991213-33037434 12:23991213-33037434 12:23991213-33037434 12:21356707-30181682	Calf size Birth index Calf size Calving ease Stillbirth

46072	27064286-27095501	QTL:5385 QTL:1695 QTL:1720	14:7423519-27796844 14:19715680-31877337 14:19715680-41376811	Gestation length Daughter pregnancy rate Twinning
48271	9437348-9441625	QTL:9924 QTL:10988 QTL:15815	15:5789135-25753553 15:5730206-10762815 15:9258332-10558420	Semen volume Scrotal circumference Male fertility
48953	31446062-31501499	QTL:30697 QTL:62409 QTL:62412 QTL:62407 QTL:62411 QTL:62410 QTL:62408 QTL:62414 QTL:62413	15:30452407-32030774 15:31321194-32320535 15:31321194-32320535 15:31321194-32320535 15:31321194-32320535 15:31321194-32320535 15:31321194-32320535 15:31321194-32320535 15:31321194-32320535	Luteinizing hormone level Inseminations per conception Interval from first to last insemination Fertility index Interval from first to last insemination Interval to first estrus after calving Inseminations per conception Non-return rate Non-return rate
48954	31558843-31590706			
48956	31595404-31599942			
48957	31619191-31621332			
48958	31647441-31665547			
49324	41764711-41772616	QTL:11000	15:36441038-51210066	Calving ease
58954	44743980-44754980	QTL:11092 QTL:11091	19:53032282-59447271 19:53032282-59447271	Calving ease Scrotal circumference
70716	34306509-34373088	QTL:4674 QTL:4675 QTL:9926 QTL:11214	25:27460729-35322340 25:27460729-35322340 25:19324607-36175299 25:27460729-35322340	Calving ease Calf size Sperm average path velocity Scrotal circumference
73665	29576824-29592610	QTL:11259 QTL:11260	27:21801052-31012979 27:22371748-31012979	Calving ease Scrotal circumference

CHAPTER 3 – USE OF HAPLOTYPES AND SNPS AS PREDICTORS FOR GENOMIC SELECTION OF REPRODUCTIVE TRAITS IN NELORE CATTLE

Abstract – The aim of this study was to compare the accuracy of genomic selection using as predictor single nucleotide polymorphism (SNP) or haplotype for reproductive traits, age at first calving (AFC) and scrotal circumference (SC) in Nelore cattle. The breeding programs provided the phenotypes and pedigree data: DeltaGen®, Cia do Melhoramento and CRV Lagoa. The dataset contained 258,305 records for AFC (h^2=0.08) and 463,919 records for SC (h^2=0.42). The phenotype was adjusted for fixed effects (Y*) and used as response variable in genomic analysis. Genotypes of 2,390 and 4,832 animals were available for AFC and SC respectively. Genotypes were obtained with the HD panel from Ilumina® (BovineHD - Illumina® bead chip - 777,962 SNPs) and GeneSeek Genomic Profiler HDi 75K (GeneSeek Inc., Lincoln, NE). Animals genotyped with lower density panel (75K) were imputed to the HD. The missing genotypes and the haplotypes phase were inferred with the Fimpute software. The confidence interval methodology and the linkage disequilibrium measurement D' were used to build the haplotypes in the software HaploView. A cross-validation with five folds was performed and GBLUP method was used for genomic prediction using haplotypes and SNPs as predictors. The accuracy and bias were used to compare the prediction models. The accuracy of prediction for AFC was 0.16 for both SNPs and haplotypes and for SC was 0.21 for SNPs and 0.22 for haplotypes. The use of haplotypes presented a slightly increment in the accuracy of prediction for SC. The regression coefficient for AFC was higher than one and for SC was less than one, showing that the prediction accuracy was underestimating for AFC and overestimating for SC. Similar regression coefficient values were found for both traits when compared haplotypes and SNPs. The heritability coefficient affected the prediction accuracies, causing slight increase using haplotypes for trait of high heritability and equal accuracy between predictors for trait of low heritability.

Keywords: age at first calving; genomic breeding values; GBLUP; scrotal circumference.

The regression coefficient for AFC was >1 and for SC was <1, showing that the genomic prediction was overestimating for AFC and underestimating for SC

1. Introduction

The profitability of beef cattle production depends on several factors, such as productive and reproductive performance of herds. Regarding reproductive factors, they have a direct impact in the success of beef cattle production (Brumatti et al., 2011), since early animals produce a faster economic return and decrease the interval of generations (Pirlo et al., 2000).

The age at first calving (AFC) is an indicator trait of sexual precocity for females. This trait has been included in breeding programs due to economic importance (Brumatti et al., 2011), for being easily measured in cows and it does not to represent an additional cost to be obtained. Decreasing the AFC reduces the production costs in raising the animals and anticipates the cow's productive life (Perotto et al., 2006). Besides the importance of AFC for beef cattle production, a direct selection for AFC presents some bottlenecks because this trait generally presents low heritability (Boligon et al., 2010; Boligon e Albuquerque, 2011; Espigolan, 2017) and it is considered a complex trait, since many Quantitative Trait Loci (QTLs) with small effects control it.

Scrotal circumference (SC) is a fertility trait and development indicator for male and an indirect indicator of fertility for females due to its favorable genetic correlation (-0.22 to -0.42) with AFC, heifer pregnancy (0.26 to 0.43) (Terakado et al., 2015) and stayability (0.19) (Van Melis et al, 2010). This trait presents medium to moderate heritability (Forni e Albuquerque, 2005; Yokoo et al., 2007; Boligon et al., 2010; Laureano et al., 2011; Marques et al., 2013) and positive association with other important traits as final weight (Yokoo et al., 2007) and conformation, precocity and muscling at yearling (Boligon et al., 2007).

The best linear unbiased prediction model (BLUP) proposed by Henderson (1975) is a traditional approach widely used in animal breeding programs to predict breeding values (EBV) of individuals candidates for selection. In these breeding programs, several economically important traits as growth, morphological and reproductive traits are included as selection criteria. However, traditional selection presents some limitations, especially for sex-limited traits or/and for traits of lately or expensive measure and with low heritability (Hayes et al., 2009). Although traits like AFC and SC can be easily collected in many animals, they are sex-limited, and AFC has low heritability.

With the development of single nucleotide polymorphisms (SNPs) distributed over the whole genome of several livestock species, the selection of the animals by their genomic breeding values (GEBVs) became possible. This approach proposed by Meuwissen et al. (2001) known as genomic selection (GS), estimates the GEBVs as a function of the average of markers effects for each individual across its whole genome.

Markers effects can be estimated by two approaches in a GS model, individually (Boichard et al., 2016) or in blocks of neighbors SNPs presenting high linkage disequilibrium (LD) among them, which are known as haplotypes. GS studies using SNP haplotypes have been widely performed in the last decade with real (Cuyabano et al., 2015; Hess et al., 2017; Karimi et al., 2018) and simulated (Calus et al., 2008; Cuyabano et al., 2014, Villumsen et al., 2009) data. Some authors have reported an increase of the prediction accuracy by using haplotypes. Higher accuracies cam be reached because haplotypes are more informative than individual SNPs, are in a higher LD with the QTLs (Meuwissen et al., 2014), capture better identical by descent relationships (Hickey et al. 2013; Ferdosi et al., 2016) and detect epistatic interaction effects between markers located in the same locus (Clark, 2004; Bardel et al., 2005).

The most GS studies for reproductive traits in beef cattle have used individual SNPs to estimate the breeding values of the animals. Nevertheless, using haplotypes to perform GS for reproductive traits in beef cattle are still scarce. Thus, the aim of this study was to compare the accuracy of genomic selection using individual SNP and haplotype for traits of AFC and SC in Nelore cattle.

2. Material and methods

2.1 Ethical approval

Procedures were approved by the Animal Care of the School of Agricultural and Veterinary Science, São Paulo State University (UNESP) Ethical Committee (protocol No. 18.340/16).

2.2. Phenotype and genotype data

DeltaGen®, Cia do Melhoramento and CRV Lagoa / PAINT® Nelore cattle breeding programs provided the phenotypes and pedigree data for AFC and SC.

According to the breeding program criteria, three mating systems were used: artificial insemination, controlled breeding and multiple-sire mating system. Regardless the body condition score and the body weight, all the heifers were exposed to reproduction with an age of 14 to 16 months in a breeding season

Two breeding seasons were performed: a first season of approximately 60 days (from February to March) and another season beginning in the second half of November and last approximately 70 days, where all the females participate. The pregnancy confirmation was made by rectal palpation, 60 days after finishing the breeding season.

The AFC, measured in days, was obtained as the difference between the female birth and her first calving data. The SC was measured in centimeter and collected at yearling.

Contemporary groups (CG) for both traits were defined by birth, weaning and yearling farm, management group at weaning and yearling and year of birth. GC with less than three animals and phenotype data outside the intervals given by the mean ±3 standard deviations were excluded from the analysis.

The phenotypes (y*) were adjusted for fixed effect of CG for AFC and SC. In addition, the age of the animal at the measurement was considered as covariable for SC. The y* and the heritability were estimated using the programs from the family BLUPF90 (Misztal et al., 2002). A descriptive statistics and heritability coefficient are summarized in Table 1.

Table 1. Descriptive statistics and heritability estimates (h^2) for age at first calving (AFC) and scrotal circumference (SC) in Nelore cattle.

Trait	h^2	Mean ± SD	Min	Max	Nphe	CG
AFC (day)	0.08	1,042 ± 111.58	634	1250	258,305	15,258
SC (cm)	0.42	26.25 ± 3 .90	15.12	38.50	463,919	17,627

h^2 = heritability; Mean = mean of phenotype; SD = standard deviation, Min= minimum; Max= Maximum; N_{phe} = number of animals with phenotypes; CG = number of contemporary groups.

A total of 1900 heifers and 1500 young bulls were genotyped with the Illumina BovineHD BeadChip assay (770k, Illumina Inc., San Diego, CA, USA) and 719 heifers and 3341 young bulls were genotyped using the GeneSeek Genomic Profiler HDi 75K (GeneSeek Inc., Lincoln, NE). Animals genotyped with lower density panel (75K) were imputed to the HD panel using FImpute software v.2.2

(Sargolzaei et al., 2014). A quality control (QC) was performed in the genomic data maintaining only autosomal SNPs, with minor allele frequency (MAF) > 0.02, Hardy-Weinberg equilibrium p-value >10-5, SNP call rate > 92% and a call rate > 90% for the samples. After quality control, 2,390 heifers genotyped with 407,158 SNP markers and 4,832 young bulls genotyped with 412,424 SNP markers remained in genomic dataset.

2.3 Haplotypes construction

The missing genotypes and the haplotype phase were determined using the software FImpute (Sargolzaei et al., 2014). The linkage disequilibrium (LD) was estimated using D' following Hill (1981):

$$D' = \begin{cases} \dfrac{D}{\min\{freq.A^{*}freq.b, freq.a^{*}freq.B\}} \; if \quad D > 0, \\ \dfrac{D}{\min\{freq.A^{*}freq.B, freq.a^{*}freq.b\}} \; if \;\; D < 0 \end{cases}$$

where D = freq.AB – (freq.A x freq.B). The freq. A, freq. a, freq. b and freq. B are the frequencies of alleles A, a, b and B, respectively, and freq. AB, are the frequencies of haplotypes AB in the population. Then, if two loci are independent, the AB haplotype frequency (freq. AB) is calculated as the product between freq. A and freq. B.

The haplotype blocks were built using the software HaploView (Barrett et al., 2005) and their definition considered that a pair of SNP markers had a strong LD, i.e. a D' ranging of 0.70 to 0.98 (Gabriel et al., 2002). An incidence matrix was constructed with the alleles (variations) of each haplotype, and the number of copies that every subject possessed of the allele *I* in the haplotype *j*. Therefore, depending on the number of copies of the allele (paternal and maternal), the alleles were encoded as 0, 1 or 2. The haplotype matrix has $n \times m$ dimension, being n the total of animals in the analysis and m the number of haplotype alleles.

2.4 Genomic prediction

The cross-validation with animals of known phenotypes and genotypes was used to evaluate the genomic prediction ability using SNPs and haplotypes as markers. The dataset was randomly divided into five folds of equal size. Four subsets were used as training population to estimate the markers effect and the other one as validation population to estimate the direct genomic value (DGV) using the information of the training population. This process was replicated five times.

The BLUPF90 family (Misztal et al.,2002) software was used for the analyses. The methodology used to estimate the markers effects (SNPs and haplotypes) was the Genomic Best Linear Unbiased Predictor (GBLUP). This model is similar to BLUP but use a genomic relationship matrix (**G**) instead of **A**, resulting in DGV prediction based on the markers effects. The model used in this approach can be shown below:

$$\boldsymbol{y}^* = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{Zg} + \boldsymbol{e},$$

Where $\boldsymbol{y}^*$ is the vector of phenotypes adjusted for fixed effects, 1 is a vector of ones, $\boldsymbol{\mu}$ is the overall mean, $\boldsymbol{Z}$ is an incidence matrix of markers effects, $\boldsymbol{g}$ is a vector of animal additive genetic effects and $\boldsymbol{e}$ is a vector of residual effects. It was assumed that $\boldsymbol{g}$ ~ N(0,$\mathbf{G}\sigma_g^2$), where σ_g^2 is the variance of markers and **G** is the genomic relationship matrix. Random residuals were assumed e ~ (0, $\mathbf{I}\sigma_e^2$), where **I** is an identity matrix and σ_e^2 is the residual variance. The **G** matrix was created according to VanRaden (2008) as:

$$G = [(M - P)(M - P)']/[2\sum_{j=1}^{m} P_j(1 - P_j)]$$

M is a *n x m* marker matrix (SNPs or haplotypes), where n is the total number of genotyped animals and m is the total number of markers; **P** is a matrix containing two times the observed frequency of the second allele P_j. Elements of **M** are 0, 1 or 2 depending on the number of copies of the alleles in each animal.

The DGVs were predicted from the estimated markers effects using the following formula:

$$DGV = \sum_{i}^{n} X_i \, g_i$$

where *n* is the number of markers, X_i is the incidence markers matrix for all individuals, and g_i is the vector of effects for each marker (SNPs or haplotypes).

The models were evaluated using the following criteria in the animals of the validation population:

1) Pearson's correlation between DGV and Y* was used to calculate the accuracy of prediction.

2) The bias in the prediction of DGVs was estimated by the regression coefficient of Y* on predicted DGVs (b). This measures the degree of inflation or deflation of the genomic prediction in relation to the Y*. A regression coefficient > 1 indicates that the model is underestimating DGVs and a value < 1 indicates an overestimation of DGVs. Estimates of regression close to 1 indicate that the predictor is in the same scale as the Y*.

3. Results and discussion

A total of 407,158 and 412,424 SNP markers remaining after the quality control, however, approximately the 84% (342,626 and 344,933) were used to construct 75,069 and 77,283 haplotypes for AFC and SC, respectively. In this study, 93% of the haplotypes length ranged between 2 and 10 SNPs, with an average of 4.56 and 4.46 SNPs per haplotype for AFC and SC respectively (Table 2). Generally, short haplotypes correspond to segments of common ancestral chromosomes shared in a population where has occurred recombination over the time (Ong et al., 2011). Small length haplotypes also could be the result of long selection processes within a line, while larger haplotypes can be generated by the re-use of particular sires or also for due recent population events such as crossing and migration (Villumsen et al., 2009).

Villumsen et al. (2009), Hess et al. (2017) and Karimi et al. (2018) previously reported a relationship between the haplotype length and the prediction accuracy. In a simulated study, Villumsen et al., (2009) tested different haplotype size (2, 5, 10, 20 and 40 SNPs) and found that 10 SNPs haplotypes achieved the betters

predicted accuracy across all traits. Karimi et al. (2018) compared the prediction accuracy employed SNPs and haplotypes, and reported slightly greater prediction accuracies using haplotypes with 5 and 10 SNPs while Hess et al. (2017) found that haplotypes constructed with less than 8 haplotypes resulted in highest prediction accuracies when compared with the SNPs approached.

A high alleles combination was found, ranging from 2 to 141 for AFC and 2 for 209 for SC. The total of alleles for AFC and SC were 483,933 and 570,064, respectively, which are higher than the 366,167 found by Cuyabano et al. (2014) and 325,269 by Feitosa (2018). The large number of alleles found in our study can indicate that the number of genotyped animals is a representative sample of all possible haplotypes of the entire population. Notwithstanding, the increase in the number of alleles leads to an increase in the number of effects that must be estimated (Hayes et al., 2007) and consequently in the computation time (Hess et al., 2017).

Table 2. Summary of the genomic information available for age at first calving (AFC) and scrotal circumference (SC) in Nelore cattle.

Trait	N_{HAP}	N_{SNPs}	N_{AL}	SNPs per haplotype			Alleles per haplotype		
				Min	Max	Mean	Min	Max	Mean
AFC	75,069	342,626	483,933	2	104	4.56	2	141	6.45
SC	77,283	344,933	570,064	2	104	4.46	2	209	7.38

N_{HAP}= number of haplotypes; N_{SNPs}= number of SNPs used to construct haplotypes; N_{AL} = number of haplotypes alleles.

The highest number of alleles was found in chromosome 1, while the smallest number of alleles for AFC and SC was harbored in chromosome 28 and 25, respectively (Table 3). The means of allele per haplotype in this study (AFC = 6.45 and SC = 7.38) were higher than found in Nordic Holsten population (5.10) (Cuyabano et al., 2014) and that reported in simulated data (4.33) by Arce (2018).

The prediction accuracies for AFC (0.16) using SNPs and haplotypes as predictor were similar, while for SC, the accuracy of prediction was slightly higher when we used haplotypes as predictor (Table 4). Past reports have also shown that the gain in accuracy in genomic predictions using haplotypes, independent of

construction method of haplotypes, were small (Edriss et al. 2013; Meuwissen et al. 2014; Jonas et al. 2016; Cuyabano et al. 2015; Karimi et al. 2018). This slightly improvent can be due SNPs are biallelic and they have lower information content (Karimi et al. 2018), while haplotypes are generally multiallelic (Ryynänen et al. 2007). Therefore, haplotype may better capture LD between multi-allelic QTL than the SNP markers.

Table 3. Number of haplotypes, SNPs and alleles for age at first calving (AFC) and scrotal circumference (SC) in a population of Nelore cattle.

Chr	AFC			SC		
	N_{HAP}	N_{SNPs}	N_{AL}	N_{HAP}	N_{SNPs}	N_{AL}
1	4,621	21,077	29,506	4,782	21,424	34,644
2	3,968	18,226	25,520	4,096	18,658	30,575
3	3,701	17,910	24,861	3,789	17,771	29,224
4	3,480	15,090	21,599	3,601	15,537	25,382
5	3,207	15,619	21,869	3,311	15,665	25,676
6	3,677	18,400	26,034	3,784	18,588	30,632
7	3,253	15,486	21,192	3,373	15,877	26,052
8	3,536	17,382	24,296	3,670	17,398	28,765
9	3,261	15,617	22,319	3,354	15,745	26,061
10	2,921	13,011	18,642	2,995	13,122	21,960
11	3,015	14,322	19,110	3,166	14,553	23,000
12	2,519	11,254	16,517	2,626	11,389	19,207
13	2,521	11,209	15,752	2,600	11,284	18,649
14	2,791	13,265	18,797	2,834	13,310	22,041
15	2,596	11,041	16,054	2,665	11,102	18,487
16	2,484	11,246	16,131	2,510	11,211	18,610
17	2,447	11,224	15,839	2,546	11,010	18,245
18	1,897	9,178	12,591	1,948	9,125	14,907
19	1,837	7,645	10,886	1,890	7,693	12,849
20	2,129	9,514	13,478	2,193	9,555	15,721
21	2,193	9,809	13,664	2,227	9,869	16,184
22	1,807	8,037	11,074	1,850	8,047	13,383
23	1,705	6,852	10,220	1,717	6,793	11,784
24	1,907	8,746	12,684	1,978	8,703	14,522
25	1,407	6,108	8,528	1,434	6,101	10,135
26	1,711	7,383	10,584	1,753	7,397	12,534
27	1,486	5,968	8,789	1,513	6,023	10,358
28	1,454	5,842	8,523	1,500	5,866	10,155
29	1,538	6,165	8,874	1,578	6,117	10,322
Total	75,069	342,626	483,933	77,283	344,933	570,064

Chrm= chromossome, N_{HAP} = number of haplotypes, N_{SNPs} = number of SNPs; N_{AL} = number of alleles

Table 4 Genomic prediction accuracy and bias of SNPs and haplotypes for age at first calving (AFC) and scrotal circumference (SC) traits in Nelore cattle

Trait	ACC_{SNPs} ± SD	ACC_{HAP} ± SD	b_{SNPs}	b_{HAP}
AFC	0.16± 0.03	0.16 ± 0.03	1.30	1.25
SC	0.21 ± 0.03	0.22 ± 0.03	0.62	0.63

ACC_{SNPs} = Genomic prediction accuracy for SNPs; ACC_{HAP} = Genomic prediction accuracy for haplotypes; b_{SNPs} = Genomic prediction bias for SNPs; b_{HAP} = Genomic prediction bias for haplotypes

Cuyabano et al. (2015) compared haplotypes and SNPs for different traits in dairy cattle and concluded that for most of the analyzed traits the prediction accuracy was only slightly better (differences of 0.01) for the based haplotype approach. However, according to the same authors, the genetic progress is linearly related to the accuracy of genetic evaluation, so a small improvement in accuracy of prediction is considered important for cattle breeding.

Regardless the haplotype construction method, training population size, heritability of the trait, number of alleles per haplotype and the estimation marker effects approach, the results found in the literature have been similar to the found in the present study. However, one important factor determining the prediction accuracy is the heritability of the trait, i.e. the trait with low heritability (AFC=0.08) did not show difference of predictive ability between predictors, while trait of higher estimation of heritability (SC=0.42) obtained a slighty increase in accuracy when haplotypes approach was used. According to Cuyabano et al. (2014) is difficult to improve prediction accuracy using haplotypes for traits with low heritability. Karimi et al. (2018) studied 57 traits with different heritability in Holstein cattle using GBLUP method and SNPs and haplotypes as predictors and they also related that there was a slight increase in predicted reliabilities for traits with medium to moderate heritability when haplotypes was compared with SNP.

Between traits, we also observed that the accuracy was affected by estimated heritability, i.e. AFC has the lowest heritability and consequently the lower prediction accuracy (0,16) when compared with the accuracy of SC (0,21 – 0,22) that has a heritability of 0,42. Costa et al. (2019), studying others reproductive traits (heifer rebreeding, age at first calving, and occurrence of early

pregnancy) in Nellore, found similar results showed that, independent of approach to estimate genomic breeding value , the prediction accuracy is directly proportional to the heritability of the trait.

Other factor that can affect the prediction accuracy is the size of training population. According to Hess et al. (2017) the haplotype is more sensitive to small training data than SNP individual. So, our results agree with this, since our largest training population was to SC trait, that obtained higher accuracy with haplotypes comparing with AFC trait, which did not differ between predictor and has the smallest training population.

The regression coefficients (Table 4) ranged from 0.62 to 1.30 for SC and AFC respectively. The results indicated that the genomic prediction for AFC was underestimated, while for SC was overestimated. Between traits, regression coefficients using SNPs or haplotype had similar values. Cuyabano et al. (2014) found similar results, not difference between SNPs and haplotypes were observed by authors. On the other hand, Calus et al (2008) and Karimi et al. (2018) using simulation and real data, respectively, observed that haplotype information could reduce bias.

With respect to computation time, there was no difference between the two analyses (haplotype and SNP) of genomic selection, but an extra time was spent on constructing haplotypes.

4. Conclusion

Prediction accuracies using SNPs and haplotypes approaches for AFC were equal and for SC, the gains in accuracy of genomic prediction using haplotypes in a GBLUP approach was small.

The heritability coefficient affected the prediction accuracies, causing slight increase using haplotypes for trait of high heritability and equal accuracy between predictors for trait of low heritability.

5. References

Arce CDS (2018) **Comparação de métodos de construção de haplótipos em estudo de associação genômica ampla com dados simulados**. 43 f. Dissertaçao (Mestrado em Genética e Melhoramento Animal) - Unesp, Jaboticabal.

Bardel C, Danjean V, Hugot JP, Darlu P, Genin E (2005) On the use of haplotype phylogeny to detect disease susceptibility loci. **BMC Genetics** 6:24.

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

Boichard D, Ducrocq V, Croiseau P, Fritz S. (2016) Genomic selection in domestic animals: Principles, applications and perspectives **C. R. Biologies** 339: 274–277

Boligon AA, Rorato PRN, Albuquerque LG (2007) Correlações genéticas entre medidas de perímetro escrotal e características produtivas e reprodutivas de fêmeas da raça Nelore. **Revista Brasileira de Zootecnia** 36:565-571.

Boligon AA, Silva JAV, Sesana RC, Sesana JC, Junqueira, JB, Albuquerque LG (2010) Estimation of genetic parameters for body weights, scrotal circumference, and testicular volume measured at different ages in Nelore cattle. **Journal of animal science** 88:1215-1219.

Boligon AA, Albuquerque LG (2011) Genetic parameters and relationships of heifer pregnancy and age at first calving with weight gain, yearling and mature weight in Nelore cattle. **Livestock Science** 141:12–16

Brumatti RC, Ferraz JBS, Eler JP, Formigonni IB (2011) Desenvolvimento de índice de seleção em gado corte sob o enfoque de um modelo bioeconômico. **Revista archivos de zootecnia** 60:205-213

Calus MMPL, Meuwissen THE, de Roos APW, Veerkamp RF (2008) Accuracy of prediction using different methods to define hapltypes. **Genetics** 178: 553–561

Clark AG (2004) The role of haplotypes in candidate gene studies. **Genetic Epidemiology** 27:321–333.

Costa R, Irano N, et al. (2019) Prediction of genomic breeding values for reproductive traits in Nelore heifers. **Theriogenology** 125: 12-17

Cuyabano BC, Su G, Lund MS (2014) Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. **BMC Genomics** 15: 1171.

Cuyabano BC, Su G, Lund MS. (2015) Selection of haplotype variables from a high-density marker map for genomic prediction. **Genetics Selection Evolution** 47:61.

Edriss V, Fernando RL, Su G, Lund MS, Guldbrandtsen B (2013) The effect of using genealogy-based haplotypes for genomic prediction. **Genetics Selection Evolution** 45: 5.

Espigolan R (2017) **Parametric and semi-parametric models for predicting genomic breeding values of complex traits in Nelore cattle**. 78 f. Tese (Doutorado em Genética e Melhoramento Animal) - Unesp, Jaboticabal

Feitosa FLB (2018) **Associação e seleção genômica para perfil de ácidos graxos utilizando haplótipos como marcadores em bovinos Nelore.** 90 f. Tese (Doutorado em Genética e Melhoramento Animal) - Unesp, Jaboticabal.

Forni S, Albuquerque LG (2005) Estimates of genetic correlations between days to calving and reproductive and weight traits in Nelore cattle. **Journal of animal science** 83:511-1515.

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

Gabriel SB, Schaffner SF, et al. (2002) The structure of haplotype blocks in the human genome. **Science** 296:2225-2229.

Hayes BJ (2009) Whole Genome Association and Genomic Selection Jaboticabal, SP: FUNEP, 120p (Course Notes).

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model **Biometrics** 31:423–447.

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 JM, Kinghorn BP, Tier B, Clark SA, van der Werf JHJ, Gorjanc G (2013) Genomic evaluations using similarity between haplotypes. **Journal of Animal Breeding and Genetics**. 130: 259–269.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. **Genetics Research** 38:209-216.

Jonas D, Ducrocq V, Fouilloux MN, Croiseau P (2016) Alternative haplotype construction methods for genomic evaluation. **Journal of Dairy Science** 99: 4537–4546.

Karimi Z, Sargolzaei M, Robinson JAB, Schenkel FS (2018) Assessing haplotype-based models for genomic evaluation in Holstein cattle. **Canadian Journal of Animal Science** 98: 750–759

Laureano MMM, Boligon AA, Costa RB, Forni S, Severo JLP, Albuquerque LGD (2011) Estimativas de herdabilidade e tendências genéticas para características de crescimento e reprodutivas em bovinos da raça Nelore: Estimates of heritability and genetic trends for growth and reproduction traits in Nelore cattle. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, 63:143-152.

Marques EG, Magnabosco CU, Lopes FB, Silva MC (2013) Estimativas de crescimento, carcaça e perímetro escrotal de animais confinamento. **Bioscience Journal** 29:159–167.

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

Meuwissen THE, Odegard J, Andersen-Ranberg I, Grindflek E (2014) On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. **Genetics Selection Evolution** 46:49.

Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH (2002) BLUPF90 and related programs (BGF90). In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier.

Ong RTH, Liu X, Poh WT, Sim X, Chia KS, Teo YY (2011). A method for identifying haplotypes carrying the causative allele in positive natural selection and genome-wide association studies. **Bioinformatics** 27: 822-828.

Perotto D, Miyagi Ap, Souza Jc, Moletta Jl, Freitas Ja (2006) Estudos de características reprodutivas de animais da raça Canchim, criados a pasto, no estado do Paraná, Brasil. **Archives of Veterinary Science** v. 11:1-6.

Pirlo G, Miglior F, Speroni M (2000) Effect of age at first calving on production traits and on difference between milk yield returns and rearing costs in Italian Holsteins. **Journal of Dairy Science** 83:603-608.

Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR (2007). A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (Salmo salar). **Journal of Heredity** 98:692–704.

Sargolzaei M, Chesnais JP, Schenkel FS (2011) FImpute - An efficient imputation algorithm for dairy cattle populations. **Journal of Animal Science** 94:421.

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives **BMC Genomics** 15:478.

Terakado APN, Boligon AA, Baldi F, Silva JAIIV, Albuquerque LG (2015) Genetic associations between scrotal circumference and female reproductive traits in Nelore cattle. **Journal of Animal Science** 93:2706–2713.

Van Melis, MH, Eler JP, Rosa GJM, Ferraz JBS, Figueiredo LGG, Mattos EC, Oliveira HND (2010). Additive genetic relationships between scrotal circumference, heifer pregnancy, and stayability in Nelore cattle. **Journal of Animal Science** 88:3809-3813.

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

Villumsen TM, Janss L, Lund MS (2009) The importance of haplotype length and heritability using genomic selection in dairy cattle. **Journal of Animal Breeding and Genetics** 126:3–13.

Yokoo MJI, Albuquerque LG, Lôbo RB, Sainz RD, Carneiro Júnior JM, Bezerra LAF, Araujo FRC (2007) Estimativas de parâmetros genéticos para altura do

posterior, peso e circunferência escrotal em bovinos da raça Nelore. **Revista Brasileira de Zootecnia** 36:1761-1768.