FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
UNIVERSIDADE ESTADUAL PAULISTA
CÂMPUS DE JABOTICABAL

GENOME-WIDE ASSOCIATION STUDY OF REPRODUCTION TRAITS IN NELORE CATTLE, INCLUDING ADDITIONAL PHENOTYPIC INFORMATION FROM NON-GENOTYPED ANIMALS

Thaise Pinto de Melo
Zootecnista

2015

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
UNIVERSIDADE ESTADUAL PAULISTA
CÂMPUS DE JABOTICABAL

GENOME-WIDE ASSOCIATION STUDY OF REPRODUCTION TRAITS IN NELORE CATTLE, INCLUDING ADDITIONAL PHENOTYPIC INFORMATION FROM NON-GENOTYPED ANIMALS

Thaise Pinto de Melo
Advisor: Dr. Roberto Carvalheiro
Co-Advisor: Prof. Dra. Lucia Galvão de Albuquerque

Dissertation presented to The Faculdade de Ciências Agrárias e Veterinárias Unesp, Câmpus de Jaboticabal, in partial fulfilment of requeriments for the degree of Mestre em Genética e Melhoramento Animal
(*Master in Animal Breeding and Genetics*)

2015

Melo, Thaise Pinto de
M528g Genome-wide association study of reproduction trais in Nelore cattle, including additional phenotypic information from non-genotyped animals / Thaise Pinto de Melo. – – Jaboticabal, 2015
xi, 73 p. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2015
Orientador: Roberto Carvalheiro
Banca examinadora: Idalmo Garcia Pereira, Henrique Nunes de Oliveira
Bibliografia

1. Age at first calving. 2. Heifer rebreeding. 3. GWAS. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.2

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação – Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

unesp UNIVERSIDADE ESTADUAL PAULISTA

CAMPUS DE JABOTICABAL

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS DE JABOTICABAL

CERTIFICADO DE APROVAÇÃO

TÍTULO: GENOME-WIDE ASSOCIATION STUDY OF REPRODUCTION TRAITS IN NELORE CATTLE, INCLUDING ADDITIONAL PHENOTYPIC INFORMATION FROM NON-GENOTYPED ANIMALS

AUTORA: THAISE PINTO DE MELO

ORIENTADOR: Prof. Dr. ROBERTO CARVALHEIRO

CO-ORIENTADORA: Profa. Dra. LUCIA GALVAO DE ALBUQUERQUE

Aprovada como parte das exigências para obtenção do Titulo de MESTRE EM GENÊTICA E MELHORAMENTO ANIMAL , pela Comissão Examinadora:

Prof. Dr. ROBERTO CARVALHEIRO
Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. HENRIQUE NUNES DE OLIVEIRA
Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. IDALMO GARCIA PEREIRA
Universidade Federal de Minas Gerais / Belo Horizonte/MG

Data da realização: 26 de fevereiro de 2015.

DADOS CURRICULARES DO AUTOR

THAISE PINTO DE MELO - was born in Natal, Rio Grande do Norte, Brazil, on February 16th 1991, as single child of Maria Necy de Melo and José Coelho Pinto. She started her undergraduate in Zootecnia (Animal Science) at Federal University of Rio Grande do Norte, Natal, on February 2008. During her undergraduate course she had involvement in university extension and scientific projects as internship or volunteer member. She started her animal breeding studies in one of these projects, which was focused in survival analyses of simulated data, under advice of Prof. Elizângela Emídio Cunha. After graduating, she started a M.Sc. within postgraduate program in Animal Breeding and Genetics, receiving financial support from Capes and FAPESP. In her second year of Master course she conducted a research project in Canada under advice of Prof. Flavio Schramm Schenkel, receiving financial support from FAPESP. In February 2015, she defended her Master dissertation under advice of Dr. Roberto Carvalheiro and Prof. Lucia Galvão de Albuquerque.

Epigraphy

Quando você partir, em direção a Ítaca,
que sua jornada seja longa
repleta de aventuras, plena de conhecimento.

Não tema Laestrigones e Cíclopes
nem o furioso Poseidon;
você não irá encontrá-los durante o caminho,
se você não carrega-los em sua alma,
se sua alma não os colocar diante de seus passos.

Espero que sua estrada seja longa.
Que sejam muitas as manhãs de verão,
e que o prazer de ver os primeiros portos
traga uma alegria nunca vista.
Procura visitar os empórios da Fenícia
e recolha o que há de melhor.
Vá as cidades do Egito,
e aprenda com um povo que tem tanto a ensinar.

Não perca Ítaca de vista,
pois chegar lá é o seu destino.
Mas não apresse os seus passos;
é melhor que a jornada demore muitos anos
e seu barco só ancore na ilha
quando você já estiver enriquecido
com o que conheceu no caminho.

Não espere que Ítaca lhe dê mais riquezas.
Ítaca já lhe deu uma bela viagem;
sem Ítaca, você jamais teria partido.
Ela já lhe deu tudo, e nada mais pode lhe dar.

Se, no final, você achar que Ítaca é pobre,
não pense que ela lhe enganou.
Porque você tornou-se um sábio, e viveu uma vida intensa,
e este é o significado de Ítaca.

Konstantinos Kaváfis

DEDICATION

To God, for the omnipresence in my life, for hearing me, advising me and being with me in all moments. To my family, specially my parents for their support and care, for understanding my decisions and for giving me support for the realization of my dreams, I dedicate this dissertation.

ACKNOWLEDGEMENTS

I would like to thank God for too many things, so I just say, thanks God for everything. You are wonderful and merciful.

I thank my parents for all support.

I thank Leandro for the patience in these two years, for hearing and helping me in many moments.

I thank my childhood friends, especially Brenda and Heloisa for their friendship and fellowship. My “Zoo friends” from UFRN, specially Alane and Thayana, who are great friends until today. My “salinha’s friends” and friends from exact sciences (FCAV), for supporting me and for the interesting discussions. My Jaboticabal’s friends, Kamila, Giovana (and family) and Luciana, for friendship and support me.

I wish to thank to all friends and colleagues I met during this master course.

I would like to express my gratitude to all professors of postgraduate program on Genetics and Animal Breeding at FCAV (Genética e Melhoramento Animal), for all transmitted teaching.

I thank to the Faculdade de Ciências Agrárias e Veterinárias - Unesp Jaboticabal (FCAV / Unesp) and all staffs for all support provided to the development of this dissertation. The support and incentive of the postgraduate program on Genetics and Animal Breeding at FCAV/Unesp is gratefully acknowledged.

I thank the financial support from Coordenação de Aperfeiçoamento de pessoal de nível superior (CAPES) and Fundação de Amparo à Pesquisa (FAPESP), process nº 2013/17396-4 and nº 2014/09603-2. Also financial supports from CNPq (nº 559631/2009-0) and FAPESP (nº 2009/16118-5) are acknowledged.

I thank DeltaGen® breeding program for providing the data used in this study.

I thank the University of Guelph and all staffs for all support given in my research internship. I’m really grateful to Prof. Flavio Schenkel, for all support, attention, suggestions to improve this work, and for spent his time transmitting me valuable knowledge.

I thank all friends that I made in Canada, specially Mrs. Elvira, Mr. Elmer and Mrs. Linda and Mr. Dave, for the care, patience and for receiving me so lovely.

I would like to thank Prof. Elizângela, my advisor in Animal Science undergraduate course, who always challenge me, since the Quantitative Genetics classes, what awoke in me the love by genetics. I'm also grateful to my forage Professor, Emerson, who warned me about my gift to teach, what encouraged me to follow the academic area.

I would like to express my gratitude to all members of the internal and external examining comitees: Dr. Henrique Nunes de Oliveira, Dr. Idalmo Pereira Garcia and Dr. Danísio Prado Munari for their valuable suggestions and contributions to improve this dissertation.

I'm thankful to my advisors, Dr. Roberto Carvalheiro for his patience, spent his valuable time teaching me the main part of genomic knowledge that I have now and for supporting me in the conduction of this work, and Dra. Lucia Galvão de Albuquerque for her valuable contribution for this work and for having accepted me in her study group.

CONTENTS

Page

GENOME WIDE ASSOCIATION OF REPRODUCTIVE TRAITS IN NELORE CATTLE, INCLUDING PHENOTYPIC INFORMATION FROM NON-GENOTYPED ANIMALS

ABSTRACT – This dissertation was divided in three chapters, the first one is a literature review about the subject that will be discussed in subsequent chapters. In the second chapter, a genome wide association study (GWAS) for age at first calving (AFC) in Nelore cattle was performed, using real and simulated data, aiming to 1) assess if additional phenotypic information from non-genotyped animals affect QTL mapping of AFC; 2) evaluate, by simulation, if this additional phenotypic information contributes to detect QTLs more precisely for a low heritable complex trait, and with few available genotypes. The third chapter presents a GWAS for heifer rebreeding (HR) in Nelore cattle. In chapter two, GWA studies were performed using Bayes C and weighted single step GBLUP (WssGBLUP) methods and the top 10 marker windows (1Mb) that explained the larger proportion of variance for AFC were identified and further explored. Two scenarios were investigated, one including all females with available phenotypic information (SI scenario, with 43,482 females), and the other including just the females with available genotype (SII scenario, with 1,813 females). Three iterations were performed in WssGBLUP, recomputing the animals and SNPs effect in each subsequent iteration. It was simulated a population mimicking the parameters and the structure of the real dataset. Two different disequilibrium linkage levels (low and high) between adjacent markers were simulated. In chapter three, the data consisted of 142,878 HR phenotypic records and 2,923 genotypes. The GWAS was performed with WssGBLUP method using three different weightings (iterations) for the SNP effects. Total genetic variances were calculated for the top 10 1Mb SNP-windows, detected by each iteration. On each subsequent iteration, the genetic variance was distributed for a smaller number of SNPs, and the SNP effects were recomputed. Genes possibly associated with HR were searched to reinforce the suggestive importance of the detected windows. The results from chapter two revealed that considering or not additional phenotypic information influenced more the SNP effect estimates than the applied method of GWAS. Although most of genomic regions indicated by different analyses were not the same, some coincidence was observed. Some identified regions presented previously reported QTLs for reproductive traits. The results of simulated data indicated that including all available phenotypic information, even of non-genotyped animals, can improve QTL detection of complex low heritability traits. In chapter three, the GWA analyses detected a total of 21 different windows that harbored 13 QTLs previously reported in the literature for different reproductive (or related) traits. The top 10 marker-windows contained 182 annotated genes. Some of them were associated with pathways of reproductive traits. Evidence was found in this study that important candidate genes affecting HR in Nelore cattle were identified.

Keywords: age at first calving, Bayes C, GWAS, heifers rebreeding, weighted single step GBLUP

ASSOCIAÇÃO GENÔMICA AMPLA DE CARACTERÍSTICAS REPRODUTIVAS EM BOVINOS DA RAÇA NELORE, INCLUINDO INFORMAÇÃO FENOTÍPICA DE ANIMAIS NÃO GENOTIPADOS

RESUMO – Esta dissertação foi dividida em três capítulos, o primeiro é uma revisão de literatura sobre o assunto que será discutido nos capítulos seguintes. No segundo capítulo, foi realizado um estudo de associação genômica ampla para a característica idade ao primeiro parto (IPP) em gado Nelore, que objetivou: 1) avaliar se a informação fenotípica adicional dos animais não genotipados afeta o mapeamento de QTLs da IPP; avaliar, por simulação, se esta informação fenotípica adicional contribui para detectar QTLs mais precisamente para uma característica complexa, com poucos genótipos disponíveis. O terceiro capítulo apresenta um estudo de associação para a característica reconcepção de primíparas (RP) em gado Nelore, cujo objetivo foi detectar importantes regiões genômicas (QTLs) associadas com esta característica. No capítulo dois, estudos de associação foram realizados utilizando as metodologias Bayes C e "Weighted single step GBLUP" (WssGBLUP) e as 10 janelas de marcadores (de 1Mb) que explicaram a maior proporção da variância para IPP foram identificadas e exploradas. Dois cenários foram investigados, um incluindo todas as fêmeas com informação fenotípica disponível (cenário SI – 43.482 fêmeas), e outro incluindo apenas as fêmeas com genótipo disponível (cenário SII – 1.813 fêmeas). Três iterações foram realizadas no método WssGBLUP, sendo recomputados os efeitos dos animais e dos SNPs a cada iteração. Foi simulada uma população com parâmetros e estrutura similares aos dos dados reais, com dois diferentes níveis de desequilíbrio de ligação (alto e baixo) entre os marcadores adjacentes. No capítulo três, os dados consistiram de 142.878 registros fenotípicos de RP e 2.923 genótipos. Os estudos de associação foram realizados com o método WssGBLUP usando três diferentes pesos (iterações) para os efeitos dos SNPs. Para cada iteração subsequente a variância genética foi distribuída para um número menor de SNPs, e os efeitos dos SNPs foram recomputados. Genes possivelmente associados com RP foram pesquisados para reforçar a importância sugestiva das janelas detectadas. Os resultados do capítulo dois indicaram que considerar ou não a informação fenotípica adicional influenciou mais os efeitos estimados dos SNPs do que o método aplicado nos estudos de associação. Embora a maioria das regiões genômicas indicadas pelas diferentes análises não terem sido as mesmas, alguma coincidência foi observada. Algumas regiões associadas com IPP apresentaram QTLs previamente reportados para características reprodutivas. Os resultados dos dados simulados indicaram que incluir toda a informação fenotípica disponível, mesmo dos animais não genotipados, pode melhorar a habilidade de detecção de QTLs de características complexas de baixa herdabilidade. Os estudos de associação do capítulo três detectaram um total de 21 diferentes janelas, as quais abrigaram 13 QTLs previamente reportados na literatura para diferentes características reprodutivas (ou associadas). As top 10 janelas de marcadores continham 182 genes descritos. Alguns deles estavam associados com vias metabólicas de algumas características reprodutivas. Foi encontrada evidência

neste estudo que importantes genes candidatos afetando a RP em gado Nelore foram identificados.

Palavras-chave: Bayes C, idade ao primeiro parto, GWAS, reconcepção de primíparas, weighted single step GBLUP

CHAPTER 1. GENERAL CONSIDERATIONS

Introduction

Brazil is one of the world's largest beef meat exporters (USDA, 2014). Most of the Brazilian meat production comes from Zebu cattle, especially Nelore breed that is the most populous breed in the country. Nelore cattle are well adapted to climatic conditions and to Brazil's production systems, but a common problem related to Zebu cattle is the low reproductive efficiency.

Reproductive traits are associated with the success of beef cattle operations (CAMPOS *et al.*, 2005). Age at first calving (AFC) is one of the reproductive traits used as selection criteria for Nelore females and affect directly the herd productivity. AFC is associated with animals' puberty, and its reduction is associated with reduction on production costs, due to cows entering in the productive age earlier (NÚÑEZ-DOMINGUEZ *et al.*, 1991, PEROTTO *et al.*, 2006). Pötter *et al.* (1998) listed the following benefits of reducing AFC: economic return is faster, increase the quantity of live weight per hectare and the number of females in reproduction stage. An important aspect is that the AFC information is regularly obtained by the farmers, i.e. there is no need of additional costs for its obtainment.

Heifers rebreeding (HR) is also an important reproductive trait because heifers that conceive in breeding season soon after the first calving allow a fast return on investments. Furthermore, this is a delicate stage of female life, since they require energy for maintenance, lactation and, in some cases, for growth or for recovering their body condition score (NAAB, 2004).

HR is been considered in different breeding programs (RILEY *et al.*, 2010; BOLIGON & ALBUQUERQUE, 2012), because heifers usually present low rebreeding rates (CORRÊA *et al.*, 2001).

AFC and HR are considered complex traits because they are most probably controlled by a large number of genes. Despite the economic importance of these reproductive traits, their selection is difficult because they

present low heritability, are expressed just in female (for which the selection intensity is lower than males) and late in female life. So, the genetic evaluations based just in phenotype and pedigree records provide expected breeding values (EBV) with low accuracy for these traits, even when considering information from correlated traits as, for example, scrotal circumference.

Trying to overcome this, genome-wide association studies (GWAS) of reproductive traits have been performed (SASAKI *et al.*, 2013) with the hope to detect genomic regions associated to the trait, and use this information as a tool to possibly increase genetic gain. GWAS consist basically in the detection of statistical associations between the trait of interest and any of the markers (usually Single Nucleotide Polymorphism - SNP) contained in high-density panels (Goddard & Hayes, 2009).

There are many methods used to perform GWAS, one of these is the GBLUP (genomic BLUP), that was initially applied to genomic selection studies by multiple step procedures (CHRISTENSEN & LUND, 2010 and MIZSTAL *et al.*, 2009). However, it has a limitation for using phenotypes or pseudo-phenotypes (e.g. deregressed EBVs) only from genotyped animals. Mizstal *et al.* (2009) proposed a single step procedure, named Single Step GBLUP (ssGBLUP), in which a relationship matrix (G) is estimated using markers information, in contrast with classic BLUP, in which a relationship matrix (A) is based on pedigree. The ssGBLUP combine information from genotyped and non-genotyped animals, which are represented by **G** and **A** matrices, respectively, in an **H** matrix (LEGARRA *et al.*, 2009 and CHRISTENSEN & LUND, 2010). The ssGBLUP has been applied in studies with different species – dairy cattle (VANRADEN, 2012); chicken (CHEN *et al.*, 2011) and swine (FORNI *et al.*, 2011) – and had presented satisfactory results (CHRISTENSEN *et al.*, 2012 and GAO *et al.*, 2012).

Since ssGBLUP method has improved the EBVs accuracy in genomic selection studies, it is expected that the inclusion of phenotypic information from non-genotyped animals could also increase the accuracy to predict the QTL regions in GWAS. Dikmen *et al.* (2013) used ssGBLUP to identify QTLs associated to rectal temperature in Hostein Cattle. Tiezzi et al. (2015) also used ssGBLUP methodology to perform GWAS for clinical mastitis in first parity of US Holstein cows. Despite the benefits of this method, it assumes equal variance

for all markers, i.e, all markers with the same weight, which is particularly limiting when dealing with traits that present markers with more pronounced effects than others.

Trying to overcome this problem, Wang *et al.* (2012) proposed the "Weighted Single Step GBLUP" (wssGBLUP) method that allows to combine pedigree, phenotypic and genotypic information in a single step, weighting the effect of the SNPs according to their supposedly relevance for the trait of interest. In WssGBLUP it is not necessary to restrict information from non-genotyped animals or to include pseudo-phenotypes. This is a useful method for complex models, as non-linear, using multiple traits or, when there are many animals with available phenotype, but few genotyped, a common situation in real dataset. Wang *et al.* (2014) used WssGBLUP in a GWA study for body weight in broiler chickens. They concluded that each GWAS tested method presented some weakness, and their efficiency would depend on the number of animals with phenotype and genotype available, and the proportion of SNPs that explains the genetic variance.

As ssGBLUP, the Bayesian methods were developed primarily for GS studies, but also have been adapted to GWAS. The Bayes A and Bayes B (MEUWISSEN *et al.*, 2001) assume that variances of each marker can vary. Gianola *et al.* (2009) verified some statistical problems with these methods and suggested some modifications. The authors suggested for Bayes A to consider the linkage disequilibrium that exists between the markers and QTL(s), to reduce the priors influence, attributing the same variance to all markers and non-informative priors for scale and degrees of freedom parameters, and/or combine these two last options in a single procedure. For Bayes B, they suggested to attribute zero value to some SNP effects, instead of attributing zero value for their variances. These suggestions were implemented by Habier *et al.* (2011) that tried to reduce the scale parameter influence, considering the same variance for all SNPs (Bayes C), and treated a priori the scale parameter as unknown (Bayes D). However, these authors suggested another modification in those original models that was the π estimation, the proportion of SNPs with zero-effect, since π is assumed known in Bayes A and Bayes B. They argued that it is better to estimate π from the data because π influences the shrinkage of SNP effects. This change formulated the Bayes Cπ and Bayes Dπ methods.

Van den Berg *et al.* (2013), working with simulated data, compared the Bayes C and Bayes Cπ methods and concluded that under certain situations, as a trait with low heritability, low number of records and/or many QTLs affecting the trait, the Bayes C can achieve better results than Bayes Cπ, because in this situations the π is not well estimated, so it is better to fix it.

With the crescent number of GWA studies in literature, several new QTLs are detected for many traits in different species. In QTLdb database (HU *et al.*, 2013) 16,919 QTLs were reported for many traits in cattle until Feb. 13, 2015. However, few of these QTLs have been validated or reproduced by other studies. Therefore, even though there is evidence of important genomic regions being identified, these results should be carefully interpreted (FRAGOMENI et al., 2014). Known genes associated with the concern trait in the region could be a good evidence of a true QTL for real data set (SAATCHI et al., 2014).

Despite advances in genomic and GWA studies, QTL detection with real data is still challenging. However, with simulated data these regions can be previously known, providing a more reliable manner to evaluate the usefulness of a method. For instance, Wang *et al.* (2012) used simulated data in GWA studies to assess the accuracy, precision and computation time for different methods in genomic predictions and QTL detection. Vitezica *et al.* (2011) also performed simulation studies to assess the predictive ability of single and multiple-step methods. Van den Berg et al. (2013), in a simulation study, evaluated the influence of different factors to detect true QTL using two methodologies. Thus, using simulated data to test the power of a method to detect true QTLs with known positions seems to be a helpful strategy.

The present study was developed with the following specific objectives: 1) to assess if additional phenotypic information from non-genotyped animals affect QTL mapping of AFC in Nelore cattle; 2) to evaluate, by simulation, if this additional phenotypic information contributes to detect QTLs more precisely for a low heritable complex trait, with few available genotypes; 3) to search for genomic regions (QTLs) associated with heifer rebreeding (HR) in Nelore cattle.

References

AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743-752, 2010.

BOLIGON, A. A.; ALBUQUERQUE L.G. Genetic parameters and relationships between heifers rebreeding and hip height in Nellore cattle**. Revista Brasileira de Zootecnia**, v.41, n.3, p.598-602, 2012

CAMPOS, W. E.; SAUERESSIG, M. G.; SATURNINO, H. M.; de SOUZA, B. M.; AMARAL, T. B.; FERREIRA, F. **Manejo reprodutivo em gado de corte**. Planaltina, DF: Embrapa Cerrados: 2005. 54 p. (Embrapa. Documentos, 134).

CHEN, C. Y.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W. M. Effect of different genomic relationship matrices on accuracy and scale. **Journal of Animal Science**, v. 89, p. 2673-2679, 2011.

CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v. 42, n. 2, p. 1-8, 2010.

CHRISTENSEN, O. F.; MADSEN, P.; NIELSEN, B.; OSTERSEN, T.; SU, G. Single-step methods for genomic evaluation in pigs. **Animal**, v. 6, n. 10, p. 1565-1571, 2012.

CORRÊA, E.S.; EUCLIDES FILHO, K.; ALVES, R.G.O., et al. **Desempenho reprodutivo em um sistema de produção de gado de corte.** Campo Grande: EMBRAPA/CNPGC, 2001. 33p. (Boletim de Pesquisa, 13).

DIKMEN, S.; COLE, J. B.; NULL, D. J.; HANSEN, P. J. Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. **PLoS ONE**, v. 8, n. 7, 2013.

DOYLE, S. P.; GOLDEN, B. L.; GREEN, R. D.; BRINKS, J. S. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. **Journal of Animal Science**, v.78, p.2091-2098, 2000.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, v. 43, n. 1, p. 1-7, 2011.

FRAGOMENI, B. O.; MISZTAL, I.; LOURENCO, D. L.; AGUILAR, I.; OKIMOTO, R.; MUIR, W. M. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. **Frontiers in Genetics**, v. 5, n. 332, p. 1-7, 2014.

GAO, H.; CHRISTENSEN, O. F.; MADSEN, P.; NIELSEN, U. S.; ZHANG, Y.; LUND, M. S.; SU, G. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. **Genetics Selection Evolution**, v. 44, n. 8, p. 1-8, 2012.

GIANOLA, D.; DE LOS CAMPOS, G.; HILL, W. G.; MANFREDI, E.; FERNANDO, R. Additive Genetic Variability and the Bayesian Alphabet. **Genetics**, v. 183, p. 347-363, 2009.

GUARINI, A. R.; NEVES, H. H. R.; SCHENKEL, F. S.; CARVALHEIRO, R.; OLIVEIRA, J. A.; QUEIROZ, S. A. Genetic relationship among reproductive traits in Nellore cattle. **Animal**, p. 1-6, 2014.

HABIER, D.; FERNANDO, R.L.; KIZILKAYA, K.; GARRICK, D. J. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, n. 186, p. 1-12, 2011.

HU, Z. H.; PARK, C. A.; WU, X. L.; REECY, J. M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, v. 41, p. 871-879, 2013.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of Dairy Science**, v. 92, n. 9, p. 4656-4663, 2009.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-1829, 2001.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, n. 9, p. 4648-4655, 2009.

NATIONAL ASSOCIATION OF ANIMAL BREEDERS. **Rebreeding First-Calf Heifers**. United States, 2004. 3 p. (Angus Beef Bulletin).

NUÑEZ-DOMINGUEZ, R.; CUNDIFF, L. V.; DICKERSON, G. E.; GREGORY, K. E.; KOCH, R. M. Lifetime production of beef heifers calving first at two vs. three years of age. **Journal of Animal Science**, v. 69, p. 3467-3479, 1991.

PEROTTO, D.; MIYAGI, A. P.; SOUZA, J. C.; MOLETTA, J. L.; FREITAS, J. A. Estudos de características reprodutivas de animais da raça Canchim, criados a pasto, no estado do Paraná, Brasil. **Archives of Veterinary Science**, v. 11, n. 2, p. 1-6, 2006.

PÖTTER, L.; LOBATO, J. F. P.; MIELITZ NETTO, C. G. A. Produtividade de um modelo de produção para novilhas de corte primíparas aos dois, três e quatro anos de idade. **Revista Brasileira de Zootecnia**, v. 27, n. 3, p. 613-619, 1998.

RILEY, D. G.; CHASE, C. C. JR.; COLEMAN, S. W.; OLSON, T. A.; RANDEL, R. D. Evaluation of tropically-adapted straightbred and crossbred beef cattle: Heifer age and size at first conception and characteristics of their first calves. **Journal of Animal Science**. v. 88, n.10, p. 3173-82, 2010.

SAATCHI, M.; SCHNABEL, R. D.; TAYLOR, J. F.; GARRICK, D. J. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. **BMC Genomics**, v. 15, n. 442, p. 1-16, 2014.

STRANGER, B.E.; STAHL, E. A.; RAJ, T. Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. **Genetics**, v. 187, n. 2, p. 367-383, 2011.

TIEZZI, F.; PARKER-GADDIS, K. L.; COLE, J. B.; CLAY, J. S.; MALTECCA, C. A Genome-Wide Association Study for Clinical Mastitis in First Parity US Holstein Cows Using Single-Step Approach and Genomic Matrix Re-Weighting Procedure. **PLoS ONE**, v. 10, n. 2, p. 1-15, 2015.

USDA, Livestock and poultry: world markets and trade. Oct, 2014. Available in: <http://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf>. Accessed in Nov. 2014.

VAN DEN BERG, I.; FRITZ, S.; BOICHARD, D. QTL fine mapping with Bayes C (π): a simulation study. **Genetics Selection Evolution**, v. 45, n. 19, 2013.

VANRADEN, P. M. Avoiding bias from genomic pre-selection in converting daughter information across countries. In INTERNATIONAL BULL EVALUATION SERVICE, 2., 2012, Verona, Italy. Proceedings... Verona, n. 45, 2012, p. 1-5.

VITEZICA, Z. G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. Bias in genomic predictions for populations under selection. **Genetics Research (Cambridge)**, v.93, p.357-366, 2011.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; FERNANDO, R. L.; VITEZICA, Z.; OKIMOTO, R.; WING, T.; HAWKEN, R.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 5, n. 134, 2014.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v. 94, n. 2 p. 73-83, 2012.

CHAPTER 2. IMPORTANCE OF INCORPORATING PHENOTYPIC INFORMATION OF NON-GENOTYPED ANIMALS IN A GENOME-WIDE ASSOCIATION STUDY OF A COMPLEX TRAIT

ABSTRACT: Despite the advances in methods applied to genome-wide association studies (GWAS), QTL detection with real data is still challenging, especially in unfavorable scenarios, like those posed by low heritable complex traits, with few available genotyped animals, as is currently the case for age at first calving (AFC) in Nelore cattle. The weighted single-step GBLUP (WssGBLUP) method for GWAS combines simultaneously different sources of information to estimate marker effects, allowing using phenotypic information even from non-genotyped animals. It is not clear, however, in which extend the inclusion of phenotypes from relatives without genotypes can contribute to detect QTLs more precisely. The objectives of this study were: 1) to assess if additional phenotypic information from non-genotyped animals affect QTL mapping of AFC in Nelore cattle; 2) to evaluate, by simulation, if this additional phenotypic information contributes to detect QTLs more precisely for a low heritable complex trait, with few available genotyped animals. The real and simulated data analyses were performed using the WssGBLUP and Bayes C methods. For the WssGBLUP method, two scenarios were tested, including (SI) or not (SII) the phenotypic information from non-genotyped animals, using different weights for the markers. The Bayes C method was tested just under SII. The number of records and the structure of the simulated population mimicked the real data. The results of the real data analyses showed that the use of phenotypes from non-genotyped animals, in addition to phenotypes from genotyped animals, affected the SNP effect estimates and the mapping of the most important genomic regions. Besides some coincidence, the most important genomic regions indicated by the analyses, either considering or ignoring phenotypes from non-genotyped animals, were not the same. The results from simulate data indicated that the inclusion of all available phenotypic information, even from non-genotyped animals, can provide improvement in the detection of QTLs in GWAS of low heritable complex traits. For populations presenting low levels of linkage disequilibrium, the improvement is expected to be milder.

Keywords: age at first calving, beef cattle, Nelore, GWAS, simulation, WssGBLUP

2.1. Introduction

Age at first calving (AFC) is closely associated with the success of beef cattle production systems and is still a bottleneck in many *Bos indicus* breeds as, for example, in Nelore cattle. Genetic gain of AFC is challenging mainly because the trait is sex limited and presents low heritability, especially if the heifers are lately exposed for reproduction (ATENCIO, 2000). Trying to overcome this, genome-wide association studies (GWAS) of AFC have been performed (SASAKI *et al.*, 2013) with the hope to detect genomic regions associated to the trait, and use this information as a tool to possibly increase genetic gain.

There are different methods available to perform GWAS as, for example, Bayesian multiple-regression methods (FERNANDO & GARRICK, 2013). Besides allowing including all markers simultaneously in the analyses, the Bayesian methods allow to consider different prior distributions for the marker effects, what is advocated to be appealing especially for traits affected by large quantitative trait loci (QTL). Van den BERG *et al.* (2013), in a simulation study, tested the adequacy of using Bayesian multiple-regression methods for QTL mapping, and observed feasibility of Bayes C (HABIER *et al.*, 2011) to detect large QTLs, especially for traits with medium to high heritability and with a large number of records.

The single-step GBLUP method (ssGBLUP) (MIZSTAL *et al.*, 2009, CHRISTENSEN & LUND, 2010), originally used in plant and animal breeding for the prediction of breeding values, has also been used in GWAS (WANG *et al.*, 2012, WANG *et al.*, 2014). Despite the benefits of ssGBLUP compared to the multiple-step approach (AGUILAR *et al.*, 2010, CHEN *et al.*, 2011, FORNI *et al.*, 2011), it is based on an infinitesimal model, i.e. a model in which all markers are assumed to have an small effect. This model is particularly limiting when working with traits that present markers with more pronounced effect than others. Given this limitation, Wang *et al.* (2012) proposed the "Weighted Single-Step GBLUP" (WssGBLUP) method, which allows combining pedigree, phenotypic and genotypic information in a single-step, weighting the marker

effects according to their importance for the trait of interest. Furthermore, in WssGBLUP, as in ssGBLUP, it is not necessary to restrict the use of phenotypes just from genotyped animals or to use pseudo-phenotypes like in multiple-step methods. Therefore, the method is expected to be advantageous when working with complex models, such as non-linear or multiple-trait; or in datasets with many phenotypes, but few available genotyped animals, a common situation in real datasets (LOURENCO *et al.*, 2014).

Despite the advances in methods applied to GWAS, QTL detection with real data is still challenging, especially in unfavorable scenarios, like those posed by low heritable complex traits, with few available genotyped animals, as is currently the case for AFC in Nelore cattle. It is not clear in which extend the inclusion of phenotypes from relatives without genotypes can contribute to detect QTLs more precisely in this unfavorable scenario. Aiming to elucidate this, the present study was developed with the following specific objectives: 1) to assess if additional phenotypic information from non-genotyped animals affect QTL mapping of AFC in Nelore cattle; 2) to evaluate, by simulation, if this additional phenotypic information contributes to detect QTLs more precisely for a low heritable complex trait, with few available genotyped animals.

2.2. Material and Methods

2.2.1. Real data

2.2.1.1. Phenotypes and pedigree

Phenotypic information of AFC (in days) was obtained from the DeltaGen® Nelore breeding program database. This database contains records from animals raised under tropical pasture in different commercial herds located in the southeast, west and central regions of Brazil. During the mating season, usually in the rainy period, the heifers are either artificially inseminated or naturally mated. In general, the first mating of heifers occurs at about 26 months of age, although some herds expose heifers earlier, at around 14-18 months of age. Two scenarios were established to perform the GWAS for AFC. The first

including all cows with observed AFC (SI), totaling 43,482 females, and another considering only cows with observed AFC and available genotype (SII), totaling 1,813 females. The pedigree file considered in both scenarios had 237,602 animals.

2.2.1.2. Genotyped animals

A total of 1,829 Nelore females were genotyped with the Illumina Bovine HD panel (Illumina®, San Diego, CA, USA). These females were born between 2007 and 2009 and were chosen among contemporary heifers that were exposed in a precocious age (14-18 months), pertaining to two main farms of the commercial breeding program. The genotyped heifers that did not get pregnant in the anticipated mating season were also exposed in the regular mating season, at about 26 months of age. The heifers that did not conceived a calf after being exposed to these two mating seasons, were not considered in this study. The minimum, maximum and average values of observed AFC of the genotyped females were equal to 698, 1,200 and 1,050 days, respectively, and the proportion of heifers considered precocious (AFC < 900 days) was equal to 28.2%. The quality control (QC) of the genotypes were performed considering the following exclusion criteria: i) for the single nucleotide polymorphism markers (SNPs): from non-autosomal regions, mapped to the same position, presenting a p-value for the Hardy-Weinberg equilibrium test lower than 10^{-5}, with GC score lower than 0.7, call rate lower than 0.95, minor allele frequency lower than 0.02, and highly correlated ($r^2 > 0.995$) with other SNPs within a sliding window containing 100 consecutive markers (only one marker from each pair of highly correlated SNPs was removed); ii) for samples: with call rate lower than 0.9 and being a replicated genotype. The remaining number of SNPs and samples after QC were 333,878 and 1,813, respectively.

2.2.2. Simulated data

The number of phenotypes and genotypes available in the simulation study mimics the information available in the real dataset. The phenotypes and genotypes were simulated using the software QMSim v.1.10 (SARGOLZAEI & SCHENKEL, 2013). Ten replicates were performed of a hypothetical trait with

heritability equal to 0.14 and phenotypic variance of 1.0. This heritability was chosen for being the estimate obtained with the real data.

2.2.2.1. Simulated population

Two different populations were simulated, differing in the number of historical generations, to produce different levels of linkage disequilibrium (LD). In both cases, a first historical population was generated from generation zero to 1,000, with constant size of 1,000 animals. Then, from generation 1,001, a second historical population was simulated with a gradual reduction in the number of animals (from 1,000 to 200), producing a "bottleneck effect" and, consequently, genetic drift and LD. This gradual reduction was made over 1,020 or 2,020 generations, resulting in a population with a lower (LLD) or a higher (HLD) level of LD, respectively. The remaining parameters of the simulation process were the same for both populations (LLD and HLD).

The 200 animals (from which 100 are females) of the latest generation of the historical population were selected for the expanded population. The size of this population mimicked the effective size of the real population (BRITO et al., 2013). For the expanded population, it was considered a mating system based on random union of gametes, absence of selection and an exponential growth of the number of females (every generation generates twice the number of females of the previous generation), with a replacement rate of 100% every generation and an average of five products per dam. Six generations of expansion were generated, resulting in 16,000 animals from which 8,000 were females. After the expansion process, 240 males and 6,000 females of the last generation were randomly selected, comprising the founder animals of the selection population. This population was span over 15 generations. At each generation of the selection population, the selected males and females were randomly mated, generating a single product with equal probability of being a male or a female. The replacement rate of sires and dams was kept constant at 20% and the selection criterion was the expected breeding values. Through the 15 generations, a total of 90,000 animals were generated, mimicking the number of available phenotypes in the real dataset, being 50% females whose phenotypic data were used for GWAS. The genotypes of 2,000 females of the last three generations (13, 14 and 15) were randomly selected for GWAS.

2.2.2.2. Simulated genome and phenotype

It was assumed that QTLs explain 100% of genetic variance. The simulated genome presented a total length of 2,333 cM, 735,293 markers and 7,000 QTLs. The number of markers and QTLs per chromosome ranged from 46,495 to 12,931 and from 121 to 438, respectively, and they were both randomly distributed over 29 autosomes. All markers were bi-allelic, mimicking SNPs present in bovine commercial panels. For QTL, the amount of alleles for loci ranged (randomly) from 2 to 4. Minor allele frequencies (MAF) were assumed equal for markers and QTL alleles. QTL allele effects were sampled from a gamma distribution with a shape parameter equal to 0.4 (HAYES & GODDARD, 2001). It was considered a mutation rate of 10^{-4} for markers and QTLs, in the historical populations. A total of 335,000 markers (with MAF greater or equal than 0.02) and 1,000 QTL were randomly selected from the last generation of the historical population to generate genotypic data for the selection population. The average distance between adjacent markers was 0.007 cM. Although the “genetic architecture” of reproduction traits is unknown, the simulation parameters used in this study aimed to mimic a polygenic complex trait affected by many genes of small effects and few with more pronounced effects. The phenotypes of the animals were comprised by the sum of the QTL effects plus an error term sampled from a normal distribution with zero mean and variance equal to 0.86.

2.2.3. Linkage disequilibrium analyses

The LD between any two loci within the same chromosome was assessed using the r^2 measure (HILL & ROBERTSON, 1968) using the software SnppldHD (SARGOLZAEI, University of Guelph, Canada). The pattern of LD decay of the real and simulated data was compared aiming to evaluate the adequacy of the simulation process (LLD population) to mimic the real scenario.

2.2.4. Statistical analyses

For real and simulated data two statistical methods were used, namely WssGBLUP (WANG *et al.*, 2012) and Bayes C (HABIER *et al.*, 2011). Although

the comparison of methods was not part of our objective, Bayes C was also used for being indicated as a feasible method for QTL detection (VAN DEN BERG et al., 2013).

For the real data, the WssGBLUP method adopted was based on the following model: $y=X\beta+Z_a a+e$, where y is a vector of phenotypic observations (AFC, in days), X is an incidence matrix relating phenotypes to fixed effects, β is a vector of fixed effects, including contemporary group (formed by concatenating the classes for herd, year, season and weaning and yearling management groups) and age of dam as covariate (linear and quadratic effects), Z_a is an incidence matrix that relates animals to phenotypes, a is the vector of additive direct genetic effects, and e is the vector of residuals. The covariance between a and e was assumed equal to zero and their variances were considered, respectively, equal to $H\sigma_a^2$ and $I\sigma_e^2$, where σ_a^2 and σ_e^2 are the additive direct and the residual variances, respectively, H is the matrix which combines pedigree and genomic information (AGUILAR *et al.*, 2010), and I is an identity matrix. As previously stated, the above model was applied considering all the available phenotypes in SI and only the phenotypes of the genotyped cows in SII. For a fairly comparison between scenarios, as fixed effects would not be as good estimated in SII as in SI, the phenotypic observations in SII were pre-adjusted for the fixed effect estimates obtained in a previous "regular" BLUP analysis considering all the available phenotypes. Thus, the β vector of the GWAS in SII contained just an intercept, and the y vector the pre-adjusted AFC. The "regular" BLUP analysis assumed the same model described above with the exception that the variance of a was assumed equal to $A\sigma_a^2$, where A is the regular numerator relationship matrix.

For the simulated data, the same WssGBLUP model was used, except that just an intercept was considered as fixed effect. As no contemporary group effect was simulated, there was no necessity to pre-adjust the phenotypes in the scenario SII, for the simulated data. .

For both scenarios (SI and SII) and datasets (real and simulated), the solutions of SNP effects ($\hat{u}$) were obtained such as in VanRaden *et al.* (2009) and Stranden & Garrick (2009): $\hat{u}=DZ'[ZDZ']^{-1}\hat{a}_g$, where D is a diagonal matrix with weights for SNPs, Z is a matrix relating genotypes of each locus, and $\hat{a}_g$ is

the vector of predicted breeding values of genotyped animals. The D matrix and the SNP and animal effects were iteratively recomputed following the method described by Wang *et al.* (2012) as "ssGBLUP/S2".

Three iterations (w1, w2 and w3) were performed for each scenario, resulting in an increasing shrinkage from w1 to w3 for the SNPs explaining lower variance. Actually, w1 represented the situation where the same weight (w1=1) was given to all SNPs, and served as the basis to initially calculate the proportion of variance explained by each SNP, that was subsequently used to calculate the weights of the next iteration. Therefore, SIw1 refers to scenario SI using w1 weight, SIw2 to scenario SI using w2 weight, and so forth. An equivalent notation was adopted for scenario SII. The BLUPF90 family programs (MISZTAL *et al.*, 2012) were used to run the analyses.

Bayes C analyses were based on the following model (HABIER *et al.*, 2011): $y=1\mu+\sum_{i=1}^{n} g_i b_i \delta_i + e$, where the vectors y and e are the same as previously described, 1 is a vector of ones, μ is the overall mean, g_i is a vector with the genotypes of the animals for the i^{th} SNP, b_i is the allele substitution effect of the i^{th} SNP, δ_i is an indicator variable (0, 1) sampled from a binomial distribution with parameters n and π, where n is the number of SNPs and π is the fraction of SNPs not included in the model. For the real dataset, a prior beta distribution with parameters $\alpha=10^8$ and $\beta=10^{10}$ was assumed for π so that, in practice, π was almost fixed to 0.99 (LEGARRA *et al.*, 2014). For the simulated datasets, the π value assumed two possible values, almost fixed to 0.99, as used in real data, or almost fixed to 0.999, assuming a prior beta distribution with parameters $\alpha=10^8$ and $\beta=10^{11}$. This strategy was adopted, instead of letting π be estimated from the data (Bayes Cπ), because it was found to give better results in QTL detection (VAN DEN BERG *et al.,* 2013). A scaled inverse chi-squared prior distribution was assumed for the variance of SNP effects (σ^2_g) and the residual variance (σ^2_e).

The SNP genotypes were coded as the number of copies of one of the SNP alleles, *i.e.*, 0, 1 or 2. For Bayes C, only the scenario SII was evaluated because the adopted implementation did not allow the inclusion of phenotypes from non-genotyped animals. For the same reason stated in scenario SII of WssGBLUP (real data), the vector y in Bayes C contained the pre-adjusted phenotypes of AFC. The Bayes C analyses were performed using the Markov

chain Monte Carlo algorithm implemented in the software GS3 (LEGARRA *et al.*, 2014), running a single chain with 550,000 iterations, a burn-in period of 50,000 and a thin interval of 50 iterations.

2.2.5 Criteria for comparison

For the real data, GWAS results were compared based on SNP effects estimates and on proportions of variance explained by SNPs within consecutive 1Mb windows. A total of 2,525 windows were considered, spanning over all the autosomes, with an average density of 132±44 SNPs per window. The top 10 windows, that captured the highest proportion of variance explained by the markers, were identified for the different scenarios and methods. The QTLdb database (HU et al., 2013) was consulted to assess the existence of previouslly described QTLs, related to reproduction traits, overlapping the top 10 windows and their neighbouring 1Mb windows (the next to the left and to the right), using the UMD3.1 bovine genome assembly (ZIMIN et al., 2009) as the reference map. The presence of a previous described QTL in the QTLdb database was double checked in the original listed references. Recent scientific publications were also inspected by hand, with the same purpose. When the study used a different reference assembly, the alignment with UMD3.1 was made using the SNPchiMp database (NICOLAZZI et al., 2014), prior to inferring.

For the simulated data, the following statistics were calculated: number of QTLs explaining 1% or more of the genetic variance (topQTL); number of top 1Mb marker windows, accounting for the highest proportion of variance explained by the markers (topMRKw) - this number was enforced to be equal to topQTL; sum of the percentages of genetic variances explained by all the topQTL (Pvar_topQTL) and top marker windows (Pvar_topMRKw); highest percentage of genetic variance explained by a topQTL (Pvar_1stQTL) and a top marker window (Pvar_1stMRKw); and the number of true QTLs (NtrueQTL) *i.e.*, the number of topQTL identified by a topMRKw distant no more than 1Mb from a true QTL position.

2.3. Results and Discussion

2.3.1. Linkage disequilibrium

The LD decay presented by the simulated population with higher level of LD (HLD) was closer to the pattern presented by taurine breeds (PÉREZ O'BRIEN *et al.*, 2014), and was a consequence of the high number of historical generations (from 1,001 to 3,020) simulated to produce the "bottleneck effect". For the LLD simulated population, the LD was slightly higher than the LD presented by the real data, but it was similar to the levels of LD observed by Espigolan et al. (2013) in another set of Nelore animals, indicating adequacy of the LLD population to represent real indicine populations.

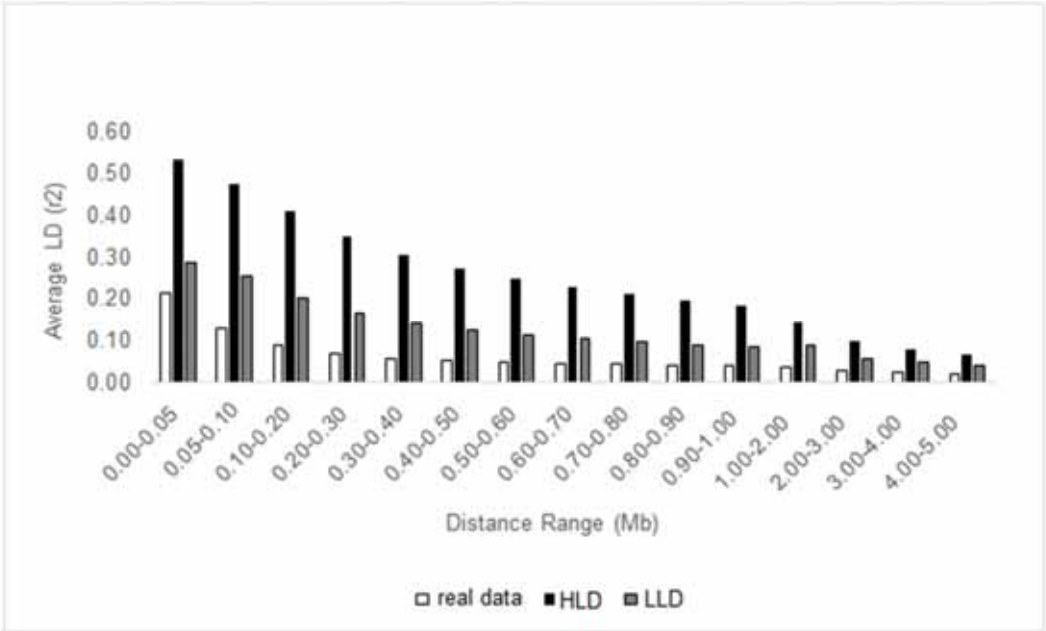

Figure 1. Linkage disequilibrium (LD) decay of real data and two simulated populations (HLD and LLD)

2.3.2. Simulated data

The simulation process resulted in an average number of QTLs explaining 1% or more of the genetic variance (topQTL) equal to 16.7, for the HLD population (Table 1). Together, the topQTL explained 29.74% of the genetic variance, with the most important QTL explaining, on average, 5.07%. As presented in Table 1, the top marker windows were able to capture a smaller

proportion of the genetic variance, in comparison with the topQTL, except for the analyses with a stronger shrinkage (using π=0.999 and w3). These results are partially related to the imperfect LD between markers and QTLs, and also due to false positive signals being captured by the markers.

As expected, a stronger shrinkage of the SNPs explaining lower variance (from w1 to w3 or, equivalently, from π=0.99 to π=0.999) resulted in a higher variance being accounted by the top marker windows. The method that produced the stronger shrinkage was the Bayes C (π=0.999), were the genetic variance captured by the SNPs was distributed among approximately 334 markers (Table 1).

In general, irrespective of the method and scenario, the GWAS presented poor ability to map the topQTLs (maximum 2.9 true QTLs – Table 1), indicating that the available phenotypic and genotypic information from the simulated population was not large enough to map the QTLs properly. These results are in agreement with Van den Berg *et al.* (2013), who also observed in a simulation study that the QTLs were poorly identified in low heritable traits, with large number of QTLs and few records.

Table 1. Average (SD), over ten replicates, of marker and QTL related statistics, for the Bayes C and weighted single step GBLUP (WssGBLUP) analyses of the simulated population presenting high level of LD.

Method/Scenario[a]	Pvar_topMRKw(%)[b]	Pvar_1st MRKw(%)[c]	NtrueQTL[d]
Bayes C (π=0.99)	7.78 (0.99)	1.19 (0.63)	2.20 (1.23)
Bayes C (π=0.999)	46.13 (14.13)	16.45 (17.86)	1.90 (1.29)
WssGBLUP/SIw1	5.16 (0.74)	0.54 (0.17)	2.90 (1.66)
WssGBLUP/SIw2	17.03 (3.11)	2.29 (0.94)	2.90 (1.79)
WssGBLUP/SIw3	38.07 (6.27)	7.30 (3.30)	1.30 (1.16)
WssGBLUP/SIIw1	5.30 (0.63)	0.59 (0.19)	2.00 (1.49)
WssGBLUP/SIIw2	18.76 (1.70)	2.65 (0.93)	1.90 (1.29)
WssGBLUP/SIIw3	39.31 (7.47)	7.82 (5.80)	0.90 (0.74)
True values	Pvar_topQTL(%)[b]	Pvar_1st QTL (%)[c]	NtopQTL[d]
	29.74 (4.88)	5.07 (2.36)	16.7 (2.83)

[a]GWAS using (SI) or ignoring (SII) phenotypic information of non-genotyped animals, applying different weights (w1, w2 and w3) for the SNP effects in the WssGBLUP method. And using π=0.99 and π=0.999 in the Bayes C method.

[b]Genetic variance (%) explained by the sum of variances accounted by top marker windows (Pvar_topMRKw) and by the NtopQTLs (Pvar_topQTL).

[c]Maximum genetic variance (%) explained by a top marker window (Pvar_1stMRKw) and by a topQTL (Pvar_1stQTL).

[d]Number of true QTLs explaining 1% or more of the genetic variance (NtopQTL), and number of NtopQTLs identified by a top marker window distant no more than 1Mb from a NtopQTL (NtrueQTL).

As can be seen in Table 1, analyses using different weights (w1 and w2) presented similar ability to map the QTLs. Also, a higher percentage of variance being explained by the top marker windows, as a result of a stronger shrinkage, was not necessarily associated to a better performance in mapping the QTLs. For the HLD population, the WssGBLUP analyses using w1 and w2 outperformed the analysis using w3 in terms of true QTL detected. Using a higher π also did not improve the ability of Bayes C to detect QTLs (Table 1).

The use of phenotypic information from non-genotyped animals contributed to detect QTLs more precisely, for the HLD population. The analysis with the best result considering additional phenotypic information was able to detect, on average, 17.4% (2.9 out of 16.7) of the topQTL, whereas the WssGBLUP analyses ignoring this information were able to detect at best 12.0% of the topQTL.

Within the methods, a relatively large variability over replicates was observed for the NtrueQTL parameter. For instance, in the HLD population, the minimum and maximum NtrueQTL values presented by the WssGBLUP/SIw2 analysis were, respectively, equal to 0 out of 14 topQTL and 6 out of 19 topQTL. These values were equal to 1/18 and 4/16 for Bayes C (π=0.99) and equal to 0/16 and 4/17 for WssGBLUP/SIIw1, respectively. Besides this variability, the number of times a method beat the others in QTL detection reinforced the superiority of WssGBLUP/SIw2 over the other methods, for the HLD population. It won 5 times, drew 3 and lost 2 out of the 10 replicates, based on the NtrueQTL criteria (data not shown).

For the LLD population (Table 2), the importance of using additional phenotypic information became less evident. This could be explained because the genotyped animals are in some extend related with the animals presenting just available phenotype. For presenting low level of LD, the size of identic by descendent haplotype segments shared is reduced and, as a consequence, the inclusion of additional phenotypic information does not help the detection of true QTLs.

For the LLD population the number of QTLs explaining 1% or more of the genetic variance (topQTL) was equal to 15.4 (Table 2). Together, the topQTL

explained 24.32% of the genetic variance, with the most important QTL explaining, on average, 3.19%.

In comparison with HLD, all methods and scenarios from the LLD population presented poorer ability to detect true QTLs. In addition, the LLD population presented a different pattern of the Bayes C results, compared to HLD population. The average NtrueQTL value of the analysis assuming π=0.999 was higher than the value presented by the analysis using π=0.99. These results reinforce that the adequacy of a method to detect QTLs is dependent on the pattern of LD presented by the population of interest.

As observed in Table 2, although slightly better on average, the "SI" analyses presented similar NtrueQTL values than their counterpart "SII" analyses. The analysis with the best result, WssGBLUP/SIw2, considering additional phenotypic information was able to detect, on average, 13.6% (2.1 out of 15.4) of the topQTL.

Table 2. Average (SD), over ten replicates, of marker and QTL related statistics, for the Bayes C and weighted single step GBLUP (WssGBLUP) analyses of the simulated population presenting low level of LD.

Method/Scenario[a]	Pvar_topMRKw(%)[b]	Pvar_1st MRKw(%)[c]	NtrueQTL[d]
Bayes C (π=0.99)	3.95 (0.58)	0.42 (0.10)	1.40 (0.97)
Bayes C (π=0.999)	36.71 (12.35)	12.39 (11.93)	1.80 (1.62)
WssGBLUP/SIw1	2.73 (0.53)	0.25 (0.08)	2.00 (1.70)
WssGBLUP/SIw2	10.58 (1.59)	1.45 (0.37)	2.10 (1.29)
WssGBLUP/SIw3	26.80 (4.93)	4.51 (1.31)	1.20 (0.63)
WssGBLUP/SIIw1	2.82 (0.35)	0.26 (0.04)	1.90 (1.20)
WssGBLUP/SIIw2	12.05 (1.72)	1.69 (0.38)	1.90 (1.10)
WssGBLUP/SIIw3	31.60 (3.76)	5.66 (2.49)	1.10 (0.88)
True values	Pvar_topQTL(%)[b]	Pvar_1st QTL (%)[c]	topQTL[d]
	24.32 (4.92)	3.19 (0.61)	15.4 (2.32)

[a]GWAS using (SI) or ignoring (SII) phenotypic information of non-genotyped animals, applying different weights (w1, w2 and w3) for the SNP effects in the WssGBLUP method. And using π=0.99 and π=0.999 in the Bayes C method.
[b]Genetic variance (%) explained by the sum of variances accounted by top marker windows (Pvar_topMRKw) and by the NtopQTLs (Pvar_topQTL).
[c]Maximum genetic variance (%) explained by a top marker window (Pvar_1stMRKw) and by a topQTL (Pvar_1stQTL).
[d]Number of true QTLs explaining 1% or more of the genetic variance (NtopQTL), and number of NtopQTLs identified by a top marker window distant no more than 1Mb from a NtopQTL (NtrueQTL).

2.3.2. Real data

As observed in simulated data, the Figures 2A to 2C indicate that adding phenotypic information of non-genotyped animals influenced the SNP effect estimates also in real data. Although correlated, the WssGBLUP SNP effect estimates were not the same among scenarios SI and SII. The influence of using or not the phenotypes from non-genotyped animals became stronger as the shrinkage on SNPs explaining lower variance was higher. For the SNPs with more pronounced effects, the WssGBLUP SNP effect estimates from SI were more similar to those from SII in w1 (Figure 2A) than in w2 (Figure 2B) and w3 (Figure 2C). This occurred because within each scenario the weights of further iterations were calculated as a function of the SNP effects estimated in the previous iteration (WANG *et al.*, 2012). Consequently, the differences became larger in each every subsequent iteration.

As observed by Wang *et al.* (2012) and in our simulation study, the total genetic variance was distributed for a smaller number of SNPs as the weights and subsequently the animal and SNP effects were recomputed in the WssGBLUP method. As a consequence, solutions from w1 (Figures 2A, 2D and 2G) were more similar to those expected from a trait following an infinitesimal model, and solutions from w3 (Figures 2C, 2F and 2I) were closer to what would be expected for an oligogenic trait. Unfortunately, when analysing real data, the WssGBLUP method does not allow inferring which weight led to better estimates. Wang et al. (2012) recognized that their proposed method calculates the weights in a suboptimal manner and suggested some refinements using, for example, Bayesian methods.

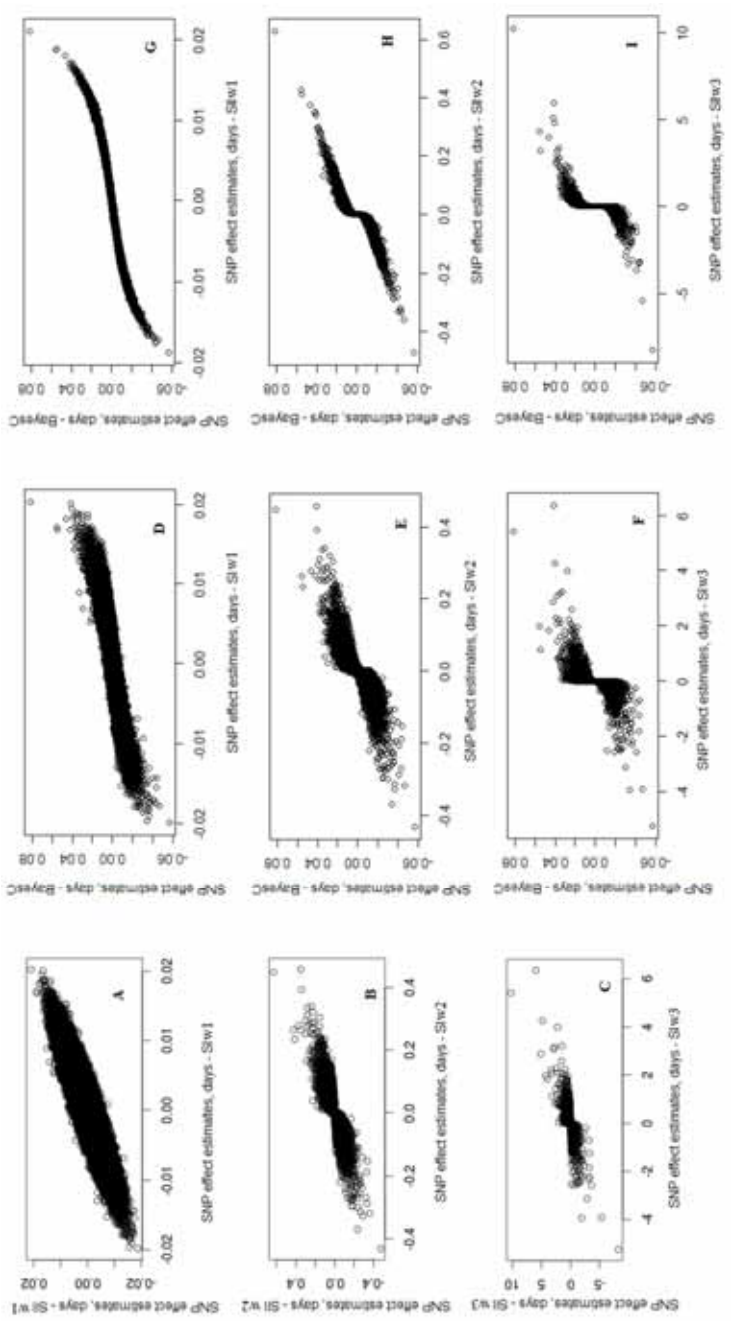

Figure 2. Scatter plots of SNP effect estimates for GWA analyses of age at first calving in Nelore cattle, using Bayes C and weighted single step GBLUP method, considering (SI) or ignoring (SII) phenotypes from non-genotyped animals, under different weights (w1, w2 and w3) for the SNPs

As stressed by Gianola *et al.* (2009), unrevealing the "genetic architecture" underlying the traits is not trivial even with the adoption of Bayesian methods, as the priors dominate the inference when we have few observations and want to estimate simultaneously a huge number of SNP effects. The simulation results from our study reinforce this statement, where the prior value assumed for π determined the shrinkage on SNP effect estimates. Letting the π be estimated from the data seems not to be a good strategy for GWAS purpose (VAN DEN BERG *et al.*, 2013).

The Bayes C solutions were more similar to the WssGBLUP solutions of scenario SII (Figures 2G to 2I), where the same phenotypic information was used, than those from the scenario SI (Figures 2D to 2F). For instance, the adherence between Bayes C and WssGBLUP SIIw1 solutions was very high (Figure 2G). This result suggests that, for the present study, considering or not additional phenotypic information influenced more SNP effect estimates than difference in the method applied for GWAS. This evidence was not so strong in the simulated data, mainly for the LLD population. The sigmoidal shape of Figures 2E, 2F, 2H and 2I indicated that the weights w2 and w3, applied in the WssGBLUP method, resulted in more SNPs with their effects shrunk to zero than Bayes C with $\pi \approx 0.99$.

As also observed in the simulation study, a stronger shrinkage (w2 and w3) on SNPs explaining lower variance redistributed the variance and resulted in larger variances being explained by the most important regions (Table 3). In SI, the proportion of variance explained by the top 10 windows was equal to 1.83%, 7.18% and 23.65% for w1, w2 and w3, respectively. These figures were equal to 1.84%, 8.07% and 34.70% in SII. For the Bayes C method, the top 10 windows explained 2.65% of the genetic variance.

Although the most important genomic regions indicated by the different analyses were not the same, some coincidence was observed. Considering all the seven analyses, 31 different windows were indicated as top 10 (Table 3). The most common windows between the analyses were: 4 of chromosome 7 (7/4), indicated as top 10 in 6 analyses; 5/115, 8/107 and 18/5, indicated as top 10 in 5 analyses; 17/50 and 23/27, indicated as top 10 in 4 out of 7 analyses. WssGBLUP/SIIw2 and Bayes C were the only analyses which indicated the six most common regions within the top 10 windows.

Based on the number of QTLs previously described in the literature, the WssGBLUP/SIw2 outperformed the other methods, presenting 5 out of the 10 top windows with a previous described QTL, within the window or in a neighbouring window, associated to reproduction traits. This result is an indicative that the use of phenotypic information from non-genotyped animals contributed to detect QTLs more precisely, also for the real data.

Table 3. Top 10 windows explaining the highest proportion of variance of age at first calving in GWAS using Bayes C and weighted single step GBLUP method considering (SI) or ignoring (SII) phenotypes from non-genotyped animals, under different weights (w1, w2 and w3) for the SNPs.

Rank	Bayes C	w1				w2				w3			
		SI		SII		SI		SII		SI		SII	
	Ch/Wi/pvar[1]	Ch/Wi/pvar		Ch/Wi/pvar		Ch/Wi/pvar		Ch/Wi/pvar		Ch/Wi/pvar		Ch/Wi/pvar	
1st	**18/005/0.40** *	**18/005/0.23** *		**18/005/0.23** *		**18/005/1.13** *		**18/005/1.17** *		13/071/4.59		**23/027/9.66**	
2nd	**08/107/0.34**	05/016/0.19		**17/050/0.21**		**08/107/0.84**		**05/115/1.16**		**23/027/3.26**		12/003/6.43	
3rd	**17/050/0.29**	**08/107/0.19**		23/025/0.20		06/108/0.80 *		**08/107/1.08**		12/003/3.24		13/071/3.42	
4th	**05/115/0.28**	**07/004/0.19**		**08/107/0.20**		02/018/0.71		**23/027/0.94**		02/018/2.94		**05/115/3.14**	
5th	**07/004/0.27**	**17/050/0.18**		**07/004/0.18**		22/002/0.68 **		**07/004/0.78**		07/026/1.79		03/012/2.72 *	
6th	03/028/0.24	23/028/0.17		23/029/0.18		**07/004/0.68**		**17/050/0.67**		04/077/1.77 *		07/041/2.48	
7th	**23/027/0.22**	05/013/0.17		03/028/0.16		24/009/0.68		06/044/0.61 *		03/012/1.75 *		02/018/2.39	
8th	23/025/0.21	05/017/0.17		01/027/0.16 **		**05/115/0.56**		14/063/0.59 **		**05/115/1.50**		**07/004/1.60**	
9th	14/063/0.20 **	10/092/0.17 **		23/028/0.16		06/016/0.56 *		12/003/0.55		06/108/1.45 *		21/023/1.57	
10th	23/029/0.20	23/029/0.17		05/013/0.15		25/033/0.55 *		03/028/0.53		11/009/1.37		01/027/1.29 **	

[1]Ch=chromosome; Wi=1 Mb window within the chromosome; pvar=proportion of variance explained by the SNPs within the window. The most common windows (ranked as top 10 in at least four analyses) are highlighted in bold. *Window (or neighbouring window) with a previous described QTL for bovine reproduction traits. **Window (or neighbouring window) with a previous described QTL for bovine sexual precocity traits

Although the simulated results indicated poor ability to detect QTLs, evidence was found in the literature that important regions were identified by the real data analyses, as presented in Table 4.

Table 4. Mainly reproductive traits reported by different authors for different breeds in the most important regions (Ch, Chromossome and wi, Window) detected by Bayes C (π=0.99) and wSSGBLUP (w1, w2 and w3) methods.

Ch/wi	Traits reported in the region/Breed	Authors
6/16	Maternal stillbirth later/ Nordic cattle	Hoglund et al. (2012)
6/108	Calving Difficult (Direct)/ Nordic cattle	Hoglund et al. (2012)
25/33	Calving easy traits/Holstein	Sahana et al. (2011)
18/5	Calving index/Nordic cattle	Hoglund et al. (2012)
3/12	Calving easy traits/Holstein	Sahana et al. (2011)
6/44	Calving easy (direct)/Maine-Anjou	Saatchi et al. (2014)
4/77	Calving easy (direct)/Angus	Saatchi et al. (2014)
	Body condition score*/Brahman and an Australian crossbreed herd	Porto-Neto et al. (2014)
14/63	Age at first calving/Hanwoo	Hyeong et al. (2014)
1/27	Age at first calving/Hanwoo	Hyeong et al. (2014)
10/92	Heifer pregnancy/Angus	Peters et al. (2013)
22/2	Semen volume/Holstein	Druet et al. (2009)
	Calving traits/Holstein	Thomasen et al. (2008)
	Non-return of daughters at 56 days after insemination/Holstein	Scrooten et al. (2004)

*Body condition score was listed for indirectly affecting females reproductive performance (HERD & SPROTT, 1996).

There was also evidence of important genes being located. For instance, Waters *et al.* (2014), studying the transcriptional regulation process in the uterine endometrium of beef heifers, under a special dietary supplementation, detected many genes, which were clustered in similar functional groups. One of these genes is the *TSPO*, which was grouped in a cluster described as “potentially co-regulated genes with reproductive function and steroid

biosynthesis". Considering the importance of steroids to reproductive female regulation process, this could be an important candidate gene associated with AFC in Nelore cattle. The *TSPO* gene is located within the window 115 of BTA5 (UMD3.1), which was identified as a top window by all the analyses, except those using w1 (Table 3).

2.4. Conclusions

Additional phenotypic information from non-genotyped animals influences the GWAS results. Besides some coincidence, the most important genomic regions for AFC in Nelore cattle, indicated by the analyses considering or ignoring phenotypes from non-genotyped animals, were not the same. The results from simulated data indicate that the inclusion of all available phenotypic information, even from non-genotyped animals, can provide a small improvement in the detection of QTLs in GWAS of low heritable complex traits. For populations presenting low levels of LD, the improvement is expected to be milder.

2.5. References

AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743-752, 2010.

ATENCIO, A.M. Predicción genética de la fertilidad em la hembra cebu. In: Congresso Internacional de Transferência Tecnológica Agropecuária, 20, 2000, Assunción. **Anales...** Assunción: Congresso Internacional de Transferência Tecnológica Agropecuária, 2000. p.29-42.

BRITO, F.V.; SARGOLZAEI, M.; BRACCINI NETO, J.; COBUCI, J.A.; PIMENTEL, C.M.; BARCELLOS, J.; SCHENKEL, F.S. In-depth pedigree analysis in a large Brazilian Nellore herd. **Genetics and Molecular Research**, v.12(4), p.5758-5765, 2013.

CHEN, C. Y.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W. M. Effect of different genomic relationship matrices on accuracy and scale. **Journal of Animal Science**, v. 89, p. 2673-2679, 2011.

CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v. 42, n. 2, p. 1-8, 2010.

DRUET, T.; FRITZ, S.; SELLEM, E.; BASSO, B.; GÉRARD, O.; SALAS-CORTES, L.; HUMBLOT, P.; DRUART, X.; EGGEN, A. Estimation of genetic parameters and genome scan for 15 semen characteristics traits of Holstein bulls. **Journal of Animal Breeding and Genetics**, v. 126, p. 269-277, 2009.

ESPIGOLAN, R.; BALDI, F.; BOLIGON, A. A.; SOUZA, F. R. P.; GORDO, D. G. M.; TONUSSI, R. L.; CARDOSO, D. F.; OLIVEIRA, H. N.; TONHATI, H.; SARGOLZAEI, M.; SCHENKEL, F. S.; CARVALHEIRO, R.; FERRO, J. A.; ALBUQUERQUE, L. G. STUDY OF WHOLE GENOME LINKAGE DISEQUILIBRIUM IN NELLORE CATTLE. **BMC Genomics**, v. 14, p. 1-8, 2013.

FERNANDO, R. L.; GARRICK, D. Bayesian methods applied to GWAS. In GONDRO, C.; van der WERF, J.; HAYES, B. **GENOME-WIDE ASSOCIATION STUDIES AND GENOMIC PREDICTIONS.** Springer: Humana Press. 2013, chap.10, p. 237-274.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, v. 43, n. 1, p. 1-7, 2011.

GIANOLA, D.; DE LOS CAMPOS, G.; HILL, W. G.; MANFREDI, E.; FERNANDO, R. Additive Genetic Variability and the Bayesian Alphabet. **Genetics**, v. 183, p. 347-363, 2009.

HABIER, D.; FERNANDO, R. L.; KIZILKAYA, K.; GARRICK, D. J. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, n. 186, p. 1-12, 2011.

HAYES, B. J.; GODDARD, M. E. The distribution of the effects of genes affecting quantitative traits in livestock. **Genetics Selection Evolution**, v. 33, p. 209-229, 2001.

HERD, D. B.; SPROTT, L. R. **Body condition, nutrition and reproduction of beef cows**. College Station, Texas: Texas Agricultural Extension Service. 1996, 12 p.

HILL, W. G.; ROBERTSON, A. Linkage Disequilibrium in Finite Populations. **Theoretical and Applied Genetics**, v. 38, n. 6, p. 226-231, 1968.

HÖGLUND, J. K.; GULDBRANDTSEN, B.; LUND, M. S.; SAHANA, G. Analyzes of genome-wide association follow-up study for calving traits in dairy cattle. **BMC Genetics**, v. 13, n. 71, 2012.

HYEONG, K.-E.; IQBAL, A.; KIM, J.J. 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**. v. 27, n.10, 2014.

HU, Z. H.; PARK, C. A.; WU, X. L.; REECY, J. M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, v. 41, p. 871-879, 2013.

LEGARRA, A.; RICARD, A.; FILANGI, O. GS3. Manual. France. [S.l.: s.n.], 2014, 24p. Available in: <http://snp.toulouse.inra.fr/~alegarra/manualgs3_last.pdf>. Accessed in Oct. 2014.

LOURENCO, D. A. L.; MISZTAL, I.; TSURUTA, S., AGUILAR, I.; EZRA, E.; RON, M.; SHIRAK, A.; WELLER, J. I. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. **Journal of Dairy Science**, v. 97, p. 1742-1752, 2014.

MISZTAL, I. **BLUPF90 - a flexible mixed model program in Fortran 90**. Manual. Animal and Dairy Science, University of Georgia, Athens, GA, USA. [S.l.: s.n.], 2012, 25p. Available in: <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90.pdf>. Accessed in Feb. 2014.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, n. 4, p. 1819-1829, 2001.

NICOLAZZI, E. L.; PICCIOLINI, M.; STROZZI, F.; SCHNABEL, R. D.; LAWLEY, C.; PIRANI, A.; BREW, F.; STELLA, A. SNPchiMp: A database to disentangle the SNPchip jungle in bovine livestock. **BMC Genomics**, v. 15, n. 123, p. 1-6, 2014.

PÉREZ O'BRIEN, A. M.; MÉSZÁROS, G.; UTSUNOMIYA, Y. T.; SONSTEGARD, T. S.; GARCIA, J. F.; TASSELL, C. P. V.; CARVALHEIRO, R.; SILVA, M. V. B.; SÖLKNER, J. Linkage disequilibrium levels in Bos indicus and Bos taurus cattle using medium and high density SNP chip data and different

minor allele frequency distributions. **Livestock Science**, v. 166, p. 121-132, 2014.

PETERS, S. O.; KIZILKAYA, K.; GARRICK, D. J.; FERNANDO, R. L.; REECY, J. M.; WEABER, R. L.; SILVER, G. A.; THOMAS, M. G. Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. **Journal of Animal Science**, v. 91, p. 605-612, 2013.

PORTO-NETO, L. R.; REVERTER, A.; PRAYAGA, K.C.; CHAN, E. K. F.; JOHNSTON, D. J.; HAWKEN, R. J.; FORDYCE, G.; GARCIA, J. F.; SONSTEGARD, T. S.; BOLORMAA, S.; GODDARD, M. E.; BURROW, H. M.; HENSHALL, J. M.; LEHNERT, S. A.; BARENDSE, W. The Genetic Architecture of Climatic Adaptation of Tropical Cattle. **PLoS ONE**, v. 9, n. 11, p. 1-22, 2014.

SAATCHI, M.; SCHNABEL, R. D.; TAYLOR, J. F.; GARRICK, D. J. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. **BMC Genomics**, v. 15, n. 442, p. 1-16, 2014.

SAHANA, G.; GULDBRANDTSEN, B.; LUND, M. S. Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. **Journal of Dairy Science**, v. 94, p. 479-486, 2011.

SARGOLZAEI, M.; SCHENKEL, F. S. **QMSim: User's Guide** [S.l.: s.n.], 2013, p.77.

SASAKI, S.; IBI, T.; IKEDA, S.; SUGIMOTO, Y. A genome-wide association study reveals a quantitative trait locus for age at first calving in delta/notch-like EGF repeat containing on chromosome 2 in Japanese Black cattle. **Animal Genetics**, v. 45, n. 2, p. 285-287, 2013.

SCHROOTEN, C.; BINK, M. C. A. M.; BOVENHUIS, H. Whole genome scan to detect chromosomal regions affecting multiple traits in dairy cattle. **Journal of Dairy Science**, v. 87, p. 3550-3560, 2004.

STRANDEN, I.; GARRICK, D. J. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. **Journal of Dairy Science**, v. 92, n. 6, p. 2971-2975, 2009.

THOMASEN, J. R.; GULDBRANDTSEN, B.; SØRENSEN, P.; THOMSEN, B.; LUND, M. S. Quantitative trait loci affecting calving traits in Danish Holstein cattle. **Journal of Dairy Science**, v. 91, p. 2098-2105, 2008.

VAN DEN BERG, I.; FRITZ, S.; BOICHARD, D. QTL fine mapping with Bayes C (π): a simulation study. **Genetics Selection Evolution**, v. 45, n. 19, 2013.

VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v. 92, n. 1, p. 16-24, 2009.

WANG, H.; MISZTAL, I.; AGUILAR, I., LEGARRA, A.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v. 94, n. 2 p. 73-83, 2012.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; FERNANDO, R. L.; VITEZICA, Z.; OKIMOTO, R.; WING, T.; HAWKEN, R.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 5, n. 134, 2014.

WATERS, S. M.; COYNE, G. S.; KENNY, D. A.; MORRIS, D. G. Effect of dietary n-3 polyunsaturated fatty acids on transcription factor regulation in the bovine endometrium. **Molecular Biology Reports**, v. 41, p. 2745-2755, 2014.

ZIMIN, A. V.; DELCHER, A. L.; FLOREA, L.; KELLEY, D. R.; SCHATZ, M. C.; PUIU, D.; HANRAHAN, F.; PERTEA, G.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; MARÇAIS, G.; ROBERTS, M.; SUBRAMANIAN, P.; YORKE, J. A.; SALZBERG, S. L. A whole-genome assembly of the domestic cow, *Bos taurus.* **Genome Biology**, v. 10, i. 4, 2009.

CHAPTER 3. GENOME WIDE ASSOCIATION STUDY OF HEIFER REBREEDING IN NELORE CATTLE

ABSTRACT: A genome wide association study (GWAS) was conducted aiming to find important genomic regions (QTLs) associated with heifer rebreeding (HR) in Nelore cattle. The data consisted of 142,878 HR records and 2,923 high-density (777K SNPs) genotypes. The GWAS was performed using the weighted single step GBLUP method, under three different weightings for the SNPs. Genetic variances explained by consecutive no-overlapping 1Mb SNP-windows were calculated, for each of the three GWAS. The top 10 windows, that captured the highest proportion of variance explained by the markers, were further investigated. To reinforce their suggestive importance, scientific papers were consulted to inspect if any identified top window overlapped with a previously described QTL related to bovine reproductive traits. A possible association of the genes, within the top 10 windows, with HR were also investigated. From the three performed GWAS, a total of 21 different windows were detected, that harbored 13 QTLs previously reported in literature for different reproductive (or related) traits. The top 10 marker-windows contained 182 annotated genes. Some of them were associated with pathways of reproductive traits. Evidence was found in the present study that important candidate genes affecting HR in Nelore cattle were identified.

Keywords: beef cattle, genes, GWAS, QTL, reproductive trait

3.1 Introduction

Reproductive traits are closely associated with the success of beef cattle productive system (CAMPOS *et al.*, 2005). Some of the reproduction traits are included in breeding programs aiming to produce a more profitable herd. Heifer rebreeding (HR) is an important reproductive trait because heifers that conceive in breeding season soon after the first calving allow a fast return on investments. Furthermore, this is a delicate stage of female life, because it requires energy to its maintenance, lactation, and in some cases to growth (NAAB, 2004) or to recover its body condition score. HR is been considered in different breeding programs (RILEY *et al*., 2010; BOLIGON & ALBUQUERQUE, 2012), because heifers usually present low rebreeding rates (CORRÊA *et al.*, 2001).

Despite the economic importance of HR, its selection is difficult because HR presents low heritability (DOYLE *et al*., 2000; GUARINI *et al.*, 2014), it is expressed late in life and just in females (which present lower selection intensity compared to males). So, the genetic evaluations based just in phenotype and pedigree records usually provide low accurate genetic proofs, even when correlated traits are considered in the analyses.

Using genomic information from high-density Single Nucleotide Polymorphism (SNP) panels have been applied to perform genome wide association studies (GWAS). A methodology that has been used to GWAS is the “Weighted Single Step GBLUP (WssGBLUP)” (WANG *et al.*, 2012), that allows to combine pedigree, phenotypes and genotypes information in a single step, in which different weights can be attributed to each marker.

With the crescent number of GWA studies in literature, several new QTLs are detected for many traits in different species. In QTLdb database (HU *et al.*, 2013) 16,919 QTLs were reported for many traits in cattle until Feb. 13, 2015. However, few of these QTLs have been validated or reproduced by other studies. Therefore, even though there is evidence of important genomic regions

being identified, these results should be carefully interpreted (FRAGOMENI et al., 2014). Known genes associated with the trait of interest in the region could be a good evidence of a true QTL (SAATCHI et al., 2014).

The purpose of this study was to find important genomic regions (QTLs) associated with heifer rebreeding (HR) in Nelore cattle.

3.2. Material and Methods

3.2.1. Phenotypic and pedigree information

Phenotypic information of HR was obtained from Alliance Nelore database. After editing, the data contained HR records of 142,878 Nelore heifers, born from 1984 to 2010, and raised in 188 different commercial farms located in the southeast, west and central regions of Brazil and in Paraguay. The feeding system adopted by these farms consisted basically of tropical pastures, mineral salt and water *ad libitum*. In the dry season, the animals usually receive mineral supplementation.

During the mating season, which lasts approximately 90 days and occurs usually in the rainy period, the heifers are either artificially inseminated or naturally mated. A total of 44.1% of the first-calves (62,996) and 44.3% of the second-calves (35,539) were born from AI-sired heifers. In general, the first mating of heifers occurs at about 26 months of age, although some herds expose heifers earlier, at around 14-18 months of age, in an anticipated mating season. In the data used, 46.6% (66,643) of the heifers were earlier exposed, from which 21.4% (14,287) conceived. Those that were exposed and did not conceive earlier had a second chance at about 26 months of age, in the regular mating season, and all the heifers that did not get pregnant at this period were culled, including those that were exposed for the first time.

HR was defined as success (1) or failure (0) for heifers that calved or not, respectively, since they had previously produced the first calf. In the editing process, were excluded heifers presenting age at first and second calving lower or greater than 21 and 40 months, and 32 and 53 months, respectively, and

calving interval lower than 11 months. Contemporary groups (CG) were defined concatenating information of herd, year and season of birth, and weaning and yearling management groups of the heifers. CG with less than five heifers and without variability for HR, *i.e.* composed by animals with the same categorical response, were excluded from the data. After editing, 78,389 (54.9%) and 64,489 (45.1%) heifers presented success and failure to re-conceive, respectively. The 142,878 heifers were daughters of 2,391 different sires and 108,440 dams. The pedigree file contained 223,195 animals distributed over five generations.

3.2.2. Genomic information

A total of 2,925 Illumina Bovine HD genotypes (Illumina®, San Diego, CA, USA) were used. Genotypes came from 2,212 Nelore heifers, from 12 different herds, and from 713 Nelore sires that had on average 73.6 progeny evaluated for HR. The genotyped heifers and sires were born from 2002 to 2009 and 1965 to 2006, respectively.

The quality control (QC) of the genotypes were performed considering the following exclusion criteria: i) for the single nucleotide polymorphism markers (SNPs): from non-autosomal regions, mapped to the same position, presenting a *p*-value for the Hardy-Weinberg equilibrium test lower than 10^{-5}, with GC score lower than 0.15, call rate lower than 0.95, and minor allele frequency lower than 0.02; ii) for samples: with call rate lower than 0.9 and being a replicate. The remaining number of SNPs and samples after QC were 409,376 and 2,923, respectively.

3.2.3. Statistical analysis

The SNP marker effects were estimated using the weighted single-step GBLUP method (WssGBLUP) proposed by Wang *et al.* (2012). This method was chosen for allowing combining pedigree, phenotypic and genomic information in a single-step, weighting the marker effects according to their importance for the trait of interest. The WssGBLUP is claimed to be

advantageous when working with datasets with many phenotypes, but few available genotypes (WANG *et al.*, 2012, 2014), as is the case in the present study. To run the analysis, the BLUPF90 family programs (MISZTAL *et al.*, 2012) were used.

Firstly, predicted breeding values were obtained based on the following threshold animal model (GIANOLA & FOULLEY, 1983): $y = X\beta + Z_a a + e$, where *y* is a vector of underlying liabilities for HR, β is a vector of fixed effects of CG, *a* is a vector of random additive direct genetic effects (breeding values), *X* and Z_a are incidence matrices relating elements in β and *a* to *y*, respectively, and *e* is the vector of random residuals. The underlying liabilities for HR were defined as follows: HR = 0 if $y \leq t_1$; and HR = 1 if $y > t_1$, where t_1 is the threshold corresponding to the discontinuity in the observed scale of HR. The covariance between *a* and *e* was assumed equal to zero and their variances were considered, respectively, equal to $H\sigma_a^2$ and $I\sigma_e^2$, where σ_a^2 and σ_e^2 are the additive direct and the residual variances, respectively, *H* is the matrix which combines pedigree and genomic information (AGUILAR *et al.*, 2010), and *I* is an identity matrix. As the variable in the underlying distribution is not observable, the parameterization $\sigma_e^2 = 1$ was adopted (SORENSEN & GIANOLA 2002).

The parameter estimates of the threshold model were obtained under a Bayesian framework using the Gibbs sampling program THRGIBBS1F90 (TSURUTA & MISZTAL, 2006). The default prior distributions were assumed for the variance components and for the fixed and random effects. The Gibbs sampler was run in a single chain of 500,000 iterations, with a burn-in of 50,000 and a thin interval of 50 iterations, totalling 9000 posterior samples for each parameter being estimated. The posterior means of the samples were used as the parameter estimates. The convergence of Monte Carlo chains of σ_a^2 and heritability were evaluated using the software postGibbsf90 (MISZTAL et al., 2012) and the R package BOA (SMITH, 2008).

The solutions of SNP effects ($\hat{u}$) were then obtained according to VanRaden et al. (2009) and Stranden & Garrick (2009) equation: $\hat{u} = DZ'\left[ZDZ'\right]^{-1}\hat{a}_g$, where *D* is a diagonal matrix with weights for SNP effects, *Z*

is a matrix relating genotypes of each locus, and $\hat{a}_g$ is the vector of predicted breeding values of genotyped animals. The *D* matrix and the SNP effects were iteratively recomputed following the method described by Wang *et al.* (2012) as "ssGBLUP/S1". In the first iteration, the diagonal elements of *D* (d_i) were assumed equal to 1 (*i.e.* the same weight for all the markers). For the subsequent iterations, d_i was calculated as: $d_i = \hat{u}_i^2 p_i(1-p_i)$, where $\hat{u}_i$ is the allele substitution effect of the i^{th} marker, estimated from the previous iteration, and p_i is the allele frequency of the second allele of the i^{th} marker. Prior to re-computing $\hat{u}$, the *D* matrix was normalized to enforce the total genetic variance to be constant across iterations. Three iterations were performed resulting in an increasing shrinkage from iteration 1 to 3 for the markers explaining lower variance and, as a consequence, in an increasing proportion of variance being explained by the remaining markers. According to Wang *et al.* (2014), 3 iterations are sufficient to reduce the noise of unimportant markers, *i.e.* to shrink their effects towards zero.

3.2.4. QTL mapping

The attempt to identify quantitative trait locus (QTL) affecting HR was made based in the proportion of variance explained by SNPs within non-overlapping consecutive 1Mb windows. As the effect of a QTL may be spread over a number of neighbouring SNPs, due to linkage disequilibrium, the presence of a previous described QTL was also inspected in the neighbouring 1Mb windows of the top ones.

A total of 2,523 windows were considered, spanning over all the autosomes, with an average density of 162±48 SNPs per window. The top 10 windows, that captured the highest proportion of variance explained by the markers, were assumed as important genomic regions. To reinforce their suggestive importance, scientific papers and the cattle QTLdb database (HU *et al.*, 2013) were consulted to inspect if any identified top window overlapped with a previously described QTL related to bovine reproductive traits. The UMD3.1 bovine genome assembly (ZIMIN *et al.*, 2009) was used as the reference to build the 1Mb windows and to consult the QTLdb database.

3.2.5. Candidate genes

Annotated genes located within the top 10 1Mb windows, from the three iterations of WssGBLUP (Table 2), were further inspected. The list of genes was provided by the NCBI Map Viewer tool (www.ncbi.nlm.nih.gov/mapview/), using the annotation release 103 and the *Bos taurus* UMD 3.1 as the primary assembly. The Panther database v.9.0 (MI et al., 2013) was then used to find biological processes related to these genes. Finally, it was investigated if there was evidence that these processes were associated with reproductive traits.

3.3 Results and Discussion

The estimated parameters, genetic additive variance and residual variance converged, as showed in supplementary material (Figure S1A and S1B). Their convergence was evaluated by the heritability pattern. Table S1 present some convergence statistics and Geweke test. The effective sample size and independent chain size for the additive variance were 149.3 and 110, respectively. All the statistics evidenced adequacy of the chain size.

It was detected 21 different top 10 marker-windows using w1, w2 and w3 weights (Table 1). There were five common windows between w1 and w2 weights, three between w3 and w1 and between w3 and w2 (in bold). Eight regions presented some association with reproductive traits (one star) and six regions presented some association with sexual precocity traits (two stars). It was observed an increasing in the proportion of variance explained by the top 10 marker windows from w1 to w3. The sum of this proportion was equal to 1.53, 3.63 and 4.66 for w1, w2 and w3, respectively. It is important to emphasize that the applied algorithm to calculate the weights in the WssGBLUP method iterated just on SNP effects (S1). As observed by Wang et al. (2012), the S1 algorithm promotes a slower “thinning” than S2, which recalculates the SNP and animal effects in each iteration. This effect can be better observed in

Figure 1, which presents the Manhattan plots of the variance explained by SNP windows for the three iterations. Although it was observed a stronger shrinkage in each subsequent iteration for the SNPs explaining lower variance, there were no clear peaks, even increasing the weights. This result indicates the polygenic nature of HR and can also be associated to the lack of sufficient information in the data to properly identify the important genomic regions affecting HR.

Table 1. Top 10 windows explaining the highest proportion of variance of heifer rebreeding in GWAS using weighted single step GBLUP method using three iterations (w1, w2 and w3).

	w1		w2		w3	
Rank	Ch/Wi[1]	Pvar[2]	Ch/Wi	Pvar	Ch/Wi	Pvar
1st	**17/050**	0.250	**17/050**	0.630	**17/050**	0.694
2nd	25/023*	0.179	11/075*	0.404	06/059	0.502
3rd	14/081	0.160	04/007	0.386	11/073*	0.486
4th	23/025**	0.151	**06/079**	0.354	26/043*	0.444
5th	**20/004***	0.146	**20/004***	0.328	**20/004***	0.440
6th	**06/079**	0.138	26/043*	0.328	14/022**	0.434
7th	14/065*	0.136	29/014*	0.313	**06/079**	0.428
8th	10/089**	0.125	25/023*	0.308	19/033*	0.419
9th	05/012**	0.124	06/111	0.297	22/043**	0.407
10th	04/007	0.122	11/072	0.280	09/068**	0.407

[1]Ch=chromosome; Wi=1Mb window within the chromosome; [2]pvar=proportion of variance explained by the SNPs within the window. The most common windows (ranked as top 10 in all analyses) are highlighted in bold. *Window (or neighbouring window) with a previous described QTL for bovine reproduction traits. **Window (or neighbouring window) with a previous described QTL for bovine sexual precocity traits.

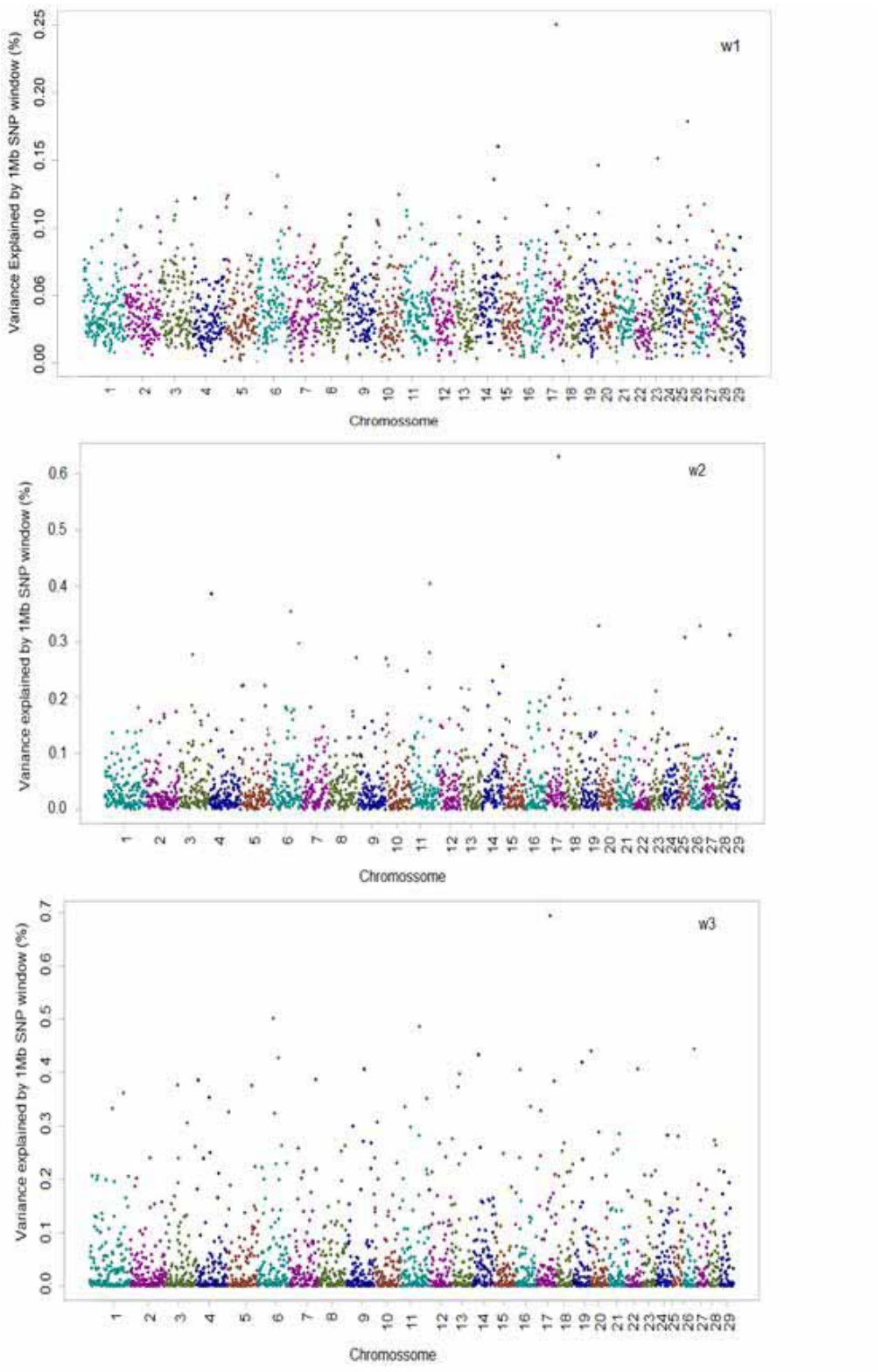

Figure 1. Manhattan Plots of a genomic wide association study on heifer rebreeding with three iterations (w1, w2 and w3).

The studies that reported QTLs associated with reproductive traits within the top 10 windows are presented in Table 2. Three of the QTLs are directly associated with HR: 56-d nonreturn rate for Holstein heifers (HOGLUND et al., 2009), age at puberty in Brahman and Tropical composite herds (HAWKEN et al., 2012) and length in days of interval from first to last insemination for Holstein cows (HOGLUND et al., 2009). Many calving traits were reported in the same important regions detected for HR. This could be explained because most of the studies involved taurine breeds, known to present calving problems. Despite fat thickness is not a reproductive trait it is associated with the female body condition, which affects the ability to conceive.

Table 2. The most important regions associated with heifer rebreeding that contained previously reported QTLs associated with reproductive traits for different breeds.

Ch/wi[1]	Traits reported in the region/Breed	Authors[2]
26/43	Direct first calving easy and maternal calving survival in first lactation/ Holstein	Sahana et al. (2011)
20/4	Calving easy direct and calving easy maternal/ Angus, Hereford, Simmental and Red Angus	Saatchi et al. (2012)
19/33	Calving traits/Holstein	Sahana et al. (2011)
	Gestation length/US Holstein and Italian Brown	Malteca et al. (2011)
	Scrotal circumference/ Angus	McClure et al. (2010)
9/68	Fertility index/ Holstein	Sahana et al. (2010)
23/25	Scrotal circumference/ Angus	McClure et al. (2010)
25/23	Maternal calving survival/Holstein	Sahana et al. (2011)
14/65	Fat thickness/ Angus	McClure et al. (2010)
5/12	Scrotal circumference/ Angus	McClure et al. (2010)
29/14	Scrotal circumference/ Angus	McClure et al. (2010)

22/43	56-d nonreturn rate for heifers/ Holstein	Hoglund et al. (2009)
14/22	Age at puberty[3]/Brahman and Tropical composite	Hawken et al. (2012)
11/75	Calving easy maternal/Hereford	Saatchi et al. (2014)
10/89	Length in days of interval from first to last insemination for cows/Holstein	Hoglund et al. (2009)

[1]Ch/Window= Chromosome/Window.
[2]Studies that detected regions or significant SNPs overlapping or neighboring the top 10 windows, using the UMD3.1 bovine genome assembly.
[3]Defined as age in days at first observed *corpus luteum* after frequently ovarian ultrasound scans.

In the 21 different top 10 marker-windows (and neighboring windows) associated with HR (Table 1) there were 271 annotated genes, including LOC genes according to NCBI Map Viewer. Excluding these LOC genes, remained 182 described genes (Supplementary material, Table S2). The software Panther (v.9.0) (Ml et al., 2013) classified these genes by biological process in 12 different categories, as showed in Figure 2. Most of the genes were associated with metabolic and cellular processes (37.9% of the total).

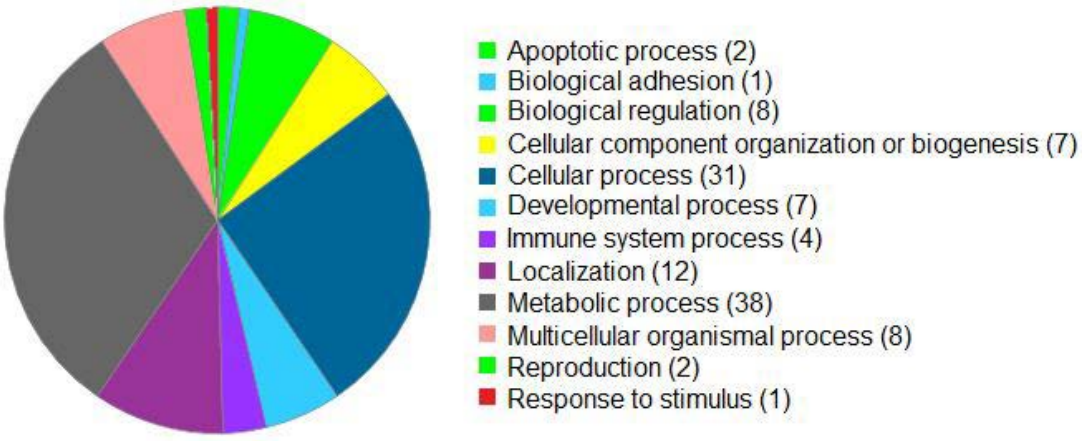

Figure 2. Biological processes of genes (and number of genes associated) located in the top 10 windows for heifer rebreeding.

Two genes were detected associated with reproduction process, the *Transforming growth factor beta 3* (TGFB3) and *von Willebrand C domain-*

containing protein (VWC2) genes. TGFB3 gene located in BTA10 at 88Mb is associated with polycystic ovary syndrome (PCOS) in human and cattle females (HATZIRODOS et al., 2011). Females with PCOS present the activity of this gene increased during ovarian fetal development. These authors concluded that when the glycoprotein fibrillin 3, that regulate TGFB (1, 2 and 3) bioactivity, is highly expressed in ovarian fetal tissue, the PCOS development in adult female life could be induced. Cows with PCOS present decrease in fertility and increase the calving interval (EYESTONE & AX, 1984). The TGFB superfamily members are also associated with immunomodulatory/inflammatory actions, some of which may be important during recognition and maintenance of pregnancy (WALKER et al., 2010). These authors observed that TGFB3 was down-regulated in pregnant animals, in accordance with Jones et al. (2006), that associated this protein with immunological functions and observed that TGFB3 inhibits potentially the trophoblast outgrowth (early embryo structure). Hoglund et al. (2009) detected a QTL in this region associated with length in days of interval from first to last insemination in Holstein cows, as presented in Table 2. VWC2 gene located in BTA4 at 60Mb was associated with fertility in hens (ELIS et al., 2009), been an ortholog gene between chicken and cattle (NCBI gene).

Uterine infections are common in postpartum period, especially in dairy cattle and cause infertility problems. In endometrial and ovarian follicle cells there are innate immune mechanisms to defense the reproductive female system against infections (SHELDON et al., 2009). These authors described the complex process that links the immune and reproductive systems in dairy cattle, and the relation of the steroids hormones with this process. Besides the immune system are in constant activity against pathogens, in embryonic implantation and development stage a complex immunological process is rolled to recognize the embryo, allowing the pregnancy success (WALKER et al., 2010). However, if this recognition fails the female could be infertile.

Four genes associated with immune system were detected, *interleukin 17A* (IL17A), *interleukin 17F* (IL17F), *FOS-like antigen* 2 (FOSL2) and *cell growth regulator with EF-hand domain 1* (CGREF1). Ozkan et al. (2014) reported lower IL17 cytokines rates in infertile women. These cytokines are

encoded by IL17 genes family, as IL17A and IL71F (BTA23 at 24Mb). Walker et al. (2010) observed an upregulating of some interleukins (IL1B, IL7, IL15, IL18) in pregnant cows, likely allowing an increase of immune tolerance promoting T-regulatory (T-reg) cells in uterus, that would allow the embryo implantation. It was not founded evidence linking directly the genes CGREF1 (BTA11 at 72Mb) and FOSL2 (BTA11 at 71Mb) with reproductive or immune processes, but Shimizu et al. (2010) observed CGREF1 gene in a cluster of genes that were differentially expressed in bovine endometrium under an estradiol treatment. Besides this gene was not discussed by the authors, this result could suggest an (indirect) association between CGREF1 and estradiol hormone. The FOSL2 gene was associated as a mediator in immune processes in mice (MURPHY *et al.*, 2013).

Others genes associated with reproductive traits, located within the top 10 windows, were also reported in literature. Messenger RNAs (mRNAs) from *solute carrier family 5 (sodium/inositol cotransporter), member 11* (SLC5A11), gene in BTA25 at 22Mb, were abundant in endometria from pregnant ewes (BAZER et al., 2008). *Leucine carboxyl methyltransferase 1* (LCMT1) located in BTA25 at 23Mb, was associated with embryo survival in mice (MACKAY et al., 2013). *G protein-coupled receptor 113* (GPR113), located in BTA11 at 73Mb, was differentially expressed in rhesus macaque (*Macaca mulatta*) *corpus luteum* during luteal phase of natural menstrual cycles (BOGAN et al., 2008). *Fibroblast growth factor 18* (FGF18), located in BTA20 at 3Mb, was associated with pathways in bovine ovarian granulosa cells (JIANG et al., 2013). Portela et al. (2010) also studied this gene and concluded that the FGF18 mRNA influences the process of atresia in ovarian follicles in cows. *DNA (cytosine-5-)-methyltransferase 3 alpha* (DNMT3A) protein, gene located in BTA11 at 73Mb, was present during embryo preimplantation development in bovine (GOLDING et al., 2003). *MicroRNA mir-206* (MIR206) protein, gene located in BTA23 at 24Mb, was associated with oocytes maturation in bovine (TESFAYE et al., 2009). Rabahi et al. (1999) concluded that *glutathione S-transferase alpha 2* (GSTA2), located in BTA23 at 24Mb, is associated with steroidogenically active cells, and is hormonally regulated by gonadotropins in the bovine ovarian

follicle. All these genes were orthologs between cattle and cited species (NCBI gene).

The existence of previously reported genes associated to reproductive traits within the top windows identified in the present study, reinforce the evidence that some of these regions can be associated with HR in Nelore cattle.

3.4 Conclusions

There was strong evidence that QTLs associated with heifer rebreeding in Nelore cattle were detected in the present study. These are candidate regions for searching causal mutations affecting the trait in further studies.

3.5 References

AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743-752, 2010.

BAZER, F. W.; BURGHARDT, R. C.; JOHNSON, G. A.; SPENCER, T. E.; WU, G. Interferons and progesterone for establishment and maintenance of pregnancy: interactions among novel cell signaling pathways. **Reproductive biology**, v. 8, n. 3, p. 179-211, 2008.

BOGAN, R. L.; MURPHY, M. J.; STOUFFER, R. L.; HENNEBOLD, J. D. Systematic Determination of Differential Gene Expression in the Primate Corpus Luteum during the Luteal Phase of the Menstrual Cycle. **Molecular Endocrinology**, v. 22, n. 5, p. 1260-1273, 2008.

BOLIGON, A. A.; ALBUQUERQUE L.G. Genetic parameters and relationships between heifers rebreeding and hip height in Nellore cattle**. Revista Brasileira de Zootecnia**, v. 41, n. 3, p. 598-602, 2012

CAMPOS, W. E.; SAUERESSIG, M. G. SATURNINO, H. M., et al; **Manejo reprodutivo em gado de corte**. Planaltina, DF: Embrapa Cerrados, 2005. 54 p. (Embrapa. Documentos, 134).

CORRÊA, E.S.; EUCLIDES FILHO, K.; ALVES, R.G.O., et al. **Desempenho reprodutivo em um sistema de produção de gado de corte.** Campo Grande: EMBRAPA/CNPGC, 2001. 33p. (Boletim de Pesquisa, 13).

DOYLE, S. P.; GOLDEN, B. L.; GREEN, R. D.; BRINKS, J. S. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. **Journal of Animal Science**, v. 78, p. 2091-2098, 2000.

ELIS, S.; BLESBOIS, E.; COUTY, I.; BALZERGUE, S.; MARTIN-MAGNIETTE, M. L.; BATELLIER, F.; GOVOROUN, M. S. Identification of Germinal Disk Region Derived Genes Potentially Involved in Hen Fertility. **Molecular Reproduction & Development**, v. 76, p.1043-1055, 2009.

EYESTONE, W. H.; AX, R. L. A review of ovarian follicular cysts in cows, with comparisons to the condition in women, rats and rabbits. **Theriogenology**, v. 22, n. 2, p. 109-125, 1984.

FRAGOMENI, B. O.; MISZTAL, I.; LOURENCO, D. L.; AGUILAR, I.; OKIMOTO, R.; MUIR, W. M. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. **Frontiers in Genetics**, v. 5, n. 332, p. 1-7, 2014.

GIANOLA, D.; FOULLEY, J. L. Sire evaluation for ordered categorical data with a threshold model. **Genetics Selection Evolution**, v. 15, p. 201-224, 1983.

GOLDING, M. C.; WESTHUSIN, M. E. Analysis of DNA (cytosine 5) methyltransferase mRNA sequence and expression in bovine preimplantation embryos, fetal and adult tissues. **Gene Expression Patterns**, v. 3, p. 551-558, 2003.

GUARINI, A. R.; NEVES, H. H. R.; SCHENKEL, F. S.; CARVALHEIRO, R.; OLIVEIRA, J. A.; QUEIROZ, S. A. Genetic relationship among reproductive traits in Nellore cattle. **Animal**, p. 1-6, 2014.

HABIER, D.; FERNANDO, R.L.; KIZILKAYA, K.; GARRICK, D. J. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, n. 186, p. 1-12, 2011.

HATZIRODOS, N.; BAYNE, R. A.; IRVING-RODGERS, H. F.; HUMMITZSCH, K.; SABATIER, L.; LEE, S.; BONNER, W.; GIBSON, M. A.; RAINEY, W. E.; CARR, B. R.; MASON, H. D.; REINHARDT, D. P.; ANDERSON, R. A.;

RODGERS, R. J. Linkage of regulators of TGF-β activity in the fetal ovary to polycystic ovary syndrome. **The FASEB Journal**, v. 25, p. 2256-2265, 2011.

HAWKEN, R. J.; ZHANG, Y. D.; FORTES, M. R. S.; COLLIS, E.; BARRIS, W. C.; CORBET, N. J.; WILLIAMS, P. J.; FORDYCE, G.; HOLROYD, R. G.; WALKLEY, J. R. W.; BARENDSE, W.; JOHNSTON, D. J.; PRAYAGA, K. C.; TIER, B.; REVERTER, A.; LEHNERT, S. A. Genome-wide association studies of female reproduction in tropically adapted beef cattle. **Journal of Animal Science**, v. 90, p. 1398-1410, 2012.

HÖGLUND, J. K.; BUITENHUIS, A. J.; GULDBRANDTSEN, B.; SU, G.; THOMSEN, B.; LUND, M. S. Overlapping chromosomal regions for fertility traits and production traits in the Danish Holstein population. **Journal of Dairy Science**, v. 92, p. 5712-5719, 2009.

HU, Z. H.; PARK, C. A.; WU, X. L.; REECY, J. M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, v. 41, p. 871-879, 2013.

JIANG, Z.; GUERRERO-NETRO, H. M.; JUENGEL, J. L.; PRICE, C. A. Divergence of intracellular signaling pathways and early response genes of two closely related fibroblast growth factors, FGF8 and FGF18, in bovine ovarian granulosa cells. **Molecular and Cellular Endocrinology**, v. 375, p. 97-105, 2013.

JONES, R. L.; STOIKOS, C.; FINDLAY, J. K.; SALAMONSEN, L. A. TGF-β superfamily expression and actions in the endometrium and placenta. **Reproduction**, v. 132, p. 217-232, 2006.

MACKAY, K. B.; TU, Y.; YOUNG, S. G.; CLARKE, S. G. Circumventing Embryonic Lethality with Lcmt1Deficiency: Generation of Hypomorphic Lcmt1 Mice with Reduced Protein Phosphatase 2A Methyltransferase Expression and Defects in Insulin Signaling. **PLoS ONE**, v. 8, n. 6, p. 1-11, 2013.

MCCLURE, M. C.; MORSCI, N. S.; SCHNABEL, R. D.; KIM, J. W.; YAO, P.; ROLF, M. M.; MCKAY, S. D.; GREGG, S. J.; CHAPPLE, R. H.; NORTHCUTT, S. L.; TAYLOR, J.F. A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. **Animal Genetics**, v. 41, p. 597-607, 2010.

MI, H.; MURUGANUJAN, A.; CASAGRANDE, J. T.; THOMAS, P. D. Large-scale gene function analysis with the PANTHER classification system. **Nature Protocols**. v. 8, p. 1551-66, 2013.

MISZTAL, I. **BLUPF90 - a flexible mixed model program in Fortran 90**. Manual. Animal and Dairy Science, University of Georgia, Athens, GA, USA.

[S.l.: s.n.], 2012, 25p. Available in: <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90.pdf>. Accessed in Feb. 2014.

MURPHY, T. L.; TUSSIWAND, R.; MURPHY, K. M. Specificity through cooperation: BATF–IRF interactions control immune-regulatory networks. **Nature reviews immunology**, v. 13, p. 499-509, 2013.

NATIONAL ASSOCIATION OF ANIMAL BREEDERS. **Rebreeding First-Calf Heifers**. United States, 2004. 3 p. (Angus Beef Bulletin).

OZKAN, Z. S.; DEVECI, D.; KUMBAK, B.; SIMSEK, M.; ILHAN, F.; SEKERCIOGLU, S.; SAPMAZ, E. What is the impact of Th1/Th2 ratio, SOCS3, IL17, and IL35 levels in unexplained infertility? **Journal of reproductive immunology**, v. 103, p. 53-58, 2014.

PORTELA, V. M.; MACHADO, M.; BURATINI JR., J.; ZAMBERLAM, G.; AMORIM, R. L.; GONCALVES, P.; PRICE, C. A. Expression and Function of Fibroblast Growth Factor 18 in the Ovarian Follicle in Cattle. **Biology of reproduction**, n. 83, p. 339-346, 2010.

RABAHI, F.; BRÛLÉ, S.; SIROIS, J.; BECKERS, J. F.; SILVERSIDES, D. W.; LUSSIER, J. G. High Expression of Bovine a Glutathione S-Transferase (GSTA1, GSTA2) Subunits Is Mainly Associated with Steroidogenically Active Cells and Regulated by Gonadotropins in Bovine Ovarian Follicles. **Endocrinology**, v. 140, n. 8, 1999.

RILEY, D. G.; CHASE, C. C. JR.; COLEMAN, S. W.; OLSON, T. A.; RANDEL, R. D. Evaluation of tropically-adapted straightbred and crossbred beef cattle: Heifer age and size at first conception and characteristics of their first calves. **Journal of Animal Science**. v. 88, n. 10, p. 3173-82, 2010.

SAATCHI, M.; SCHNABEL, R. D.; TAYLOR, J. F.; GARRICK, D. J. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. **BMC Genomics**, v. 15, n. 442, p. 1-16, 2014.

SAHANA, G.; GULDBRANDTSEN, B.; LUND, M. S. Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. **Journal of Dairy Science**, v. 94, p. 479-486, 2011.

SHELDON, I. M.; CRONIN, C.; GOETZE, L.; DONOFRIO, G.; SCHUBERTH, H. J. Defining Postpartum Uterine Disease and the Mechanisms of Infection and Immunity in the Female Reproductive Tract in Cattle. **Biology of reproduction**, v. 81, n. 6, p. 1025-1032, 2009.

SHIMIZU, T.; KREBS, S.; BAUERSACHS, S.; BLUM, H.; WOLF, E.; MIYAMOTO, A. Actions and interactions of progesterone and estrogen on transcriptome profiles of the bovine endometrium. **Physiological genomics**, v. 42, p. 290-300, 2010.

SMITH, B. J. Bayesian Output Analysis Program (BOA) for MCMC version 1.1 user's manual. [S.l.: s.n.], 2015, p. 40.

SORENSEN, D.; GIANOLA, D. **Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics**. New York: Springer; 2002.

STRANDEN, I.; GARRICK, D. J. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. **Journal of Dairy Science**, v. 92, n. 6, p. 2971-2975, 2009.

TESFAYE, D.; WORKU, D.; RINGS, F.; PHATSARA, C.; THOLEN, E.; SCHELLANDER, K.; HOELKER, M. Identification and Expression Profiling of MicroRNAs During Bovine Oocyte Maturation Using Heterologous Approach. **Molecular Reproduction & Development**, v. 76, p. 665-677, 2009.

TSURUTA, S.; MISZTAL, I. THRGIBBS1F90 for estimation of variance component with threshold and linear models. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8., 2006, Belo Horizonte. **Proceedings...** Belo Horizonte: UFMG, 2006. p.27-31.

VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v. 92, n. 1, p. 16-24, 2009.

WALKER, C. G.; MEIER, S.; LITTLEJOHN, M. D.; LEHNERT, K.; ROCHE, J. R.; MITCHELL, M. D. Modulation of the maternal immune system by the pre-implantation embryo. **BMC genomics**, v. 11, p. 1-13, 2010.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; FERNANDO, R. L.; VITEZICA, Z.; OKIMOTO, R.; WING, T.; HAWKEN, R.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 5, n. 134, 2014.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v. 94, n. 2 p. 73-83, 2012.

ZIMIN, A. V.; DELCHER, A. L.; FLOREA, L.; KELLEY, D. R.; SCHATZ, M. C.; PUIU, D.; HANRAHAN, F.; PERTEA, G.; VAN TASSELL, C. P.; SONSTEGARD, T. S.; MARÇAIS, G.; ROBERTS, M.; SUBRAMANIAN, P.; YORKE, J. A.; SALZBERG, S. L. A whole-genome assembly of the domestic cow, *Bos taurus.* **Genome Biology**, v. 10, i. 4, 2009.

APPENDIX
(Supplemental Material)

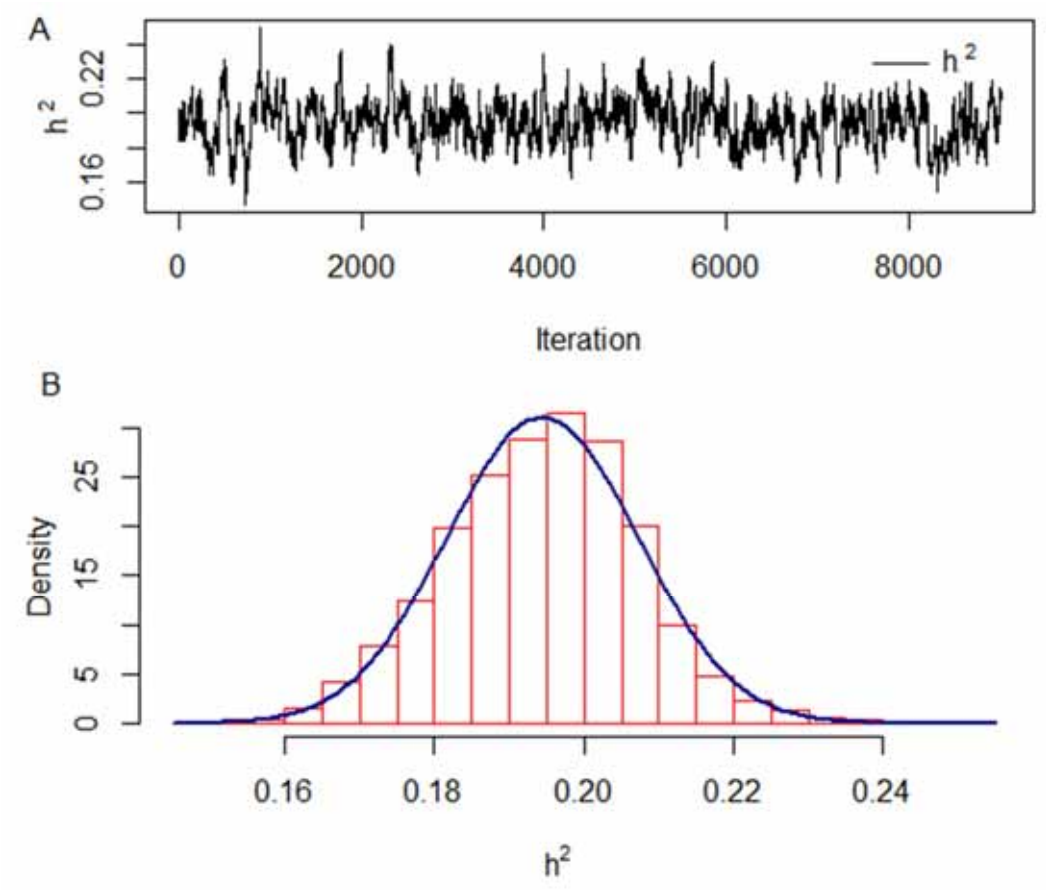

Figure S1. Trace (**A**) and histogram (**B**) plots of h^2 posterior samples.

Table S1. Estimates of marginal posterior distributions of heritability for heifers rebreeding of Nellore Cattle.

Marginal posterior distributions	
Mean±s.d.	0.194 ± 0.013
Mode	0.190
Median	0.195
HPD* 95%	0.168; 0.218
Geweke Test	
z-score	-0.576
p-value	0.565

*HPD 95%= 95% highest posterior density interval

Table S2. Annotated genes located in the most important regions.

Region Ch:wi[1]	Position (Kb)		Gene Symbol	Gene ID	Description
	start	stop			
25:22..23	21759577	22137203	PRKCB	282325	*protein kinase C, beta*
	22178045	22278775	CACNG3	538496	*calcium channel, voltage-dependent, gamma subunit 3*
	22458298	22493834	RBBP6	282034	*retinoblastoma binding protein 6*
	22506484	22506555	TRNAC-GCA	100170902	*transfer RNA cysteine (anticodon GCA)*
	22514641	22746792	TNRC6A	541118	*trinucleotide repeat containing 6A*
	22780656	22838003	SLC5A11	539084	*solute carrier family 5 (sodium/inositol cotransporter), member 11*
	22843172	22900633	ARHGAP17	504245	*Rho GTPase activating protein 17*
	22964717	23034012	LCMT1	618021	*leucine carboxyl methyltransferase 1*
	22980790	22980862	TRNAG-CCC	100170915	*transfer RNA glycine (anticodon CCC)*
17:49..50	48198810	49092160	TMEM132D	100297935	*transmembrane protein 132D*
	49196029	49313907	GLT1D1	516510	*glycosyltransferase 1 domain containing 1*
	49306959	49307031	TRNAE-UUC	100170899	*transfer RNA glutamic acid (anticodon UUC)*
	49336390	49366012	SLC15A4	510499	*solute carrier family 15 (oligopeptide transporter), member 4*
	49443475	49739775	TMEM132C	512602	*transmembrane protein 132C*
14:80..81	80029572	80865916	RALYL	516989	*RALY RNA binding protein-like*
20:3..4	3064510	3066676	TLX3	618473	*T-cell leukemia homeobox 3*
	3079143	3079213	TRNAK-UUU	100170895	*transfer RNA lysine (anticodon UUU)*
	3111166	3124100	NPM1	614028	*nucleophosmin (nucleolar phosphoprotein B23, numatrin)*
	3132655	3195037	FGF18	533929	*fibroblast growth factor 18*
	3497942	3503110	SMIM23	101902619	*small integral membrane protein 23*
	3582475	3715174	FBXW11	540642	*F-box and WD repeat domain containing 11*
	3758358	3881011	STK10	526376	*serine/threonine kinase 10*
	3878916	3898485	EFCAB9	618333	*EF-hand calcium binding domain 9*

	3903960	3968828	UBTD2	541134	*ubiquitin domain containing 2*
6:78..79	78020861	78020932	TRNAC-GCA	100170902	*transfer RNA cysteine (anticodon GCA)*
	78375230	79314449	LPHN3	282651	*latrophilin 3*
11:71..76	71116626	71242728	PLB1	506508	*phospholipase B1*
	71326652	71349193	FOSL2	509889	*FOS-like antigen 2*
	71402758	71876930	BRE	614152	*brain and reproductive organ-expressed (TNFRSF1A modulator)*
	71827005	71921665	RBKS	513276	*ribokinase*
	71923198	71931587	MRPL33	614711	*mitochondrial ribosomal protein L33*
	71945747	72021177	SLC4A1AP	533732	*solute carrier family 4 (anion exchanger), member 1, adaptor protein*
	72021273	72033987	SUPT7L	539152	*suppressor of Ty 7 (S. cerevisiae)-like*
	72034549	72058245	GPN1	508522	*GPN-loop GTPase 1*
	72058294	72061512	CCDC121	782407	*coiled-coil domain containing 121*
	72066251	72097295	ZNF512	533884	*zinc finger protein 512*
	72147750	72220365	GCKR	100140421	*glucokinase (hexokinase 4) regulator*
	72163773	72183229	FNDC4	781806	*fibronectin type III domain containing 4*
	72185339	72222960	IFT172	100848219	*intraflagellar transport 172*
	72223034	72224689	KRTCAP3	508550	*keratinocyte associated protein 3*
	72224811	72237416	NRBP1	532919	*nuclear receptor binding protein 1*
	72252676	72273095	PPM1G	286880	*protein phosphatase, Mg2+/Mn2+ dependent, 1G*
	72273511	72277103	ZNF513	100138621	*zinc finger protein 513*
	72277209	72283024	SNX17	529972	*sorting nexin 17*
	72283196	72287773	EIF2B4	521926	*eukaryotic translation initiation factor 2B, subunit 4 delta, 67kDa*
	72291215	72311947	GTF3C2	782752	*general transcription factor IIIC, polypeptide 2, beta 110kDa*
	72315228	72324568	MPV17	505763	*MpV17 mitochondrial inner membrane protein*
	72326391	72326953	UCN	518336	*urocortin*
	72327350	72349010	TRIM54	535320	*tripartite motif containing 54*
	72349455	72367155	DNAJC5G	616608	*DnaJ (Hsp40) homolog, subfamily C, member 5 gamma*

	72360262	72373191	SLC30A3	512803	*solute carrier family 30 (zinc transporter), member 3*
	72383235	72404916	CAD	504261	*carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase*
	72405112	72409446	ATRAID	100125942	*all-trans retinoic acid-induced differentiation factor*
	72418186	72425535	SLC5A6	516021	*solute carrier family 5 (sodium/multivitamin and iodide cotransporter), member 6*
	72470283	72474442	TCF23	616841	*transcription factor 23*
	72489673	72492238	PRR30	782932	*proline rich 30*
	72493261	72497358	PREB	525256	*prolactin regulatory element binding*
	72497318	72504362	ABHD1	510774	*abhydrolase domain containing 1*
	72508048	72524708	CGREF1	507586	*cell growth regulator with EF-hand domain 1*
	72525752	72538457	KHK	614868	*ketohexokinase (fructokinase)*
	72539296	72547268	EMILIN1	540451	*elastin microfibril interfacer 1*
	72554350	72555599	OST4	617040	*oligosaccharyltransferase 4 homolog (S. cerevisiae)*
	72555450	72573569	AGBL5	538585	*ATP/GTP binding protein-like 5*
	72573765	72573837	TRNAA-AGC	100170883	*transfer RNA alanine (anticodon AGC)*
	72574188	72574276	TRNAY-GUA	10017088	*transfer RNA tyrosine (anticodon GUA)*
	72582288	72590697	TMEM214	514683	*transmembrane protein 214*
	72598138	72653485	MAPRE3	528839	*microtubule-associated protein, RP/EB family, member 3*
	72671399	72760988	DPYSL5	100126171	*dihydropyrimidinase-like 5*
	72779997	72788067	CENPA	613898	*centromere protein A*
	72793629	72809693	SLC35F6	525690	*solute carrier family 35, member F6*
	72831888	72868438	KCNK3	519188	*potassium channel, subfamily K, member 3*
	72868317	72971983	CIB4	617324	*calcium and integrin binding family member 4*
	72918581	72918653	TRNAE-UUC	100170899	*transfer RNA glutamic acid (anticodon UUC)*

	72973675	72991252	C11H2orf70	767918	*chromosome 11 open reading frame, human C2orf70*
	72995726	73083691	OTOF	617299	*otoferlin*
	73084125	73122618	DRC1	509524	*dynein regulatory complex subunit 1*
	73134169	73173557	EPT1	506381	*ethanolaminephosphotransferase 1 (CDP-ethanolamine-specific)*
	73173796	73210096	GPR113	522893	*G protein-coupled receptor 113*
	73215423	73246308	HADHB	281811	*hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), beta subunit*
	73251925	73288804	HADHA	281810	*hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit*
	73290662	73307412	GAREML	783429	*GRB2 associated, regulator of MAPK1-like*
	73335433	73400455	RAB10	783373	*RAB10, member RAS oncogene family*
	73388934	73389006	TRNAC-ACA	100170901	*transfer RNA cysteine (anticodon ACA)*
	73451221	73487060	KIF3C	777770	*kinesin family member 3C*
	73533277	73646859	ASXL2	100140419	*additional sex combs like transcriptional regulator 2*
	73693621	73935363	DTNB	534119	*dystrobrevin, beta*
	73839206	73839277	TRNAC-GCA	100170902	*transfer RNA cysteine (anticodon GCA)*
	73975870	73993397	DNMT3A	359716	*DNA (cytosine-5-)-methyltransferase 3 alpha*
	73977000	73977079	MIR1301	100313397	*microRNA mir-1301*
	74109268	74116894	POMC	281416	*proopiomelanocortin*
	74118767	74216478	EFR3B	517590	*EFR3 homolog B (S. cerevisiae)*
	74319425	74360335	DNAJC27	540033	*DnaJ (Hsp40) homolog, subfamily C, member 27*
	74381532	74465806	ADCY3	535603	*adenylate cyclase 3*
	74465013	74480160	CENPO	767930	*centromere protein O*
	74480125	74483870	PTRHD1	617088	*peptidyl-tRNA hydrolase domain containing 1*
	74494566	74710538	NCOA1	525346	*nuclear receptor coactivator 1*

	74856277	74982506	ITSN2	521706	*intersectin 2*
	74981143	74996869	FAM228A	768001	*family with sequence similarity 228, member A*
	74997149	75032065	FAM228B	100125925	*family with sequence similarity 228, member B*
	75023805	75027668	PFN4	508874	*profilin family, member 4*
	75032015	75039194	TP53I3	508875	*tumor protein p53 inducible protein 3*
	75039681	75047806	SF3B6	508876	*splicing factor 3b, subunit 6, 14kDa*
	75047687	75065261	FKBP1B	785179	*FK506 binding protein 1B, 12.6 kDa*
	75049603	75080553	C11H2orf44	785211	*chromosome 11 open reading frame, human C2orf44*
	75084498	75098468	MFSD2B	539221	*major facilitator superfamily domain containing 2B*
	75099783	75130859	UBXN2A	613578	*UBX domain protein 2A*
	75141071	75263986	ATAD2B	518785	*ATPase family, AAA domain containing 2B*
	75290644	75623195	KLHL29	537141	*kelch-like family member 29*
23:24..25	23795613	24254448	PKHD1	537895	*polycystic kidney and hepatic disease 1 (autosomal recessive)*
	24308042	24308127	MIR206	100313017	*microRNA mir-206*
	24312330	24312413	MIR133B	100312996	*microRNA mir-133b*
	24346988	24350509	IL17A	282863	*interleukin 17A*
	24350987	24351059	TRNAG-CCC	100170915	*transfer RNA glycine (anticodon CCC)*
	24390361	24398587	IL17F	506030	*interleukin 17F*
	24438946	24456316	MCM3	281302	*minichromosome maintenance complex component 3*
	24567875	24571512	PAQR8	531018	*progestin and adipoQ receptor family member VIII*
	24588032	24627769	EFHC1	510124	*EF-hand domain (C-terminal) containing 1*
	24637958	24702830	TRAM2	617287	*translocation associated membrane protein 2*
	24803694	24814246	TMEM14A	510383	*transmembrane protein 14A*
	24823645	24837368	GSTA2	281805	*glutathione S-transferase alpha 2*
	24875605	24890466	GSTA3	777644	*glutathione S-transferase alpha 3*

	24894756	24916117	GSTA5	521685	*glutathione S-transferase alpha 5*
	24935503	24961752	GSTA3	768055	*glutathione S-transferase, alpha 3*
	24963224	24977372	GSTA4	533917	*glutathione S-transferase alpha 4*
	24982762	25027197	ICK	521685	*glutathione S-transferase alpha 5*
14:64..65	64025334	64034795	ODF1	281366	*outer dense fiber of sperm tails 1*
	64169767	64308334	UBR5	515582	*ubiquitin protein ligase E3 component n-recognin 5*
	64315036	64353050	RRM2B	528960	*ribonucleotide reductase M2 B (TP53 inducible)*
	64568850	64899889	NCALD	281940	*neurocalcin delta*
	64902765	65076564	GRHL2	519918	*grainyhead-like 2 (Drosophila)*
10:88..89	87921827	88238777	TTLL5	538422	*tubulin tyrosine ligase-like family, member 5*
	88242178	88275212	TGFB3	538957	*transforming growth factor, beta 3*
	88272278	88379908	IFT43	513228	*intraflagellar transport 43*
	88428551	88527778	GPATCH2L	509393	*G patch domain containing 2-like*
	88763706	88818476	ESRRB	528437	*estrogen-related receptor beta*
	11972438	11978864	CCDC59	615189	*coiled-coil domain containing 59*
	11978341	12088741	METTL25	513885	*methyltransferase like 25*
4:6..7	5908453	6046089	VWC2	784924	*von Willebrand factor C domain containing 2*
26:42..43	42038699	42199345	ATE1	534601	*arginyltransferase 1*
	42214937	42228682	NSMCE4A	527312	*non-SMC element 4 homolog A (S. cerevisiae)*
	42250855	42485658	TACC2	533768	*transforming, acidic coiled-coil containing protein 2*
	42482649	42482717	MIR2396	100313189	*microRNA mir-2396*
	42493403	42546248	BTBD16	518515	*BTB (POZ) domain containing 16*
	42579541	42635461	PLEKHA1	513040	*pleckstrin homology domain containing, family A (phosphoinositide binding specific) member 1*
	42659747	42717176	HTRA1	282326	*HtrA serine peptidase 1*
	42775664	42813474	DMBT1	404104	*deleted in malignant brain tumors 1*
	42838876	42845746	SPADH2	527722	*spermadhesin 2*

	42871224	42879183	SPADH1	282373	*spermadhesin 1*
	42882755	42885754	C26H10orf120	617773	*chromosome 26 open reading frame, human C10orf120*
	42965045	42965117	TRNAG-CCC	100170915	*transfer RNA glycine (anticodon CCC)*
	13003058	13004913	FAM181B	613623	*family with sequence similarity 181, member B*
	13578501	13578572	TRNAC-GCA	100170902	*transfer RNA cysteine (anticodon GCA)*
	13649345	13649417	TRNAC-GCA	100170902	*transfer RNA cysteine (anticodon GCA)*
6:110..111	109886766	110062393	SLC2A9	100337051	*solute carrier family 2 (facilitated glucose transporter), member 9*
	110000858	110000930	TRNAG-CCC	100170915	*transfer RNA glycine (anticodon CCC)*
	110136745	110178567	WDR1	533223	*WD repeat domain 1*
	110338920	110353711	ZNF518B	539006	*zinc finger protein 518B*
	110394181	110603541	CLNK	100298588	*cytokine-dependent hematopoietic cell linker*
	110935376	110935447	TRNAA-UGC	100170885	*transfer RNA alanine (anticodon UGC)*
	58241488	58480596	NWD2	539460	*NACHT and WD repeat domain containing 2*
	58486550	58565111	C6H4orf19	511424	*chromosome 6 open reading frame, human C4orf19*
	58547169	58547240	TRNAC-GCA	100170902	*transfer RNA cysteine (anticodon GCA)*
	58581624	58660258	RELL1	768210	*RELT-like 1*
	58778059	58816669	PGM2	506980	*phosphoglucomutase 2*
	58838060	59065295	TBC1D1	282704	*TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1*
14:21..22	20740959	21019777	SPIDR	512910	*scaffolding protein involved in DNA repair*
	21033276	21034329	H3F3C	512741	*H3 histone, family 3C*
	21037822	21164607	PRKDC	512740	*protein kinase, DNA-activated, catalytic polypeptide*
	21164877	21192689	MCM4	504367	*minichromosome maintenance complex component 4*
	21225303	21245523	UBE2V2	286803	*ubiquitin-conjugating enzyme E2 variant 2*
	21272523	21273328	MCM4	504367	*minichromosome maintenance complex component 4*

	21450645	21464764	EFCAB1	505272	*EF-hand calcium binding domain 1*
	21577308	21580910	SNAI2	520631	*snail family zinc finger 2*
	21617777	21622268	C14H8orf22	100126183	*chromosome 14 open reading frame, human C8orf22*
	21933264	22350056	SNTG1	517353	*syntrophin, gamma 1*
	32339870	32427059	HS3ST3A1	521959	*heparan sulfate (glucosamine) 3-O-sulfotransferase 3A1*
	32696437	32804797	COX10	511440	*COX10 homolog, cytochrome c oxidase assembly protein, heme A: farnesyltransferase (yeast)*
	32829907	32869783	HS3ST3B1	617069	*heparan sulfate (glucosamine) 3-O-sulfotransferase 3B1*
22:42..43	40603979	42134601	FHIT	692183	*fragile histidine triad*
	42888731	43159638	C22H3orf67	534800	*chromosome 22 open reading frame, human C3orf67*
9:67:68	67019071	67594209	PTPRK	509657	*protein tyrosine phosphatase, receptor type, K*
	67860085	67860156	TRNAY-AUA	100170887	*transfer RNA tyrosine (anticodon AUA)*

[1]Ch=Chromossome, wi= start window..end window in Mb.