UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

PREDICTION OF GENOMIC-ENABLED BREEDING VALUES AND GENOME-WIDE ASSOCIATION STUDY FOR FEEDLOT AVERAGE DAILY WEIGHT GAIN IN NELORE CATTLE

Adriana Luiza Somavilla
Zootecnista

2015

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

PREDICTION OF GENOMIC-ENABLED BREEDING VALUES AND GENOME-WIDE ASSOCIATION STUDY FOR FEEDLOT AVERAGE DAILY WEIGHT GAIN IN NELORE CATTLE

Adriana Luiza Somavilla

Orientador: Prof. Dr. Danisio Prado Munari

Co-orientadoras: Dra. Luciana Correia de Almeida Regitano

Dra. Fabiana Barichello Mokry

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

2015

Somavilla, Adriana Luiza
S693p Prediction of genomic-enabled breeding values and genome-wide association study for feedlot average daily weight gain in Nelore cattle / Adriana Luiza Somavilla. – – Jaboticabal, 2015
v, 71 p. : il. ; 29 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2015
Orientador: Danísio Prado Munari
Coorientadora: Luciana Correia de Almeida Regitano
Coorientadora: Fabiana Barichello Mokry
Banca examinadora: Rogério Abdallah Curi, Nedenia Bonvino Stafuzza, Luiz Lehmann Coutinho, Marcos Vinicius Gualberto Barbosa da Silva
Bibliografia

1. Bayesian inference. 2. *Bos taurus indicus*. 3. Functional annotation. 4. Growth rate. 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: PREDICTION OF GENOMIC-ENABLED BREEDING VALUES AND GENOME-WIDE ASSOCIATION STUDY FOR FEEDLOT AVERAGE DAILY WEIGHT GAIN IN NELORE CATTLE

AUTORA: ADRIANA LUIZA SOMAVILLA

ORIENTADOR: Prof. Dr. DANÍSIO PRADO MUNARI
CO-ORIENTADORA: Profa. Dra. FABIANA BARICHELLO MOKRY
CO-ORIENTADORA: Profa. Dra. LUCIANA CORREIA DE ALMEIDA REGITANO

Aprovada como parte das exigências para obtenção do Título de DOUTOR EM GENÉTICA E MELHORAMENTO ANIMAL , pela Comissão Examinadora:

Prof. Dr. DANÍSIO PRADO MUNARI
Departamento de Ciências Exatas / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. LUIZ LEHMANN COUTINHO
Universidade de São Paulo / Piracicaba/SP

Prof. Dr. ROGERIO ABDALLAH CURI
Departamento de Melhoramento e Nutrição Animal / Faculdade de Medicina Veterinaria e Zootecnia de Botucatu

Prof. Dr. MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA
Empresa Brasileira de Pesquisa Agropecuária / Juiz de Fora/MG

Profa. Dra. NEDENIA BONVINO STAFUZZA
Pós-doutoranda / Departamento de Ciências Exatas / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Data da realização: 26 de junho de 2015.

DADOS CURRICULARES DO AUTOR

Adriana Luiza Somavilla – nascida em 5 de maio de 1985, na cidade de Curitiba – PR, filha de Francisco Carlos Somavilla e Maria Heleno Somavilla. Iniciou em março de 2005 o curso de graduação em Zootecnia na Universidade Federal do Paraná, obtendo o título em novembro de 2010. Durante a graduação exerceu atividades de monitora e de iniciação científica em Melhoramento Genético Animal, sob orientação da Profª. Drª. Laila Talarico Dias. Em março de 2011, ingressou no Programa de Pós-graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus de Jaboticabal – SP, como bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), obtendo o grau de mestre em 30 de julho de 2012, sob orientação do Prof. Dr. Mauricio Mello de Alencar e co-orientação da Dra. Luciana Correia de Almeida Regitano. Em agosto de 2012, iniciou o doutorado em Genética e Melhoramento Animal na mesma universidade, como bolsista da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e posteriormente da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Durante o ano de 2014 realizou o doutorado sanduíche na Universidade de Wisconsin-Madison, USA, sob supervisão do Prof. Dr. Guilherme Jordão de Magalhães Rosa, por meio da Bolsa Estágio de Pesquisa no Exterior (BEPE) fornecida pela FAPESP. Obteve o grau de doutora em 26 de junho de 2015, sob orientação do Prof. Dr. Danísio Prado Munari e co-orientação da Dra. Luciana Correia de Almeida Regitano e da Dra. Fabiana Barichello Morky.

O MUNDO DESTAMANHO

Quando me ócio
de mim mesmo
desensimesmado

Um ser em fotossíntesse
carne, osso, pele e sol

Me desanuvio:

O que sou
está sido,
vai-se-sendo,
entre o pouco já sabido
e o muito que nem desconfio.

Luiz Carlos Heleno

Dedico esse trabalho à memória do meu orientador, **Prof Marcílio Dias Silveira da Mota.**

"..E até lá, vamos viver
Temos muito ainda por fazer..."

AGRADECIMENTOS

Aos meus pais, Maria e Francisco, e meus irmãos, Jefferson e Ramilio, pelo amor e apoio incondicionais às minhas escolhas. Aos demais familiares, pelo apoio emocional, e muitas vezes financeiro. Em especial às minhas primas-irmãs Sariana, Fabiula, Gislaine e Cecília, por esses 30 anos de amizade.

Ao meu namorado Rafael, pelo amor, amizade, apoio e paciência. Obrigada por estar ao meu lado nos momentos bons e por me fazer ver o lado bom das coisas nos momentos difíceis.

Aos meus ex-orientadores, hoje colegas de profissão e amigos, Laila e Rodrigo, pela confiança, amizade e apoio.

Ao meu orientador, Prof Danisio, por ter me adotado no decorrer do doutorado, pelo apoio profissional e pessoal. À minha orientadora Dra Luciana Regitano, pela confiança depositada em mim quando eu ainda estava no final da graduação, pelo apoio, conversas e oportunidades ao longos dos últimos 5 anos. À minha co-orientadora Dra Fabiana Mokry, pelas longas conversas, conselhos e orientações profissionais e pessoais, apoio e palavras de incentivo.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos no início do curso.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pela concessão das bolsas de estudo no país e no exterior, que permitiram a realização desse projeto, do doutorado sanduíche e a participação em cursos e congressos, que contribuíram imensamente para minha formação profissional.

Ao pessoal da Embrapa: Flávia, Marina, Suelen, Andressa, Priscila, Polyana, Dani, Marcela, Wellison, Wil, Giba e Maurício, pelos anos de convivência e apoio.

Ao Professor Guilherme Rosa, da Universidade de Wisonsin-Madison, EUA, por me receber no período do Doutorado Sanduíche, pelas conversas e ensinamentos transmitidos. Ao grupo do Departamento de "Animal Science" da UW, pelos cafés e conversas interessantes, que fizeram com que o ano de 2014 fosse um marco na minha vida, profissional e pessoal. Em especial aos amigos que fiz em Madison: Angelina, Isabela, Mariana, Gerardo, Tom, Afshin, Yaodong, Keiichi, Pancho, Kathy e Lei, que tornaram a experiência mais valiosa.

CONTENTS

PREDICTION OF GENOMIC-ENABLED BREEDING VALUES AND GENOME-WIDE ASSOCIATION STUDY FOR FEEDLOT AVERAGE DAILY WEIGHT GAIN IN NELORE CATTLE

ABSTRACT: Selection for fast growth rates using number of days to achieve specific weights or average weight gain would result in shorter production periods. Maintaining the rate of productivity increasing will demand, among other factors, genetically improved animals in both pasture and feedlot systems. Besides, genomic information could be used to predict genomic-enabled breeding values (GEBVs) earlier in animals' life, which would reduce generation intervals and increase productivity gains. Numerous studies have been conducted in order to identify appropriate methodologies to specific breeds and traits, which will result in more accurate GEBVs. The aim of this study was to compare the prediction accuracy of GEBVs and the ability to identify genomic regions and genes related to average weight daily gain in Nelore cattle, by applying different regression models and genotypes densities datasets. Genomic and phenotypic information of 804 steers born in three season, offspring of 34 bulls, were used to predict GEBVs through three models (Bayesian GBLUP, BayesA and BayesCπ), four genotypic densities (Illumina BovineHD BeadChip, TagSNPs, GeneSeek Genomic Profiler High (HDi) and Low (LDi) density indicus) and two adjusted phenotypes. Family structure was accounted by using principal component analysis. Animals were assigned either to training (seasons 1 and 2) or testing (season 3) subsets to perform the cross-validation analysis. Estimates of Pearson correlation, regression coefficients and mean squared errors were used to access accuracy, inflation and bias of the estimated GEBVs, respectively. Genome-wide association study (GWAS) was also performed on above datasets, however, results were compared based on model complexity (pD) and deviance information criterion (DIC). SNP effects greater that upper limit of the 99.9% highest posterior interval were considered significant and genes located 25kb up or downstream were assigned to them. These genes were then, submitted to functional annotation clustering using DAVID database. Overall, accounting for family structure did not affect accuracy, but it seems having an effect on bias and inflation of predictions. Regarding to prediction of GEBVs, BayesCπ

models produced the less biased and inflated results for the considered adjusted phenotypes. Furthermore, results from *Bos taurus indicus*-based chips were as informative as those from Illumina BovineHD, which could be an alternative in implementing genomic selection at lower costs. In relation to GWAS, BayesCπ model fitted better to the data, except for TagSNPs. A total of 11, 32 and 236 genes were found for LDi, HDi and 770k datasets, respectively, and the functional annotation assigned 3 genes to 1 cluster (LDi), 4 genes to 1 cluster and 35 genes to 3 (770k) enriched clusters. Five genes in common were found through all datasets: SYT1 (BTA 5), EYA2 (BTA 13), ZFHX4, MATN2 and CPQ (BTA 14). In general, the phenotype is a complex trait controlled by many genes and the model BayesCπ, which accounts for this performs better. Further investigations should account for the influence of the relationship between training and testing dataset.

Keywords: Bayesian inference, *Bos taurus indicus*, functional annotation, growth rate

PREDIÇÃO DE VALORES GENÉTICOS GENÔMICOS E ESTUDO DE ASSOCIAÇÃO AMPLA DO GENOMA PARA GANHO DE PESO MÉDIO DIÁRIO DE BOVINOS NELORE EM CONFINAMENTO

RESUMO: A seleção para taxa de crescimento utilizando o número de dias para atingir determinado peso ou ganho de peso médio resultaria em menores ciclos de produção. Manter o aumento da produtividade exige, entre outros fatores, a utilização de animais melhorados, tanto nos sistemas de pastagem quanto de confinamento. Além disso, as informações genômicas podem ser usadas para predizer os valores genéticos genômicos (GEBVs) mais cedo na vida dos animais, o que reduziria os intervalos de geração e aumentaria os ganhos de produtividade. Inúmeros trabalhos tem sido conduzidos para identificar metodologias apropriadas à determinadas raças e características, o que irá resultar em GEBVs mais acurados. Os objetivos deste estudo foram comparar a acurácia de predição dos GEBVs e a habilidade de identificar regiões genômicas e genes relacionados ao ganho de peso médio diário em bovinos da raça Nelore, pela aplicação de diferentes modelos de regressão e densidades genotípicas. Informações genômica e fenotípica de 804 novilhos nascidos em três safras, filhos de 34 touros, foram utilizadas para predizer GEBVs por meio de três modelos (Bayesian GBLUP, BayesA e BayesCπ), quatro densidades genotípicas (Illumina BovineHD BeadChip, TagSNPs, GeneSeek indicus de alta (HDi) e baixa (LDi) densidades) e dois fenótipos ajustados. A estrutura de família foi considerada por meio da análise de componentes principais. Os animais foram distribuídos em subconjunto de treinamento (safras 1 e 2) ou validação (safra 3) para realização da análise de validação cruzada. Estimativas de correlação de Pearson, coeficientes de regressão e erro quadrado médio foram usados para avaliar acurácia, inflação e viés dos GEBVs estimados, respectivamente. O estudo de associação ampla do genoma (GWAS) também foi realizado nos mesmos conjuntos de dados, entretanto, os resultados foram comparados com base na complexidade do modelo (pD) e no critério de informação do desvio (DIC). Os SNPs com efeitos maiores que o limite superior do intervalo de maior densidade a posteriori, com 99,9% de probabilidade, foram considerados significativos e, genes

distantes 25kb de ambos os lados do SNP, foram relacionados. Esses genes foram submetidos ao agrupamento de anotação funcional no banco de dados DAVID. Em geral, a inclusão de informação de estrutura de família no modelo não afetou a acurácia, mas parece ter tido efeitos no viés e inflação das predições. Em relação a predição dos GEBVs, o modelo BayesCπ produziu os resultados menos viesados e inflacionados para os fenótipos ajustados. Além disso, os resultados obtidos com painéis baseados em *Bos taurus indicus* foram tão informativos quanto aqueles provenientes do painel Illumina BovineHD, o que pode ser uma alternativa para implementar a seleção genômica com menores custos de genotipagem. Em relação a GWAS, o modelo BayesCπ foi o que se ajustou melhor aos dados, com exceção do conjunto de TagSNPs. O total de 11, 32 e 236 genes foram identificados usando os conjuntos de dados LDi, HDi e 770k, respectivamente, e a anotação funcional resultou em 1 grupo com 3 genes (LDi), 1 grupo com 4 genes (HDi) e 35 genes distribuídos em 3 grupos (770k). Foram observados cinco genes em comum entre os três conjuntos de dados: SYT1 (BTA 5), EYA2 (BTA 13), ZFHX4, MATN2 e CPQ (BTA 14). De maneira geral, o fenótipo é uma característica complexa controlada por muitos genes e o modelo BayesCπ, que leva isso em consideração, teve melhores resultados. Estudos futuros poderiam considerar a influência do grau de relacionamento entre os animais das populações de treinamento e teste.

Palavras-chave: Anotação funcional, *Bos taurus indicus*, inferência Bayesiana, taxa de crescimento

1. INTRODUCTION

Agriculture industry represents about 23% of the gross national product (GNP) in Brazil (CEPEA, 2014a), wherein 31% of its revenue was derived from livestock production in 2014 (CEPEA, 2014b) and 21.5% of total beef production was exported in 2013 (ANUALPEC, 2014). Brazil has the world's second largest cattle herd with over 200 million heads (IBGE, 2013) and it is predominantly composed of *Bos taurus indicus* breeds which are adapted to tropical climates mainly due to their heat and parasites resistance (BRADLEY; CUNNINGHAM, 1999). Among indicine breeds, Nelore is the most widespread and economically important in Brazil.

Since 1985, total pasture area has decreased and productivity gains became an important factor of beef production growth in Brazil, wherein 65% was due to improving animal performance (MARTHA JR; ALVES; CONTINI, 2012). Besides pasture and nutritional improvements, Nelore breed has been largely selected for growth rate traits, instead only for weaning and yearling weights for example, in order to avoid increasing mature body size. The concept of higher growth rate allowing shorter production periods (PANETO et al., 2002) is well known in beef cattle production, where animals with higher growth rates, which spend less time in the pasture or feedlots, are desired (BOLIGON et al., 2009). The number of days to achieve mature weight (MACHADO; AQUINO; GONÇALVES, 1999) or average weight gain (SARMENTO et al., 2003) should be used as selection criteria.

Studies about genetic trends have identified positive results over the years, such as genetic improvements on days to gain 160 kg and 240 kg (MALHADO et al., 2005), average daily gain and female reproductive traits (LAUREANO et al., 2011). Although these factors have contributed to productivity gains, there is still much to explore. For example, the Brazilian beef feedlot industry has grown about 50% in the last decade, however the system is mainly utilized to finish cattle for external markets (MILLEN et al., 2011). In 2013, about 9.3% of slaughtered animals were feedlot-finished (BRAZILIAN BEEF EXPORTERS ASSOCIATION, 2013), and it is known that the majority of remaining cattle has been raised on extensive grazing system (FERRAZ; FELICIO, 2010).

The current challenge is to increase beef productivity and this will demand pasture maintenance and investments in technologies to improve animal performance (MARTHA JR; ALVES; CONTINI, 2012). In order to combine these remarkable factors, genetically improved animals must be used in both pasture and feedlot systems. Nowadays, available genotyping technologies are being used to predict genomic-enabled breeding values (GEBVs) earlier in animals' life, as proposed by Meuwissen, Hayes and Goddard (2001) and known as genomic selection, and to identify single nucleotide polymorphisms (SNPs) and genomic regions associated to economically important traits, called quantitative trait loci (QTL).

Genomic selection has several advantages when compared to the traditional BLUP, highlighting the decrease in inbreeding rates, since GEBV becomes less correlated among related individuals and, over generations, the breeding value of a specific individual would have less influence in selection of its offspring, thus the genetic variability would be maintained in the long term (DAETWYLER et al., 2007). Furthermore, it is possible to reduce generation intervals, as a consequence of more accurate predictions of GEBVs from young, even newborn, animals.

Genomic selection is already a reality in dairy cattle production in many countries, such as the USA (VANRADEN et al., 2009) and New Zealand (WINKELMAN; JOHNSON; HARRIS, 2015). There are also collaborative projects among the USA, Canada and European countries (BERRY et al., 2014). Recently, numerous studies have been conducted in order to identify appropriate methodologies to specific breeds and traits, which will result in more accurate GEBVs. For beef cattle industry initiatives are recent, mainly in the USA (BERRY et al., 2014; SAATCHI et al., 2014) and there is still many studies being conducted without national genotypes databases being arranged (GARRICK, 2011).

Based on the importance of the Nelore (*Bos taurus indicus*) cattle in Brazil, it is necessary to identify appropriate methodologies that allow genomic selection of animals with higher growth rates. Besides that, the improvement of hard-to-measure and/or sex-limited traits could be favored by genomic selection. However, the implementation of genomic selection must be planned according to the production system and breeding objectives, since these information will determine the ideal

number of genotyped and phenotyped animals, and how the population should be splitted for cross-validation.

2. LITERATURE REVIEW

2.1. Achievements in Nelore breeding programs

In classical genetic improvement programs, observable phenotypes of the candidates for selection and their relatives are combined in order to estimate genetic parameters such as variances and heritability (WELLER, 1994). It is assumed that the number of loci controlling a quantitative or complex trait is infinitely large, thus the genetic variance of individual loci are so small that they should be investigated by phenotypic resemblance between relatives and environmental influence (LYNCH; WALSH, 1998).

Consequently, traditional breeding programs are based on estimated breeding values (EBVs), which can be defined as the sum of additive effects of the genes an individual carries (GODDARD, 2009), and the most common applied methodology throughout the world is the best linear unbiased prediction (BLUP) (HENDERSON, 1975).

Brazilian Nelore has been selected over the last fifty years and, according to the main breeding programs (e.g. Programa de Seleção da Estação Experimental de Zootecnia de Sertãozinho, Programa Nelore Brasil, Nelore Qualitas, Aliança Nelore, Conexão Delta G, PAINT) it has been used selection criteria such as growth, reproductive and sexual precocity, carcass and morphological traits, which can also be combined into selection indexes. Besides, many genetic trends studies has been conducted in order to quantify the effects of selection.

At Institute of Animal Science (IZ), a researcher center in Sao Paulo state in Brazil, it was observed that, after 16 years (from 1981 to 1996) of selection for growth traits, EBVs of yearling weight had increased without compromising days to calving and calving success (MERCADANTE et al., 2003), female reproductive traits related to economic efficiency of beef production.

Based on the database of Nelore Brasil coordinated by the National Association of Breeders and Researchers (ANCP), Garnero et al. (2006) observed that EBVs of asymptotic weight (predicted weight when time tends to infinite) of females born from 1985 to 1995, in seven Brazilian states, had dropped an average of -0.12 kg/year, which means cows' adult weight were not increasing. Grossi et al. (2008) also analyzed animals from the same database, but those born from 1976 to 2002, and found that EBVs for accumulated productivity, an index that accounts for frequency of calving and offspring's weaning weight, had a significant increase of 0.166 kg/year.

Animals born from 1975 to 2001 in Midwestern and Southeastern Brazil were also evaluated and an increment of 0.341, 0.311 and 0.267% per year in weaning, year and yearling weights were estimated (SOUZA et al., 2011). Similar results were found by Boligon et al. (2013) while analyzing data from more than 600,000 animals born in Brazil and Paraguay between 1984 and 2010 from Aliança Nelore. The authors found positive genetic trends for weaning and yearling weights without increase of individuals' mature weight.

The majority of recent studies have emphasized the existence of genetic variance for economically important traits. As animal breeders are constantly looking for novel ways to explore these variability, molecular genetics knowledge has been extensively studied as new information for genetic evaluation.

2.2. Overview of animal breeding in postgenomic era

The availability of dense panels, containing hundreds of thousands of single nucleotide polymorphisms (SNPs) distributed across the genome, has allowed researchers to explore old concepts in quantitative genetics, such as linkage disequilibrium (LD). The two most common measures of LD, D' and r^2, were proposed by Lewontin (1964) and Hill and Robertson (1968), respectively, and they can be applied in order to quantify statistical association between pair of loci, based on their allelic frequencies.

In the 1980-90 decade, Fernando and Grossman (1989) proposed a marker-assisted selection (MAS) using BLUP in which marker information would be used to

construct a second relationship matrix, in addition to the pedigree matrix, A. Later, Lande and Thompson (1990) stated that MAS would perform better for traits with low heritability and if a moderate to large proportion of the additive variance was associated with the marker loci. In order to increase the fraction of variance explained, Goddard (1991) and Meuwissen and Goddard (1996) suggested the use of haplotypes based on several closely linked markers. Even those were pioneering studies, they were not genome-wide approaches since previous QTL identification studies were necessary and molecular genetics technology was a limitation.

The sequence of the human genome (INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001; VENTER et al., 2001) and the advent of genotyping through DNA chips became basis for future studies direction even in animal breeding and, Meuwissen, Hayes and Goddard (2001) proposed a novel approach, known as genomic selection (GS), in order to efficiently use all available marker information. Briefly, they compared least squares versus BLUP and Bayesian estimation of genome-assisted breeding values (GEBVs), considering SNPs as random effects for the last two. Although using simulations, the authors could conclude that by using SNPs panels covering the entire genome, it was possible to accurately estimate the GEBV of individuals without their own phenotypic records or progeny.

Thereby studies have been encouraged by the attention on marker-assisted prediction of breeding values. Many researchers have worked with simulated data to develop new statistical methodologies, such as Gianola, Perez-Enciso and Toro (2003) and Xu (2003) which included SNPs genotype information on Bayesian framework, using Markov Chain Monte Carlo (MCMC) for drawing samples of the unobservables (e.g., SNPs effects, genetic and errors variances) from joint posterior distribution.

Since the publication of the *Bos taurus taurus* cattle genome (THE BOVINE GENOME SEQUENCING ANALYSIS CONSORTIUM et al., 2009), investigations using low and high density SNP panels have been conducted in real data, not only to estimate GEBVs, but also in genome-wide association studies (GWAS). The Illumina Bovine 50k and Illumina BovineHD BeadChip (Illumina Inc., San Diego, CA) have been the most used commercial SNP chips for large scale analysis in cattle. Both of

them were built with SNP markers evenly distributed across the genome, based on various beef and dairy breeds information, mainly from *Bos taurus taurus* (MATUKUMALLI et al., 2009). The Illumina BovineHD BeadChip has become one of the most important genomic tool because it has almost 800k. However, because of the small populations available for genomic evaluation, many studies have been conducted in order to select more informative SNPs and, consequently, to increase the statistical power of the prediction and association analysis. Among these efforts, the Neogen Corporation have developed two new custom SNP chips (GGP indicus 20k and 90k), based on information of commercial available panels and three Bos taurus indicus breeds information (Nelore, Brahman and Gir).

Thus, current and future whole-genome approaches are possible due to the development of SNPs arrays and large scale analysis (ANDERSSON, 2009), once plenty of data have been generated by genetic companies and academic studies.

2.3. Genomic Selection

A common procedure to estimate additive genetic effects of QTLs using a limited number of genetic markers available was multiple regression of observed phenotypes, y, on the number of copies of a particular allele (e.g., 0, 1, or 2) at the marker loci (LANDE; THOMPSON, 1990). However, as the constant improvement in molecular genetics technologies would eventually results in a greater number of available markers, Meuwissen, Hayes and Goddard (2001) proposed to combine close markers into haplotypes and to estimate their effects on quantitative traits. The authors came up with application of Bayesian framework as an alternative to least squares model (LS) in order to estimate many chromosome segment effects at the same time, including them as random effects into a mixed model. Using a similar notation, the linear model suggested by the authors can be represented as:

$$y = X\beta + e$$
$$e \sim N(0, I\sigma_e^2)$$

where y is the vector of phenotypes, X is a matrix of genotypes (e.g., 0, 1, 2 for aa, Aa and AA, respectively), β is the marker or haplotype genetic effect, e is the vector of random errors, and σ_e^2 is the residual variance. The vector y is sometimes

replaced by pseudo-phenotypes, such as EBVs, deregressed EBVs (dEBVs) (GARRICK; TAYLOR; FERNANDO, 2009) or observed values corrected for fixed effects (e.g., year, herd, season). However, preadjusting phenotypes could have consequences, such as bias and inconsistency of estimates of SNP effects (DE LOS CAMPOS et al., 2013).

Bayesian approach in animal breeding was proposed by Gianola and Fernando (1986) and the basic idea can be summarized as a function of the data (y) and the unobservables (θ) as follows:

$$f(\theta|y) \propto f(y|\theta) \times f(\theta)$$

where $f(\theta|y)$ is the joint posterior density of θ and y, $f(y|\theta)$ is the likelihood function, which actually is a function of the observed data under various possible values of θ (CASELLA; BERGER, 2002), and $f(\theta)$ is the prior density of θ, where all inferences about the unknowns are made from. Monte Carlo sampling methods have been used to simulate a Markov chain in order to obtain a sequence of samples from the posterior distribution of unknowns (HASTINGS, 1970). For many years, the lack of computational power was a major limitation towards more widespread implementation of these techniques.

One of the most common MCMC methods for generating random variables from a marginal distribution without having to calculate the density is the Gibbs sampler. Further explanation can be found, for example, in Casella and Berger (2002), but briefly, let X, Y, Z be random variables and the marginal distribution $f_x(x)$:

$$f_x(x) = \int \left[\int \int f_{X|YZ}(x|y,z) f_{YZ|X}(y,z|t) dy dz \right] f_x(t) dt$$

The process would sample iteratively from $f_{X|YZ}$, $f_{Y|XZ}$ and $f_{Z|WY}$ as follows:

$$X'_j \sim f\left(x \middle| Y'_j = y'_j, Z'_j = z'_j\right)$$

$$Y'_{j+1} \sim f\left(y \middle| X'_j = x'_j, Z'_j = z'_j\right)$$

$$Z'_{j+1} \sim f\left(z \middle| X'_j = x'_j, Y'_{j+1} = y'_{j+1}\right)$$

and this sampling produces the sequence: $X'_0, Y'_0, Z'_0, X'_1, Y'_1, Z'_1$ and so on. After the chain converges, one can extract information such as mean and variance from the posterior distribution.

The prior marginal density of SNP markers effects characterizes the various Bayesian regression models in a genomic scenario. On the BayesA model

(MEUWISSEN; HAYES; GODDARD, 2001) for example, the prior marginal distribution of marker effects is a scaled-t density and it is implemented on a hierarchical way such that in each iteration samples of SNP effects ($\beta_j|\sigma^2_{\beta j} \sim N(0, \sigma^2_{\beta j})$) and SNP effect variances ($\sigma^2_{\beta j} \sim \chi^{-2}(S_\beta, df_\beta)$) are obtained. These authors also presented the BayesB method, which has the same implementation as BayesA, except for the inclusion of a density peak at $\sigma^2_{\beta j} = 0$. The most appropriated implementation would be including the mixture with a density equals to zero at the level of effects instead of at the level of variances (GIANOLA et al., 2009).

Due to the seemingly endless options of priors, there is today a so called Bayesian alphabet (HABIER et al., 2011; GIANOLA, 2013). For example, in BayesC method (VERBYLA et al., 2009) SNP effects are sampled from a mixture of two-student t distributions and the introduction of a latent variable, γ, allows the SNP being included as a significant effect in the model or not. In BayesCπ (HABIER et al., 2011) the prior for marker effects is mixture with a point of mass at zero and a slab Gaussian distribution. The additional parameter π represents the prior proportion of non-zero effects and is treated as unknown, and effects of SNPs not fitted into the model are shrunk to zero.

Gianola et al. (2009) and Gianola (2013) have criticized the grow of Bayesian alphabet under an incorrect perception, because by applying these regression models in "small n, large p" situation results are being driven by prior densities and not by the data, which means there is no Bayesian learning. These concerns should be taken into account when concluding about genetic architecture of complex traits.

VanRaden (2008) suggested an approach where it implements BLUP in a genomic scenario, which requires the use of a genomic-build relationship matrix, G, and the model become known as GBLUP. The authors also presented three different ways to obtain the G matrix. Let M_{nxm} be a genotype matrix with n (number of samples) rows and m (number of SNPs) columns, and Z_{nxm} be the centered M matrix, which means all columns having mean equal to zero.

1. $G = ZZ'/2\sum p_i(1-p_i)$, where the denominator is the total variance across loci and the ratio scales G to be analogous to the numerator matrix, A.

2. $G = ZDZ'$, where D is a diagonal matrix of markers' reciprocal expected values used as weighing factors.

3. $G = MM' - g_o 11'/g_1$, where g_o and g_1 are the intercept and slope of the regression of MM' on A, respectively. This procedure adjusts G for mean homozygosity.

Other than in 3, results from the first to ways are positive semidefinite matrices, unless the number of markers are much smaller than the number of animals or there are individuals with identical genotypes, such as twins or clones.

GBLUP can also be implemented as a particular case of the semiparametric model called Reproducing Kernel Hilbert Spaces (RKHS) (WAHBA, 1990). Its Bayesian implementation using whole-genome SNP markers information was suggested (GIANOLA; FERNANDO; STELLA, 2006; GIANOLA; VAN KAAM, 2008) and the motivation was to avoid multiple testing, strong dependence of priors, and ambiguous interpretation of effects due to collinearity (GIANOLA; VAN KAAM, 2008). According to de los Campos, Gianola and Rosa, (2009), SNP marker information is used to build the kernel, which is a prior (co)variance matrix between individuals. The authors highlighted that choosing a specific kernel could be the main problem of its implementation in either a non or semiparametric context. However, A or G matrices can be used as kernels, such that BLUP and GBLUP are considered special cases of RKHS.

Bayesian regression models have been known as multiple-step methods, and their main advantage is the possibility of incorporating prior information into the model. Disadvantages of multiple-steps methods, such as the requirement of prior variances and hyperparameters into estimation of SNPs effects and GEBVs (MISZTAL; LEGARRA; AGUILAR, 2009) should be considered.

In order to simplify this procedure, a single-step method, which performs a joint analysis including pedigree, phenotype and genomic information has been proposed (LEGARRA; AGUILAR; MISZTAL, 2009; MISZTAL; LEGARRA; AGUILAR, 2009). Legarra, Aguilar and Misztal (2009) suggested the construction of a new matrix, H, which would incorporate information from A and G matrices, and the basic idea was to replace the co-variances between genotyped individuals by their realized relationship calculated in G. Recently, Fernando, Dekkers and Garrick (2014)

presented the single-step Bayesian regression model, which combines the advantages of both previous class of models.

Independently of which Bayesian regression one have chosen, animals' GEBVs will be a function of SNP markers' genotypes and effects, as proposed by Meuwissen, Hayes and Goddard, (2001). Let m be the number of SNPs, g_{ij} the genotype of animal i at locus j and $\widehat{\beta}_j$ the estimated additive effect of locus j, then:

$$GEBV_i = \sum_{j=1}^{m} g_{ij}\widehat{\beta}_j$$

When performing GBLUP, single-step GBLUP or RKHS using a G matrix as kernel, solution will be the vector of random effects, $\hat{u}$, or the GEBVs.

Hence, many models have been proposed and studies have extensively compared different models using simulated and real data, mainly BLUP versus Bayesian regression, in an attempt to find the most appropriated one. An in-depth review can be found in de los Campos et al. (2013), in which after gathering results from theoretical and empirical studies, the authors observed that prediction accuracies from GS has been greater than pedigree-based BLUP and almost no differences between models for GS have been reported. However, using data from Nelore breed, Neves et al. (2014) reported an advantage of Bayesian regression models over multiple-step GBLUP for gestation length, weight gain from birth to weaning, conformation at weaning and birth weight, meanwhile the opposite was found for weight gain from weaning to yearling and conformation at yearling.

Unfortunately, implementing GS is not a trivial task and there are several critical points to care about. According to Goddard and Hayes (2007), GS should be ideally implemented in three steps through independent populations: model training, validation and selection. However in practice only the first two groups are actually used. For model training, an estimation method is applied to a discovery dataset containing animals with phenotypes and genotypes information in order to obtain prediction equations for important traits. The next step is testing these prediction equations in an independent sample of animals with genotypic and phenotypic information, so that the accuracy of prediction could be accessed. Lastly, tested prediction equations would be applied to another independent group of selection candidates with genotypic information in order to estimate their GEBVs.

One of these critical points is how to choose training and validation population and factors such as cost of genotyping and phenotyping are often considered. Among plenty available approaches, some of them are: progeny tested sires, animals chosen randomly across the whole population and within or across families. (USAI; GODDARD; HAYES, 2009). A widely used approach is setting animals to either training or testing group according to their generation or year of birth (SOLBERG et al., 2009; VANRADEN et al., 2009; HABIER et al., 2011).

Beyond that, simulation studies have shown that less biased predictions could be accessed if randomly chosen animals (JIMÉNEZ-MONTERO; GONZÁLEZ-RECIO; ALENDA, 2012) or those with extreme yield deviation values (BOLIGON et al., 2012) were included in the reference population. However, if the validation dataset is more closely related to the training population than it is to the selection candidates, the prediction accuracy will be overestimated (WRAY et al., 2013).

It has been claimed that genomic selection could lead to a great change on generation interval and rate of genetic gain, which could reduce logistical costs in animal production (SCHAEFFER, 2006). Besides that, one of the most expected changes is the control of inbreeding rates, since capturing the Mendelian sampling information could decrease the number of selected sibs or high related animals (DAETWYLER et al., 2007). Furthermore, increasing accuracy of GEBVs for phenotypes that are hard and/or expensive to measure, such as fertility or feed efficiency traits, would balance breeding objectives (HAYES et al., 2009).

2.4. Genome-wide association study

This genomic approach has become a routine analysis in animal breeding studies, in order to understand the genetic background of complex traits. The basic idea is similar to GS, but the goal is detecting statistical associations between the trait and SNP markers (GODDARD; HAYES, 2009), instead of summing over their effects to predict the breeding value. This is possible because using thousands of SNP markers widely distributed across the genome would have a good chance of including at least one that are in strong LD with a common causal variant (BALDING, 2006). The majority of studies in the literature has been conducted in human

populations, due to its power in discovering and understanding genetic variation in common diseases (NEALE; PURCELL, 2008).

Results from human studies emboldened GWAS in domestic animals, and they were first conducted based on relatively few markers distributed across the genome (ABASHT; LAMONT, 2007; BARENDSE et al., 2007; LIU et al., 2007) and for identification of genes underlying monogenic traits, such as some congenital defects and coat color (GEORGES, 2007). Meanwhile GWAS has also been carried out for complex traits, thus potential genomic regions has been detected (DAETWYLER et al., 2008) and trying to unveil their genetic architecture has motivated scientific investigations (HAYES et al., 2010).

The advent of a high-density SNP panel (MATUKUMALLI et al., 2009) brought a new perspective for GWAS in dairy and beef cattle. One can notice that in the first moment results were being compared to previous QTL reports, as a way to corroborate them. For instance, Mai et al. (2010) reported seven new QTLs for milk production traits in Danish Jersey cattle in a total of 33 identified. Furthermore, some studies has detected significant signals at the same chromosome but in different positions from previous QTL analysis (SAHANA; GULDBRANDTSEN; LUND, 2011), which was expected since QTLs real position are unknown (VIGNAL et al., 2002) and would have a large genetic interval (e.g., 15cM) (GEORGES et al., 1995).

As well as in GS studies, many GWAS were conducted based on mixed models equations including SNP markers as fixed effects or a single simple regression for each SNP. These procedures could lead to false positive associations, thus in order to avoid them in milk production traits, Jiang et al. (2010) and Bolormaa et al. (2010) applied adjustments for multiple testing. A problem remaining is that, apart from reducing spurious genetic associations, introducing the conservative multiple testing adjustment could lead to a great number of false negatives (KIM et al., 2011).

Nevertheless, Bayesian methods proposed by Meuwissen, Hayes and Goddard (2001) were also being applied to fit all SNPs simultaneously, which clearly reduced the false discovery rates (PRYCE et al., 2010) and has become a common practice. In dairy cattle numerous studies have identified SNPs and/or genomic regions related to milk composition (BOUWMAN et al., 2011), body conformation

(COLE et al., 2011), fertility (BERRY et al., 2012; PENAGARICANO; WEIGEL; KHATIB, 2012) or health traits (FINLAY et al., 2012) for example.

Regarding to beef cattle, SNPs significantly associated to growth (SNELLING et al., 2010), meat and carcass (BOLORMAA et al., 2011), fertility (PETERS et al., 2012) and puberty (FORTES et al., 2012b) traits have been found. Also, as many different breeds are used around the world for beef production, some authors have studied them jointly, aiming at increasing the sample sizes and the power of detecting associations (BOLORMAA et al., 2013a). Another concern has been detecting and correcting data for population stratification (LU et al., 2013), which could be an important source of bias leading to false positive associations (GODDARD; HAYES, 2009). Diverse approaches have been proposed in GWAS including constructions of association weight matrix, and gene networks, which has shown how complex the "complex traits" could be (FORTES et al., 2010; FORTES et al., 2012a).

In Brazil, many association studies have been conducted in Nelore cattle genotyped using the Illumina BovineHD BeadChip and have identified QTLs and some genes related to meat quality traits (TIZIOTO et al., 2013), birth weight (UTSUNOMIYA et al., 2013), intramuscular fat and fatty acid composition (CESAR et al., 2014), feed efficiency (OLIVEIRA et al., 2014; SANTANA et al., 2014a), growth traits (SANTANA et al., 2014b), and even for muscle mineral content (TIZIOTO et al., 2015a). Canchim, a composite breed developed in Brazil has also been evaluated for backfat thickness (MOKRY et al., 2013) and growth traits (BUZANSKAS et al., 2014).

One can notice there are plenty of GS and GWAS studies available in the literature, which have applied several regression models in different breeds and phenotypes. There is no consensus of which is the most appropriated model or SNP density to be used. It is up to the researcher or animal breeder to evaluate its options, often a balance between cost and benefit.

In this study, the aims were to investigate whether different regression models and density of SNP chips could affect the prediction accuracies of genomic estimated breeding values and genome-wide association for average daily gain in

feedlot-finished Nelore steers. Besides that, we identified potential genomic regions and genes related to feedlot average daily weight gain in Nelore cattle.

3. MATERIAL AND METHODS

3.1. Samples

During the mating seasons of 2007 through 2009, 804 steers, offsprings of 34 Nelore bulls from 17 lineages, were generated through fixed-time artificial insemination in five farms, where they were raised to around 21 months of age, and then moved to either Embrapa Southeast Livestock (São Carlos, SP) or Embrapa Beef Cattle (Campo Grande, MS) for three seasons of feedlot experiment periods (2009/2010/2011). Animals were fed corn or sorghum silage with at least 35% dry matter, and concentrate containing corn, soybean meal, soybean hull, cotton seed, limestone, mineral mixture, urea and monensin (Rumensin®), accounting around 13% crude protein and 71% total digestible nutrients. The diet was provided twice a day in which the feed offered (total mixture composed by concentrate + silage, in a proportion of 40:60) was adjusted daily to ensure *ad libitum* consumption. The animals were weighed every 14 days without fasting, for an average period of 91 days.

3.2. Phenotype and genotype datasets

Initial dataset consisted in 7,236 weighting records from the 804 steers, but only those from the 15^{th} up to 77^{th} days in feedlot were considered, disregarding the first weight and also because after this period more than 30% of the animals had been slaughtered. A linear regression analysis of live weight over time was performed in the remained 3,523 records from 803 steers. The slope was used as the average daily gain during the feedlot period (ADG) and the purpose of it was to consider only the linear weight gain and to avoid comparison among different feedlot period lengths.

Three bulls with small progeny, as well as animals with ADG ≤ 0.15 kg or ADG ≥ 2.50 kg were removed from phenotype and genotype datasets. Steers were assigned to 39 contemporary groups (CG) containing from 5 to 42 animals, which combined information of season (3), farm of feedlot (2) and slaughter group (32). Thereafter, phenotype and genotype datasets were merged to ensure they had exactly the same individuals, so 718 steers remained into the dataset. Table 1 shows the summary of age at feedlot entry, starting weight, ADG and days in feedlot on remaining animals.

Table 1. Summary of age and weight at feedlot entry, feedlot average daily gain (ADG) and days in feedlot for the remaining Nelore cattle

	Age (days)	Weight (kg)	ADG (kg)	Days in feedlot
Min	542	226	0.193	48
Mean (±sd)	649 (±45)	361 (±51)	1.235 (±0.407)	92 (±20)
Max	745	510	2.457	119

sd: standard deviation

The Illumina BovineHD BeadChip (Illumina, San Diego, CA) was used for genotyping 780 steers and 34 bulls. The initial dataset contained 742,906 single nucleotide polymorphisms (SNP). Unplaced SNPs, mitochondrial and sex-linked SNPs were first discarded, as well as duplicated markers (e.g.: two different names and positions for the same SNP). We also filtered SNPs based on two other chips: GeneSeek® Genomic Profiler (GGP) HDi 80K and GGP LDi 20K (Gene Seek Inc., Lincoln, NE), which were built specifically for *Bos taurus indicus* breeds. Originally, the GGP HDi 80k/LDi 20k contained 74,085/19,721 markers, but only 69,942/18,464 were available in the main dataset. The following described procedures were applied to all datasets.

Paternity correction and quality control were performed to avoid bias results (LAURIE et al., 2010). There were some concerns about SNPs with significant Hardy-Weinberg Equilibrium (HWE) deviation and how this could reflect an error during genotyping procedure (ZIEGLER, 2009). To deal with this question we checked plots of HWE deviation versus percentage of heterozygous (%hetero)

and 17 SNPs with %hetero ≥ 0.8 were also excluded. Quality control (QC) was performed using R "snpstats" package (CLAYTON, 2012) and criteria to keep markers were: SNP call rate > 98% and minor allele frequency (MAF) > 1%. The MAF filter excluded 20.00, 1.90 and 7.28% of the total SNPs from 770k, HDi and LDi, respectively.

After quality control, Beagle v.3.3.2 (BROWNING; BROWNING, 2009) software was used for phase inference and imputation of missing genotypes for each SNP panel, without using pedigree information. Finally, Tagger tool (BAKKER et al., 2005), implemented in Haploview v4.1 (BARRETT et al., 2005) and based on correlation coefficients between markers (r^2), was used to select TagSNPs, a minimal set of SNPs that were correlated at an $r^2 \geq 0.3$ with at least one another marker on the same chromosome. This threshold was chosen because it is the overall average r^2 at the distance of 10kb to 25kb in a previous analysis using the same animals. The final number of SNPs per dataset and their distribution across chromosomes are shown in Table 1A. The average spacing between markers is presented in Table 2A.

3.3. Principal components analysis (PCA)

In order to explore the impact of family structure, PCA analysis was performed using all datasets. Table 3A shows the size of half-sib families and the summary description of ADG within each one. The main goal of this analysis is to set each individual's coordinates along axes of variation (PATTERSON; PRICE; REICH, 2006), which are the eigenvectors of a covariance matrix between individuals (PRICE et al., 2006), and it might have some genealogical interpretation (MCVEAN, 2009). The procedure was conducted as follows. Let C be a rectangular matrix with rows indexed by n individuals and columns indexed by p SNP marker genotypes $c_{ij} \in \{0,1,2\}$. The data was centered to guarantee that the mean of each column was equal to zero:

$$\mu_j = \frac{\sum_{i=1}^{n} c_{ij}}{n} \quad ; \quad c_{ij} = c_{ij} - \mu_j.$$

Then each row was normalized dividing the new entries by $\sqrt{p_j(1-p_j)}$, which is related to the rate of frequency change of a SNP due to genetic drift (PATTERSON; PRICE; REICH, 2006), resulting a new matrix $M_{(nxp)}$ in which columns presented mean zero and the same variance. Using M we computed a realized covariance matrix $G_{(nxn)}$ of individuals, wherein using this procedure, A matrix was the expected value of G $(E(G) = A)$. PCA was then performed on this G matrix using some functions (eigen, princomp and SVD) in R (R CORE TEAM, 2013). Eigenvectors corresponding to the top eigenvalues could be related to nonrandom population structure (PATTERSON; PRICE; REICH, 2006). The R package nFactors (RAICHE, 2010) was used to investigate the number of component to retain in order to explain the greater proportion of variance among individuals. However, it is known that the projection space could be distort if sample sizes between populations (or groups) are much different (MCVEAN, 2009). As PCA apply a strong correction to SNPs with large differences in allele frequencies (PRICE et al., 2010), results were used to verify whether the amount of variance among individuals could be explained by the largest principal components and how they were clustered.

3.4. Fixed effects modeling and adjusted phenotypes

Systematic effects analysis for ADG was conducted through linear models using packages "car" (FOX; WEISBERG, 2011), "MASS" (VENABLES; RIPLEY, 2002) and "fmsb" (NAKAZAWA, 2013) using R software (R CORE TEAM, 2013), and it was used to adjust the phenotype. The model can be represented as:

$$y = W\alpha + \varepsilon$$

where y = vector of observations (average daily gain in feedlot period); W = incidence matrix for fixed effects (CG and age at feedlot entry (AF)); α= vector of fixed effects; ε = random error associated with each record. We also added the first two eigenvectors as covariates to verify whether or not there was some influence on predictions.

A residual analysis was also performed to guarantee the assumption of normality for the residuals. The Shapiro-Wilk test of normality (SHAPIRO; WILK, 1965) was performed and |studentized residuals|>3.5 were excluded, so 718 animals remained in the dataset. The adjusted phenotype is represented as $\hat{y} = \bar{y} + (y - W\alpha)$, which is the mean of observed phenotypes added to the residual of the model. There were five adjusted phenotypes: without principal components and considering the two largest principal components as covariates from 770k, TagSNPs, HDi and LDi dataets.

3.5. Models for genomic-enabled prediction

The BGLR package (DE LOS CAMPOS; RODRIGUEZ, 2014) was used to fit the Bayesian models. The basic model can be represented in matrix notation as $y = \eta + \varepsilon$, where y is the adjusted phenotype $(y_i; i = 1, \ldots, n)$, η is the linear predictor and ε is the independent random residual, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. The linear predictor is the conditional expectation function:

$$\eta = 1\mu + \sum_j^J X_j \beta_j \text{ , for BayesA and BayesC}\pi$$

$$\eta = 1\mu + \sum_l^L u_l \text{ ,for Bayesian GBLUP}$$

where μ is an intercept, X_j are design matrices for predictors, β_j are the regression coefficients associated with the columns of X_j, and u_l are vectors of random effects (e.g. breeding values). After collecting the unknowns (θ), the following likelihood (conditional distribution of the data) is a multivariate normal density: $p(y|\theta) = \prod_{i=1}^n N(y_i|\, \mu + \sum_j^J x_{ij}\beta_j, \sigma_\varepsilon^2)$ (BayesA and BayesC π) and $p(y|\theta) = \prod_{i=1}^n N(y_i|\, \mu + \sum_{l=1}^L u_{li}, \sigma_\varepsilon^2)$ (Bayesian GBLUP). The prior distribution of θ is assumed as $p(\theta) = p(\mu)p(\sigma_\epsilon^2)\prod_{j=1}^J p(\beta_j)$, (BayesA and BayesC π) and $p(\theta) = p(\mu)p(\sigma_\epsilon^2)\prod_{l=1}^L p(u_l)$ (Bayesian GBLUP). It is assigned a flat (non-informative) prior to intercept and a scaled-inverse chi-square density prior to variance parameters. Thus $p(\sigma_\cdot^2) = \chi^{-2}(\sigma_\cdot^2|S_\cdot, df_\cdot)$, where degree of freedom $(df_\cdot)$ and scale parameter $(S_\cdot)$ must be > 0.

Prior of the regression coefficients (β_j) of marker effects need to be informative, so the choice of the prior would determine the extent of shrinkage of their estimates and hence the Bayesian method. The vectors of random effects u_l

are assigned multivariate-normal priors $\prod_{l=1}^{L} N(0, K_l\sigma_{ul}^2)$, where K_l is the (co)variance matrix between individuals and σ_{ul}^2 is the variance parameter.

3.5.1. BayesA (MEUWISSEN; HAYES; GODDARD, 2001)

In a first level of hierarchy a normal distribution is assigned to the marker effects (β_j), with mean zero and marker-specific variance ($\sigma_{\beta j}^2$). In a second level $\sigma_{\beta j}^2$ is assigned independent and identically distributed (iid) Scaled-inverse Chi-square density, with df_β degree of freedom set to 5 (to guarantee a finite prior variance) and scale parameter S_β treated as unknown, following a Gamma distribution with shape (s) and rate (r) parameters. On this model the prior marginal distribution of marker effects is a scaled-t density, with parameters df_β and S_β, but this hierarchic implementation is due to computational convenience:

$$\begin{cases} \beta_j|\sigma_{\beta j}^2 \sim N\left(0, \sigma_{\beta j}^2\right) \\ \sigma_{\beta j}^2 \sim \chi^{-2}\left(S_\beta, df_\beta\right) \\ S_\beta \sim Gamma(s, r) \end{cases}$$

The variance of the additive effects of SNP markers can be represented as the sum of posterior variances along all markers ($\sigma_g^2 = \sum_j^J \sigma_{\beta j}^2$).

3.5.2. BayesCπ (HABIER et al., 2011)

Prior for marker effects are assigned as an iid mixture of point of mass at zero and a slab that is a Gaussian distribution. So BayesCπ is an extend of Bayesian ridge regression, by introducing an additional parameter π, which represents the prior proportion of non-zero effects and is treated as an unknown with a Beta prior distribution, which is parameterized in a way that the $E(\pi) = \pi_0$ and p_0 is the number of priors "successes" plus "failures". The parameters were set as $p_0 = 2$ and $\pi_0 = 0.5$, which gives an uniform prior in the interval $[0,1]$. Thus, differently from BayesA, the effects of SNPs not fitted to the model are shrunk to zero.

$$\begin{cases} \beta_j|\sigma^2_{\beta j} \sim N(0, \sigma^2_{\beta j})\pi \\ \pi \sim Beta(p_0, \pi_0), for\ p_0 > 0\ and\ \pi_0 \in [0,1] \\ \sigma^2_{\beta j} \sim \chi^{-2}(S_\beta, df_\beta) \\ S_\beta \sim Gamma(s, r) \end{cases}$$

On this model, the variance of the additive effects of markers is defined as:

$$\sigma^2_g = \frac{\sigma^2_\beta}{(1-\pi)\sum_j^J 2p_j(1-p_j)}$$

where σ^2_β is the posterior additive variance explained by SNPs and p_j is the allele frequency of SNP j. In BGLR package, $\sum_j^J 2p_j(1-p_j)$ is calculated as the sum of the sample variances of the columns of the genotype matrix.

3.5.3. Bayesian GBLUP (BGBLUP)

The BGLR package implements Bayesian Reproducing Kernel Hilbert Spaces (RKHS) regression (WAHBA, 1990) using a user-defined (co)variance matrix. Despite this method have been used for long time, the suggestion of incorporating information of whole-genome SNP markers is recent (GIANOLA; FERNANDO; STELLA, 2006; GIANOLA; VAN KAAM, 2008).

Briefly, the regression function is a linear combination of the basis function provided by the reproducing kernel (RK), which is a solution to find a nonparametric function that minimizes the sum of squares of a particular model associated with a Hilbert space of real-valued functions and it allows for using information on pedigree and SNP markers to proceed genetic evaluation of quantitative traits (DE LOS CAMPOS; GIANOLA; ROSA, 2009).

$$\begin{cases} u_l|\sigma^2_{ul} \sim N(0, K_l\sigma^2_{ul}) \\ \sigma^2_{ul} \sim \chi^{-2}(S_l, df_l) \end{cases}$$

The (co)variance matrix, K, must be a symmetric positive semi-definite with dimension equal to the number of animals. In order to implement Bayesian GBLUP, K was set as a marker-derived relationship matrix - G (VANRADEN, 2008) calculated as described in section 3.1.5 (Principal components analysis)

The Gibbs sampler parameters and the length of final chain are shown in Table 2. Posterior means of any estimated effects are always drawn from the final chain, which is supposed to be stable.

For the three models, the narrow sense heritability was estimated as the proportion of genetic and phenotypic variances: $h^2 = \frac{\sigma_g^2}{\sigma_g^2+\sigma_e^2}$. The 95% highest posterior density interval (HPD95%) of each analyses were estimated using the R package "boa" (SMITH, 2007).

Table 2. Parameters of Gibbs sampler for each applied method

	Method		
Parameter	BayesA	BayesCπ	BGBLUP
Iterations	400,000	600,000	160,000
Burn-in	150,000	200,000	60,000
Thin	10	20	10
Final chain	25,000	20,000	10,000

3.6. Validation

The dataset was divided into training (animal from seasons 1 and 2) and testing (animals from season 3) subgroups which contained 568 and 150 animals, respectively. For BayesA and BayesCπ, genomic estimated breeding value (GEBV) of each individual was defined as the sum of posterior predicted effects ($\hat{\beta}_{trn}$) over all p SNPs estimated in the training set: $GEBV_{i(tst)} = \sum_{j=1}^{p} g_{ij}\hat{\beta}_{trn}$. For Bayesian GBLUP, phenotypes of testing subgroup was set as missing and samples of u_l were obtained in each iteration from the posterior distribution $[u, \sigma_u^2, \sigma_e^2|\hat{y}]$.

The correlation between GEBV and adjusted phenotype of animals on testing subgroup, $r(GEBV_{i(tst)}, \widehat{y_{i(tst)}})$, was used as an estimation of prediction accuracy and a independent t-test was applied to ensure that $r \neq 0$. The slope of regressing adjusted phenotypes on GEBVs for animals in testing subgroup $\left(b_{\widehat{y_{tst}},GEBV_{tst}}\right)$ was evaluated, which is a measure of bias and can be used to verify whether genomic

predictions are inflated or deflated (VERBYLA et al., 2009; SAATCHI et al., 2011), and it is expected to be close to one and its significance was accessed by a F test. The last comparison criteria was the mean square errors (CASELLA; BERGER, 2002), $MSE = \frac{\sum_1^n (GEBV_i - \hat{y}_i)}{n}$, was also used as a measure of precision and bias of the point estimator.

3.7. Genome-wide association study

Instead of assigning the animals either for training or testing groups, the previously mentioned regression models (Section 3.5) were applied to the whole dataset, wherein the Bayesian GBLUP was no longer being used and it was replaced by the equivalent Bayesian ridge regression model (BRR), also implemented in BGLR package. The hierarchical model is similar to BayesCπ, however there is no parameter π, thus regression coefficients β_j are assigned normal iid Gaussian distributions, mean zero and common variance σ_β^2. In a second level of the hierarchy, σ_β^2 is assigned a scaled-inverse Chi-squared density. By using BRR the effects of all SNPs are equally shrunk.

$$\begin{cases} \beta_j | \sigma_\beta^2 \sim N(0, \sigma_\beta^2) \\ \sigma_\beta^2 \sim \chi^{-2}(S_\beta, df_\beta) \\ S_\beta \sim Gamma(s, r) \end{cases}$$

If the MCMC for all SNPs were available, their 95% highest posterior density (HPD) could be used to accept as significant those which 95% HPD does not contain the zero value, which means their posterior probability is over zero (AMILLS et al., 2005). However, with only posterior mean and standard deviation of SNP effects available, for all twenty-four scenarios (3 regression models, 4 SNP densities and 2 adjusted phenotypes, with or without PCs as covariates), the absolute posterior SNP effects were divided by their posterior standard deviation, and the distribution of these standardized SNP effects were evaluated.

Before looking for significant SNPs, we compared two goodness of fit criteria available in BGLR package and proposed by Spiegelhalter et al. (2002), aside from the residual variance. According to the authors, the effective number of parameters

(pD) is a measure of the complexity of the model which considers posterior likelihoods and level of hierarchy, and negative values could indicate the posterior mean is a poor estimator. Moreover, the deviance information criterion (DIC), penalized by pD, is the expected posterior loss when adopting that model, thus lower values are desired. The most appropriate model was chosen based on these criteria.

As HPD can be applied to non-symmetric marginal distributions and for a given probability it will be the shortest length possible (CHEN; SHAO, 1999), then the 99.9% HPD of the estimated SNP effects were calculated through the R package BOA (SMITH, 2007) and marker effects above the upper limit were considered significant. Then significant SNPs were assigned to genes located within 25kb up or downstream. This distance was chosen to have a moderate LD extension ($r^2 \cong 0.3$), based on previous analysis performed on this population and results found by Espigolan et al. (2013).

Assigned genes were submitted to functional classification using the DAVID ("Database for Annotation, Visualization, and Integrated Discovery") database (DENNIS JR et al., 2003), which is able to cluster genes according to their functional resemblance, which means that genes sharing similar functional annotations would be related (HUANG et al., 2007). The clustering is based on the corrected probability of co-occurrence between the annotated terms of two genes, and ranged from 0 (co-occurrence by chance) to 1 (related co-occurrence). The following parameters were applied: a pair of genes should have, at least, 3 terms in common, the probability of co-occurrence should be higher than 0.90, genes in a cluster must have a minimum of 3 terms in common and, groups sharing more than 50% of the genes were combined together.

According to Hosack et al. (2003), DAVID applies the EASE score for accessing the importance of a cluster of genes, which is a modification of the Fisher's exact test and favors categories having more terms in common and classifies the clusters according to their importance in the set of available genes. In order to access the enrichment score of each cluster, the full list of genes located within 25kb up or downstream of all SNPs in the dataset was used as a background.

4. RESULTS

Regarding to PCA, we focused on two results: proportion of variance explained by eigenvectors associated with largest eigenvalues, and how animals were clustered based on the eigenvectors. The top 10 PCs explained 49.7, 48.6, 50.5 and 50.2% of the variance among individuals on 770k, TagSNPs, HDi and LDi datasets, respectively. Besides that, results from nFactors R package (RAICHE, 2010) suggested that 28 principal components should be retained in order to capture the variance among individuals. This is almost the number of half-sib families (31), which means the PCA is mainly clustering them. These findings corroborate Habier, Fernando and Dekkers (2007), who have demonstrated that irrespective of the amount of LD in a population, markers act as a proxy of pedigree.

According to Daetwyler et al. (2012), larger groups are distinguished by the top PCs, thus to avoid the bias that information from the divergent largest families could bring to the analysis, the eigenvectors associated with the two top eigenvalues were added to the models as covariates (PRICE et al., 2006). Figure 1 shows how PCs 1 and 2 clustered the individuals and it is possible to notice that clusters from 770k and TagSNPs were better defined than those from *Bos taurus indicus*-based datasets, which means information from closer relatives could have greater impact on prediction accuracies (LEGARRA et al., 2008). In addition, the more SNPs available, the more additive-genetic relationships are captured by them and the accuracy of GEBVs is expected to be linked to the level of relationship between training and testing sets (HABIER et al., 2010). It is worth to mention that, in the present study, both training and testing groups, for all datasets, contained individuals from every seeming cluster from Figure 1 (data not shown).

Adjusted phenotypes were the residuals of the regression models based on the 718 remained animals. The mean (SE) of the five adjusted ADG was the same, 1.24 kg (0.01), and based on t-test results (data not shown), for all datasets, only mean from half-family labeled as six (6) in Figure 1 was statistically different from the other families.

Estimates of variance components and heritabilities are shown in Table 3 and were also used to access predictive ability of different models and densities. Additive

variance (σ_g^2), residual variance (σ_e^2) and heritability (h^2) ranged from 0.023 to 0.065, 0.046 to 0.063 and 0.283 to 0.534, respectively.

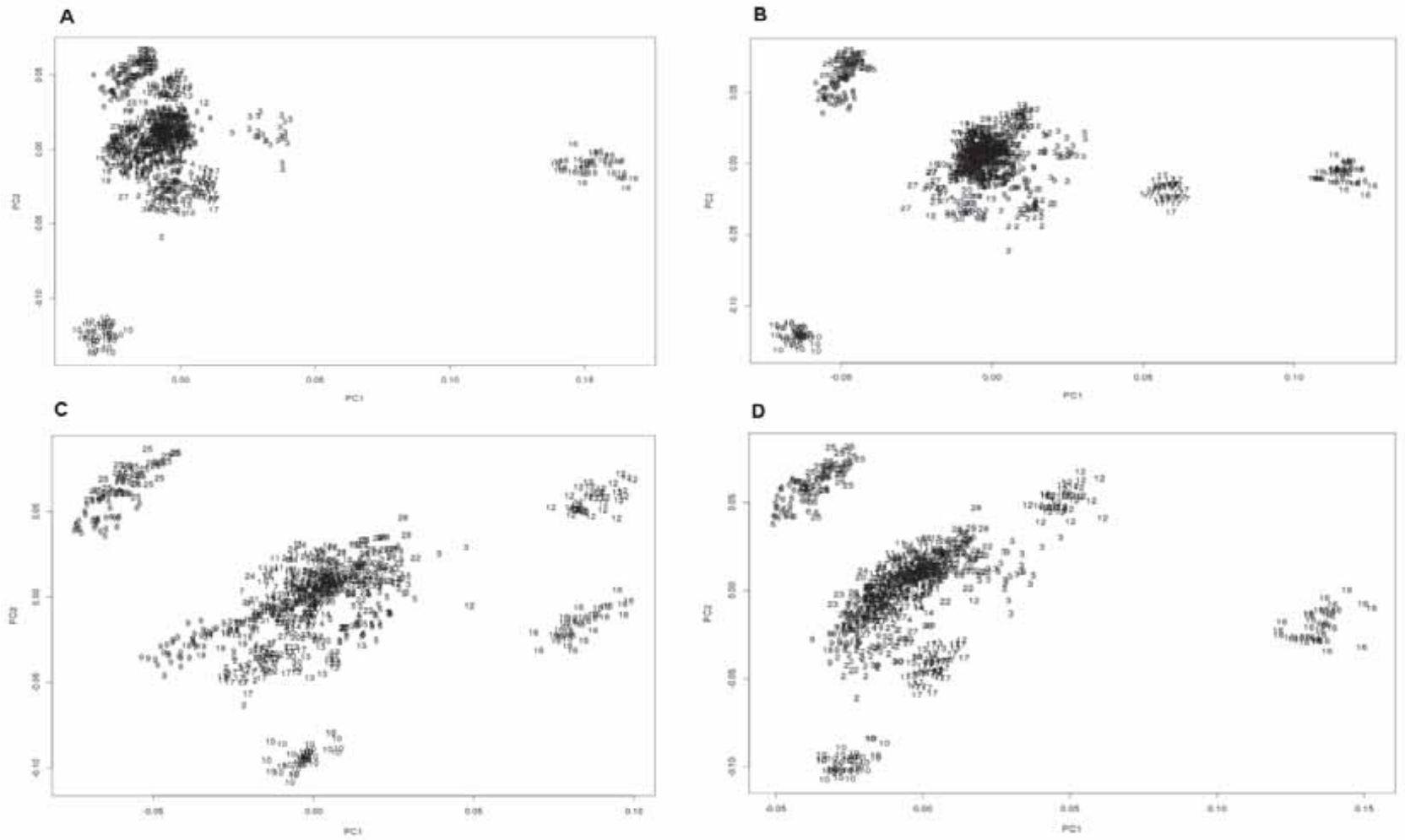

Figure 1. Plot of the first two principal components for 770k (A), TagSNPs (B), HDi (C) and LDi (D) datasets. Individuals are labeled by their half-family number.

Table 3. Estimates of variance components, heritability and proportion of non-zero effects for all models

Density	Parameters	BGBLUP	BGBLUP[1]	BayesA	BayesA[1]	BayesCπ	BayesCπ[1]
770k	σ_e^2	0.052	0.052	0.056	0.063	0.047	0.048
	(HPD95%)	0.042-0.063	0.041-0.063	0.047-0.066	0.054-0.072	0.037-0.058	0.036-0.059
	σ_g^2	0.024	0.024	0.065	0.033	0.033	0.032
	(HPD95%)	0.014-0.035	0.013-0.036	—	—	—	—
	h^2	0.314	0.313	0.534	0.345	0.413	0.395
	(HPD95%)	0.188-0.445	0.183-0.447	—	—	—	—
	π	—	—	—	—	0.983	0.981
	(HPD95%)	—	—	—	—	0.963-0.999	0.960-0.999
TagSNPs	σ_e^2	0.054	0.053	0.062	0.063	0.048	0.048
	(HPD95%)	0.043-0.065	0.043-0.065	0.052-0.074	0.053-0.073	0.036-0.059	0.036-0.059
	σ_g^2	0.025	0.025	0.042	0.038	0.035	0.035
	(HPD95%)	0.014-0.037	0.014-0.037	—	—	—	—
	h^2	0.316	0.317	0.402	0.375	0.424	0.425
	(HPD95%)	0.194-0.461	0.189-0.454	—	—	—	—
	π	—	—	—	—	0.983	0.983
	(HPD95%)	—	—	—	—	0.965-0.999	0.964-0.999

Density	Parameters	BGBLUP	BGBLUP[1]	BayesA	BayesA[1]	BayesCπ	BayesCπ[1]
HDi	σ_e^2	0.052	0.052	0.061	0.058	0.047	0.046
	(HPD95%)	0.041-0.063	0.041-0.063	0.052-0.071	0.045-0.069	0.036-0.058	0.035-0.058
	σ_g^2	0.025	0.023	0.028	0.038	0.034	0.035
	(HPD95%)	0.014-0.036	0.014-0.035	—	—	—	—
	h^2	0.320	0.307	0.313	0.391	0.422	0.431
	(HPD95%)	0.190-0.459	0.180-0.436	—	—	—	—
	π	—	—	—	—	0.978	0.979
	(HPD95%)	—	—	—	—	0.956-0.998	0.956-0.999
LDi	σ_e^2	0.052	0.052	0.063	0.060	0.046	0.047
	(HPD95%)	0.041-0.062	0.041-0.063	0.051-0.073	0.049-0.0712	0.035-0.057	0.036-0.058
	σ_g^2	0.025	0.024	0.025	0.032	0.036	0.034
	(HPD95%)	0.014-0.036	0.014-0.036	—	—	—	—
	h^2	0.319	0.317	0.283	0.349	0.439	0.424
	(HPD95%)	0.192-0.453	0.186-0.445	—	—	—	—
	π	—	—	—	—	0.981	0.980
	(HPD95%)	—	—	—	—	0.961-0.999	0.959-0.998

σ_e^2: residual variance; σ_g^2: genetic variance; h^2: heritability; π: proportion on non-zero effects HPD95%: 95% highest posterior density, [1]:adding the top two PCs as covariates in the model.

4.1. Genomic-enabled prediction

Table 4 shows the Pearson correlation coefficients between adjusted phenotypes and GEBVs which were considered as proxies for accuracies of genomic predictions. In addition, regression coefficients of adjusted phenotypes on GEBVs are presented in Table 5 and Figures 1A-4A, and were used to measure the extent of prediction bias and values greater (lower) than 1 are related to deflated (inflated) GEBVs. The last applied criterion was the MSE (Table 5, Figures 1A-4A) wherein lower values are associated with better overall fit.

Table 4. Estimates of prediction accuracies of genomic estimated breeding values for feedlot average daily gain of animals in testing subgroup

	Dataset			
Model	770k	TagSNPs	HDi	LDi
BGBLUP	0.26**	0.24**	0.25**	0.26**
BGBLUP[1]	0.26**	0.24**	0.18*	0.25**
BayesA	0.26**	0.25**	0.26**	0.27**
BayesA[1]	0.26**	0.25**	0.18*	0.25**
BayesCπ	0.26**	0.25**	0.25**	0.26**
BayesCπ[1]	0.26**	0.24**	0.18*	0.24**

[1]adding the two largest PCs as covariates, **P<0.01, *P<0.05 (independent test-t).

Table 5. Regression coefficients (b) of GEBVs on adjusted phenotype and mean squared errors (MSE) of predictions for animals in testing subgroup

	Dataset							
	770k		TagSNPs		HDi		LDi	
Model	b	MSE	b	MSE	b	MSE	b	MSE
BGBLUP	1.149**	1.58	0.465^{ns}	1.59	**1.099****	1.58	**1.114****	1.59
BGBLUP[1]	1.152**	1.58	0.472 ^{ns}	1.59	**0.896***	1.59	**1.094****	1.59
BayesA	1.289**	1.09	0.687 ^{ns}	1.24	1.684**	1.32	1.999***	1.37
BayesA[1]	1.828**	1.33	0.688 ^{ns}	1.31	1.187*	1.38	1.649**	1.29
BayesCπ	**0.997****	1.12	0.446 ^{ns}	1.12	**0.941****	**0.94**	**0.927****	**0.94**
BayesCπ[1]	**1.022****	**0.90**	0.446 ^{ns}	**1.11**	0.717*	1.08	**0.911****	1.06

[1]adding the two largest PCs as covariates, ***P<0.001, **P<0.01, *:P<0.05, ^{ns}non-significant (F test). The closest regression coefficients to the value 1 (one) and the lowest MSE are in boldface.

4.2. Genome-wide association study

Table 6 shows the residual variance and the measures of model fit (DIC) and complexity (pD) for each model in all datasets. Comparing these parameters, it seems the most appropriated models to the data were: HDi-BayesCπ, LDi-BayesCπ and 770k-BayesCπ, all of them considering PCs as covariates. It was expected that family structure would be more important to association than prediction analysis, because differences in allele frequencies among subgroups could lead to false positive associations (PRICE et al., 2006).

BayesC π is also the most complex model, because it has an extra hierarchical level, π. At the same time, it presented lower DIC than the other models within each SNP density, indicating they probably fitted better to the data. It is also clear that TagSNPs dataset did not fit well to the data, perhaps because the chosen method of SNPs selection, which was based exclusively on pairwise LD and could have selected SNPs explaining the same amount of variance from a QTL, for example.

Estimates of π, additive variance and heritability were 0.98, 0.03 and 0.40, respectively, for the three previous models. Even though all individuals were used for GWAS analysis, the estimates were very similar from those in the genomic prediction section. However, in the case of GWAS, the main goal was to search for genomic regions associated to the phenotype. Figures 2 to 4 show the Manhattan plots of the absolute standardized SNP effects of each dataset. It is possible to notice that, as more density the dataset, smaller are the SNP effects, which means that many close markers are sharing the additive effect of the region. One possibility to explore this information would be selecting TagSNPs from phased haplotypes, instead of using only pairwise LD information.

The number of significant SNPs for LDi $(n = 17)$ and HDi $(n = 64)$ were much lower than the 770k $(n = 535)$, then results from exploring genes interactions or combining them in functional clusters would be more robust using genes from 770k dataset, but the following procedures were applied to the three datasets, in order to verify if the less density SNP panels would also be useful to GWAS. The significant SNPs were then assigned to 11, 32 and 236 genes located within 25kb up or

downstream apart, for LDi, HDi and 770k datasets, respectively. The functional annotation assigned 3 genes to 1 cluster (LDi), 4 genes to 1 cluster (HDi) and 35 genes to 3 ranked clusters (770k), and their information are presented in Table 7. The other genes are possible related to the trait, however, because ADG is a complex trait controlled by many genomic regions, investigating genes that share annotation terms and that are part of enriched clusters seems plausible.

Table 6. Measures of fit and model complexity for all datasets

Dataset	Model	Criteria		
		σ_e^2	DIC	pD
LDi	BRR	0.05(0.005)	50.7	197
	BRR[1]	0.05(0.005)	51.0	192
	BayesA	0.05(0.005)	71.1	119
	BayesA[1]	0.06(0.004)	74.0	95
	BayesCπ	0.04(0.005)	36.3	237
	BayesCπ[1]	0.05(0.005)	**35.7**	236
HDi	BRR	0.05(0.004)	60.7	179
	BRR[1]	0.05(0.004)	58.5	175
	BayesA	0.05(0.005)	68.4	135
	BayesA[1]	0.06(0.004)	74.6	90
	BayesCπ	0.05(0.005)	40.8	234
	BayesCπ[1]	0.05(0.005)	**35.6**	237
TagSNPs	BRR	0.05(0.005)	72.1	192
	BRR[1]	0.05(0.005)	73.0	195
	BayesA	0.06(0.005)	95.1	101
	BayesA[1]	0.06(0.005)	99.1	76
	BayesCπ	0.05(0.005)	59.7	232
	BayesCπ[1]	0.05(0.005)	56.9	238
770k	BRR	0.05(0.005)	49.2	206
	BRR[1]	0.05(0.005)	56.3	192
	BayesA	0.05(0.005)	67.2	146
	BayesA[1]	0.05(0.005)	70.7	138
	BayesCπ	0.04(0.004)	42.1	234
	BayesCπ[1]	0.04(0.005)	**36.5**	245

σ_e^2 (sd): residual variance (standard deviation); DIC: deviance information criterion; pD: effective number of parameters; [1]adding the two largest PCs as covariates.

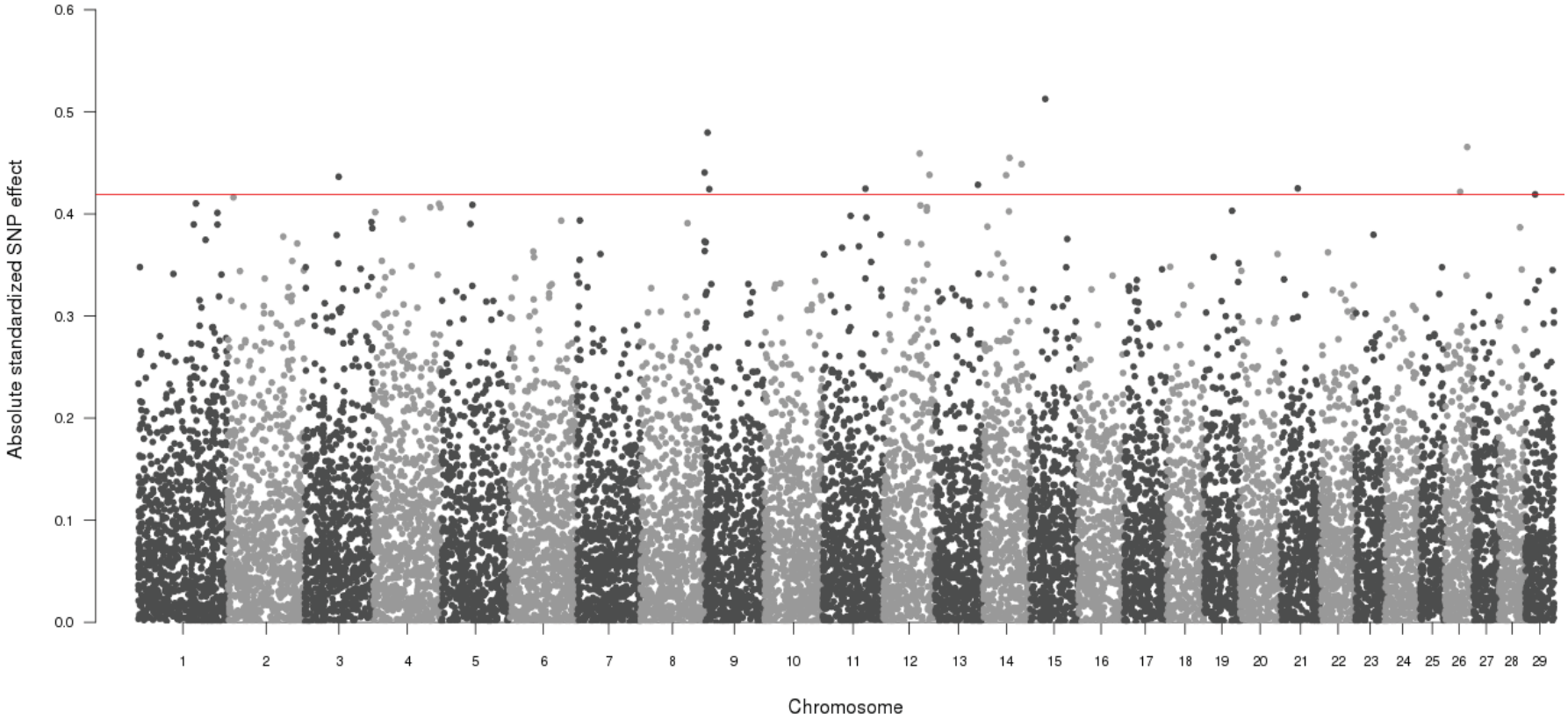

Figure 2. Manhattan plot of absolute standardized SNP effects from LDi-BayesCπ[1] model for average daily gain in Nelore cattle. The line represents the upper limit of the 99.9% highest posterior density of SNP effects. [1]adding the two largest PCs as covariates.

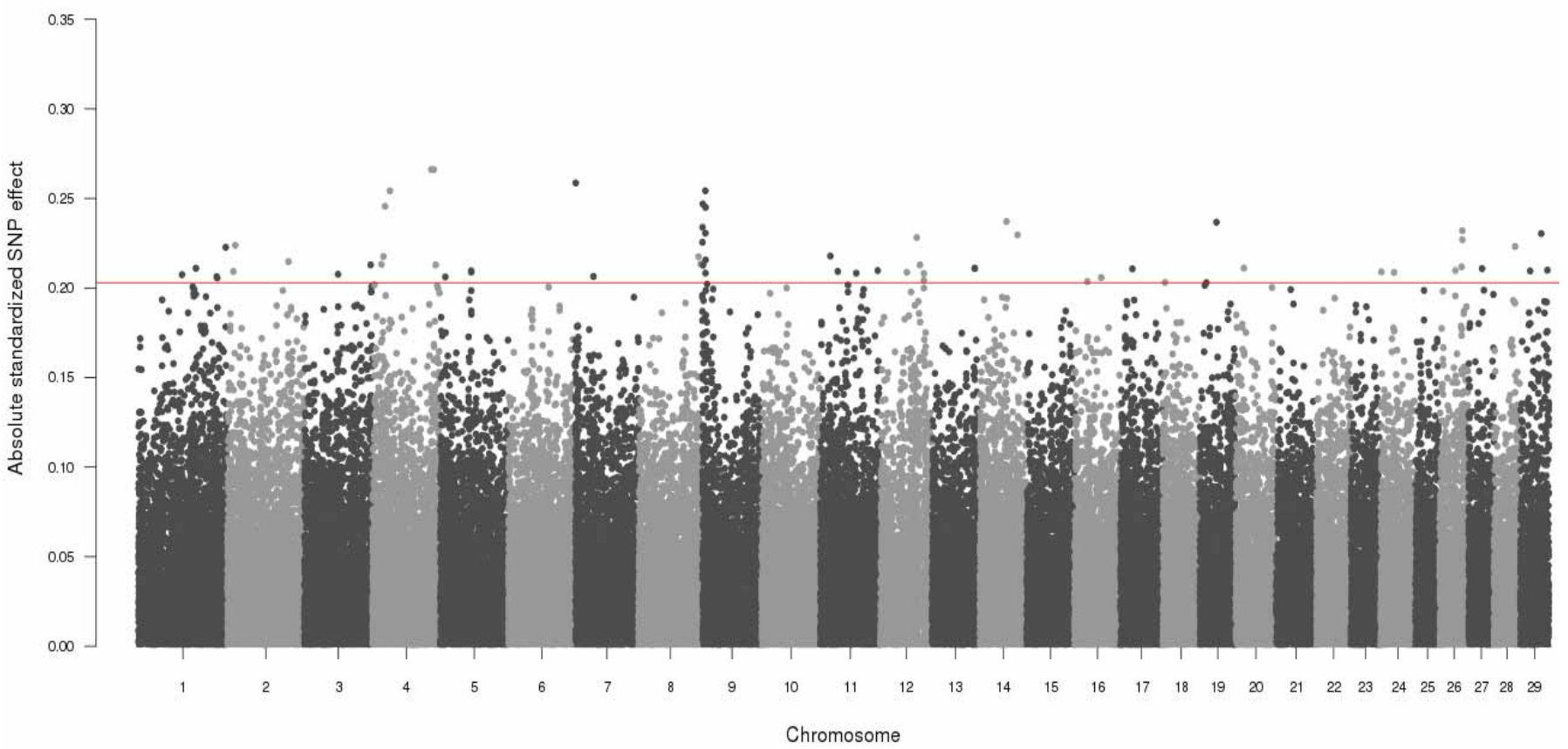

Figure 3. Manhattan plot of absolute standardized SNP effects from HDi-BayesCπ[1] model for average daily gain in Nelore cattle. The line represents the upper limit of the 99.9% highest posterior density of SNP effects. [1]adding the two largest PCs as covariates.

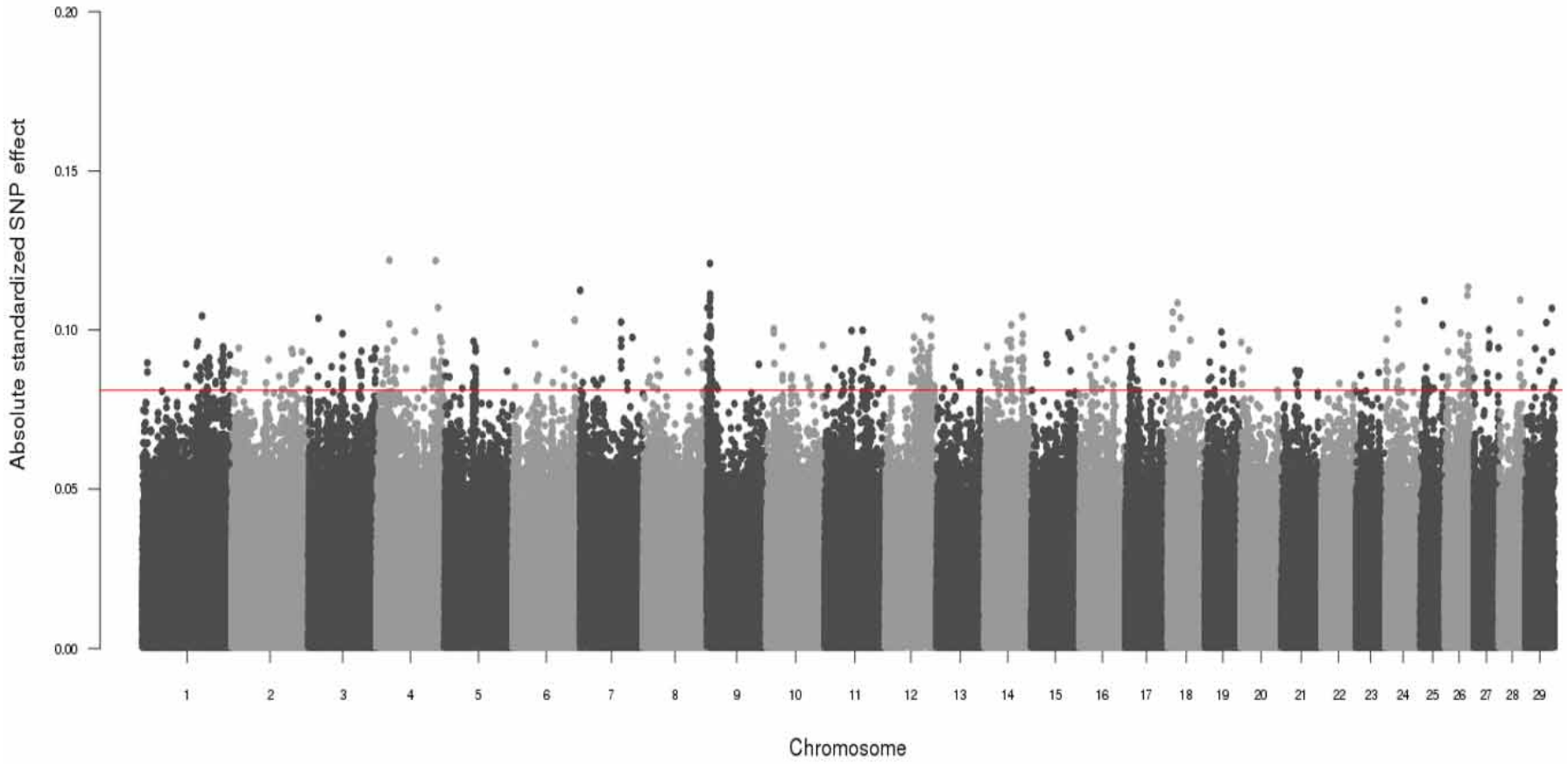

Figure 4. Manhattan plot of absolute standardized SNP effects from 770k-BayesCπ[1] model for average daily gain in Nelore cattle. The line represents the upper limit of the 99.9% highest posterior density of SNP effects. [1]adding the two largest PCs as covariates.

Table 7. Functional annotation clusters from genes assigned to SNPs significantly associated to average daily gain

Dataset	Cluster	Gene symbol	Gene Name	BTA:Position(bp)[3]
LDi-BayesCπ[1]	**1 - Enrichment Score: 0.76**			
	[2]ion binding	**EYA2**	EYA transcriptional coactivator and phosphatase 2	13:76311439-76487993
	cation binding	**ZFHX4**	zinc finger homeobox 4	14:41988047-42189775
	metal ion binding	**CPQ**	carboxypeptidase Q	14:69287300-69893722
HDi-BayesCπ[1]	**1 - Enrichment Score: 0.48**			
	[2]ion binding	ILKAP	Integrin-linked kinase-associated serine/ threonine phosphatase	3:118132617-118158586
	cation binding	**SYT1**	synaptotagmin I	5:8836101-9062297
	metal ion binding	**EYA2**	EYA transcriptional coactivator and phosphatase 2	13:76311439-76487993
		MATN2	matrilin 2	14:68478180-68651684
770k-BayesCπ[1]	**1 - Enrichment Score: 1.50**			
	[2]microtubule motor activity	DNAH7	dynein axonemal heavy chain 7	2:84773343-85028058
	dynein complex	DNAH11	dynein heavy chain Bv1	4:30420509-30799675
	microtubule associated complex	DYNLL2	dynein light chain LC8-type 2	19:9118008-9125251
	2 - Enrichment Score: 1.18			
	[2]ion binding	KCNAB1	potassium voltage-gated channel, shaker-related subfamily, beta member 1	1:112013084-112455492
	cation binding	MBNL1	muscleblind-like splicing regulator 1	1:116238831-116394390
	metal ion binding	ATP2C1	ATPase, Ca++ transporting, type 2C, member 1	1:140368053-140522639
		PADI2	peptidyl arginine deiminase, type II	2:136049644-136103406
		AGMO	alkylglycerol monooxygenase	4:23499626-23891286
		SYT1	synaptotagmin I	5:8836101-9062297
		SOD3	superoxide dismutase 3, extracellular	6:45909842-45930828
		WHSC1	Wolf-Hirschhorn syndrome candidate 1	6:109784642-109826922
		CDHR2	cadherin-related family member 2	7:39392482-39419245
		RYR3	ryanodine receptor 3	10:28789149-29100901

Dataset	Cluster	Gene symbol	Gene name	BTA:Position(bp)[3]
	2 - Enrichment Score: 1.18 (cont)			
		USP39	ubiquitin specific peptidase 39	11:49190710-49227244
		RNF181	ring finger protein 181	11:49244485-49246356
		CIB4	calcium and integrin binding family member 4	11:72920571-72971981
		LMO7	LIM domain 7	12:51141568-51245841
		EYA2	EYA transcriptional coactivator and phosphatase 2	13:76311439-76487993
		NSMCE2	non-SMC element 2, MMS21 homolog	14:16424535-16619178
		CPA6	carboxypeptidase A6	14:33559384-33644209
		ZFHX4	zinc finger homeobox 4	14:41988047-42189775
		MATN2	matrilin 2	14:68478180-68651684
		CPQ	carboxypeptidase Q	14:69287300-69893722
770k-BayesCπ[1]		ACCN1	acid-sensing (proton-gated) ion channel 2	19:16353233-17562209
		RCVRN	recoverin	19:29653157-29661163
		KCNIP1	Kv channel interacting protein 1	20:2283783-2544502
		GMPR	guanosine monophosphate reductase	23:40616330-40677006
		CYB5A	cytochrome b5 type A (microsomal)	24:4413368-4445873
		DTNA	dystrobrevin, alpha	24:22445691-22767026
		TNP2	transition protein 2 (during histone to protamine replacement)	25:9988822-9990357
		ZNF316	zinc finger protein 316	25:39006966-39016262
		IKZF5	IKAROS family zinc finger 5 (Pegasus)	26:43118090-43134286
	3 - Enrichment Score: 1.08			
	[2]respiratory system development	TAB1	TGF-beta activated kinase 1/MAP3K7 binding protein 1	5:111346362-111372218
	lung development	FGF18	fibroblast growth factor 18	20:3132655-3195021
	respiratory tube development	FGFR2	fibroblast growth factor receptor 2	26:41823653-41926635

[1] adding the two largest PCs as covariates, BTA: chromosome, [2]gene ontology identification related to the cluster,[3] UMD3.1. Genes in boldface were identified in, at least, two datasets.

5. DISCUSSION

5.1. Genomic-enabled prediction

For the purpose of comparing estimation efficiency using different SNP densities datasets and Bayesian models, we considered some of the most common criteria recommended to evaluate accuracy, precision and bias of the genome-enabled predictions. In a general overview all SNP densities and models provided similar accuracies, however *Bos taurus indicus* based SNP chips (HDi and LDi) and methods that shrink a proportion of the SNP effects towards zero, such as BayesCπ, seems to result in more reliable predictions. It was expected these marker panels being at least as informative as 770k dataset, since they have been built to capture indicine breeds information. These results corroborate the statement that there is a lack of conserved phase in divergent populations (GIBBS et al., 2009).

We noticed the first divergence among datasets while comparing MAF and the proportion of markers excluded by this criterion. Besides being close to that reported recently (0.25) for Nelore cattle (ESPIGOLAN et al., 2013), the mean MAF of the *Bos taurus indicus* based chips (0.30) are higher than the two other datasets (0.20 and 0.21, for 770k and TagSNPs, respectively). It is also known that MAF may influences the extent of LD at short distances, probably due to an increment in the number of SNP pairs with similar allele frequencies (KHATKAR et al., 2008), thus it was expected that LD extent would be greater on HDi and LDi datasets and their markers would provide results as good as more density SNP chips.

According to the clusters presented in Figure 1, there are, at least, two explanations why most of half-sib families were not so easily differentiated. First, even thought the chosen sires were the less related among Nelore sires available in artificial insemination centers in Brazil, it was expected they were somewhat related and so their offspring. Secondly, as we do not have information about the dams, it is possible that the relationship among them, and even to the sires, could have influenced the observed grouping. Thus, it was expected some level of relationship within and even between clusters, which means it is not simple to point out how much predictions were influenced by either linkage equilibrium or disequilibrium.

The first and second PCs, for all datasets, explained up to 8 and 7% of the variance among animals, respectively. It was not expected a large explanation of the variance because they were all from the same breed and, mainly, from 31 half-sib families. On the contrary, Bolormaa et al. (2013b) observed a much larger proportion explained by the first PC (87.8%) while analyzing individuals from divergent breeds (*Bos taurus taurus* x *Bos taurus indicus*). Besides that, based on the size of training subset in this study $(\mathrm{n} = 568)$, it was expected a large influence of the family relationship (WIENTJES; VEERKAMP; CALUS, 2013). Thereby, it is important to make clear that results found in this study could not persist in a more divergent sample. In a previous analysis at the same dataset used in the present study, effective population size (N_e) was estimated to be 214 animals (data not shown), much higher than the 68 reported by Faria et al. (2009), thus we would expect to find less extent LD.

Regarding to the estimation of variance components (Table 3), there was a clear divergence among regression models within SNP densities. For example, additive variance estimated using BGBLUP were the lowest (~0.02) in this study. This model was implemented as a special case of RKHS using the realized G matrix as the kernel, which was built based on probability of identity by state (IBS) alleles. GEBVs were predicted as function of the covariance among individuals due to additive inheritance captured by the SNPs genotypes and even non-related animals presented some level of relationship. While comparing results obtained using IBS and identity by descent (IBD) matrices, Luan et al. (2012) observed that IBS tends to explain a lower proportion of variation. A possible explanation is that SNP markers cannot capture all the genetic relatedness that matters to access the additive genetic variance for a given trait, simply because they are not QTLs (GIANOLA; ROSA, 2015).

On the other hand, BayesA and BayesCπ are marker-based linear models, so the additive genetic variance is a function of SNP effects and their uncertainty variances and allelic frequencies (GIANOLA et al., 2009). Results from BayesA models were not consistent among SNP densities and, based on this previous statement, we hypothesized that by modeling a great number of markers, some of the major effects would not be shrunk towards zero, thus their unique variances

could be associated to the larger additive genetic variances presented in Table 3. Moreover, it seems the family structure had a stronger effect on the estimation of additive variances by using BayesA, probably because differences in allelic frequencies of SNPs with larger effects would have greater influence on the results.

Besides that, it has been reported that models that allow unequal variances for SNP effects (e.g.: BayesA) perform better for traits which appear to have genes or QTLs with large effects (DE LOS CAMPOS et al., 2013; ROLF et al., 2015), which is the opposite of what has been found recently for ADG in different feedlot-raised animals. For example, for *Bos taurus taurus* and crossbred animals, Lu et at. (2013) identified 39 significant SNPs explained from 3.04% to 7.12% of the phenotypic variance. In Nelore cattle, Santana et al. (2014b) reported that the ten most significant SNPs explained up to 6.28% of the phenotypic variance, and by analyzing a subset of the same population of the present study, Oliveira et al. (2014) found six 1-Mb windows of SNPs explained from 1.01 to 4.94% of the additive genetic variance. These studies support the assumption that ADG is probably controlled by a large number of genes or QTLs with small effect. Furthermore, ADG might be viewed as a combination of other traits, such as feed intake, feed efficiency, rate of muscle and fat deposition.

Similar to what was observed for BGBLUP, there was almost no variation on results from BayesCπ among all SNP densities. However, the lower residual variances estimated by BayesCπ models indicate they fitted better to the data than BGBLUP. Beyond that, nearly the same additive genetic variance could be accessed even through less density panels. The need to estimate π from a relatively small dataset could be a concern and, in fact, the MCMC chains for this parameter have not mixed well as the residual variance chain (Figure 5A). A viable option would be setting the parameter π as a fixed value, for which the number of SNPs considered in the model would be less than the number of individuals, similar to what was done by Saatchi et al. (2011) and Neves et al. (2014).

Coefficients of heritability estimated by BayesCπ models ranged from 0.395 to 0.439 (Table 3). Recently, Oliveira et al. (2014) analyzed a subset of the present population $(\mathrm{n} = 593)$ for feed efficiency traits and reported a heritability of 0.42 for ADG. This was slightly lower than reported by Marques et al. (2013) for Nelore cattle

up to 426 days of age in feedlot performance test, using only pedigree information ($h^2 = 0.55$). Grion et al. (2014) analyzed data from animals in feedlot performance test, in which males and females with initial age from 223 to 391 days were evaluated up to 112 days. Even though they were much younger than the animals in the present study and the author have not used SNP marker information, their estimate of heritability ($h^2 = 0.42$) was similar to our results. Although heritability is a population parameter, it is known that magnitudes of heritability estimates of similar traits are also often similar in other populations (VISSCHER; HILL; WRAY, 2008).

Estimates of Pearson's correlation coefficients between adjusted phenotypes and GEBVs were used as proxy of genome-enabled prediction accuracies (Table 4). All estimates were similar (from 0.24 to 0.27), except for HDi dataset with PCs as covariates, which were about 18% lower that other all. It is clear that PCs were mainly clustering half-sib families without any other apparent standard, such as sires lineages. Nevertheless, in the case of HDi dataset, the second PC was significantly associated to the raw phenotype (data not shown). We hypothesized that, in this case, PCs could capture noise information and despite this may have influenced the phenotype adjustment, it remains unclear even by comparing the summary statistics of ADG within families (Table 3A). Bolormaa et al. (2013b) reported even lower accuracies (from 0.13 to 0.24) of GEBVs for ADG in feedlot estimated using GBLUP in *Bos taurus taurus* and *Bos taurus indicus* animals.

In order to investigate whether the difference from HDi dataset with PCs as covariates was due to noise information, analyses based on training and testing groups formed by different half-sib families, for example, could be performed. There are also techniques that may determine which PCs are related to population structure as well (PATTERSON; PRICE; REICH, 2006). Another possibility is clustering the most related individuals together and performing predictions between less related groups (SAATCHI et al., 2011), in order to avoid overestimated accuracies of GEBVs. However, it is important to keep in mind whether the selection candidates will be related to training dataset in commercial herds in which genomic selection is intended to be applied.

It is known that the success of genomic selection depends on accuracy of GEBVs, which in turn is a function of heritability, size of training population and N_e

(GODDARD; HAYES, 2009). Given the actual size of training subset of the present study ($\mathrm{n} = 568$), it was not expected much larger accuracies, and it was probably the reason why results from different models were all similar. Based on the simulation presented by van der Werf (2013), who considered a population with $N_e = 250$ and a trait with $h^2 = 0.5$, a training population of 500 animals would reach an accuracy of 0.2, similar to our results. Besides that, it seems that a training population of more than 2000 individuals would be required to achieve an accuracy of 0.4, thus increasing the number of reference animals is a crucial point. In general, the author showed that for the same heritability, the smaller the N_e, the smaller the size of reference population needed to achieve the same prediction accuracy.

On the other hand, ADG in feedlot finished steers could be view as a new selection criteria for Nelore cattle, and if accuracies of the GEBVs allow selection of young animals and genetic gains at a reduced genotyping costs, it would be widely used. Accuracy of GEBVs and potential genetic gains by using genomic enabled-prediction should be compared to the results from progeny tests, for example. It is well known that progeny tests increase the generation interval, since selection candidates must have a great number of progenies in order to have greater accuracy, thus finding a balance between these two factor remains as a critical point in animal breeding.

The prediction bias was measured by regressing the adjusted phenotypes on GEBVs (Table 5). The LDi dataset had more slopes closer to the value 1, indicating that less bias could be accessed using few representative SNPs. For 770k, only results from BayesCπ models would be considered no biased. Also it is clear that results from BayesA models (except for TagSNPs) were deflated, which means that the variance of GEBV was generally underestimated (CALUS; MULDER; VEERKAMP, 2011). The opposite was observed for all models applied to TagSNPs dataset, thus it seems that selecting markers based only on their pairwise r^2 resulted in overestimated predictors, although the results were non-significant.

Disregarding HDi dataset with PCs as covariates, differences among prediction accuracies were minimum, thus information of slopes and MSE (Table 5) were combined to determine the most appropriated model and SNP density. The models which better adjusted to the data were 770k-BayesC π with PCs as

covariates, HDi-BayesCπ and LDi-BayesCπ. The current average cost of genotyping can easily reach $150.00, $100.00 and $50.00 per animal, for 770k, HDi and LDi, respectively, therefore if it would be possible to predict accurate GEBVs using less density panels of SNPs at lower costs, the implementation and application of genomic selection would be better accepted for beef cattle industry.

5.2. Genome-wide association study

Association analysis between SNPs from four different SNP densities (770k, TagSNPs, HDi and LDi) and feedlot average daily gain were performed by applying three Bayesian regression models. To evaluate how the models fitted to the data, information of residual variance, DIC and pD were used. As in genomic-enabled prediction, *Bos taurus indicus* based SNP chips (HDi and LDi) performed as well as 770k dataset and BayesCπ regression models were considered more appropriated. However, the number of significant SNPs and, consequently, the number of genes associated to the phenotype were much higher for 770k.

All the three genes from LDi cluster and three out of four genes from HDi were also identified in the second cluster from 770k dataset (Table 7). The SYT1 (s*ynaptotagmin I*) gene is located on chromosome 5 and is a synaptic vesicle protein important for Ca2+ dependent neurotransmitter (UBACH et al., 2001). It has been reported as associated to eating behavior (DO et al., 2013) and growth traits (PUIG-OLIVERAS et al., 2014) in pigs. The EYA2 (*EYA transcriptional coactivator and phosphatase 2*) gene, which is located on chromosome 13, is a member of homeobox subfamily, that are involved in proliferation of muscle precursor cells in embryonic stages (MUNTONI et al., 2002). It has also been identified as having a function on myogenesis in mouse (GRIFONE et al., 2007) and has been associated with muscle tissue development in pigs (PÉREZ-MONTARELO et al., 2012).

The other three genes are located on chromosome 14, which has been reported containing QTLs for ADG and growth traits in *Bos taurus taurus* cattle (KNEELAND et al., 2004; MIZOSHITA et al., 2004), and recently, SNPs associated to birth weight (UTSUNOMIYA et al., 2013) and ADG (OLIVEIRA et al., 2014) in Nelore cattle. The ZFHX4 (*zinc finger homeobox 4*) gene may be related to muscle

differentiation in mice (HEMMI et al., 2006) and to genes which regulate energy balance in *Bos taurus indicus* cattle (FORTES et al., 2011). Besides that, it seems associated with lipid and carbohydrate metabolism (RAMAYO-CALDAS et al., 2014). The MATN2 (*matrilin 2*) was first identified in mouse and human tissues by Deák et al. (1997) and was found expressed in developing bones in mouse (ASZÓDI et al., 1999). Furthermore, Piecha et al. (1999) observed that matrilin-2 mRNA was also present in myoblasts, epithelial and endothelial cells, however, the protein was apparently transported to connective tissue. Also known as PGCP (*plasma glutamate carboxypeptidase*), the CPQ gene (*carboxypeptidase Q*) was found to be significantly expressed more highly in muscle with low intramuscular fat in Hanwoo animals, a *Bos taurus taurus* Korean cattle (LEE et al., 2011). The authors also stated that marbling in the muscle would be a combination of metabolic pathways of adipocytes inside de muscle and also in the liver. Recently, analyzing the global liver gene expression of Nelore cattle, actually a subset of the animals in this study, Tizioto et al. (2015b) found that CPQ was significantly more expressed in livers of steers with low residual feed intake, labeled as inefficient.

Back to the first cluster from 770k dataset is related to dynein proteins complex and was identified only by genes from this dataset. Dyneins are ATP-dependent cytoskeletal motor proteins, related with cell motility, which generate force and movement on microtubules by biological processes, such as meiosis, mitosis and intracellular transport (ROBERTS et al., 2013). This family of proteins have been described as part of adipose tissue differentiation in humans (VIGNAL et al., 2002), and muscle fiber length, function and myonuclear position in *Drosophila* (FOLKER; SCHULMAN; BAYLIES, 2012). Besides that, Saatchi et al. (2014) have found a SNP on chromosome 19 associated to residual feed intake in a *Bos taurus taurus* x *Bos taurus indicus* composite cattle, next to a dynein family gene (DNAH17 - dynein, axonemal, heavy chain 17).

Regarding to the third cluster from 770k dataset, the gene FGF18 (*fibroblast growth factor 18*) have been reported having function in postnatal development of lung in rats (CHAILLEY-HEU et al., 2005) and ovarian follicle growth in cattle (PORTELA et al., 2010). Based on the information that oxygen consumption and heart rate play a role in the estimation of energy expenditure in beef cattle (BROSH

et al., 2004), the lung development could be somehow related to growth and efficiency traits.

Genomic association has a different goal from genomic prediction and, even though it is possible to obtain part of the information from the 770k by using less density SNPs panels, it might be senseless. The central idea of an association analysis is to find genes or genomic regions associated to the phenotype and, it is expected that complex traits will be related to many of these regions. Some of the genes associated to the phenotype seem to be part of some growth-related pathway, however, we should be careful in making inferences about results from small samples, such as used in the present study.

6. CONCLUSION

Accuracy of genomic-enabled predictions and genome-wise association analysis of feedlot average gain were affected by regression models and density of SNP chips. In a general overview, the phenotype seems to be a complex traits controlled by many genes with small effects, thus the model which accounts for this performed better for both approaches. However, two factors were observed to affect the accuracy and should be further investigated, such as the relationship between animals from training and testing subsets and sample size. Furthermore, as expected with the sample size used in this study $(\mathrm{n} = 718)$, results from less density marker panels based on *Bos taurus indicus* information were as good as the high density panel, and at lower genotyping costs. Meanwhile, for association analysis, results from the high density panel would be more useful, since information about a higher number of genes could be reached.

It was possible to identify genes associated to feedlot average daily gain, cluster and classified them by importance on the dataset, which makes the interpretation easier. Based on previous studies about these genes, they seem to be somehow related to feedlot average daily weight gain in Nelore cattle. However, it is not reasonable to make inferences about them. First because, even though family structure was considered in the model, it is important to make sure that results were

not highly influenced by large differences in allele frequencies. Secondly, as well as in the prediction, the sample size is still a major concern.

Further investigations on this dataset, or any other which is formed by half-sib families, should test the influence of relationship between training and testing set on predictions and associations. It is important to keep in mind that, in order to implement genomic selection, analysis should mimic the actual production system.

7. REFERENCES

ABASHT, B.; LAMONT, S.J. Genome-wide association analysis reveals cryptic alleles as an important factor in heterosis for fatness in chicken F2 population. **Animal Genetics,** v. 38, n. 5, p. 491-8, 2007.

AMILLS, M.; VIDAL, O.; VARONA, L.; TOMAS, A.; GIL, M.; SANCHEZ, A.; NOGUERA, J.L. Polymorphism of the pig 2,4-dienoyl CoA reductase 1 gene (DECR1) and its association with carcass and meat quality traits. **Journal of Animal Science,** v. 83, n. 3, p. 493-8, 2005.

ANDERSSON, L. Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. **Genetica,** v. 136, n. 2, p. 341-9, 2009.

ASZÓDI, A.; BATEMAN, J.F.; HIRSCH, E.; BARANYI, M.; HUNZIKER, E.B.; HAUSER, N.; BÖSZE, Z.; FÄSSLER, R. Normal Skeletal Development of Mice Lacking Matrilin 1: Redundant Function of Matrilins in Cartilage? **Molecular and Cellular Biology,** v. 19, n. 11, p. 7841-7845, 1999.

BAKKER, P.I.W.; YELENSKY, R.; PE'ER, I.; GABRIEL, S.B.; DALY, M.J.; ALTSHULER, D. Efficiency and power in genetic association studies. **Nature Genetics,** v. 37, n. 11, p. 1217-1223, 2005.

BALDING, D.J. A tutorial on statistical methods for population association studies. **Nature Review Genetics,** v. 7, n. 10, p. 781-791, 2006.

BARENDSE, W.; REVERTER, A.; BUNCH, R.J.; HARRISON, B.E.; BARRIS, W.; THOMAS, M.B. A validated whole-genome association study of efficient food conversion in cattle. **Genetics,** v. 176, n. 3, p. 1893-905, 2007.

BARRETT, J.C.; FRY, B.; MALLER, J.; DALY, M.J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics,** v. 21, n. 2, p. 263-265, 2005.

BERRY, D.P.; BASTIAANSEN, J.W.; VEERKAMP, R.F.; WIJGA, S.; WALL, E.; BERGLUND, B.; CALUS, M.P. Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries. **Animal,** v. 6, n. 8, p. 1206-15, 2012.

BERRY, D.P.; COFFEY, M.P.; PRYCE, J.E.; DE HAAS, Y.; LØVENDAHL, P.; KRATTENMACHER, N.; CROWLEY, J.J.; WANG, Z.; SPURLOCK, D.; WEIGEL, K.; MACDONALD, K.; VEERKAMP, R.F. International genetic evaluations for feed intake in dairy cattle through the collation of data from multiple sources. **Journal of Dairy Science,** v. 97, n. 6, p. 3894-3905, 2014.

BOLIGON, A.A.; ALBUQUERQUE, L.G.; MERCADANTE, M.E.Z.; LÔBO, R.B. Herdabilidades e correlações entre pesos do nascimento à idade adulta em rebanhos da raça Nelore. **Revista Brasileira de Zootecnia,** v. 38, p. 2320-2326, 2009.

BOLIGON, A.A.; CARVALHEIRO, R.; ALBUQUERQUE, L.G. Evaluation of mature cow weight: Genetic correlations with traits used in selection indices, correlated responses, and genetic trends in Nelore cattle. **Journal of Animal Science,** v. 91, n. 1, p. 20-28, 2013.

BOLIGON, A.A.; LONG, N.; ALBUQUERQUE, L.G.; WEIGEL, K.A.; GIANOLA, D.; ROSA, G.J.M. Comparison of selective genotyping strategies for prediction of breeding values in a population undergoing selection. **Journal of Animal Science,** v. 90, n. 13, p. 4716-4722, 2012.

BOLORMAA, S.; NETO, L.R.; ZHANG, Y.D.; BUNCH, R.J.; HARRISON, B.E.; GODDARD, M.E.; BARENDSE, W. A genome-wide association study of meat and carcass traits in Australian cattle. **Journal of Animal Science,** v. 89, n. 8, p. 2297-309, 2011.

BOLORMAA, S.; PRYCE, J.; KEMPER, K.; HAYES, B.; ZHANG, Y.; TIER, B.; BARENDSE, W.; REVERTER, A.; GODDARD, M. Detection of quantitative trait loci in Bos indicus and Bos taurus cattle using genome-wide association studies. **Genetics Selection Evolution,** v. 45, n. 1, p. 43-54, 2013a.

BOLORMAA, S.; PRYCE, J.E.; HAYES, B.J.; GODDARD, M.E. Multivariate analysis of a genome-wide association study in dairy cattle. **Journal of Dairy Science,** v. 93, n. 8, p. 3818-33, 2010.

BOLORMAA, S.; PRYCE, J.E.; KEMPER, K.; SAVIN, K.; HAYES, B.J.; BARENDSE, W.; ZHANG, Y.; REICH, C.M.; MASON, B.A.; BUNCH, R.J.; HARRISON, B.E.; REVERTER, A.; HERD, R.M.; TIER, B.; GRASER, H.U.; GODDARD, M.E. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus, and composite beef cattle. **Journal of Animal Science,** v. 91, n. 7, p. 3088-104, 2013b.

BOUWMAN, A.C.; BOVENHUIS, H.; VISKER, M.H.; VAN ARENDONK, J.A. Genome-wide association of milk fatty acids in Dutch dairy cattle. **BMC Genetics,** v. 12, p. 43-54, 2011.

BRADLEY, D.G.; CUNNINGHAM, E.P. Genetic Aspects of Domestication In: FRIES, R. e RUVINSKY, A. (Ed.). **The Genetics of Cattle**. 1. New York, NY: CABI Publishing, 1999.

BRAZILIAN BEEF EXPORTERS ASSOCIATION. **Brazilian beef profile 2013.** Disponível em: <http://www.brazilianbeef.org.br/texto.asp?id=9>. Acesso em: 08 mar. 2015

BROSH, A.; AHARONI, Y.; SHARGAL, E.; SHARIR, B.; GUTMAN, M.; CHOSHNIAK, I. Energy balance of grazing beef cattle in Mediterranean pasture, the effects of stocking rate and season: 2. Energy expenditure as estimated from heart rate and oxygen consumption, and energy balance. **Livestock Production Science,** v. 90, n. 2, p. 101-115, 2004.

BROWNING, B.L.; BROWNING, S.R. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. **The American Journal of Human Genetics,** v. 84, n. 2, p. 210-223, 2009.

BUZANSKAS, M.E.; GROSSI, D.A.; VENTURA, R.V.; SCHENKEL, F.S.; SARGOLZAEI, M.; MEIRELLES, S.L.C.; MOKRY, F.B.; HIGA, R.H.; MUDADU, M.A.; DA SILVA, M.V.G.B.; NICIURA, S.C.M.; JÚNIOR, R.A.A.T.; ALENCAR, M.M.; REGITANO, L.C.A.; MUNARI, D.P. Genome-Wide Association for Growth Traits in Canchim Beef Cattle. **PLoS ONE,** v. 9, n. 4, p. e94802, 2014.

CALUS, M.; MULDER, H.; VEERKAMP, R. Estimating genomic breeding values and detecting QTL using univariate and bivariate models. **BMC Proceedings,** v. 5, n. S3, p. 5-9, 2011.

CASELLA, G.; BERGER, R.L. Point Estimation. In: CASELLA, G. e BERGER, R.L. (Ed.). **Statistical inference**. 2nd. Australia ; Pacific Grove, CA: Thomson Learning, 2002.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). **Perspectivas para o agronegócio em 2015**. Piracicaba: Centro de Estudos Avançados em Economia Aplicada (CEPEA), 2014a.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA (CEPEA). **Relatório PIBagro-Brasil**. Piracicaba: Centro de Estudos Avançados em Economia Aplicada (CEPEA), 2014b.

CESAR, A.; REGITANO, L.; MOURAO, G.; TULLIO, R.; LANNA, D.; NASSU, R.; MUDADO, M.; OLIVEIRA, P.; DO NASCIMENTO, M.; CHAVES, A.; ALENCAR, M.; SONSTEGARD, T.; GARRICK, D.; REECY, J.; COUTINHO, L. Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. **BMC Genetics,** v. 15, n. 1, p. 39-54, 2014.

CHAILLEY-HEU, B.; BOUCHERAT, O.; BARLIER-MUR, A.-M.; BOURBON, J.R. FGF-18 is upregulated in the postnatal rat lung and enhances elastogenesis in myofibroblasts. **American Journal of Physiology,** v. 288, n. 1, p. L43-L51, 2005.

CHEN, M.-H.; SHAO, Q.-M. Monte Carlo Estimation of Bayesian Credible and HPD Intervals. **Journal of Computational and Graphical Statistics,** v. 8, n. 1, p. 69-92, 1999.

CLAYTON, D. **snpStats: SnpMatrix and XSnpMatrix classes and methods**: R package version 1.10.0, 2012.

COLE, J.B.; WIGGANS, G.R.; MA, L.; SONSTEGARD, T.S.; LAWLOR, T.J., JR.; CROOKER, B.A.; VAN TASSELL, C.P.; YANG, J.; WANG, S.; MATUKUMALLI, L.K.; DA, Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. **BMC Genomics,** v. 12, p. 408, 2011.

DAETWYLER, H.D.; KEMPER, K.E.; VAN DER WERF, J.H.J.; HAYES, B.J. Components of the accuracy of genomic prediction in a multi-breed sheep population. **Journal of Animal Science,** v. 90, n. 10, p. 3375-3384, 2012.

DAETWYLER, H.D.; SCHENKEL, F.S.; SARGOLZAEI, M.; ROBINSON, J.A. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. **Journal of Dairy Science,** v. 91, n. 8, p. 3225-36, 2008.

DAETWYLER, H.D.; VILLANUEVA, B.; BIJMA, P.; WOOLLIAMS, J.A. Inbreeding in genome-wide selection. **Journal of Animal Breeding and Genetics,** v. 124, n. 6, p. 369-376, 2007.

DE LOS CAMPOS, G.; GIANOLA, D.; ROSA, G.J.M. Reproducing kernel Hilbert spaces regression: A general framework for genetic evaluation. **Journal of Animal Science,** v. 87, n. 6, p. 1883-1887, 2009.

DE LOS CAMPOS, G.; HICKEY, J.M.; PONG-WONG, R.; DAETWYLER, H.D.; CALUS, M.P. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics,** v. 193, n. 2, p. 327-45, 2013.

DE LOS CAMPOS, G.; RODRIGUEZ, P.P. **BGLR: Bayesian Generalized Linear Regression**: R package version 1.0.3. http://CRAN.R-project.org/package=BGLR 2014.

DEÁK, F.; PIECHA, D.; BACHRATI, C.; PAULSSON, M.; KISS, I. Primary Structure and Expression of Matrilin-2, the Closest Relative of Cartilage Matrix Protein within the von Willebrand Factor Type A-like Module Superfamily. **Journal of Biological Chemistry,** v. 272, n. 14, p. 9268-9274, 1997.

DENNIS JR, G.; SHERMAN, B.T.; HOSACK, D.A.; YANG, J.; GAO, W.; LANE, H.C.; LEMPICKI, R.A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. **Genome Biology,** v. 4, n. 5, p. 3-7, 2003.

DO, D.N.; STRATHE, A.B.; OSTERSEN, T.; JENSEN, J.; MARK, T.; KADARMIDEEN, H.N. Genome-Wide Association Study Reveals Genetic Architecture of Eating Behavior in Pigs and Its Implications for Humans Obesity by Comparative Mapping. **PLoS ONE,** v. 8, n. 8, p. e71509, 2013.

ESPIGOLAN, R.; BALDI, F.; BOLIGON, A.; SOUZA, F.; GORDO, D.; TONUSSI, R.; CARDOSO, D.; OLIVEIRA, H.; TONHATI, H.; SARGOLZAEI, M.; SCHENKEL, F.; CARVALHEIRO, R.; FERRO, J.; ALBUQUERQUE, L. Study of whole genome linkage disequilibrium in Nellore cattle. **BMC Genomics,** v. 14, p. 305-312, 2013.

FARIA, F.J.C.; FILHO, A.E.V.; MADALENA, F.E.; JOSAHKIAN, L.A. Pedigree analysis in the Brazilian Zebu breeds. **Journal of Animal Breeding and Genetics,** v. 126, n. 2, p. 148-153, 2009.

FERNANDO, R.L.; DEKKERS, J.C.; GARRICK, D.J. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. **Genetics Selection Evolution,** v. 46, p. 50-63, 2014.

FERNANDO, R.L.; GROSSMAN, M. Marker Assisted Selection Using Best Linear Unbiased Prediction. **Genetics Selection Evolution,** v. 21, n. 4, p. 467-477, 1989.

FERRAZ, J.B.; FELICIO, P.E. Production systems - an example from Brazil. **Meat Science,** v. 84, n. 2, p. 238-43, 2010.

FINLAY, E.K.; BERRY, D.P.; WICKHAM, B.; GORMLEY, E.P.; BRADLEY, D.G. A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. **PLoS ONE,** v. 7, n. 2, p. e30545, 2012.

FNP CONSULTORIA E COMÉRCIO. **ANUALPEC 2014: Anuário da pecuária brasileira**. São Paulo: 2014.

FOLKER, E.S.; SCHULMAN, V.K.; BAYLIES, M.K. Muscle length and myonuclear position are independently regulated by distinct Dynein pathways. **Development,** v. 139, n. 20, p. 3827-3837, 2012.

FORTES, M.R.; SNELLING, W.M.; REVERTER, A.; NAGARAJ, S.H.; LEHNERT, S.A.; HAWKEN, R.J.; DEATLEY, K.L.; PETERS, S.O.; SILVER, G.A.; RINCON, G.; MEDRANO, J.F.; ISLAS-TREJO, A.; THOMAS, M.G. Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. **Journal of Animal Science,** v. 90, n. 9, p. 2894-906, 2012a.

FORTES, M.R.S.; LEHNERT, S.A.; BOLORMAA, S.; REICH, C.; FORDYCE, G.; CORBET, N.J.; WHAN, V.; HAWKEN, R.J.; REVERTER, A. Finding genes for economically important traits: Brahman cattle puberty. **Animal Production Science,** v. 52, n. 3, p. 143-150, 2012b.

FORTES, M.R.S.; REVERTER, A.; NAGARAJ, S.H.; ZHANG, Y.; JONSSON, N.N.; BARRIS, W.; LEHNERT, S.; BOE-HANSEN, G.B.; HAWKEN, R.J. A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. **Journal of Animal Science,** v. 89, n. 6, p. 1669-1683, 2011.

FORTES, M.R.S.; REVERTER, A.; ZHANG, Y.; COLLIS, E.; NAGARAJ, S.H.; JONSSON, N.N.; PRAYAGA, K.C.; BARRIS, W.; HAWKEN, R.J. Association weight matrix for the genetic dissection of puberty in beef cattle. **Proceedings of the National Academy of Sciences,** v. 107, n. 31, p. 13642-13647, 2010.

FOX, J.; WEISBERG, S. **An {R} Companion to Applied Regression**. Thousand Oaks CA: Sage. 2011.

GARNERO, A.D.V.; MARCONDES, C.R.; GUNSKI, R.J.; OLIVEIRA, H.N.D.; LÔBO, R.B. Genetic trends in the expected progeny difference of the asymptotic weight of Nelore females. **Genetics and Molecular Biology,** v. 29, p. 648-652, 2006.

GARRICK, D.; TAYLOR, J.; FERNANDO, R. Deregressing estimated breeding values and weighting information for genomic regression analyses. **Genetics Selection Evolution,** v. 41, n. 1, p. 55-63, 2009.

GARRICK, D.J. The nature, scope and impact of genomic prediction in beef cattle in the United States. **Genetics Selection Evolution,** v. 43, p. 17-28, 2011.

GEORGES, M. Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. **Annual Review of Genomics and Human Genetics,** v. 8, p. 131-62, 2007.

GEORGES, M.; NIELSEN, D.; MACKINNON, M.; MISHRA, A.; OKIMOTO, R.; PASQUINO, A.T.; SARGEANT, L.S.; SORENSEN, A.; STEELE, M.R.; ZHAO, X.; ET AL. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. **Genetics,** v. 139, n. 2, p. 907-20, 1995.

GIANOLA, D. Priors in Whole-Genome Regression: The Bayesian Alphabet Returns. **Genetics,** v. 194, n. 3, p. 573-596, 2013.

GIANOLA, D.; DE LOS CAMPOS, G.; HILL, W.G.; MANFREDI, E.; FERNANDO, R. Additive Genetic Variability and the Bayesian Alphabet. **Genetics,** v. 183, n. 1, p. 347-363, 2009.

GIANOLA, D.; FERNANDO, R.L. Bayesian Methods in Animal Breeding Theory. **Journal of Animal Science,** v. 63, n. 1, p. 217-244, 1986.

GIANOLA, D.; FERNANDO, R.L.; STELLA, A. Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. **Genetics,** v. 173, n. 3, p. 1761-1776, 2006.

GIANOLA, D.; PEREZ-ENCISO, M.; TORO, M.A. On marker-assisted prediction of genetic value: beyond the ridge. **Genetics,** v. 163, n. 1, p. 347-65, 2003.

GIANOLA, D.; ROSA, G.J. One hundred years of statistical developments in animal breeding. **Annual Review of Animal Biosciences,** v. 3, p. 19-56, 2015.

GIANOLA, D.; VAN KAAM, J.B.C.H.M. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. **Genetics,** v. 178, n. 4, p. 2289-2303, 2008.

GIBBS, R.A.; TAYLOR, J.F.; VAN TASSELL, C.P.; BARENDSE, W.; EVERSOIE, K.A.; GILL, C.A.; GREEN, R.D.; HAMERNIK, D.L.; KAPPES, S.M.; LIEN, S.; MATUKUMALLI, L.K.; MCEWAN, J.C.; NAZARETH, L.V.; SCHNABEL, R.D.; TAYLOR, J.F.; WEINSTOCK, G.M.; WHEELER, D.A.; AJMONE-MARSAN, P.; BARENDSE, W.; BOETTCHER, P.J.; CAETANO, A.R.; GARCIA, J.F.; HANOTTE, O.; MARIANI, P.; SKOW, L.C.; WILLIAMS, J.L.; CAETANO, A.R.; DIALLO, B.; GREEN, R.D.; HAILEMARIAM, L.; HANOTTE, O.; MARTINEZ, M.L.; MORRIS, C.A.; SILVA, L.O.C.; SPELMAN, R.J.; TAYLOR, J.F.; MULATU, W.; ZHAO, K.Y.; ABBEY, C.A.; AGABA, M.; ARAUJO, F.R.; BUNCH, R.J.; BURTON, J.; GILL, C.A.; GORNI, C.; OLIVIER, H.; HARRISON, B.E.; LUFF, B.; MACHADO, M.A.; MARIANI, P.; MORRIS, C.A.; MWAKAYA, J.; PLASTOW, G.; SIM, W.; SKOW, L.C.; SMITH, T.; SONSTEGARD, T.S.; SPELMAN, R.J.; TAYLOR, J.F.; THOMAS, M.B.; VALENTINI, A.; WILLIAMS, P.; WOMACK, J.; WOOLIAMS, J.A.; LIU, Y.; QIN, X.; WORLEY, K.C.; GAO, C.; GILL, C.A.; JIANG, H.Y.; LIU, Y.; MOORE, S.S.; NAZARETH, L.V.; REN, Y.R.; SONG, X.Z.; BUSTAMANTE, C.D.; HERNANDEZ, R.D.; MUZNY, D.M.; NAZARETH, L.V.; PATIL, S.; LUCAS, A.S.; FU, Q.; KENT, M.P.; MOORE, S.S.; VEGA, R.; ABBEY, C.A.; GAO, C.; GILL, C.A.; GREEN, R.D.; MATUKUMALLI, L.K.; MATUKUMALLI, A.; MCWILLIAM, S.; SCHNABEL, R.D.; SCLEP, G.; AJMONE-MARSAN, P.; BRYC, K.; BUSTAMANTE, C.D.; CHOI, J.; GAO, H.; GREFENSTETTE, J.J.; MURDOCH, B.; STELLA, A.; VILLA-ANGULO, R.; WRIGHT, M.; AERTS, J.; JANN, O.; NEGRINI, R.; SONSTEGARD, T.S.; WILLIAMS, J.L.; TAYLOR, J.F.; VILLA-ANGULO, R.; GODDARD, M.E.; HAYES, B.J.; BARENDSE, W.; BRADLEY, D.G.; BOETTCHER, P.J.; BUSTAMANTE, C.D.; DA SILVA, M.B.; LAU, L.P.L.; LIU, G.E.; LYNN, D.J.; PANZITTA, F.; SCLEP, G.; WRIGHT, M.; DODDS, K.G. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. **Science,** v. 324, n. 5926, p. 528-532, 2009.

GODDARD, M. Mapping genes for quantitative traits using linkage disequilibrium. **Genetics Selection Evolution,** v. 23, n. S1, p. 131-134, 1991.

GODDARD, M. Genomic selection: prediction of accuracy and maximisation of long term response. **Genetica,** v. 136, n. 2, p. 245-257, 2009.

GODDARD, M.; HAYES, B. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nature Review Genetics,** v. 10, p. 381 - 391, 2009.

GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics,** v. 124, n. 6, p. 323-30, 2007.

GRIFONE, R.; DEMIGNON, J.; GIORDANI, J.; NIRO, C.; SOUIL, E.; BERTIN, F.; LACLEF, C.; XU, P.-X.; MAIRE, P. Eya1 and Eya2 proteins are required for hypaxial somitic myogenesis in the mouse embryo. **Developmental Biology,** v. 302, n. 2, p. 602-616, 2007.

GRION, A.L.; MERCADANTE, M.E.Z.; CYRILLO, J.N.S.G.; BONILHA, S.F.M.; MAGNANI, E.; BRANCO, R.H. Selection for feed efficiency traits and correlated genetic responses in feed intake and weight gain of Nellore cattle. **Journal of Animal Science,** v. 92, n. 3, p. 955-965, 2014.

GROSSI, D.A.; FRIZZAS, O.G.; PAZ, C.C.P.; BEZERRA, L.A.F.; LÔBO, R.B.; OLIVEIRA, J.A.; MUNARI, D.P. Genetic associations between accumulated productivity, and reproductive and growth traits in Nelore cattle. **Livestock Science,** v. 117, n. 2–3, p. 139-146, 2008.

HABIER, D.; FERNANDO, R.; KIZILKAYA, K.; GARRICK, D. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics,** v. 12, p. 186-198, 2011.

HABIER, D.; FERNANDO, R.L.; DEKKERS, J.C.M. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. **Genetics,** v. 177, n. 4, p. 2389-2397, 2007.

HABIER, D.; TETENS, J.; SEEFRIED, F.-R.; LICHTNER, P.; THALLER, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics Selection Evolution,** v. 42, n. 1, p. 5-17, 2010.

HASTINGS, W.K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika,** v. 57, n. 1, p. 97-109, 1970.

HAYES, B.; BOWMAN, P.; CHAMBERLAIN, A.; GODDARD, M. Invited review: Genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science,** v. 92, p. 433-443, 2009.

HAYES, B.J.; PRYCE, J.; CHAMBERLAIN, A.J.; BOWMAN, P.J.; GODDARD, M.E. Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. **PLoS Genetics,** v. 6, n. 9, p. e1001139, 2010.

HEMMI, K.; MA, D.; MIURA, Y.; KAWAGUCHI, M.; SASAHARA, M.; HASHIMOTO-TAMAOKI, T.; TAMAOKI, T.; SAKATA, N.; TSUCHIYA, K. A Homeodomain-Zinc Finger Protein, ZFHX4, Is Expressed in Neuronal Differentiation Manner and Suppressed in Muscle Differentiation Manner. **Biological and Pharmaceutical Bulletin,** v. 29, n. 9, p. 1830-1835, 2006.

HENDERSON, C.R. Best linear unbiased estimation and prediction under a selection model. **Biometrics,** v. 31, n. 2, p. 423-47, 1975.

HILL, W.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics,** v. 38, p. 226 - 231, 1968.

HOSACK, D.A.; DENNIS, G., JR.; SHERMAN, B.T.; LANE, H.C.; LEMPICKI, R.A. Identifying biological themes within lists of genes with EASE. **Genome Biology,** v. 4, n. 10, p. R70, 2003.

HUANG, D.W.; SHERMAN, B.T.; TAN, Q.; COLLINS, J.R.; ALVORD, W.G.; ROAYAEI, J.; STEPHENS, R.; BASELER, M.W.; LANE, H.C.; LEMPICKI, R.A. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. **Genome Biology,** v. 8, n. 9, p. R183, 2007.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Anuário Estatístico do Brasil**. 2013.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Initial sequencing and analysis of the human genome. **Nature,** v. 409, n. 6822, p. 860-921, 2001.

JIANG, L.; LIU, J.; SUN, D.; MA, P.; DING, X.; YU, Y.; ZHANG, Q. Genome wide association studies for milk production traits in Chinese Holstein population. **PLoS ONE,** v. 5, n. 10, p. e13661, 2010.

JIMÉNEZ-MONTERO, J.A.; GONZÁLEZ-RECIO, O.; ALENDA, R. Genotyping strategies for genomic selection in small dairy cattle populations. **Animal,** v. 6, n. 08, p. 1216-1224, 2012.

KHATKAR, M.; NICHOLAS, F.; COLLINS, A.; ZENGER, K.; CAVANAGH, J.; BARRIS, W.; SCHNABEL, R.; TAYLOR, J.; RAADSMA, H. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. **BMC Genomics,** v. 9, n. 1, p. 187-205, 2008.

KIM, Y.; RYU, J.; WOO, J.; KIM, J.B.; KIM, C.Y.; LEE, C. Genome-wide association study reveals five nucleotide sequence variants for carcass traits in beef cattle. **Animal Genetics,** v. 42, n. 4, p. 361-365, 2011.

KNEELAND, J.; LI, C.; BASARAB, J.; SNELLING, W.M.; BENKEL, B.; MURDOCH, B.; HANSEN, C.; MOORE, S.S. Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of. **Journal of Animal Science,** v. 82, n. 12, p. 3405-3414, 2004.

LANDE, R.; THOMPSON, R. Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. **Genetics,** v. 124, n. 3, p. 743-756, 1990.

LAUREANO, M.M.M.; BOLIGON, A.A.; COSTA, R.B.; FORNI, S.; SEVERO, J.L.P.; ALBUQUERQUE, L.G. Estimativas de herdabilidade e tendências genéticas para características de crescimento e reprodutivas em bovinos da raça Nelore: Estimates of heritability and genetic trends for growth and reproduction traits in Nelore cattle. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia,** v. 63, p. 143-152, 2011.

LAURIE, C.C.; DOHENY, K.F.; MIREL, D.B.; PUGH, E.W.; BIERUT, L.J.; BHANGALE, T.; BOEHM, F.; CAPORASO, N.E.; CORNELIS, M.C.; EDENBERG, H.J.; GABRIEL, S.B.; HARRIS, E.L.; HU, F.B.; JACOBS, K.B.; KRAFT, P.; LANDI, M.T.; LUMLEY, T.; MANOLIO, T.A.; MCHUGH, C.; PAINTER, I.; PASCHALL, J.; RICE, J.P.; RICE, K.M.; ZHENG, X.; WEIR, B.S.; INVESTIGATORS, F.T.G. Quality control and quality assurance in genotypic data for genome-wide association studies. **Genetic Epidemiology,** v. 34, n. 6, p. 591-602, 2010.

LEE, S.; VAN DER WERF, J.H.J.; KIM, N.; LEE, S.; GONDRO, C.; PARK, E.; OH, S.; GIBSON, J.P.; THOMPSON, J.M. QTL and gene expression analyses identify genes affecting carcass weight and marbling on BTA14 in Hanwoo (Korean Cattle). **Mammalian Genome,** v. 22, n. 9, p. 589-601, 2011.

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.

LEGARRA, A.; ROBERT-GRANIÉ, C.; MANFREDI, E.; ELSEN, J.-M. Performance of Genomic Selection in Mice. **Genetics,** v. 180, n. 1, p. 611-618, 2008.

LEWONTIN, R.C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. **Genetics,** v. 49, n. 1, p. 49-67, 1964.

LIU, G.; JENNEN, D.G.; THOLEN, E.; JUENGST, H.; KLEINWACHTER, T.; HOLKER, M.; TESFAYE, D.; UN, G.; SCHREINEMACHERS, H.J.; MURANI, E.; PONSUKSILI, S.; KIM, J.J.; SCHELLANDER, K.; WIMMERS, K. A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. **Animal Genetics,** v. 38, n. 3, p. 241-52, 2007.

LU, D.; MILLER, S.; SARGOLZAEI, M.; KELLY, M.; VOORT, G.V.; CALDWELL, T.; WANG, Z.; PLASTOW, G.; MOORE, S. Genome-wide association analyses for growth and feed efficiency traits in beef cattle. **Journal of Animal Science,** v. 91, n. 8, p. 3612-3633, 2013.

LUAN, T.; WOOLLIAMS, J.; ODEGARD, J.; DOLEZAL, M.; ROMAN-PONCE, S.; BAGNATO, A.; MEUWISSEN, T. The importance of identity-by-state information for the accuracy of genomic selection. **Genetics Selection Evolution,** v. 44, n. 1, p. 28-35, 2012.

LYNCH, M.; WALSH, J. Long-term response: Finite populations. In: LYNCH, M. e WALSH, J. (Ed.). **Genetics and Analysis of Quantitative Traits**: Sunderland, MA: Sinauer Associates, 1998.

MACHADO, P.F.A.; AQUINO, L.H.; GONÇALVES, T.M. Estimativas de parâmetros genéticos e critérios de seleção em características ponderais de bovinos Nelore. **Ciência e Agrotecnia,** v. 23, n. 1, p. 197-204, 1999.

MAI, M.D.; SAHANA, G.; CHRISTIANSEN, F.B.; GULDBRANDTSEN, B. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. **Journal of Animal Science,** v. 88, n. 11, p. 3522-8, 2010.

MALHADO, C.H.M.; MARTINS FILHO, R.; LÔBO, R.N.B.; FACÓ, O.; AZEVEDO, D.M.M.R.; SOUZA, J.C.D.; OLIVEIRA, S.M.P. Tendências genéticas para características relacionadas à velocidade de crescimento em bovinos Nelore na região nordeste do Brasil. **Revista Brasileira de Zootecnia,** v. 34, p. 60-65, 2005.

MARQUES, E.G.; MAGNABOSCO, C.U.; LOPES, F.B.; SILVA, M.C. Estimate the Genetic Parameters of Growth Characteristics, Carcass and Scrotal Perimeter in Nellore Evaluated in Weight Gain Performance Tests in Feedlot. **Bioscience Journal,** v. 29, n. 1, p. 159-167, 2013.

MARTHA JR, G.B.; ALVES, E.; CONTINI, E. Land-saving approaches and beef production growth in Brazil. **Agricultural Systems,** v. 110, p. 173-177, 2012.

MATUKUMALLI, L.K.; LAWLEY, C.T.; SCHNABEL, R.D.; TAYLOR, J.F.; ALLAN, M.F.; HEATON, M.P.; O'CONNELL, J.; MOORE, S.S.; SMITH, T.P.L.; SONSTEGARD, T.S.; VAN TASSELL, C.P. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. **PLoS ONE,** v. 4, n. 4, 2009.

MCVEAN, G. A Genealogical Interpretation of Principal Components Analysis. **PLoS Genetics,** v. 5, n. 10, p. e1000686, 2009.

MERCADANTE, M.E.Z.; PACKER, I.U.; RAZOOK, A.G.; CYRILLO, J.N.S.G.; FIGUEIREDO, L.A. Direct and correlated responses to selection for yearling weight onreproductive performance of Nelore cows. **Journal of Animal Science,** v. 81, n. 2, p. 376-384, 2003.

MEUWISSEN, T.H.E.; GODDARD, M.E. The use of marker haplotypes in animal breeding schemes. **Genetics Selection Evolution,** v. 28, n. 2, p. 161-176, 1996.

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.

MILLEN, D.D.; PACHECO, R.D.L.; MEYER, P.M.; RODRIGUES, P.H.M.; DE BENI ARRIGONI, M. Current outlook and future perspectives of beef production in Brazil. **Animal Frontiers,** v. 1, n. 2, p. 46-52, 2011.

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.

MIZOSHITA, K.; WATANABE, T.; HAYASHI, H.; KUBOTA, C.; YAMAKUCHI, H.; TODOROKI, J.; SUGIMOTO, Y. Quantitative trait loci analysis for growth and carcass traits in a half-sib family of purebred Japanese Black (Wagyu) cattle. **Journal of Animal Science,** v. 82, p. 3415 - 3420, 2004.

MOKRY, F.; HIGA, R.; DE ALVARENGA MUDADU, M.; OLIVEIRA DE LIMA, A.; MEIRELLES, S.L.; BARBOSA DA SILVA, M.V.; CARDOSO, F.; MORGADO DE OLIVEIRA, M.; URBINATI, I.; MEO NICIURA, S.; TULLIO, R.; MELLO DE ALENCAR, M.; CORREIA DE ALMEIDA REGITANO, L. Genome-wide association study for backfat thickness in Canchim beef cattle using Random Forest approach. **BMC Genetics,** v. 14, n. 1, p. 47-58, 2013.

MUNTONI, F.; BROWN, S.; SEWRY, C.; PATEL, K. Muscle development genes: their relevance in neuromuscular disorders. **Neuromuscular Disorders,** v. 12, n. 5, p. 438-446, 2002.

NAKAZAWA, M. **Functions for medical statistics book with some demographic data**: R package version 0.4.1, 2013.

NEALE, B.M.; PURCELL, S. The Positives, Protocols, and Perils of Genome-Wide Association. **American Journal of Medical Genetics Part B-Neuropsychiatric Genetics,** v. 147B, n. 7, p. 1288-1294, 2008.

NEVES, H.H.; CARVALHEIRO, R.; O'BRIEN, A.M.; UTSUNOMIYA, Y.T.; DO CARMO, A.S.; SCHENKEL, F.S.; SOLKNER, J.; MCEWAN, J.C.; VAN TASSELL, C.P.; COLE, J.B.; DA SILVA, M.V.; QUEIROZ, S.A.; SONSTEGARD, T.S.; GARCIA, J.F. Accuracy of genomic predictions in Bos indicus (Nellore) cattle. **Genetics Selection Evolution,** v. 46, p. 17, 2014.

OLIVEIRA, P.S.N.; CESAR, A.S.M.; NASCIMENTO, M.L.; CHAVES, A.S.; TIZIOTO, P.C.; TULLIO, R.R.; LANNA, D.P.D.; ROSA, A.N.; SONSTEGARD, T.S.; MOURAO, G.B.; REECY, J.M.; GARRICK, D.J.; MUDADU, M.A.; COUTINHO, L.L.; REGITANO, L.C.A. Identification of genomic regions associated with feed efficiency in Nelore cattle. **BMC Genetics,** v. 15, 2014.

PANETO, J.C.D.C.; LEMOS, D.C.; BEZERRA, L.A.F.; MARTINS FILHO, R.; LÔBO, R.B. Estudo de Características Quantitativas de Crescimento dos 120 aos 550 Dias de Idade em Gado Nelore. **Revista Brasileira de Zootecnia,** v. 31, p. 668-674, 2002.

PATTERSON, N.; PRICE, A.L.; REICH, D. Population structure and eigenanalysis. **PLoS Genetics,** v. 2, n. 12, p. e190, 2006.

PENAGARICANO, F.; WEIGEL, K.A.; KHATIB, H. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. **Animal Genetics,** v. 43, p. 65-71, 2012.

PÉREZ-MONTARELO, D.; HUDSON, N.J.; FERNÁNDEZ, A.I.; RAMAYO-CALDAS, Y.; DALRYMPLE, B.P.; REVERTER, A. Porcine Tissue-Specific Regulatory Networks Derived from Meta-Analysis of the Transcriptome. **PLoS ONE,** v. 7, n. 9, p. e46159, 2012.

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, n. 2, p. 605-612, 2012.

PIECHA, D.; MURATOGLU, S.; MÖRGELIN, M.; HAUSER, N.; STUDER, D.; KISS, I.; PAULSSON, M.; DEÁK, F. Matrilin-2, a Large, Oligomeric Matrix Protein, Is Expressed by a Great Variety of Cells and Forms Fibrillar Networks. **Journal of Biological Chemistry,** v. 274, n. 19, p. 13353-13361, 1999.

PORTELA, V.M.; MACHADO, M.; BURATINI, 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,** v. 83, n. 3, p. 339-346, 2010.

PRICE, A.L.; PATTERSON, N.J.; PLENGE, R.M.; WEINBLATT, M.E.; SHADICK, N.A.; REICH, D. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics,** v. 38, n. 8, p. 904-909, 2006.

PRICE, A.L.; ZAITLEN, N.A.; REICH, D.; PATTERSON, N. New approaches to population stratification in genome-wide association studies. **Nature Reviews Genetics,** v. 11, n. 7, p. 459-463, 2010.

PRYCE, J.E.; BOLORMAA, S.; CHAMBERLAIN, A.J.; BOWMAN, P.J.; SAVIN, K.; GODDARD, M.E.; HAYES, B.J. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. **Journal of Dairy Science,** v. 93, n. 7, p. 3331-45, 2010.

PUIG-OLIVERAS, A.; BALLESTER, M.; COROMINAS, J.; REVILLA, M.; ESTELLÉ, J.; FERNÁNDEZ, A.I.; RAMAYO-CALDAS, Y.; FOLCH, J.M. A Co-Association Network Analysis of the Genetic Determination of Pig Conformation, Growth and Fatness. **PLoS ONE,** v. 9, n. 12, p. e114862, 2014.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2013.

RAICHE, G. **nFactors: an R package for parallel analysis and non graphical solutions to the Cattell scree test**: R package version 2.3.3 2010.

RAMAYO-CALDAS, Y.; BALLESTER, M.; FORTES, M.; ESTEVE-CODINA, A.; CASTELLO, A.; NOGUERA, J.; FERNANDEZ, A.; PEREZ-ENCISO, M.; REVERTER, A.; FOLCH, J. From SNP co-association to RNA co-expression: Novel insights into gene networks for intramuscular fatty acid composition in porcine. **BMC Genomics,** v. 15, n. 1, p. 232-247, 2014.

ROBERTS, A.J.; KON, T.; KNIGHT, P.J.; SUTOH, K.; BURGESS, S.A. Functions and mechanics of dynein motor proteins. **Nature Reviews Molecular Cell Biology,** v. 14, n. 11, p. 713-726, 2013.

ROLF, M.M.; GARRICK, D.J.; FOUNTAIN, T.; RAMEY, H.R.; WEABER, R.L.; DECKER, J.E.; POLLAK, E.J.; SCHNABEL, R.D.; TAYLOR, J.F. Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. **Genetics Selection Evolution,** v. 47, n. 1, p. 23-37, 2015.

SAATCHI, M.; BEEVER, J.E.; DECKER, J.E.; FAULKNER, D.B.; FREETLY, H.C.; HANSEN, S.L.; YAMPARA-IQUISE, H.; JOHNSON, K.A.; KACHMAN, S.D.; KERLEY, M.S.; KIM, J.; LOY, D.D.; MARQUES, E.; NEIBERGS, H.L.; POLLAK, E.J.; SCHNABEL, R.D.; SEABURY, C.M.; SHIKE, D.W.; SNELLING, W.M.; SPANGLER, M.L.; WEABER, R.L.; GARRICK, D.J.; TAYLOR, J.F. QTLs associated with dry matter intake, metabolic mid-test weight, growth and feed efficiency have little overlap across 4 beef cattle studies. **BMC Genomics,** v. 15, p. 1004-1018, 2014.

SAATCHI, M.; MCCLURE, M.C.; MCKAY, S.D.; ROLF, M.M.; KIM, J.; DECKER, J.E.; TAXIS, T.M.; CHAPPLE, R.H.; RAMEY, H.R.; NORTHCUTT, S.L.; BAUCK, S.; WOODWARD, B.; DEKKERS, J.C.; FERNANDO, R.L.; SCHNABEL, R.D.; GARRICK, D.J.; TAYLOR, J.F. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. **Genetics Selection Evolution,** v. 43, n. 1, p. 40-56, 2011.

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, n. 1, p. 479-86, 2011.

SANTANA, M.; UTSUNOMIYA, Y.; NEVES, H.; GOMES, R.; GARCIA, J.; FUKUMASU, H.; SILVA, S.; OLIVEIRA JUNIOR, G.; ALEXANDRE, P.; LEME, P.; BRASSALOTI, R.; COUTINHO, L.; LOPES, T.; MEIRELLES, F.; ELER, J.; FERRAZ, J. Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. **BMC Genetics,** v. 15, n. 1, p. 21-29, 2014a.

SANTANA, M.H.A.; UTSUNOMIYA, Y.T.; NEVES, H.H.R.; GOMES, R.C.; GARCIA, J.F.; FUKUMASU, H.; SILVA, S.L.; LEME, P.R.; COUTINHO, L.L.; ELER, J.P.; FERRAZ, J.B.S. Genome-wide association study for feedlot average daily gain in Nellore cattle (Bos indicus). **Journal of Animal Breeding and Genetics,** v. 131, n. 3, p. 210-216, 2014b.

SARMENTO, J.L.R.; PIMENTA FILHO, E.C.; RIBEIRO, M.N.; MARTINS FILHO, R. Efeitos ambientais e genéticos sobre o ganho em peso diário de bovinos Nelore no Estado da Paraíba. **Revista Brasileira de Zootecnia,** v. 32, p. 325-330, 2003.

SCHAEFFER, L.R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics,** v. 123, n. 4, p. 218-223, 2006.

SHAPIRO, S.S.; WILK, M.B. An Analysis of Variance Test for Normality (Complete Samples). **Biometrika,** v. 52, n. 3/4, p. 591-611, 1965.

SMITH, B.J. boa: An R package for MCMC output convergence assessment and posterior inference. **Journal of Statistical Software,** v. 21, n. 11, p. 1-37, 2007.

SNELLING, W.M.; ALLAN, M.F.; KEELE, J.W.; KUEHN, L.A.; MCDANELD, T.; SMITH, T.P.L.; SONSTEGARD, T.S.; THALLMAN, R.M.; BENNETT, G.L. Genome-wide association study of growth in crossbred beef cattle. **Journal of Animal Science,** v. 88, n. 3, p. 837-848, 2010.

SOLBERG, T.R.; SONESSON, A.K.; WOOLLIAMS, J.A.; MEUWISSEN, T.H. Reducing dimensionality for prediction of genome-wide breeding values. **Genetics Selection Evolution,** v. 41, p. 29-37, 2009.

SOUZA, J.C.; SILVA, L.O.C.; GONDO, A.; FREITAS, J.A.; MALHADO, C.H.M.; FILHO, P.B.F.; SERENO, J.R.B.; WEABER, R.L.; LAMBERSON, W.R. Parâmetros e tendência genética de peso de bovinos criados á pasto no Brasil. **Archivos de Zootecnia,** v. 60, p. 457-465, 2011.

SPIEGELHALTER, D.J.; BEST, N.G.; CARLIN, B.P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology),** v. 64, n. 4, p. 583-639, 2002.

THE BOVINE GENOME SEQUENCING ANALYSIS CONSORTIUM; ELSIK, C.G.; TELLAM, R.L.; WORLEY, K.C. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. **Science,** v. 324, n. 5926, p. 522-528, 2009.

TIZIOTO, P.; TAYLOR, J.; DECKER, J.; GROMBONI, C.; MUDADU, M.; SCHNABEL, R.; COUTINHO, L.; MOURAO, G.; OLIVEIRA, P.; SOUZA, M.; REECY, J.; NASSU, R.; BRESSANI, F.; THOLON, P.; SONSTEGARD, T.; ALENCAR, M.; TULLIO, R.; NOGUEIRA, A.; REGITANO, L. Detection of quantitative trait loci for mineral content of Nelore longissimus dorsi muscle. **Genetics Selection Evolution,** v. 47, n. 1, p. 15-24, 2015a.

TIZIOTO, P.C.; COUTINHO, L.L.; DECKER, J.E.; SCHNABEL, R.D.; ROSA, K.O.; OLIVEIRA, P.S.; SOUZA, M.M.; MOURAO, G.B.; TULLIO, R.R.; CHAVES, A.S.; LANNA, D.P.; ZERLOTINI-NETO, A.; MUDADU, M.A.; TAYLOR, J.F.; REGITANO, L.C. Global liver gene expression differences in Nelore steers with divergent residual feed intake phenotypes. **BMC Genomics,** v. 16, p. 242-256, 2015b.

TIZIOTO, P.C.; DECKER, J.E.; TAYLOR, J.F.; SCHNABEL, R.D.; MUDADU, M.A.; SILVA, F.L.; MOURÃO, G.B.; COUTINHO, L.L.; THOLON, P.; SONSTEGARD, T.S.; ROSA, A.N.; ALENCAR, M.M.; TULLIO, R.R.; MEDEIROS, S.R.; NASSU, R.T.; FEIJÓ, G.L.D.; SILVA, L.O.C.; TORRES, R.A.; SIQUEIRA, F.; HIGA, R.H.; REGITANO, L.C.A. Genome scan for meat quality traits in Nelore beef cattle. **Physiological Genomics,** v. 45, n. 21, p. 1012-1020, 2013.

UBACH, J.; LAO, Y.; FERNANDEZ, I.; ARAC, D.; SÜDHOF, T.C.; RIZO, J. The C2B Domain of Synaptotagmin I Is a Ca2+-Binding Module. **Biochemistry,** v. 40, n. 20, p. 5854-5860, 2001.

USAI, M.; GODDARD, M.; HAYES, B. LASSO with cross-validation for genomic selection. **Genetics Research,** v. 91, p. 427 - 436, 2009.

UTSUNOMIYA, Y.; DO CARMO, A.; CARVALHEIRO, R.; NEVES, H.; MATOS, M.; ZAVAREZ, L.; PEREZ O'BRIEN, A.; SOLKNER, J.; MCEWAN, J.; COLE, J.; VAN TASSELL, C.; SCHENKEL, F.; DA SILVA, M.; PORTO NETO, L.; SONSTEGARD, T.; GARCIA, J. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. **BMC Genetics,** v. 14, n. 1, p. 52, 2013.

VAN DER WERF, J. Genomic Selection in Animal Breeding Programs. In: GONDRO, C.;WERF, J.V.D. e HAYES, B. (Ed.). **Genome-Wide Association Studies and Genomic Prediction**. 1st: Humana Press, 2013.

VANRADEN, P. Efficient methods to compute genomic prediction. **Journal of Dairy Science,** v. 91, p. 4414 - 4423, 2008.

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.

VENABLES, W.N.; RIPLEY, B.D. **Modern Applied Statistics with S.** Fourth Edition. New York: Springer, 2002.

VENTER, J.C.; ADAMS, M.D.; MYERS, E.W.; LI, P.W.; MURAL, R.J.; SUTTON, G.G.; SMITH, H.O.; YANDELL, M.; EVANS, C.A.; HOLT, R.A.; GOCAYNE, J.D.; AMANATIDES, P.; BALLEW, R.M.; HUSON, D.H.; WORTMAN, J.R.; ZHANG, Q.; KODIRA, C.D.; ZHENG, X.H.; CHEN, L.; SKUPSKI, M.; SUBRAMANIAN, G.; THOMAS, P.D.; ZHANG, J.; GABOR MIKLOS, G.L.; NELSON, C.; BRODER, S.; CLARK, A.G.; NADEAU, J.; MCKUSICK, V.A.; ZINDER, N.; LEVINE, A.J.; ROBERTS, R.J.; SIMON, M.; SLAYMAN, C.; HUNKAPILLER, M.; BOLANOS, R.; DELCHER, A.; DEW, I.; FASULO, D.; FLANIGAN, M.; FLOREA, L.; HALPERN, A.; HANNENHALLI, S.; KRAVITZ, S.; LEVY, S.; MOBARRY, C.; REINERT, K.; REMINGTON, K.; ABU-THREIDEH, J.; BEASLEY, E.; BIDDICK, K.; BONAZZI, V.; BRANDON, R.; CARGILL, M.; CHANDRAMOULISWARAN, I.; CHARLAB, R.; CHATURVEDI, K.; DENG, Z.; FRANCESCO, V.D.; DUNN, P.; EILBECK, K.;

EVANGELISTA, C.; GABRIELIAN, A.E.; GAN, W.; GE, W.; GONG, F.; GU, Z.; GUAN, P.; HEIMAN, T.J.; HIGGINS, M.E.; JI, R.-R.; KE, Z.; KETCHUM, K.A.; LAI, Z.; LEI, Y.; LI, Z.; LI, J.; LIANG, Y.; LIN, X.; LU, F.; MERKULOV, G.V.; MILSHINA, N.; MOORE, H.M.; NAIK, A.K.; NARAYAN, V.A.; NEELAM, B.; NUSSKERN, D.; RUSCH, D.B.; SALZBERG, S.; SHAO, W.; SHUE, B.; SUN, J.; WANG, Z.Y.; WANG, A.; WANG, X.; WANG, J.; WEI, M.-H.; WIDES, R.; XIAO, C.; YAN, C.; YAO, A.; YE, J.; ZHAN, M.; ZHANG, W.; ZHANG, H.; ZHAO, Q.; ZHENG, L.; ZHONG, F.; ZHONG, W.; ZHU, S.C.; ZHAO, S.; GILBERT, D.; BAUMHUETER, S.; SPIER, G.; CARTER, C.; CRAVCHIK, A.; WOODAGE, T.; ALI, F.; AN, H.; AWE, A.; BALDWIN, D.; BADEN, H.; BARNSTEAD, M.; BARROW, I.; BEESON, K.; BUSAM, D.; CARVER, A.; CENTER, A.; CHENG, M.L.; CURRY, L.; DANAHER, S.; DAVENPORT, L.; DESILETS, R.; DIETZ, S.; DODSON, K.; DOUP, L.; FERRIERA, S.; GARG, N.; GLUECKSMANN, A.; HART, B.; HAYNES, J.; HAYNES, C.; HEINER, C.; HLADUN, S.; HOSTIN, D.; HOUCK, J.; HOWLAND, T.; IBEGWAM, C.; JOHNSON, J.; KALUSH, F.; KLINE, L.; KODURU, S.; LOVE, A.; MANN, F.; MAY, D.; MCCAWLEY, S.; MCINTOSH, T.; MCMULLEN, I.; MOY, M.; MOY, L.; MURPHY, B.; NELSON, K.; PFANNKOCH, C.; PRATTS, E.; PURI, V.; QURESHI, H.; REARDON, M.; RODRIGUEZ, R.; ROGERS, Y.-H.; ROMBLAD, D.; RUHFEL, B.; SCOTT, R.; SITTER, C.; SMALLWOOD, M.; STEWART, E.; STRONG, R.; SUH, E.; THOMAS, R.; TINT, N.N.; TSE, S.; VECH, C.; WANG, G.; WETTER, J.; WILLIAMS, S.; WILLIAMS, M.; WINDSOR, S.; WINN-DEEN, E.; WOLFE, K.; ZAVERI, J.; ZAVERI, K.; ABRIL, J.F.; GUIGÓ, R.; CAMPBELL, M.J.; SJOLANDER, K.V.; KARLAK, B.; KEJARIWAL, A.; MI, H.; LAZAREVA, B.; HATTON, T.; NARECHANIA, A.; DIEMER, K.; MURUGANUJAN, A.; GUO, N.; SATO, S.; BAFNA, V.; ISTRAIL, S.; LIPPERT, R.; SCHWARTZ, R.; WALENZ, B.; YOOSEPH, S.; ALLEN, D.; BASU, A.; BAXENDALE, J.; BLICK, L.; CAMINHA, M.; CARNES-STINE, J.; CAULK, P.; CHIANG, Y.-H.; COYNE, M.; DAHLKE, C.; MAYS, A.D.; DOMBROSKI, M.; DONNELLY, M.; ELY, D.; ESPARHAM, S.; FOSLER, C.; GIRE, H.; GLANOWSKI, S.; GLASSER, K.; GLODEK, A.; GOROKHOV, M.; GRAHAM, K.; GROPMAN, B.; HARRIS, M.; HEIL, J.; HENDERSON, S.; HOOVER, J.; JENNINGS, D.; JORDAN, C.; JORDAN, J.; KASHA, J.; KAGAN, L.; KRAFT, C.; LEVITSKY, A.; LEWIS, M.; LIU, X.; LOPEZ, J.; MA, D.; MAJOROS, W.; MCDANIEL, J.; MURPHY, S.; NEWMAN, M.; NGUYEN, T.; NGUYEN, N.; NODELL, M.; PAN, S.; PECK, J.; PETERSON, M.; ROWE, W.; SANDERS, R.; SCOTT, J.; SIMPSON, M.; SMITH, T.; SPRAGUE, A.; STOCKWELL, T.; TURNER, R.; VENTER, E.; WANG, M.; WEN, M.; WU, D.; WU, M.; XIA, A.; ZANDIEH, A.; ZHU, X. The Sequence of the Human Genome. **Science,** v. 291, n. 5507, p. 1304-1351, 2001.

VERBYLA, K.L.; HAYES, B.J.; BOWMAN, P.J.; GODDARD, M.E. Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. **Genetics Research,** v. 91, n. 05, p. 307-311, 2009.

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution,** v. 34, n. 3, p. 275-305, 2002.

VISSCHER, P.; HILL, W.; WRAY, N. Heritability in the genomics era-concepts and misconceptions. **Nature Review Genetics,** v. 9, p. 255 - 266, 2008.

WAHBA, G. **Spline models for observational data**. Philadelphia, Pa.: Society for Industrial and Applied Mathematics, 1990. xii, 169 p.

WELLER, J.I. Basic Concepts in Quantitative Genetics. In: WELLER, J.I. (Ed.). **Economic Aspects of Animal Breeding**. 1st. Padstow: Chapman & Hall, 1994.

WIENTJES, Y.C.; VEERKAMP, R.F.; CALUS, M.P. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. **Genetics,** v. 193, n. 2, p. 621-31, 2013.

WINKELMAN, A.M.; JOHNSON, D.L.; HARRIS, B.L. Application of genomic evaluation to dairy cattle in New Zealand. **Journal of Dairy Science,** v. 98, n. 1, p. 659-675, 2015.

WRAY, N.R.; YANG, J.; HAYES, B.J.; PRICE, A.L.; GODDARD, M.E.; VISSCHER, P.M. Pitfalls of predicting complex traits from SNPs. **Nature Reviews Genetics,** v. 14, n. 7, p. 507-515, 2013.

XU, S. Estimating Polygenic Effects Using Markers of the Entire Genome. **Genetics,** v. 163, n. 2, p. 789-801, 2003.

ZIEGLER, A. Genome-wide association studies: quality control and population-based measures. **Genetic Epidemiology,** v. 33, n. S1, p. S45-S50, 2009.

8. APPENDIX

Table 1A. Distribution of SNPs along autosomes after quality control for all datasets

Chromosome	LDi	HDi	TagSNPs	770k
1	964	3,928	5,085	34,667
2	800	3,302	4,361	28,470
3	738	3,046	3,956	25,991
4	775	2,989	4,029	25,327
5	672	2,949	3,537	23,888
6	770	2,935	3,387	26,880
7	685	2,828	3,551	24,000
8	701	2,845	3,594	25,017
9	671	2,657	3,464	23,472
10	656	2,585	3,583	21,372
11	698	2,703	3,677	23,361
12	555	2,213	2,755	17,721
13	550	2,140	2,693	16,641
14	547	2,140	2,739	19,670
15	543	2,193	2,830	18,308
16	499	2,100	2,728	17,666
17	479	1,913	2,522	16,115
18	404	1,750	2,163	14,284
19	396	1,709	2,340	12,860
20	461	1,914	2,471	15,666
21	442	1,882	2,414	15,988
22	429	1,645	2,030	12,873
23	368	1,481	2,001	11,378
24	413	1,655	2,115	13,490
25	308	1,185	1,723	8,621
26	357	1,374	1,891	11,426
27	311	1,227	1,779	9,506
28	318	1,275	1,532	9,845
29	353	1,382	1,983	10,284
Total	15,863	63,945	82,933	534,787

Table 2A. Average spacing (kb) between SNPs along autosomes after quality control for all datasets

Chromosome	LDi	HDi	TagSNPs	770k
1	164.14	40.28	31.11	4.56
2	168.49	40.39	30.09	4.61
3	160.77	38.32	29.63	4.52
4	154.34	38.58	28.91	4.59
5	168.83	39.03	31.07	4.76
6	153.70	38.69	31.50	4.31
7	156.48	37.63	29.34	4.46
8	154.66	37.68	29.34	4.36
9	150.08	37.70	28.95	4.38
10	150.08	37.88	27.69	4.55
11	148.64	37.68	27.90	4.44
12	151.55	38.20	29.56	4.71
13	144.80	36.95	28.72	4.64
14	144.12	36.77	28.30	4.27
15	146.70	36.86	28.05	4.46
16	147.64	36.58	28.01	4.43
17	142.99	35.96	27.79	4.40
18	143.37	34.63	28.00	4.38
19	141.80	34.26	26.81	4.55
20	141.62	35.15	27.46	4.38
21	143.34	35.31	27.75	4.32
22	133.99	34.33	28.08	4.40
23	129.62	33.03	26.53	4.31
24	135.88	34.42	27.79	4.37
25	128.51	33.36	26.08	4.39
26	129.89	33.57	27.08	4.32
27	127.78	32.95	26.55	4.35
28	128.18	32.85	28.03	4.34
29	129.85	33.40	26.49	4.44
Mean	145.58	36.29	28.37	4.45

Table 3A. Size of half-sib families and summary description of feedlot average daily gain (ADG), according to sire labels

Sire label	Half-sibs	ADG - minimum	ADG - mean	ADG - maximum
1	17	0.35	1.00	1.69
2	31	0.33	1.12	1.73
3	25	0.22	1.16	1.70
4	19	0.34	1.15	1.81
5	35	0.31	1.33	2.19
6	37	0.32	1.13	2.17
7	9	0.24	1.05	1.77
8	19	0.30	1.20	1.97
9	36	0.27	1.28	2.00
10	37	0.27	1.22	1.87
11	23	0.43	1.08	1.60
12	38	0.46	1.44	2.24
13	12	0.53	1.22	1.64
14	14	0.41	1.13	1.71
15	11	0.54	1.15	1.59
16	36	0.37	1.23	1.88
17	35	0.29	0.98	1.66
18	18	0.59	1.22	1.72
19	31	0.48	1.29	2.28
20	12	0.44	1.31	1.78
21	19	0.70	1.25	1.59
22	22	0.87	1.35	1.84
23	25	0.74	1.40	1.98
24	14	0.81	1.33	1.70
25	41	0.20	1.08	2.10
26	17	0.89	1.31	1.85
27	15	0.80	1.36	1.95
28	21	0.57	1.29	2.46
29	16	0.95	1.39	2.32
30	10	0.64	1.32	1.82
31	23	0.57	1.61	2.17

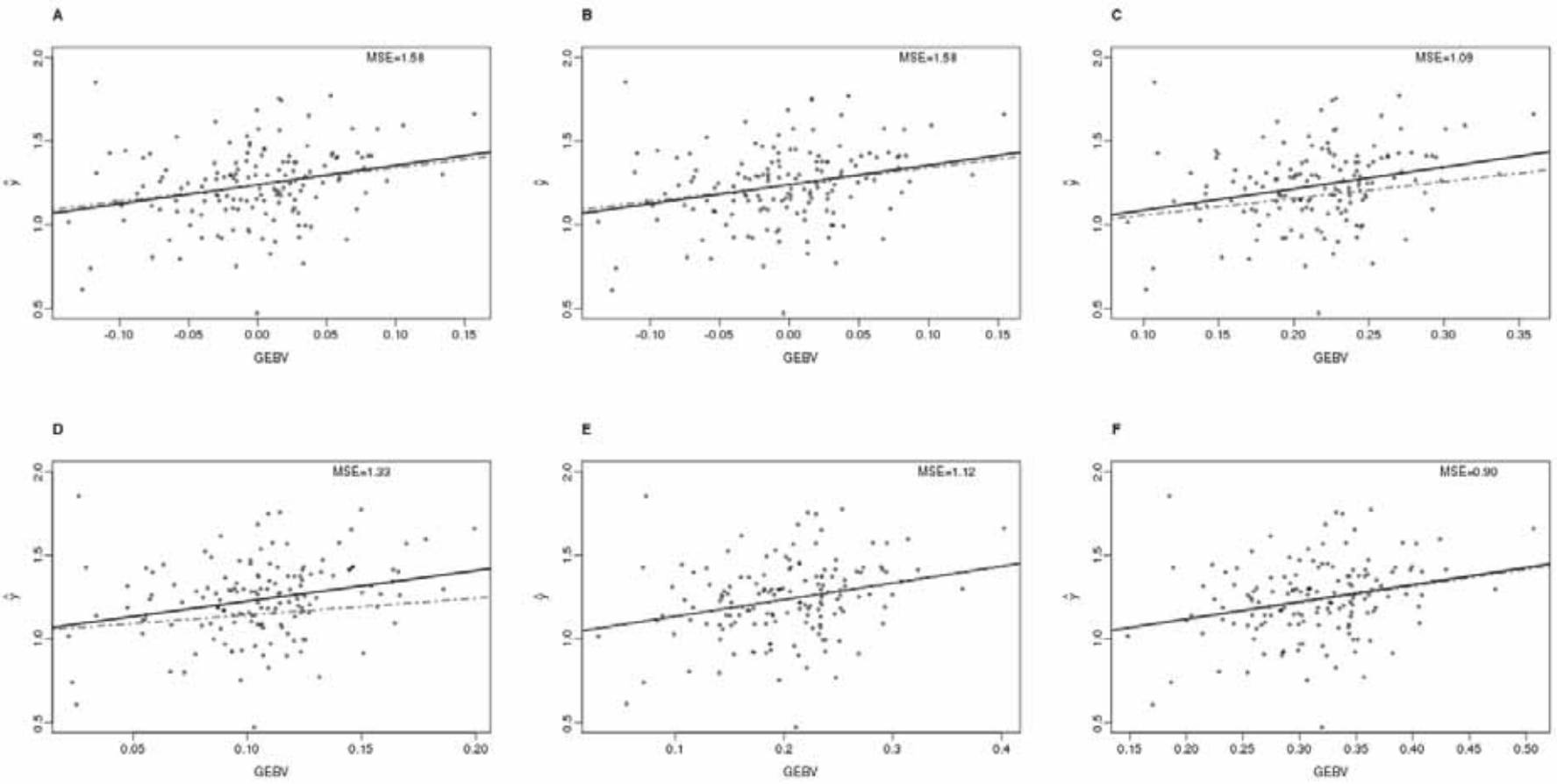

Figure 1A. Regression plots of adjusted phenotypes on GEBVs in 770k dataset. (A) BGBLUP, (C) BayesA, (E) BayesCπ. (B), (D) and (F) are the same models including PCs as covariates. Dashed grey and solid black lines represent observed and expected slopes, respectively. MSE: mean squared error.

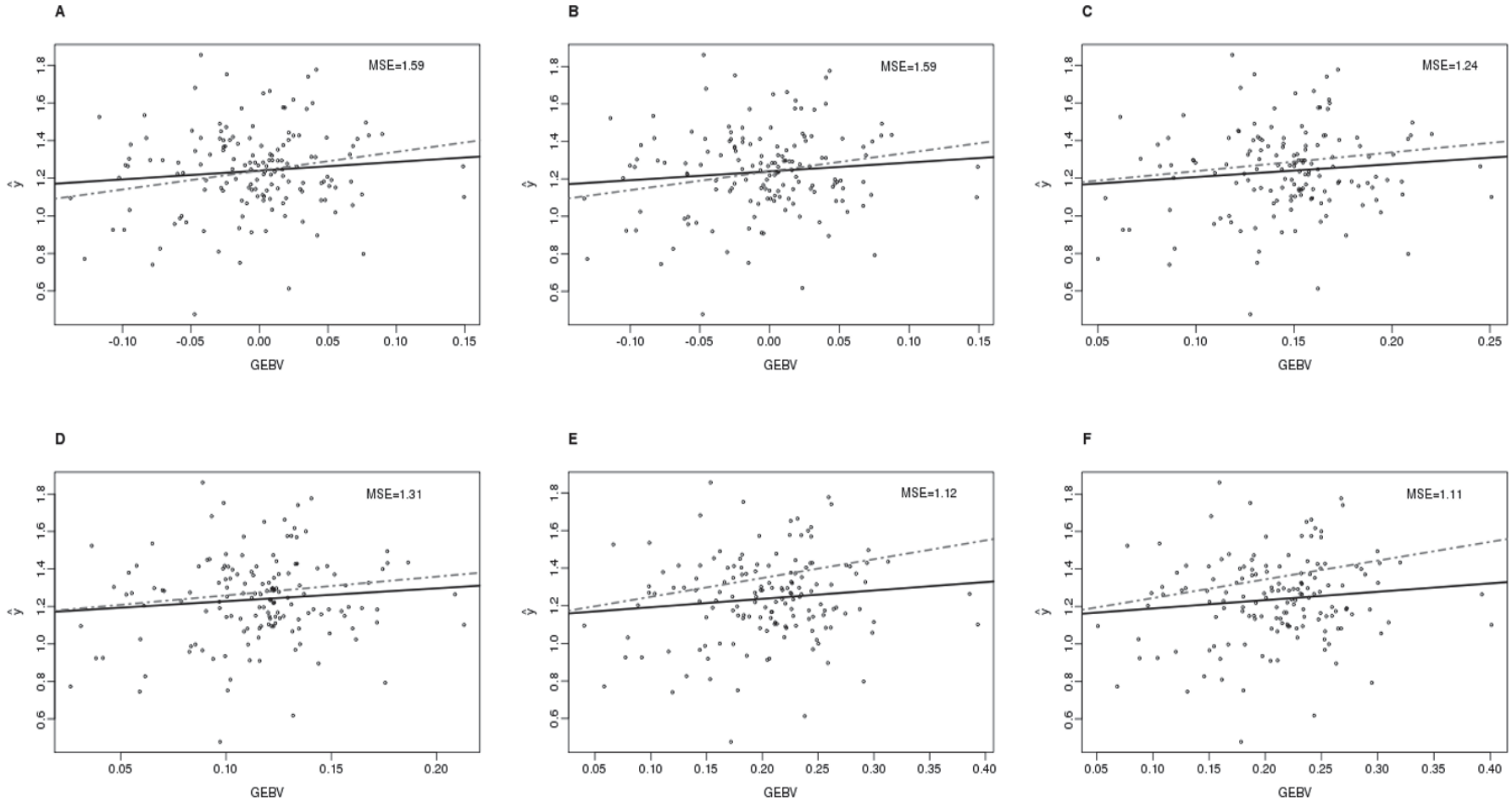

Figure 2A. Regression plots of adjusted phenotypes on GEBVs in TagSNPs dataset. (A) BGBLUP, (C) BayesA, (E) BayesCπ. (B), (D) and (F) are the same models including PCs as covariates. Dashed grey and solid black lines represent observed and expected slopes, respectively. MSE: mean squared error.

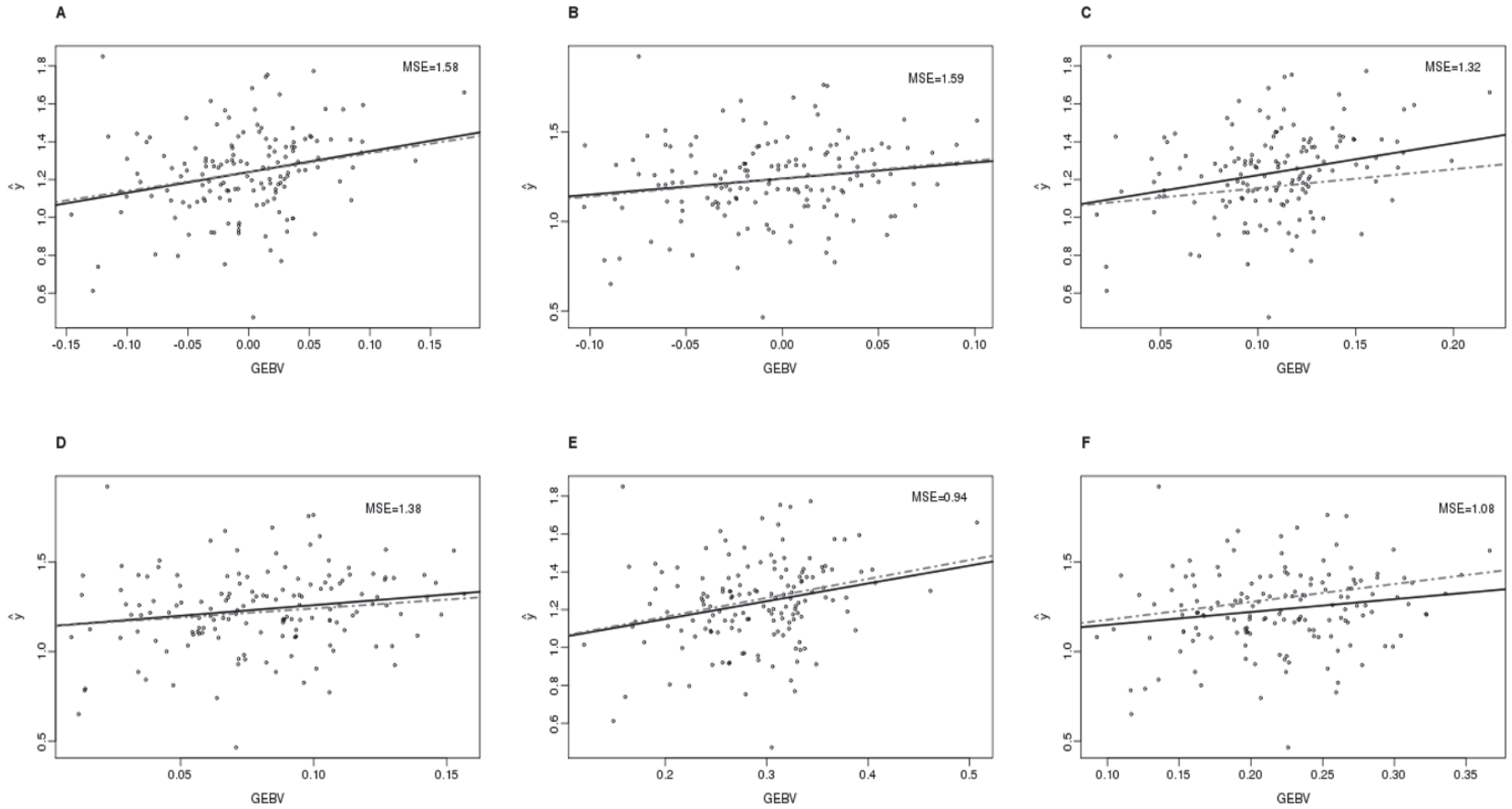

Figure 3A. Regression plots of adjusted phenotypes on GEBVs in HDi dataset. (A) BGBLUP, (C) BayesA, (E) BayesCπ. (B), (D) and (F) are the same models including PCs as covariates. Dashed grey and solid black lines represent observed and expected slopes, respectively. MSE: mean squared error.

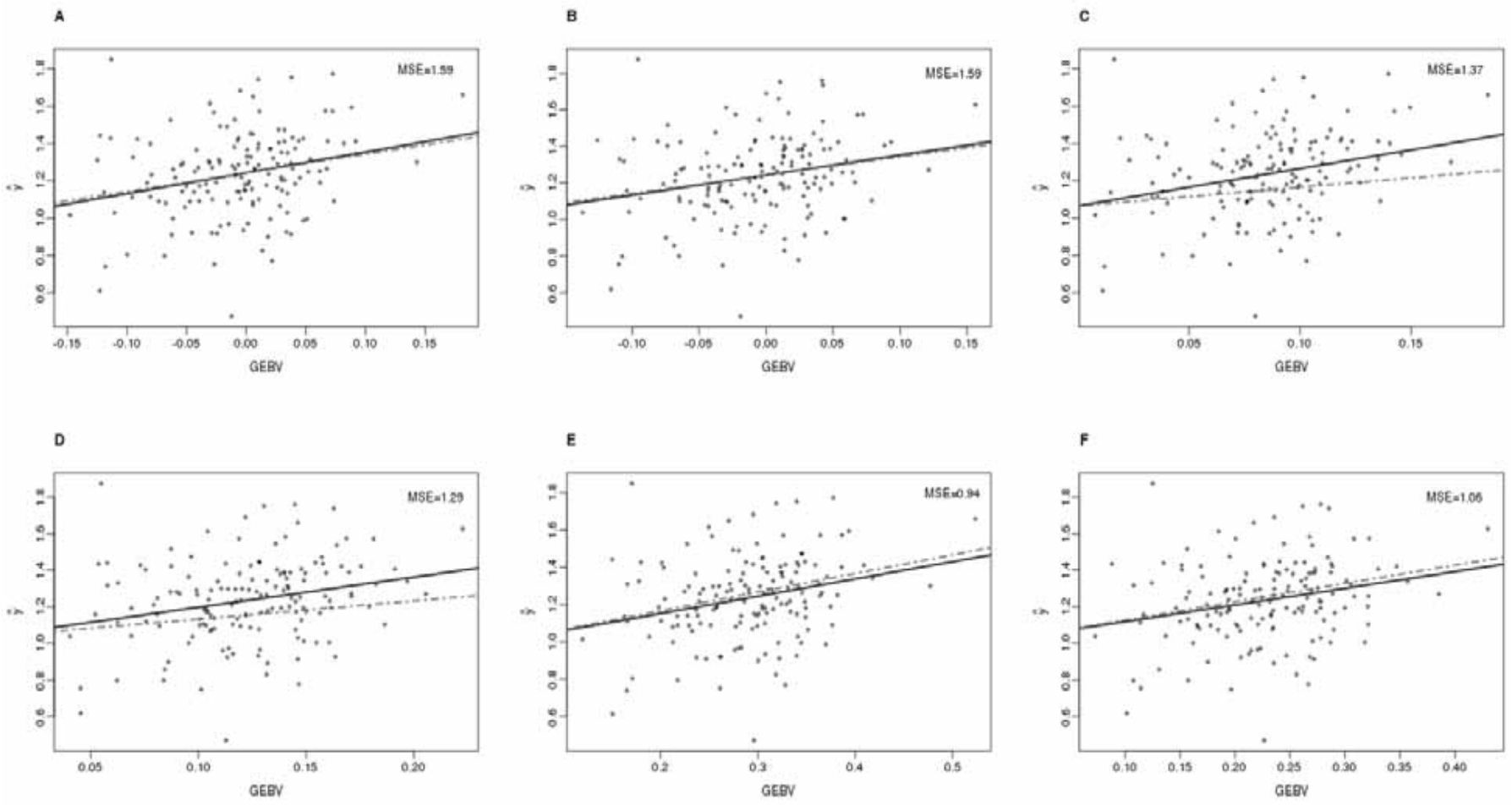

Figure 4A. Regression plots of adjusted phenotypes on GEBVs in LDi dataset. (A) BGBLUP, (C) BayesA, (E) BayesCπ. (B), (D) and (F) are the same models including PCs as covariates. Dashed grey and solid black lines represent observed and expected slopes, respectively. MSE: mean squared error.

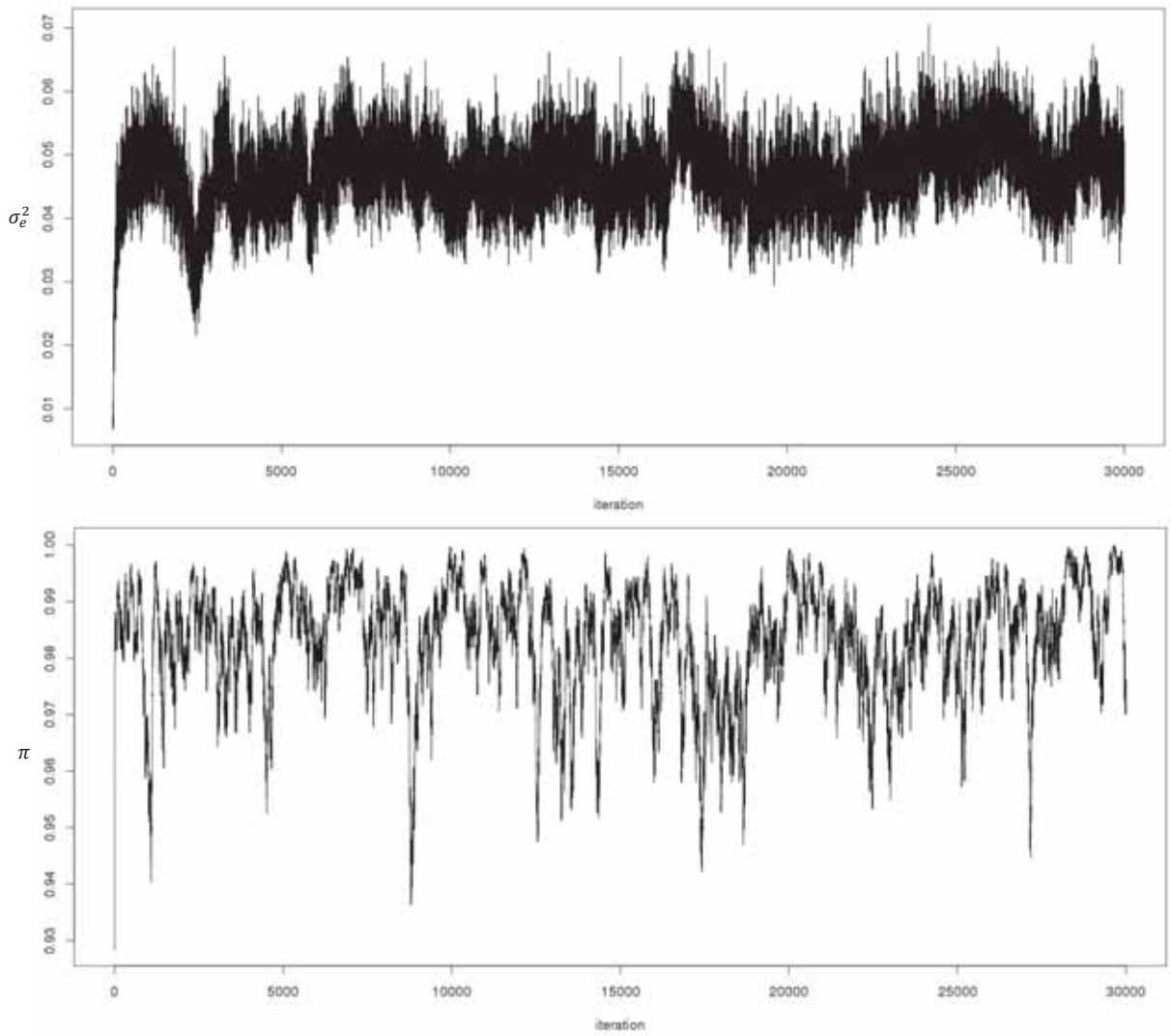

Figure 5A. Residual variance ($\boldsymbol{\sigma_e^2}$) and $\boldsymbol{\pi}$ parameter MCMC chains for 770k dataset without considering PCs as covariates.