

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**SELEÇÃO GENÔMICA E ESTUDO DE ASSOCIAÇÃO EM UM
REBANHO EXPERIMENTAL DA RAÇA NELORE**

Rafael Medeiros Oliveira Silva

Zootecnista

2015

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**SELEÇÃO GENÔMICA E ESTUDO DE ASSOCIAÇÃO EM UM
REBANHO EXPERIMENTAL DA RAÇA NELORE**

Rafael Medeiros de Oliveira Silva

Orientadora: Prof^a Dr^a Lucia Galvão de Albuquerque

Co-orientadoras: Prof^a Dr^a Maria Eugênia Mercadante

Prof^a Dr^a Arione Augusti Boligon

Tese de doutorado apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal como parte das exigências para obter o título de Doutor em Genética e Melhoramento Animal

2015

Silva, Rafael Medeiros de Oliveira
S586s Seleção genômica e estudo de associação em um rebanho experimental da raça Nelore / Rafael Medeiros de Oliveira Silva. – – Jaboticabal, 2015
xii, 75 p. : il. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2015
Orientadora: Lucia Galvão de Albuquerque
Co-orientadoras: Arione Augusti Boligon, Maria Eugênia Zerlotti Mercadante
Banca examinadora: Daniela Andressa Lino Lourenço, Lenira El Faro Zadra, Fernando Sebastián Baldi Rey, Roberto Carvalheiro
Bibliografia

1. Características de carcaça. 2. Consumo alimentar residual. 3. Gwas. 4. Seleção genômica. 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.



UNIVERSIDADE ESTADUAL PAULISTA

CAMPUS DE JABOTICABAL

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS DE JABOTICABAL

CERTIFICADO DE APROVAÇÃO

**TÍTULO: "SELEÇÃO GENÔMICA E ESTUDO DE ASSOCIAÇÃO
EM UM REBANHO EXPERIMENTAL DA RAÇA
NELORE"**

AUTOR: RAFAEL MEDEIROS DE OLIVEIRA SILVA

ORIENTADORA: Profa. Dra. LUCIA GALVAO DE ALBUQUERQUE

CO-ORIENTADORA: Profa. Dra. ARIONE AUGUSTI BOLIGON

CO-ORIENTADORA: Profa. Dra. MARIA EUGENIA ZERLOTTI MERCADANTE

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



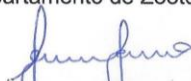
Profa. Dra. MARIA EUGENIA ZERLOTTI MERCADANTE
Instituto de Zootecnia / Sertãozinho/SP



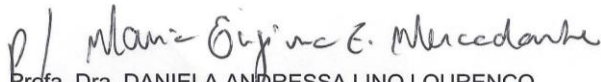
Prof. Dr. FERNANDO SEBASTIÁN BALDI REY
Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal



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



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



Profa. Dra. DANIELA ANDRESSA LINO LOURENÇO
University of Georgia / Athens / GA/USA
Participação por videoconferência

Data da realização: 03 de novembro de 2015.

DADOS CURRICULARES DO AUTOR

Rafael Medeiros de Oliveira Silva nasceu em Maceió – AL, em 20 de Novembro de 1986, filho de Antônio Batista da Silva Filho e Maria José Medeiros de Oliveira Silva. Em Março de 2005, ingressou no curso de Zootecnia na Universidade Federal de Alagoas – UFAL. Graduou-se em Janeiro de 2011 sob a orientação da Profª Drª Angelina Bossi Fraga. Em Março de 2010 ingressou no Programa de pós-graduação em Zootecnia (área de concentração: Genética e Melhoramento Animal) na Universidade Federal de Alagoas, orientado pela Profª Drª Angelina Bossi Fraga. Em Março de 2011, iniciou o estágio na Universidade Estadual Paulista – UNESP, campus de Jaboticabal, onde permaneceu por oito meses sob tutoria da Profª Drª Lucia Galvão de Albuquerque e obteve o título de mestre em Fevereiro de 2012. Em Março de 2012, iniciou como aluno do Programa de pós-graduação em Genética e Melhoramento Animal da Universidade Estadual “Júlio de Mesquita Filho” campus de Jaboticabal (UNESP – FCAV) sob a orientação da Profª Drª Lucia Galvão de Albuquerque. Em Fevereiro de 2014, iniciou o estágio sanduíche no exterior, permanecendo 12 meses na *University of Georgia* – USA, orientado pelo Prof. Dr. Ignacy Misztal.

EPÍGRAFE

“A educação moral, pois, consiste menos em dar a decorar listas de certo e errado do que em criar um ambiente moral propício ao autoexame, à seriedade interior, à responsabilidade de cada um saber o que fez quando não havia ninguém olhando.”

Olavo de Carvalho

DEDICATÓRIA

Dedico àqueles que sempre me apoiaram e nunca mediram esforços para que meus objetivos fossem alcançados, meus pais Antônio Batista de Silva Filho e Maria José Medeiros de Oliveira Silva.

Dedico à amiga Laíza Acioli (in memoriam).

AGRADECIMENTOS

Agradeço a Deus por me permitir o êxito naquilo que me propus a fazer, por estar presente em todos os momentos da minha vida, me protegendo e guiando segundo a tua vontade;

Agradeço à minha orientadora Prof^a Dr^a Lucia Galvão de Albuquerque pela confiança, orientação e exemplo de profissional responsável e dedicada.

Às minhas co-orientadoras Dr^a Arione Augusti Boligon e Dr^a Maria Eugênia Mercadante pelo apoio e ensinamentos compartilhados.

Agradeço ao Prof Dr Fernando Baldi pelo total apoio e por estar sempre disponível a ajudar em qualquer situação.

Agradeço aos membros da banca de exame de qualificação Prof Dr Henrique Nunes de Oliveira, Prof Dr Danísio Prado Munari, Prof Dr Fernando Baldi, Dr Roberto Carneiro. Agradeço também aos membros da banca de defesa, Prof^a Dr^a Daniela Lourenço, Dr^a Lenira El Faro, Prof Dr Fernando Baldi e Dr Roberto Carneiro. Obrigado pelas valiosas sugestões.

Ao Prof Dr Ignacy Misztal por me receber na *University of Georgia* e pelas valiosas sugestões para elaboração deste trabalho. Agradeço também ao pesquisador Dr Shogo Tsuruta pelo apoio e contribuição neste trabalho.

Aos amigos Daniela Lourenço e Breno Fragomeni pelo apoio e recepção em *Athens* e pelas importantes contribuições para realização deste trabalho. Agradeço também aos amigos Jefferson, El Hamidi, Mariana, Dennis, professores e funcionários da *University of Georgia*.

Aos caros amigos e companheiros de trabalho, Fabricia, Tonussi, Espigolan, Diogo, Luis Gabriel, Rodrigo, Denise, Arione, Natalia, Gordo, Raphael Costa, Tiago,

Gregório, Guilherme, Daiane, Diércles, Willian, Lucas, Ana Cristina, Dani, Lúcio, Andrés e todos os amigos que fizeram parte dessa caminhada.

À Nedenia Stafuzza pelo total apoio, companheirismo e singular contribuição na elaboração deste trabalho. Agradeço também por me mostrar que a vida pode ser bela mesmo quando tudo parece querer mostrar o contrário. Estar com você me desafia, inquieta e me traz paz.

Agradeço ao meu pai, Antônio Batista da Silva Filho, minha mãe, Maria José Medeiros de Oliveira Silva, às minhas irmãs, Mariana Medeiros de Oliveira Silva e Heloisa de Oliveira Silva pelo carinho, cuidado e compreensão em relação às minhas decisões e nunca me deixarem sentir sozinho apesar da distância. Agradeço também à minha família que torce e fica feliz com minhas conquistas.

À Gabriela Bezerra pelo carinho e ensinamentos compartilhados nos últimos anos. Você é responsável por muito daquilo que eu me orgulho em ser hoje. Minha admiração e respeito por você serão eternos.

Ao CNPq/CAPES pela concessão de bolsa de doutorado.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela concessão da bolsa de doutorado no Brasil (Processo nº: 2013/01228-5), no exterior (Processo: 2013/20796-4), e pelo apoio financeiro do projeto temático (Processo nº: 2009/16118-5).

Ao Centro Apta de gado de corte - Instituto de Zootecnia, Sertãozinho-SP pela concessão dos dados.

A todos os professores e funcionários do programa de Pós-Graduação em Genética e Melhoramento Animal da FCAV/UNESP Jaboticabal.

A todos que, direta ou indiretamente, contribuíram para o êxito deste trabalho.

SUMÁRIO

CHAPTER 1 – General Considerations	12
1. INTRODUCTION	12
2. OBJECTIVES	14
2.1 General Objective	14
2.2 Specific Objectives	14
3. USE OF GENOMIC INFORMATION	14
3.1 Genomic Selection	14
3.1.1 Accuracy of Genomic Estimate Breeding Value (GEBV)	16
3.2 Genome-wide association	16
4. TRAITS	18
4.1 Feed Efficiency traits	18
4.2 Carcass Quality Traits	20
5. FINAL CONSIDERATIONS	22
6. REFERENCES	22
CHAPTER 2 - Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population.....	31
1. INTRODUCTION	32
2. OBJECTIVES	34
3. MATERIAL AND METHODS	34
3.1 Data	34
3.2 (Co) variance component estimation	36
3.3 Methods of Genomic Analysis.....	38
3.3.1 Multistep	38
3.3.1.1 – Genomic BLUP (GBLUP).....	38
3.3.1.2 – BayesC π	40
3.3.2 Single-step genomic BLUP (ssGBLUP)	40
3.4 Cross Validation	40
3.5 Regression of phenotype on breeding value (EBV, GEBV, or DGV)	42

4. RESULTS AND DISCUSSION.....	42
5. CONCLUSIONS	48
6. REFERENCES	48
CHAPTER 3 - Genome-wide association study for carcass traits in an experimental Nelore cattle population	54
1. INTRODUCTION	55
2. OBJECTIVE	56
3. MATERIAL AND METHODS	57
3.1 Data	57
3.2 Genotyping procedure	58
3.3 (Co) variance component estimation	59
3.4 Single Step Genome Wide Association (ssGWAS).....	60
3.5 Search for Associated Genes	61
4. RESULTS AND DISCUSSION.....	62
5. CONCLUSION	71
6. ACKNOWLEDGMENTS.....	71
7. REFERENCES	71
APPENDIX.....	78

GENOMIC SELECTION AND GENOME-WIDE ASSOCIATION STUDY IN A NELORE EXPERIMENTAL POPULATION

ABSTRACT - The growing global demand for safe and sustainable food production has motivated a restructuring in the beef production sector aiming the production of better quality products without increasing the productive cost. Animal feeding is the most important economic component of beef production systems. Due the conventional processing of cattle carcasses refrigeration after slaughter, an adequate quality carcass must have enough fat covering to guarantee its preservation and desirable quality for consume. Selection for feed efficiency has not been effective mainly due to difficult and high costs to obtain the phenotypes. The application of genomic selection using single nucleotide polymorphisms (SNPs) can decrease the cost of animal evaluation as well as the generation interval. However, there is no consensus among researches about the best methodology to obtain genomic prediction for each trait. The objective of this study was to compare methods of genomic evaluation using high-density SNP panel (BovineHD *BeadChip* - Illumina) for feed efficiency traits and to identify genomic regions associated to carcass traits in a small beef cattle population. After quality control, a total of 437,197 SNP genotypes were available for 761 Nelore animals from Institute of Animal Science, Sertãozinho, SP, Brazil. The data set contained 896 records for efficiency traits, such as residual feed intake (RFI), feed conversion ratio (FCR), average daily gain (ADG), and dry matter intake (DMI), 2,306 ultrasound records for *longissimus* muscle area (LMA), 1,832 for backfat thickness (BF), and 1,830 for rump fat thickness (RF). Methods of analysis were traditional BLUP, single step genomic BLUP (ssGBLUP), genomic BLUP (GBLUP), a Bayesian regression method (BayesC π) and single-step genome-wide association (ssGWAS). Average accuracies ranged from 0.10 to 0.58 using BLUP, from 0.09 to 0.48 using GBLUP, from 0.06 to 0.49 using BayesC π and from 0.22 to 0.49 using ssGBLUP. The most accurate and consistent predictions were obtained using ssGBLUP for all analyzed traits. The single step genomic BLUP seems to be more suitable to obtain genomic predictions for feed efficiency traits on a small population of genotyped animals. The results found in this study should help to better understand the genetic and physiologic mechanism associated with LMA, BF, and RF in Zebu animals. Even though LMA, BF, and RF have a polygenic nature, genes like HSD17B12, PLAG1, and XKR4 could be considered candidates for genetic selection for *longissimus* muscle area, backfat thickness and rump fat thickness, respectively. In addition, studies focusing on identifying the role of many uncharacterized genes associated to studied traits are required to better understand the genetic architecture of those traits.

Keywords: carcass traits, genomic selection, gwas, residual feed intake.

SELEÇÃO GENÔMICA E ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA EM UMA POPULAÇÃO EXPERIMENTAL DA RAÇA NELORE

RESUMO – A crescente demanda por produção alimentar sustentável tem motivado uma re-estruturação no setor de produção de carne objetivando a obter produtos de melhor qualidade sem aumentar os custos de produção. Alimentação animal é o componente econômico mais importante dos sistemas de produção de carne bovina. Devido ao tradicional processo de refrigeração após o abate, se faz necessário uma adequada cobertura de gordura na carcaça para garantir a sua preservação e qualidade desejável para consumo. A seleção para eficiência alimentar não tem sido eficaz, principalmente devido aos custos elevados e dificuldade para obter os fenótipos. A aplicação da seleção genômica utilizando polimorfismos de nucleotídeo único (SNPs) pode diminuir o custo de avaliação animal, bem como o intervalo de geração. No entanto, ainda não há consenso entre os pesquisadores sobre a melhor metodologia para a obtenção predição genômica para cada característica. O objetivo deste estudo foi comparar métodos de avaliação genômica usando o painel de marcadores do tipo SNP de alta densidade (BovineHD BeadChip - Illumina) para eficiência alimentar e identificar regiões genômicas associadas à características de carcaça em uma pequena população de bovinos de corte. Após o controle de qualidade, um total de 437197 genótipos ficou disponível para 761 animais da raça Nelore do Instituto de Zootecnia, Sertãozinho, SP, Brasil. O conjunto de dados continha 896 informações fenotípicas para as características de eficiência, como a consumo residual de alimentar (CAR), taxa de conversão alimentar (TCA), ganho em peso médio diário (GMD), e consumo de matéria seca (CMS), 2.306 informações de ultrassom para a área de olho de lombo (AOL), 1832 para a espessura de gordura na região lombar (EGL), e 1830 para a espessura de gordura na região da garupa (EGG). Os métodos analisados foram: BLUP tradicional, *single step genomic* BLUP (ssGBLUP), *genomic* BLUP (GBLUP), abordagem Bayesiana (BayesC π) e de associação genômica ampla (ssGWAS). Acurácias das predições genômicas variaram de 0,10 a 0,58 utilizando BLUP tradicional, de 0,09 a 0,48 utilizando GBLUP, de 0,06 a 0,49 utilizando BayesC π e de 0,22 a 0,49 utilizando ssGBLUP. As predições mais acuradas e consistentes foram obtidos utilizando ssGBLUP para todas as características analisadas. O single-step GBLUP parece ser mais adequado para se obter predições genômicas para eficiência alimentar em uma pequena população de animais genotipados. Os resultados encontrados neste estudo devem ajudar a compreender melhor o mecanismo genético e fisiológico associados com AOL, EGL, e EGG em animais zebuínos. Apesar da natureza poligênica das características estudadas, genes como HSD17B12, PLAG1, e XKR4 podem ser considerados candidatos para a seleção genética para a AOL, EGL e EGG, respectivamente. Além disso, estudos com foco em identificar a função de genes que ainda não tiveram suas respectivas funções caracterizadas e foram associados com os fenótipos estudados são necessários para melhor entendimento da arquitetura genética dessas características.

Palavra-chave: características de carcaça, consumo alimentar residual, gwas, seleção genômica.

CHAPTER 1 – General Considerations

1. INTRODUCTION

The beef cattle production in tropical and subtropical regions is predominantly based on *Bos indicus* (Zebu) breeds and their crosses with *Bos taurus*. In the last years, Brazil has been on top five of largest beef exporters (USDA, 2015), and around 80% of its herd is constituted of Zebu breeds. Although the Zebu breeds have adaptive advantages to tropical conditions over Taurine breeds, they have some productivity and quality carcass limitations. The growing global demand for safe and sustainable food production has motivated a restructuring in the beef production sector aiming improvement in reproductive and productive efficiency (EUCLIDES FILHO et al., 2003; LOPES et al., 2012).

In this sense, according to Crowley et al. (2011) selection for feed efficiency could actually lead to an improvement in performance for some traits, such as muscularity, animal price, and carcass conformation. DiLorenzo and Lamb (2012) reported that selection for feed efficiency also decreases the environmental impact of the beef industry. According to these authors selecting for feed efficiency based on residual feed intake (RFI) can reduce 29% of fresh manure output and excretions of phosphorous and nitrogen, while methane emissions can be reduced by as much as 28%.

The residual feed intake is favorable correlated to carcass traits, such as *longissimus* muscle area (RETALLICK et al., 2014) which may be related to carcass weight, fat, and muscle traits in steers (BERGEN et al., 2005). Favorable genetic correlations between subcutaneous fat and reproductive traits have been reported, indicating that high subcutaneous fat deposition could denote early finishing and could result in animals more sexually precocious (CAETANO et al., 2013).

The improvement of feed efficiency and carcass traits using traditional evaluation methods is limited by the difficulty and costs to access the phenotypes of interest. Measurement of feed intake might occur in central test stations or on farms, which requires a significant investment that makes it expensive to measure. Some recent breakthroughs about the use of DNA information has been applied for the

improvement of quantitative traits, and may be especially helpful for traits that are hard or expensive to measure and because of that are not routinely recorded. Genomic prediction may combine phenotypic, pedigree and genotype information to increase the accuracy of animal evaluation and to reduce the generation interval, which increases the genetic gain (VANRADEN et al., 2009). Also, the use of genomic information may determine regions on the genome that are associated to phenotypes and may help to better understand the genetic architecture of some important traits.

Many studies have focused on determining the most suitable method to obtain the genomic prediction equation; however, there is no consensus among researches about the most adequate methodology for each trait. In the United States, dairy cattle genetic evaluation is performed by multistep methods if genomic information is available (VANRADEN, 2008; VANRADEN et al., 2009). This approach consists of predicting the genomic estimated breeding value (GEBV) by an index combining EBV and direct genomic value (DGV). On the other hand, if phenotypes, pedigrees, and genotypes are available all together, a simple way to incorporate genomic information into evaluations is by the single-step procedure (MISZTAL et al., 2009). In additionally, several studies of livestock breeding have focused on predicting genomic value or identifying regions in the genome that is associated to phenotypes of interest (KIM et al., 2011; FONTANESI et al., 2012; FRAGOMENI et al., 2014; LIU et al., 2014). The mainly reason for this is to explore those associations and identify genes that are controlling the expression of traits and selecting the superior animals.

The use of the most suitable method for genomic evaluation could provide accurate predictions for feed efficiency traits and the selection for those traits could reduce the environmental impact and the beef production cost. Also, the genome-wide association study of carcass traits can help to better understand the genetic architecture and the main genes affecting those traits, as well as the mechanism of functionality.

2. OBJECTIVES

2.1 General Objective

The objective of this study was to compare methodologies to predict genomic breeding values for feed efficiency traits as well as to detect SNP associated to carcass traits in an experimental Nelore cattle population.

2.2 Specific Objectives

- To compare the prediction ability for feed conversion efficiency related traits obtained using BLUP, genomic BLUP, and BayesC π using different cross-validation designs;
- To perform genome-wide association between single nucleotides polymorphism (SNP) markers with *longissimus* muscle area (LMA), subcutaneous backfat (BF), and subcutaneous rump fat thickness (RF) obtained by ultrasonography.

3. USE OF GENOMIC INFORMATION

3.1 Genomic Selection

Genomic selection is a form of marker-assisted selection on a genome-wide scale (MEUWISSEN et al., 2013). The effect of thousands of DNA markers are simultaneously estimated combining phenotypic information in a training population and are used for the estimation of breeding values of selection candidates. Several methods to estimate the marker effects have been strongly studied and compared in order to find the most suitable for predicting genomic values for different traits.

Marker effects can be obtained assuming that all markers contributed equally to genetic variation (no major gene effect), or assuming that the prior distribution of marker or QTL effects is not normal. The Bayesian methodology assumes different genetic variance across SNP with major effect SNP. The dairy cattle genetic evaluation in the United States is currently performed by multistep methods if genomic information is available (VANRADEN, 2008; VANRADEN et al., 2009). This approach consists of predicting the GEBV by an index combining EBV and DGV. The

genomic BLUP estimates all marker effects at the same time and assumes the same variance for all SNP (MEUWISSEN et al., 2013). A disadvantage of this method is the overestimation of the variance of markers with no effect and underestimation of the variance of high effect markers that can harm accuracy of prediction.

Genomic predictions can also be obtained considering different presuppositions in the model. The assumption that all SNP effects are normally distributed with a constant variance may be unrealistic (MEUWISSEN et al., 2013). The Bayesian approach may or not assume different variances on all segments of chromosomes considering that very few SNPs have very high effect and the majority of SNPs have very small or null effect (VANRADEN, 2008). Frequently the number of animals is much smaller than the number of marker effects to estimate, so the final marker effect estimates are strongly influenced by the prior information (MEUWISSEN et al., 2013). Pryce et al. (2012) found an advantage in the accuracy of genomic predictions for RFI obtained using Bayesian models over GBLUP, in Australian heifers. Also, Neves et al. (2014) reported that Bayesian regression models were more accurate than GBLUP in a Nelore population.

On the other hand, if phenotypes, pedigrees, and genotypes are available, a simple way to incorporate genomic information into evaluations is by the single-step genomic BLUP (MISZTAL et al., 2009). This approach consists of incorporating phenotypes, pedigrees and genomic information into only one step of evaluation. With this procedure, the relationship matrix based on pedigree (**A**) is combined with a genomic relationship matrix (**G**) based on information from SNP markers, into a single matrix of realized relationships (**H**).

Comparing GBLUP and ssGBLUP, Aguilar et al. (2010) concluded that genomic evaluations using ssGBLUP were as accurate as those using multistep procedure. According to Lourenco et al. (2014), ssGBLUP has an advantage over multistep methods mainly because it uses phenotypes rather than pseudo-phenotypes and accounts for the entire population structure to estimate GEBV. Onogi et al. (2014) also concluded that the implementation of genomic selection by ssGBLUP provided more accurate predictions than traditional BLUP even using only genotyped sires.

3.1.1 Accuracy of Genomic Estimate Breeding Value (GEBV)

The accuracy of genomic prediction is the key to the successful application of genomic selection. To validate the model and assess the accuracy of prediction, a statistical procedure named cross-validation is commonly used. Indeed, the genomic prediction equation cannot be validated in the same animals used for obtaining it. So, the evaluation of a model by cross-validation is done under the assumption that random partitioning of the data results in independent training and testing sets. In practice, an important use of genomic selection is to predict the genetic merit of the next generation without their phenotype, using only genomic information.

Accuracy is strongly dependent on many factors such as linkage disequilibrium (MEUWISSEN et al., 2001), allele frequency distribution (LETTRE, 2011), number of genotyped animals (VANRADEN et al., 2009; CALUS, 2011; DAETWYLER et al., 2010), heritability of the traits (GODDARD, 2009), effective population size (GODDARD, 2009), marker density (MOSER et al., 2009), and the method used to estimate marker effects (LOURENCO et al., 2014). Likewise, for real and simulated populations with both high and low heritabilities, Habier et al. (2007; 2010) observed that prediction accuracy decreased when the relationship between training and testing populations decreased. Many authors have showed concerns about validating the model in an unrelated population (PÉREZ-CABAL et al., 2012, SAATCHI et al., 2013) especially for traits difficult to measure. As reported by Pérez-Cabal et al. (2012), the genetic relationships between individuals have strong effect on accuracy of prediction.

Chen et al. (2013) reported the most accurate genomic prediction for RFI when the validation was done within breed (e.g., accuracy of 0.58 for Angus and 0.64 for Charolais). On the other hand, when the data set was pooled from the two breeds to form the training population the accuracies decreased to 0.31 and 0.43, respectively, for Angus and Charolais.

3.2 Genome-wide association

In recent years, many studies in livestock breeding have focused on showing the viability of using the genomic information to identify genomic regions associated with productive traits (BOLORMAA et al., 2011; FRAGOMENI et al., 2014; SANTANA

et al., 2014). The main reason of this is to explore the associations to better understand the genetic architecture of those traits and posterior selection of superior animals. Since the regions associated to the traits were identified, the selection can be performed with simple tests for a few SNPs. Thus, the success of genetic selection using major SNPs is dependent on the proportion of additive genetic variance explained by each marker or region of genome (FRAGOMENI et al., 2014).

Genome-wide association analysis (GWAS) has been applied to find regions on the genome that are associated to economically important traits. A common used method for GWAS is based on estimation of one marker at time as fixed effect (HIRSCHHORN; DALY, 2005). An alternative to simplify traditional GWAS consists in integrating all genotypes, pedigree, and phenotypic information available (from genotyped and ungenotyped animals) in one-step procedure (single-step GWAS) that allows the use of any model, and all relationships simultaneously (WANG et al., 2012; 2014; FRAGOMENI et al., 2014).

Several genomic studies including GWAS have been focused on Taurine subspecies. Santana et al. (2014) has reported the first GWAS of DMI to identify genomic regions associated with feed intake and efficiency in Nelore cattle. Recently, Olivieri (2015) has reported a total of 24 genes directly associated with RFI that were founded in 10 windows of 10 adjacent SNPs. Nkrumah et al. (2007) identified quantitative traits loci (QTL) located on chromosomes 1 (90 cM), 5 (129 cM), 7 (22 cM), 8 (80 cM), 12 (89 cM), 16 (41 cM), 17 (19 cM), and 26 (48 cM) associated to residual feed intake. Pryce et al. (2012) identified eight SNPs with big effect on the expression of RFI in the chromosome 14 that probably could be associated to NCOA2 gene, which has important function on the control of energetic metabolism.

According to Moore et al. (2009), even though many makers have been associated to RFI (BARENDSE et al., 2007; SHERMAN et al., 2008) no important genes have been described with high influence on this trait. This information is supported by the polygenic nature of RFI given that its expression is influenced by many genes with minor effect. However, a combination of genetic markers could explain a significant part of additive genetic variance. In this sense, selection based on genomic prediction considering high and low effect chromosome regions would be

suitable to increase the annual genetic gain of quantitative traits such as residual feed intake.

Similarly for feed efficiency traits, many genomic association studies for carcass traits have been conducted in Taurine breeds. Bolormaa et al. (2011) found 64 SNPs associated to *longissimus* muscle area in an Australian beef population. Kim et al. (2011) reported markers located on chromosome 3, 11, 13 and 6 associated to quality traits including marbling score, backfat thickness, and *longissimus* muscle area in indigenous beef cattle called Korean cattle (*Bos Taurus*). The SNP associated to *longissimus* muscle area was located within the DVL1 gene which has functions of muscle development. In a study evaluating twelve carcass traits, such as *longissimus* dorsi muscle area, and backfat thickness obtained by Real-time ultrasound, Lu et al. (2013) reported that many SNPs acted pleiotropically to affect carcass quality.

According to Fragomeni et al. (2014) one common issue in genome association studies is the large number of false positive gene discovery. Because many genes have been associated to several different traits, these authors have recommended carefully interpreting GWAS results before referring it as a causative effect.

4. TRAITS

4.1 Feed Efficiency traits

The costs associated with feeding represent around 50-70% of the total cost of the beef cattle industry, and are even greater in feedlot systems. According to Arthur et al. (2005) feed costs have been decisive in the economic efficiency of meat industry. Different parameters are used to evaluate the growth of animals and their efficiency to convert feed in meat. In young animals, some of the most common measures used are: feed conversion rate (FCR) and residual feed intake (RFI). Among the productive traits, the average daily gain (ADG) is the most studied and is more directly related to productivity in beef cattle (FERNANDES et al., 2004). In addition, Fox et al. (2001), comparing weight gain rate (WGR) and efficiency of metabolizable energy (EME), concluded that the increase of EME provides profit

25% bigger than increasing WGR. Thus, an efficient animal is efficient to convert feed mass into increases of the desired output, but it is not necessarily more productive (FERREIRA; SANDOVAL, 2012).

There are concerns if selection for efficiency traits might cause unfavorable effects over other important traits. For Bullock et al. (1993), selection for FCR can increase average adult weight, increasing maintenance energy requirements and, consequently, the cost of production. There are evidences that selection for other feed efficiency traits can improve some carcass traits in Taurine breeds without an undesirable correlated response with others traits (CROWLEY et al., 2011). DiLorenzo and Lamb (2012) reported that selection for RFI also decreases the environmental impact of beef industry. According to these authors, selection for feed efficiency can impact the amount of nutrients consumed and excreted per cow without interfering on animal performance. On the other hand, Santana et al. (2012) reported that selection based on RFI would decrease the proportion of subcutaneous fat tissue in the carcass. However, these authors also reported favorable genetic correlation between FCR to *longissimus* muscle area.

The RFI is an alternative trait to study the feed conversion efficiency, once, by definition it is not correlated with ADG. According to the definition of Koch et al. (1963), RFI is the difference between the real feed intake and the estimated consumption. Thus, most efficient animals are those with negative RFI which consume less food than they were supposed to have keeping the same productive performance. Mathematically RFI is considered as the residue of a regression equation of dry matter intake (DMI) on the metabolic weight ($MW^{0.75}$) and weight gain (ADG) during a given period. So, this trait is not correlated with the traits used in the equation (BASARAB et al., 2003). This definition makes RFI one of the best ways to evaluate feed efficiency, since it takes into account the daily weight gain of the animal (production) as well as to adjust the metabolic weight (maintenance) of the individual (BARENDSE et al., 2007). Selection based on RFI can reduce 29% of fresh manure output and excretions of phosphorous and nitrogen and 28% of methane emissions (DILorenzo; LAMB, 2012).

Given the economic importance of feed efficiency traits for livestock industry there is a need to use the most suitable method for genomic evaluation focusing in

increasing the accuracy. Also, considering that individual feed efficiency traits are hard and expensive to measure and there is not always phenotype available for them, it is important to measure how accurate the genomic evaluation would be when it is applied in an unrelated population.

4.2 Carcass Quality Traits

The animal carcass is composed by different muscles and fat depots, which according to their localization and characteristic may determine the commercial price of different cuts. The ratio between muscle and fat tissues depends of the age of animal with the fat proportion indicating maturity. The adipose tissue is located in different regions on the animal body and its principal function is to store fat. Furthermore, the fat of ruminants is a natural source of conjugated isomers of linoleic acid (CLA) which have a positive effect on human health, related to anticancer activity, immune functions, and potential beneficial effects on coronary heart disease (IP, 1997; DUGAN et al., 2011). According to the location of adipose tissue the fat is called intermuscular, subcutaneous or intramuscular.

The production of fat cover is an important point in the meat industry especially to protect the carcass after slaughter. An adequate quality carcass must have enough fat covering to guarantee its preservation and desirable quality for consume (CUNDIFF et al., 1993). The conventional processing of cattle carcasses refrigeration after slaughter consists to down the temperature to around 7°C, which may result in excessive contraction of the sarcomeres, resulting in tougher meat. According to Hedrick (1983) the subcutaneous fat covering the *longissimus* muscle is an efficient indicator of finishing carcass.

The *longissimus* muscle area (LMA) is one of the most used regions to evaluate quality carcass in meat production. According to Boggs and Merkel (1990) the LMA is an indicator of carcass composition and it is related to carcass muscularity. In additionally, several studies reported that LMA is favorable genetically correlated to scrotal circumference (SC) (TURNER et al., 1990; JOHNSON et al., 1993; YOKOO et al., 2014). Considering that SC is an indicator for sexual precocity (TOELLE; ROBISON, 1985; PEREIRA et al., 2001), the selection to increase LMA should provide correlated gain for reproductive traits. However, considering that both

traits (LMA and SC) are genetically associated to growing traits, this association should also be about the size of animal.

Guindolin et al. (2010) reported that age at first calving (AFC) had negative genetic correlation with back fat thickness, LMA and body weight at 210 days, which indicated that it is possible to improve meat and fat deposition beyond a consequently correlated genetic gain for AFC in a Zebu population. Still, favorable genetic correlations between subcutaneous fat and reproductive traits have been reported for Nelore populations, indicating that high subcutaneous fat deposition could indicate early finishing and could result in animals more sexually precocious (CAETANO et al., 2013; YOKOO et al., 2015). However, it is important to be considered that the excessive fat increases the cost of meat for the consumer and requires more cleaning of carcasses prior to weighing and paying the producer (RESENDE et al., 2014).

There are several ways to evaluate the quality of carcass to improve organoleptic characteristics of meat. Some methods have the disadvantage to need to cull the animal besides the long time and high price for evaluation. According to Stouffer (2004) the ultrasound procedure is considered as a low cost technology and easy for application. This is a not invasive procedure that allows evaluation without leaving residue in the meat. It has been used to evaluate some carcass traits as *longissimus* muscle area and subcutaneous fat thickness (STOUFFER et al., 1961; SIMM, 1983; WILSON, 1992; FIGUEIREDO, 2001; YOKOO, 2009; BARBOSA et al., 2010).

The Zebu meat is not well accepted in most demanding markets mainly because their production has been focused on quantitative level without standardization. Thus, the growing global demand for safe and sustainable food production has motivated a re-structure in the beef production sector (EUCLIDES FILHO et al., 2003; LOPES et al., 2012). Trying to supply the international consumer demand genetic evaluations has been also focused on feed efficiency and quality carcass traits aiming the production of better quality meat without increasing the productive cost.

5. FINAL CONSIDERATIONS

In this sense, the estimate of accurate genomic predictions for feed efficiency and the selection for those traits could reduce the environmental impact and the beef production cost. Considering the high price and difficulty to obtain the feed efficient phenotypes, the genomic selection would be especially helpful if the genomic merit could be estimated for less related select candidates. The use of genomic information to better understand the genetic architecture of some carcass traits would help to increase the genetic gain of those traits.

Considering that most of the genomic studies for feed efficiency and carcass traits has been done on Taurine breeds in temperate regions (MUJIBI et al., 2011; PRYCE et al., 2012; ELZO et al., 2012; CHEN et al., 2013; PRYCE et al., 2014) there is a need to study those traits in Zebu breeds in tropical areas to figure the genetic architecture of those traits in different breeds.

6. REFERENCES

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

ARTHUR, P. F.; HERD, R. M.; WILKINS, J. F.; ARCHER J. A. Maternal productivity for Angus cows divergently selected for post-weaning residual feed intake. **Australian Journal Experimental Agriculture**, v. 45, n. 8, p. 985-993, 2005.

BARBOSA, V.; MAGNABOSCO, C. U.; TROVO, J. B. F.; FARIA, C. U.; LOPES, D. T.; VIU, M. A. O.; LOBO, R. B.; MAMEDE, M. M. S. Quantitative genetic study of carcass traits and scrotal perimeter, using Bayesian Inference in Nelore young bulls. **Bioscience Journal**, v. 26, n. 5, p. 789-797, 2010.

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, p. 1893-1905, 2007.

BASARAB, J. A.; PRICE, M. A.; AALHUS, J. L.; OKINE, E. K.; SNELLING, W. M.; LYLE, K. L. Residual feed intake and body composition in young growing cattle. **Canadian Journal of Animal Science**, v. 83, p. 189-204, 2003.

BERGEN, R.; MILLER, S. P.; WILTON, J. W. Genetic correlations among indicator traits for carcass composition measured in yearling beef bulls and finished feedlot steers. **Canadian Journal of Animal Science**, v. 85, n. 4, p. 463-473, 2005.

BOGGS, D. L.; MERKEL, A. R. **Live animal carcass evaluation and selection manual**. 3 ed. Dubuque, Iowa, Kendall/Hunt Publishing Co., 1990. 211p.

BOLORMAA, S.; PORTO NETO, L. R.; ZHANG, Y. D.; BUNCH, R. J.; HARRISON, B. E.; GODDARD, M. E. and BARENDSE, W. A genome wide association study of meat and carcass traits in Australian cattle. **Journal of Animal Science**, v. 89, n. 8, p. 2297-2309, 2011.

BULLOCK, K. D.; BERTRAND J. K.; BENYSHEK, L. L. Genetic and environmental parameters for mature weight and other growth measures in Polled Hereford cattle. **Journal of Animal Science**, v. 71, p. 1737-1741, 1993.

CAETANO, S. L.; SAVEGNAGO, R. P.; BOLIGON, A. A.; RAMOS, S. B.; CHUD, T. C. S.; LÔBO, R. B.; MUNARI, D. P. Estimates of genetic parameters for carcass, growth and reproductive traits in Nelore cattle. **Livestock Science**, v. 155, n. 1, p. 1-7, 2013.

CALUS, M. P. L.; VEERKAMP, R. F. Accuracy of multi-trait genomic selection using different methods. **Genetic Selection Evolution**, v. 43, n. 26, 2011.

CHEN, L.; SCHENKEL, F.; VINSKY, M.; CREWS, D. H.; LI, C. Accuracy of predicting genomic breeding values for residual feed intake in Angus and Charolais beef cattle. **Journal of Animal Science**, v. 91, p. 4669-4678, 2013.

CROWLEY, J. J.; EVANS, R. D.; MC HUGH, N.; PABIOU, T.; KENNY, D. A.; MCGEE, M.; CREWS JR., D. H.; BERRY, D. P. Genetic associations between feed efficiency measured in a performance test station and performance of growing cattle in commercial beef herds. **Journal of Animal Science**, v. 89, p. 3382-3393, 2011.

CUNDIFF, L.V.; KOCH, R.M.; GREGORY, K.E.; CROUSE J.D.; DIKEMAN, M.E. Characteristics of diverse breeds in cycle IV of the cattle germoplasm evaluation program. **Beef Research-Progress Report**, v. 4, p. 63-71, 1993.

DAETWYLER, H. D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J. A. The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, p. 1021-1031, 2010.

DILORENZO, N.; LAMB, G. C. Environmental and Economic Benefits of Selecting Beef Cattle for Feed Efficiency. UF/IFAS Extension, AN276 Gainesville- FL, 2012.

DUGAN, M. E. R., ALDAI, N., AALHUS, J. L., ROLLAND, D. C.; KRAMER, J. K. G. Review: Trans-forming beef to provide healthier fatty acid profiles. **Canadian Journal of Animal Science**, v. 91, p. 54-56, 2011.

ELZO, M. A.; LAMB, G. C.; JOHNSON, D. D.; THOMAS, M. G.; MISZTAL, I.; RAE, D. O.; MARTINEZ, C. A.; WASDIN, J. G.; DRIVER, J. D. Genomic-polygenic evaluation of Angus-Brahman multibreed cattle for feed efficiency and postweaning growth using the Illumina 3K chip. **Journal of Animal Science**. v. 90, p. 2488-2497, 2012.

EUCLIDES FILHO, K.; FIGUEIREDO, G. R.; EUCLIDES, V. P. B.; SILVA, L. O. C.; ROCCO, V.; BARBOSA, R. A.; JUNQUEIRA, C. E. Desempenho de diferentes grupos genéticos de bovinos de corte em confinamento. **Revista Brasileira de Zootecnia**, v. 32, n. 5, p. 1114-1122, 2003.

FERNANDES, H. J; PAULINO, M. F.; MARTINS, R. G. R.; VALADARES FILHO, S. C.; TORRES, R. A.; PAIVA, L. M.; Moraes, G. F. B. Ganho de peso, conversão alimentar, ingestão diária de nutrientes e digestibilidade de garrotes não-castrados de três grupos genéticos em recria e terminação. **Revista Brasileira de Zootecnia**, v. 33, n. 6, p. 2403-2411, 2004.

FERREIRA, F. O. B.; SANDOVAL, G. Eficiência alimentar em Bovinos de Corte - Importância na seleção. Available at <<http://www.altagenetics.com.br/novo/noticias/Ler.aspx?nID=796>>. Accessed in: 26 June 2012.

FIGUEIREDO, L. G. G. **Estimativas de parâmetros genéticos de características de carcaças feitas por ultra-sonografia em bovinos da raça Nelore**. 2001. 52 p. Dissertação (Mestrado em Qualidade e Produtividade Animal) - Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2001.

FONTANESI, L.; SCHIAVO, G.; GALIMBERTI, G.; CALÒ, D. G.; SCOTTI, E.; MARTELLI, P. L.; BUTTAZZONI, L.; CASADIO, R.; RUSSO, V. A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. **BMC Genomics**, v. 15, 2012.

FOX, D.G.; GUUIROY, P.J.; TEDESCHI, L.O. Determining Feed Intake and Feed Efficiency of Individual Cattle Fed in Groups. Proc. 33rd Beef Improvement Federation meeting, San Antonio, TX. 2001.

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

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

GUINDOLIN, D.G.F., GRUPIONI, N.V., CHUD, T.C.S., URBINATI, I., LÔBO, R.B., BEZERRA, L.A.F., PAZ, C.C.P., MUNARI, D.P. Genetic association for growth, reproductive and carcass traits in Guzera Beef Cattle. In: World Congress on Genetics Applied to Livestock Production, 9th. 2010, Leipzig. **Proceedings...** Leipzig, Germany, 2010, p. 640.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, p. 2389-2397, 2007.

HABIER, D.; TETENS, J.; SEEFRIED, F.; LICHTNER, P.; THALLER, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics Selection Evolution**, v. 425, n. 5, 2010.

HEDRICK, H.B. Methods of estimating live animal and carcass composition. **Journal of Animal Science**, Champaign. v.57, n.5, p.1316-26, 1983.

HIRSCHHORN, J. N.; DALY, M. J. Genome-wide association studies for common diseases and complex traits. **Nature Reviews Genetics**, v. 6, p. 95-108, 2005.

IP, C. Review of the effects of trans fatty acids, oleic acid, n-3 polyunsaturated fatty acids, and conjugated linoleic acid on mammary carcinogenesis in animals. **American Journal of Clinical Nutrition**, v. 66, p. 1523-1529, 1997.

JOHNSON, M. Z.; SCHALLES, R. R.; DIKEMAN, M. E.; GOLDEN, B. L. Genetic parameter estimates of ultrasound-measured longissimus muscle area and 12th rib fat thickness in Brangus cattle. **Journal of Animal Science**, v. 71, p. 2623-2630, 1993.

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, p. 361-365, 2011.

KOCH, R. M.; SWIGER, L. A.; DOYLE, C.; GREGORY, K. E. Efficiency of feed use in beef cattle. **Journal of Animal Science**, v. 22, p. 486-494, 1963.

LETTRE, G. Recent progress in the study of the genetics of height. **Human Genetics**, v. 129, p. 465-472, 2011.

LIU, X.; WANG, L.; LIANG, J.; YAN, H.; ZHAO, K.; LI, N.; ZHANG, L.; WANG, L. Genome-Wide Association Study for Certain Carcass Traits and Organ Weights in a Large White×Minzhu Intercross Porcine Population. **Journal of Integrative Agriculture**, v. 13, p. 2721-2730, 2014.

LOPES, B. C.; MATARIM, D. L.; FRANÇA, M. G. B.; MIZIARA, M. N.; LOPES, P. A. and FRANCO, T. Genética bovina brasileira: mercado internacional e mapeamento das competências e tecnologias mineiras. Polo de Excelência em Genética Bovina / SECTES, Uberaba-MG, 113 p. 2012.

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

LU, D.; SARGOLZAEI, M.; KELLY, M.; VOORT, G. V.; WANG, Z.; MANDELL, I.; MOORE, S.; PLASTOW, G. AND MILLER, S. P. Genome-wide association analyses for carcass quality in crossbred beef cattle. **BMC Genetics**, v. 14, n. 80, 2013.

MEUWISSEN, T.; HAYES, B.; GODDARD, M. Accelerating Improvement of Livestock with Genomic Selection. *Annual Review of Animal Biosciences*. v 1, p. 221–237, 2013.

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

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

MOORE, S. S.; MUJIBI, F. D.; SHERMAN, E. L. Molecular basis for residual feed intake in beef cattle. **Journal of Animal Science**, v. 87, p. E41-E47, 2009.

MOSER, G.; TIER, B.; CRUMP, R. E.; KHATKAR, M. S.; RAADSMA, H. W. A comparison of five methods to predict genomic breeding values of dairy bulls from genome wide SNP markers. **Genetics Selection Evolution**, v. 41, n. 56, 2009.

MUJIBI, F. D. N.; NKRUMAH, J. D.; DURUNNA, O. N.; STOTHARD, P.; MAH, J.; WANG, Z.; BASARAB, J.; PLASTOW, G.; CREWS, D. H.; MOORE, S. S. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. **Journal of Animal Science**, v. 89, n. 11, p. 3353-3361, 2011.

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

NKRUMAH, J.D; BASARAB, J.A.; WANG, Z.; LI, C.; PRICE, M.A.; OKINE, E.K.; CREWS, D.H.; MOORE, S.S. Genetic and phenotypic relationship of feed intake and measures of efficiency with growth and carcass merit of beef cattle. **Journal of Animal Science**, v. 85, p. 2711-2720, 2007.

OLIVIERI, B. Estudo de Associação entre polimorfismos de base única com característica de eficiência de conversão e consumo em bovinos da raça nelore. 2015. Dissertação (Mestrado em Genética e Melhoramento Animal) - Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista, Jaboticabal, SP, 2015.

ONOGI, A.; OGINO, A.; KOMATSU, T.; SHOJI, N.; SIMIZU, K.; KUROGI, K.; YASUMORI, T.; TOGASHI, K. and IWATA, H. Genomic prediction in Japanese Black cattle: application of a single-step approach to beef cattle. **Journal of Animal Science**, v. 92, n. 5, p.1931-1938, 2014.

PEREIRA, E.; ELER, J. P.; COSTA, F. A. A.; FERRAZ, J. B. S. Análise genética da idade ao primeiro parto e do perímetro escrotal em bovinos da raça Nelore. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 53, p. 116-121, 2001.

PÉREZ-CABAL, M. A.; VAZQUEZ, A. I.; GIANOLA, D.; ROSA, G. J. M.; WIEGEL, K. A. Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. **Frontiers in Genetics**, v. 3, n. 27, 2012.

PRYCE, J.E.; WALES, W. J. HAAS, Y., VEERKAMP, R. F. AND HAYES B. J. Genomic selection for feed efficiency in dairy cattle. **Animal**, v. 8, n. 1, p. 1-10, 2014.

PRYCE, J. E.; ARIAS, J.; BOWMAN, P. J.; DAVIS, S. R.; MACDONALD, K. A.; WAGHORN, G. C.; WALES, W. J. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. **Journal of Dairy Science**, v. 95, p. 2108-2119, 2012.

RESENDE, F. D.; GESUALDI JÚNIOR, A.; QUEIROZ, A. C.; FARIA, M. H.; VIANA, A. P. Carcass characteristics of feedlot-finished Zebu and Caracu cattle. **Revista Brasileira de Zootecnia**, v. 43, n. 2, p. 67-72, 2014.

RETALLICK, K. M.; FAULKNER, D. B.; RODRIGUEZ-ZAS, S. L.; NKRUMAH, J. D. AND SHIKE, D. W. Relationship among performance, carcass, and feed efficiency characteristics, and their ability to predict economic value in the feedlot. **Journal of Animal Science**, v. 91, p. 5954–5961, 2014.

SAATCHI, M.; WARD, J.; GARRICK, D. J. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. **Journal of Animal Science**, v. 91, p. 1538-1551, 2013.

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

SANTANA, M. H. A.; ROSSI JR, P.; ALMEIDA, R.; CUCCO, D. C. Feed efficiency and its correlations with carcass traits measured by ultrasound in Nellore bulls. **Livestock Science**, v. 145, p. 252-257, 2012.

SHERMAN, E. L.; NKRUMAH, J. D.; LI, C.; BARTUSIAK, R.; MURDOCH, B.; MOORE, S. S. Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. **Journal of Animal Science**, v. 87, p. 37-45, 2008.

SIMM, G. The use of ultrasound to predict the carcass composition of live cattle – a review. **Animal Breeding Abstracts**, v. 51, n. 12, p. 853-875, 1983.

STOUFFER, J. R. History of ultrasound in animal science. **Journal of Ultrasound in Medicine**, v. 23, p. 577-584, 2004.

STOUFFER, J.R.; WALLENTINE, M.V.; WELLINGTON, G.A. Development and application of ultrasonic methods for measuring fat thickness and rib-eye area in cattle and rib-eye in cattle and hogs. **Journal of Animal Science**, v. 18, n. 4, p. 759-767, 1961.

TOELLE, V. D.; ROBISON; O. W. Estimates of genetic correlations between testicular measurements and female reproductive traits in cattle. **Journal of Animal Science**, v. 60, p. 89-100, 1985.

TURNER, J. W.; PELTON, L. S; CROSS, H. R. Using live animal ultrasound measures of ribeye area and fat thickness in yearling Hereford bulls. **Journal of Animal Science**, v. 68, p. 3502-3506, 1990.

USDA - UNITED STATES DEPARTMENT OF AGRICULTURE. Foreign Agricultural Service – Australia Livestock and Products Semi-Annual 2015. Available at <http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Livestock%20and%20Products%20Semi-annual_Canberra_Australia_3-19-2015.pdf>. Access: 04 September 2015.

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, p.16-24, 2009.

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

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

WANG, P.; DRACKLEY, J. K.; STAMEY-LANIER, J. A.; KEISLER, D.; LOOR, J. J. Effects of level of nutrient intake and age on mammalian target of rapamycin, insulin, and insulin-like growth factor-1 gene network expression in skeletal muscle of young Holstein calves. **Journal of Dairy Science**, v. 97, p. 383-391, 2014.

WILSON, D. E. Application on ultrasound for genetic improvement. **Journal of Animal Science**, v. 70, n. 3, p. 973-983, 1992.

YOKOO, M.J.; LÔBO, R.B.; MAGNABOSCO, C.U.; ROSA, G.J.M.; FORNI, S.; SAINZ, R.D. ANDALBUQUERQUE, L.G. Genetic correlation of traits measured by ultrasound at yearling and 18 months of age in Nellore beef cattle. **Livestock Science**, v. 180, p. 34-40, 2015.

YOKOO, M.J.; ORTELAN, A. A.; SARMENTO, J. L. R.; ROSA, G. J. M.; CARDOSO, F. F.; ALBUQUERQUE, L. G. Medidas repetidas no estudo de características de crescimento e carcaça avaliadas por ultrassom em novilhas de corte cruzadas. **Boletim de Indústria Animal** (Online), v. 71, p. 200-210, 2014.

YOKOO, M. J. **Análise bayesiana da área de olho do lombo e da espessura de gordura obtidas por ultrassom e suas associações com outras características de importância econômica na raça Nelore**. 2009. 84 p. Tese (Doutorado em Genética e Melhoramento Animal) - Faculdade de Ciências Agrárias e Veterinária, Universidade Estadual Paulista, Jaboticabal, SP, 2009.

CHAPTER 2 - Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population

ABSTRACT - Animal feeding is the most important economic component of beef production systems. Selection for feed efficiency has not been effective mainly due to difficult and high costs to obtain the phenotypes. The application of genomic selection using single nucleotide polymorphisms (SNPs) can decrease the cost of animal evaluation as well as the generation interval. However, there is no consensus among researches about the best methodology to obtain genomic prediction for each trait. The objective of this study was to compare methods for genomic evaluation of feed efficiency traits using different cross-validation (CV) layouts in a small beef cattle population, genotyped for a high-density SNP panel (BovineHD *BeadChip* - Illumina). After quality control, a total of 437,197 SNP genotypes were available for 761 Nelore animals from Institute of Animal Science, São Carlos, SP, Brazil. The studied traits were residual feed intake (RFI), feed conversion ratio (FCR), average daily gain (ADG), and dry matter intake (DMI). Methods of analysis were traditional BLUP, single step genomic BLUP (ssGBLUP), genomic BLUP (GBLUP), and a Bayesian regression method (BayesC π). Direct genomic values (DGV) from the last two methods were compared directly or in an index that combines DGV with parent average. Three cross-validation approaches were used to validate the models: 1) YOUNG – the partition into training and testing sets was based on year of birth and testing animals were born after 2010; 2) UNREL – the data set was split into three less related subsets and the validation was done in each subset a time; and 3) RANDOM – the data set was randomly divided into four subsets and the validation was done in each subset at a time. On average, the RANDOM design provided the most accurate predictions. Average accuracies ranged from 0.10 to 0.58 using BLUP, from 0.09 to 0.48 using GBLUP, from 0.06 to 0.49 using BayesC π and from 0.22 to 0.49 using ssGBLUP. The most accurate and consistent predictions were obtained using ssGBLUP for all analyzed traits. The single step genomic BLUP seems to be more suitable to obtain genomic predictions for feed efficiency traits on a small population of genotyped animals.

Keywords: cross-validation, genomic selection, residual feed intake, single nucleotide polymorphisms.

1. INTRODUCTION

The costs associated with feeding represent around 50-70% of the total cost of beef cattle industry, and are even greater in feedlot systems. DiLorenzo and Lamb (2012) reported that selection for feed efficiency decreases the environmental impact of beef industry. According to these authors selecting for feed efficiency based on residual feed intake can reduce 29% of fresh manure output and excretions of phosphorous and nitrogen, while methane emissions can be reduced by as much as 28%.

Selection for feed efficiency traits, using traditional BLUP, is limited by the difficulty and costs to access the phenotypes of interest (CROWLEY et al., 2011). This is one of the reasons for the small improvement observed for these traits in the last years (ZHANG et al., 2011). Genomic selection, using single nucleotide polymorphisms (SNPs), has been used for the improvement of quantitative traits, and may be especially helpful for traits that are hard or expensive to measure and because of that are not routinely recorded (e.g. feed efficiency).

Genomic predictions may combine phenotypic, pedigree, and genotypic information to increase the accuracy of animal evaluation and to reduce the generation interval, which increases the genetic gain (VANRADEN et al., 2009). Once a large number of markers are available, it is very likely to find quantitative trait loci (QTL) on linkage disequilibrium with at least one of them (RESENDE et al., 2008). Recent studies have indicated that there is not a suitable method for predicting genomic values for all traits and populations. Thus, there is a need to conduct studies focused on evaluating which method would be the most suitable for each scenario.

The marker effects can be obtained assuming that all markers contributed equally to genetic variation (no major gene effect), or using a Bayesian approach which assumes different variances on all SNPs. With Bayesian methodology it is assumed that very few SNPs have very high effect and the majority of SNPs have very small or null effect (VANRADEN, 2008). In dairy cattle in the United States, genetic evaluation is performed by multistep methods if genomic information is available (VANRADEN, 2008; VANRADEN et al., 2009). This approach consists of predicting the genomic estimated breeding value (GEBV) by an index combining

parent average (PA) of EBV and direct genomic value (DGV). On the other hand, if phenotypes, pedigrees, and genotypes are available all together, a simple way to incorporate genomic information into evaluations is by the single-step genomic BLUP (MISZTAL et al., 2009). In this procedure, the relationship matrix based on pedigree (**A**) is combined with a genomic relationship matrix (**G**), based on information from SNP markers, into a single matrix of realized relationships (**H**).

The accuracy of genomic prediction is the key to the successful application of genomic selection. However, accuracy is strongly dependent on many factors such as linkage disequilibrium (MEUWISSEN; HAYES; GODDARD, 2001), allele frequency distribution (LETTRE, 2001), number of genotyped animals (VANRADEN et al., 2009; CALUS, 2010; DAETWYLER et al., 2010), heritability of the traits (GODDARD, 2009), effective population size (GODDARD, 2009), marker density (MOSER et al., 2010) and the method used to estimate marker effects (LOURENCO et al., 2014).

The genomic prediction equation cannot be validated in the same animals used for obtaining it. In practice, an important point of genomic selection is to predict the genetic merit of the next generation without their phenotype, using only genomic information. According to Saatchi et al. (2010) and Habier et al. (2010), the number of generations separating training and validation subsets may have influence on the accuracy of prediction. Likewise, many authors have showed concerns about validating the model in a less related population (PÉREZ-CABAL et al., 2012; SAATCHI; WARD; GARRICK, 2013) especially for traits difficult and expensive to measure. As reported by Pérez-Cabal et al. (2012), the genetic relationships between individuals have strong effect on accuracy of prediction. For real and simulated populations with high and low heritabilities, Habier et al. (2007; 2010) observed that when the relationship between training and testing populations decreased prediction accuracy also decreased.

Given the economic importance of feed efficiency traits for livestock industry there is a need to use the most suitable method for genomic evaluation focusing in increasing the accuracy. Also, considering that those traits are hard and expensive to measure and there is not always phenotypes available for them, it is important to

measure how accurate the genomic evaluation would be when it is applied to a less related population.

2. OBJECTIVES

The objective of this study was to compare cross-validation designs and methodologies to predict genomic breeding values for feed efficiency traits in an experimental Nelore cattle population.

3. MATERIAL AND METHODS

3.1 Data

The analyzed Nelore cattle data set was provided by the APTA Beef Cattle Center - Institute of Animal Science (IZ), Sertãozinho, SP, Brazil. This farm has three experimental herds: a selection herd (NeS) which is a closed herd selected for yearling weight since 1978; the traditional herd (NeT) which is submitted to the same selection criterion as NeS but, eventually, receives animals from other herds; and a control herd (NeC) selected for average yearling weight. The estimated annual genetic gain is 0.73% of the phenotypic average in NeS.

The data set contained pedigree information on 9,551 animals (Table 1), of which 896 had phenotypes for all studied traits and 788 (born from 2004 to 2012) of those were genotyped with a high-density SNP chip (Illumina High-Density BovineBeadChip, 777k). Table 1 shows the description of pedigree information which has more than 95% of animals with known sire and dam. The SNP markers with the minor allele frequency (MAF) and call rate less than 5% and 98%, respectively, were deleted. Also, samples with a call rate less than 90% were not considered in the analyses. After genomic data quality control, there were available 437,197 SNP and 761 animals.

Table 1. Structure of pedigree information.

Category	Number of Animals
Animals in total	9551
Sires in total	320
Dams in total	2163
Animals with progeny	2483
Animals with no progeny	7068
Animals with only known sire	16
Animals with only known dam	0
Animals with known sire and dam	9128

Besides the weight gain test, which has been obtained for more than 30 years, the IZ has also been conducting a performance test for feed efficiency since 2005, which made possible to measure many others efficiency traits. Additionally to 80 individual troughs, there are 10 paddocks equipped with GrowSafe® feeding system. The GrowSafe paddocks measure the individual feed intake and feeding behavior even when the animals are kept in groups. In the performance test, the animals were evaluated for individual feed efficiency for at least 56 days (with average of 83.14 ± 14.66 days) preceded by an adaptation period of 28 days in individual ($n=683$) and collective pens equipped with GrowSafe System® ($n=213$). According to Archer and Bergh (1999) feed intake requires approximately 56-70 days to measure accurately, while feed conversion ratio and residual feed intake both required around 70-84 days. The groups of animals that come into the test were separated by sex, with an average of 286.48 ± 38.89 days of age (just after weaning), initial weight of 233.56 ± 48.71 kg, and final weight of 314.16 ± 58.34 . The weight of animals was measured of every 14 days with no previous fasting. The diet is based on corn silage, hay of *Brachiaria*, soy bran, corn bran, salt and urea, with 66.8% of NDT, 13.2% of CP, which allows average daily gain of 1.1Kg/day. The analyzed traits were average daily gain (ADG), dry matter intake (DMI), residual feed intake (RFI), and feed conversion ratio (FCR).

After the performance test, the ADG was obtained by the linear regression on days in test (DIT):

$$y_i = \alpha + \beta * DIT_i + \varepsilon$$

where y_i is weight of i -th animal; α is the intercept of regression equation which represent the initial weight; β is the linear regression coefficient which represent the

ADG; DIT_i is day in the performance test of i^{th} observation and ε is the error associated to each observation. The average metabolic weight ($MW^{0.75}$) was given by:

$$MW^{0.75} = [\alpha + \beta * (DIT)/2]^{0.75}$$

where α and β were already defined above.

The RFI was considered as the error of the linear regression equation of dry matter intake on ADG and metabolic weight within each contemporary group (CG: sex, year of birth, and pen), as showed below:

$$\text{Dry matter intake} = \beta_w * MW^{0.75} + \beta_G * \text{ADG} + \text{error (i.e. RFI)}$$

where β_w and β_G are the linear regression coefficient of metabolic weight and average daily gain, respectively.

The FCR was expressed as the ratio of DMI by ADG.

3.2 (Co) variance component estimation

Variance components were estimated for the feed efficiency traits using an animal model under Bayesian inference. Model for RFI and FCR included fixed effects of contemporary groups, month of birth, age of animal and age of dam as covariable (linear and quadratic effects), and random additive animal effect. Also, the linear effect of two principal components calculated based on genomic relationship matrix (**G**) were considered as covariable to correct for sub-structure of population as suggested by Price et al. (2006). Figure 1 shows the principal components analysis with the sub-structure of analyzed population. The animals showed in blue are from NeC, in red are from NeS, and animals in green are from NeT. The model used for ADG and DMI was the same as used for RFI and FCR, plus the quadratic effect of age of animal as covariable.

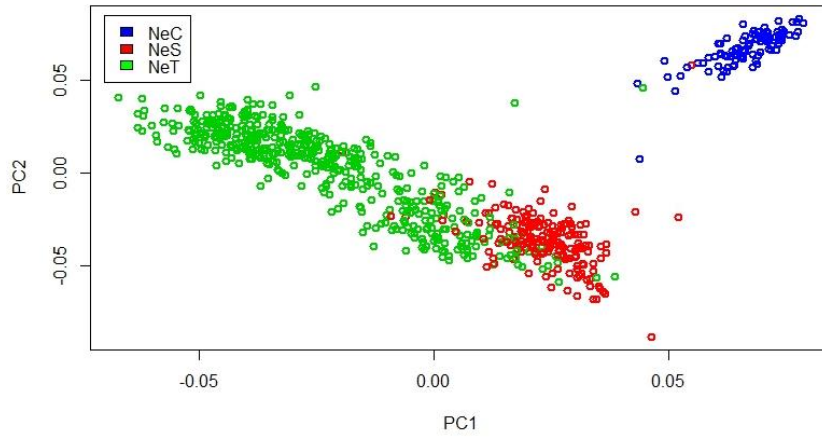


Figure 1. Distribution of animals by herd, provided by principal component analysis using genomic relationship matrix.

Phenotypes, pedigree, and genotypes were used for variance component estimation under single-step genomic BLUP. Thus, in the animal model, the inverse of the numerator relationship matrix ($\mathbf{A}^{-1}$) was replaced by $\mathbf{H}^{-1}$, which combines pedigree and genomic information. Matrix $\mathbf{H}^{-1}$ can be obtained as follows (AGUILAR et al. 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{G}^{-1}$ is the inverse of genomic relationship matrix and $\mathbf{A}_{22}^{-1}$ is the inverse of pedigree-based numerator relationship matrix for genotyped animals.

The general model can be represented as follows:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}.$$

where $\mathbf{Y}$ is the vector of phenotypic observations, $\mathbf{X}$ is a incidence matrix of phenotypes and fixed effects, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{Z}$ is a incidence matrix that relates animals to phenotypes, $\mathbf{a}$ is the vector of direct additive genetic effect, and $\mathbf{e}$ is a vector of residual effects. Assumptions were:

$$E[\mathbf{y}] = \mathbf{Xb}, \text{ var}[\mathbf{y}] = \mathbf{Z}\Sigma\mathbf{Z}' + \mathbf{R}$$

with $\Sigma = \text{var}(\mathbf{a}) = \mathbf{H}\sigma_a^2$ and $\mathbf{R} = \mathbf{I}\sigma_r^2$ in the single-trait model, where $\mathbf{H}\sigma_a^2$ is the additive genetic variance and $\mathbf{I}\sigma_r^2$ the residual variance, $\mathbf{H}$ is the numerator relationship matrix among animals and $\mathbf{I}$ is the appropriate identity matrix. An inverted qui-square distribution was used for the prior values of the direct and residual genetic variances.

The a posteriori conditional distributions of **b**, **a**, and **e** effects were sampled from a multivariate normal distribution.

The analysis consisted of a single chain of 500,000 cycles with a "burn-in" of 100,000 cycles, taking a sample every 10 iterations. Thus, 40,000 samples were used to obtain the parameters. Chain convergence was assessed by visual examination. Analyses were performed using GIBBS2f90 (MISZTAL et al., 2002; AGUILAR et al., 2010). The a posteriori estimates were obtained using the application POSTGIBBSF90 (MISZTAL et al., 2002).

Table 2 shows the additive variances and heritability estimates of the analyzed traits. The estimated variance components indicate that the studied traits are moderately to highly heritable.

Table 2. Additive genetic variance and heritability estimates with standard errors for residual feed intake (kg dry matter/day), feed conversion ratio (Kg dry matter), average daily gain (kg/day), and dry matter intake (kg).

Traits	Mean	SD	Additive genetic variance	Heritability
RFI	0.00	0.58	0.29	0.17±0.07
FCR	7.04	1.77	0.14	0.11±0.06
ADG	1.00	0.26	0.01	0.39±0.08
DMI	6.69	1.24	0.31	0.43±0.08

Mean= average of each trait; SD= Standard deviation; RFI= Residual Feed Intake; FCR= Feed conversion; ADG= Average daily gain; DMI= Dry Matter Intake.

3.3 Methods of Genomic Analysis

The studied methods for genomic analysis were genomic BLUP (GBLUP), single-step genomic BLUP (ssGBLUP) and BayesC π , as described below.

3.3.1 Multistep

3.3.1.1 – Genomic BLUP (GBLUP)

For this multistep analysis, firstly (step **a**) a traditional genetic evaluation was run using a single-trait animal model (the same used to estimate variance components), in order to obtain EBV and fixed effect solutions to estimate adjusted phenotypes. The model can be represented as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where $\mathbf{y}$ is the vector of phenotype, $\boldsymbol{\beta}$ is the vector of fixed effects, $\mathbf{u}$ is the vector of direct additive genetic effect. Considering an infinitesimal model, $\text{var}(\mathbf{u}) = \mathbf{A}\sigma_u^2$, where $\mathbf{A}$ is the numerator relationship matrix obtained from pedigree information, $\text{var}(\mathbf{e}) = \mathbf{I}\sigma_e^2$, and $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices.

The next step (**b**) consisted of obtaining direct genomic value (DGV) based on SNP effects converted from GBLUP solutions by the model showed below:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of phenotype adjusted for fixed effects, μ is the overall mean, $\mathbf{1}$ is a vector of ones, $\mathbf{Z}$ is a incidence matrix of markers effects, $\mathbf{g}$ is a vector of marker effects, and $\mathbf{e}$ is a vector of residual effects. It was assumed $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$, where σ_g^2 is the variance of markers and $\mathbf{G}$ is the genomic relationship matrix. Random residuals were assumed $\mathbf{e} \sim N(0, \mathbf{D}\sigma_e^2)$, where $\mathbf{D}$ is a diagonal matrix and σ_e^2 is the residual variance.

According to VanRaden (2008), the $\mathbf{G}$ matrix can be obtained from at least three ways; for this study we chose the following:

$$\mathbf{G} = \frac{(\mathbf{M} - \mathbf{P})(\mathbf{M} - \mathbf{P})'}{2 \sum_{j=1}^m \mathbf{P}_j (\mathbf{1} - \mathbf{P}_j)}$$

where $\mathbf{M}$ is a matrix of marker alleles with n lines (n =total number of genotyped animals) and m columns (m = total number of markers), and $\mathbf{P}$ is a matrix containing two times the observed frequency of the second allele ($\mathbf{P}_j$). Elements of $\mathbf{M}$ are set to 0 and 2 for both homozygous and to 1 for the heterozygous.

The DGV were calculated for each animal using the following formula:

$$\text{DGV}_i = \sum_{j=1}^p z_{ij}g_j,$$

where g_j is the estimated effect of marker j . The GEBVs of all validation animals were calculated by an index combining parent average and DGV (VANRADEN et al., 2009):

$$\text{GEBV}_i = b_{\text{DGV}}\text{DGV} + b_{\text{PA}}\text{PA}$$

the weights (b) for DGV and PA were obtained as showed by Guo et al. (2010).

3.3.1.2 – BayesC π

Habier et al. (2011) presented a methodology called BayesC π , which assumes that a SNP effect is zero with probability π and this probability could be estimated from the analyzed data. BayesC π assumes a mixed distribution to marker effects and specify a common variance for all *loci* using the same model equation as used in GBLUP, but considering the elements of u as $\sum_{i=1}^N (z_i a_i I_i)$, where z_i is the genotype of i^{th} marker, coded as the number of copies of the reference allele, a_i is the effect of marker i , and I_i is an indicator variable that is equal to 1 if the i^{th} marker has a non-zero effect on the trait and 0 otherwise. In this study a binomial distribution with probability π was assumed for I_i and an informative beta distribution was assigned for π (implying that this parameter was estimated from the analyzed data set, $\alpha=0.10E-01$, $\beta=0.50$).

The prediction equations obtained using GBLUP and BayeC π method were implemented in the GS3 software developed by Legarra et al. (2010), which is available at <http://snp.toulouse.inra.fr/~alegarra>.

3.3.2 Single-step genomic BLUP (ssGBLUP)

The model used in ssGBLUP was the same as used in traditional evaluations. The single-step procedure consists in combining **A** and **G** into a single matrix (**H**) as already described above. The analyses with ssGBLUP were performed using BLUPF90 software available at <http://nce.ads.uga.edu/wiki/doku.php>.

3.4 Cross Validation

Three cross-validation approaches were used to validate the models: 1) RANDOM: the data set was randomly divided into four subsets and the validation was done in each subset at a time. 2) YOUNG: the partition into training and testing sets was based on year of birth and testing animals were born after 2010. This approach was designed mainly to simulate the interest to figure how accurate the prediction of next generation will be. 3) UNREL: the data set was split into three less related subsets and the validation was done in each subset at time. For this design, the training and validation subsets were designed based on K-means approach (DING, HE 2004), which divides the data into less related groups. In this case the

principal component analysis of **G** was used to determine how the folders would be divided. Figure 2 shows which animals were used as training and testing by all folders of cross-validation. The animals in black were in training subset and the animals in gray were in testing subset.

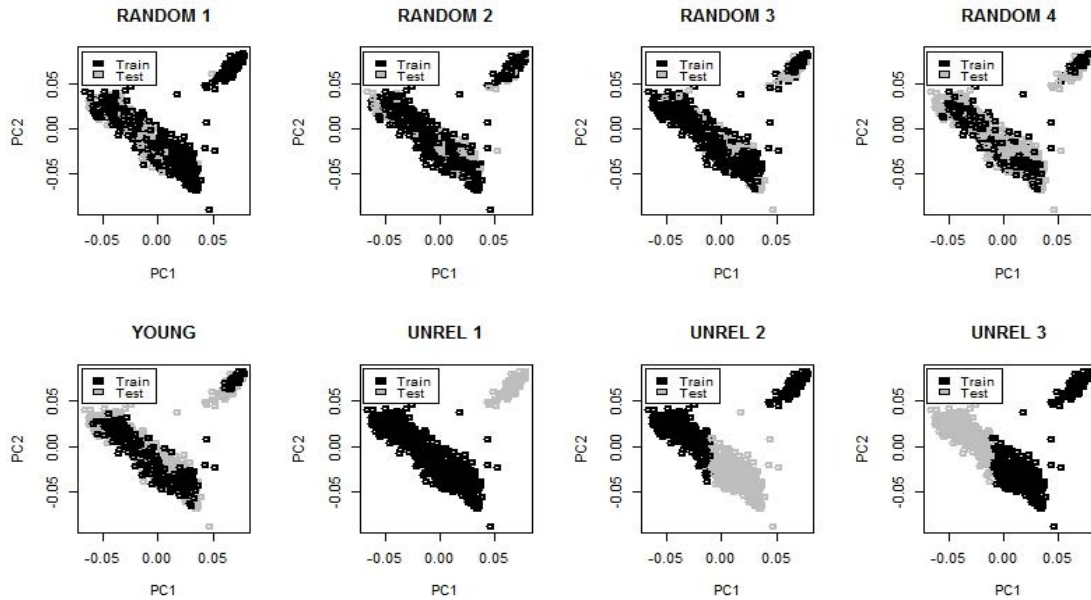


Figure 2. Principal components analysis based on genomic matrix on each subset on UNREL cross-validation design.

As expected, the average relationships between test and training subsets were smaller on UNREL followed by RANDOM and YOUNG (Table 3). The Table 3 shows the number of animals in each cross-validation layout and the proportion of relationship coefficients between the animals in each testing fold. Even though this study used only animals from one experimental farm, the average of all relationship coefficients between training and testing population was not high (around 0.06 for RANDOM and YOUNG).

The relationship coefficients between animals were calculated by CFC software (SARGOLZAEI; IWAISAKI; COLLEAU, 2006), which uses the A matrix.

Table 3. Descriptive statistics of data set used for training and validation, and relationship coefficients (f) on each fold of cross-validation (CV) layout.

CV Layout	Nt	Nv	Relationship coefficients (%)				
			f<0.10	0.10<f<0.25	0.25<f<0.50	f>0.50	Within
RANDOM_1	617	144	86.02	11.39	2.50	0.09	0.09
RANDOM_2	562	199	85.30	12.59	2.01	0.10	0.07
RANDOM_3	592	169	87.37	10.65	1.89	0.09	0.07
RANDOM_4	512	249	85.12	12.63	2.16	0.09	0.07
YOUNG	500	261	85.83	12.85	1.17	0.15	0.07
UNREL_1	670	91	99.58	0.35	0.07	-	0.18
UNREL_2	424	337	95.74	3.47	0.77	0.03	0.10
UNREL_3	428	333	95.75	3.45	0.77	0.03	0.11

Nt=number of animals on training set; Nv=number of animals on validation subset; f= proportion of relationship coefficient between animals in training and validation; Within= average of relationship coefficient within each fold of validation subset.

The accuracy of DGV/GEBV (or EBV for BLUP) was calculated as the Pearson correlation between phenotype adjusted for fixed effect (aY) and the genomic breeding value, divided by square root of heritability (h):

$$\text{acc} = \frac{\text{corr}(\text{aY}, \text{GEBV/DGV})}{h}$$

This adjustment was made to account for the fact that adjusted phenotypes were used instead of true breeding value (PRYCE et al., 2012).

3.5 Regression of phenotype on breeding value (EBV, GEBV, or DGV)

An alternative to evaluate the extent of prediction bias is to compare the regression of aY on predicted breeding value (EBV, GEBV, or DGV), with its expected value of 1 for each trait (SAATCHI et al., 2011). Hence, the regression coefficients were calculated for each trait using simple linear regression of adjusted phenotype on DGV/GEBV.

4. RESULTS AND DISCUSSION

Figure 3 shows accuracies for the feed efficiency traits for all tested genomic methods using the RANDOM validation design. Among the studied methods, ssGBLUP provided more accurate predictions than multistep procedures for all studied traits (Figure 3). The improvements on accuracy of predictions provided by using ssGBLUP were more effective for low heritability traits. It probably means that the inclusion of more than 17% of phenotypic information from ungenotyped animals

added to genomic and phenotypic information from genotyped animals is more effective for those traits. This result could also be due the fact that the information from animals relatives are take in account priority rather than the individual information for low heritability traits. According to Lourenco et al. (2014), ssGBLUP has an advantage over multistep methods mainly because it uses phenotypes rather than pseudo-phenotypes and accounts for the entire population structure to estimate GEBV. Onogi et al. (2014) also concluded that the implementation of genomic selection by ssGBLUP provided more accurate predictions than traditional BLUP for carcass traits even using only genotyped sires of Japanese Black cattle breed. Comparing GBLUP and ssGBLUP in a Holstein population, Aguilar et al. (2010) concluded that genomic evaluations using ssGBLUP were as accurate as those using multistep procedure and its advantage over other methods should increase in the future when the animals are pre-selected by genotypes information.

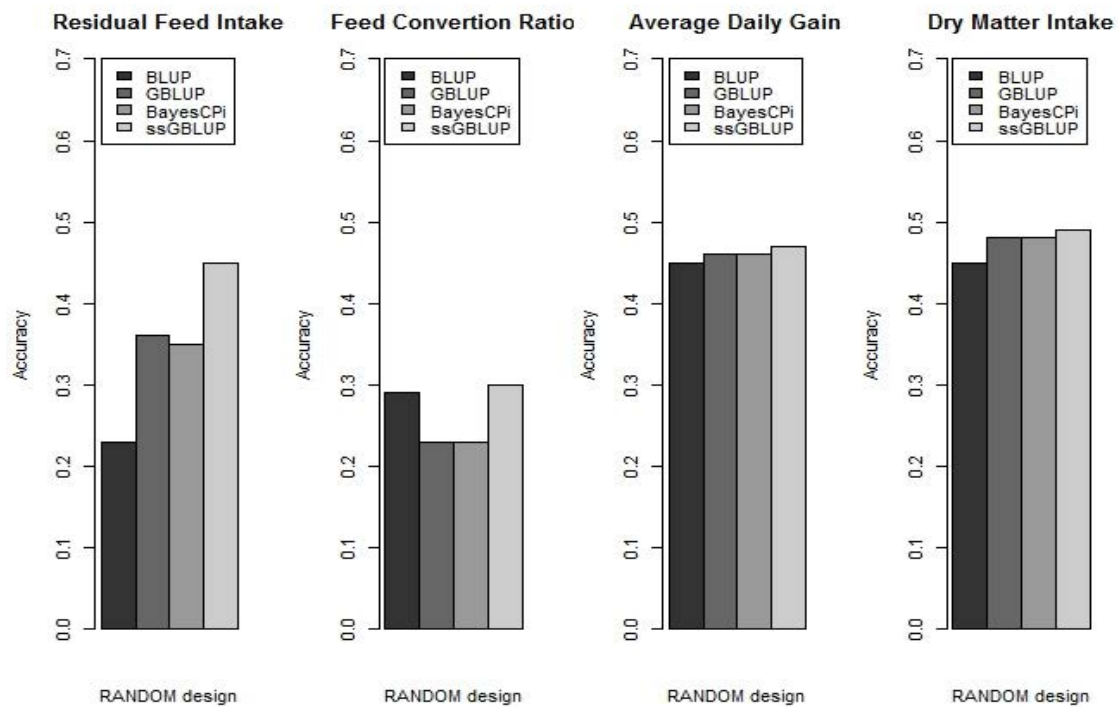


Figure 3. Average accuracies for each studied trait using the RANDOM cross-validation design obtained using BLUP (EBV), genomic BLUP (DGV), BayesCπ (DGV) and single step genomic BLUP (GEBV).

The results also showed that the inclusion of marker information can increase the accuracy of predictions for all studied traits, especially for residual feed intake that had the highest increase in accuracy over traditional BLUP. Higher prediction accuracies were observed for ADG and DMI, which have the highest heritabilities among studied traits ($h^2=0.39$, $h^2=0.43$, respectively), with accuracies ranging from 0.45 to 0.47 and from 0.45 to 0.49, respectively. Similar results were reported by Bolormaa et al. (2013), with most accurate predictions obtained for the highest heritable traits. Also, studying traits with similar heritabilities, Lourenco et al. (2015) reported lower accuracy for the trait that was under strong selection.

An alternative to improve accuracy of genomic prediction is to calculate the GEBV using an index composed by DGV and PA (VANRADEN et al., 2012). Thus, predictions using multiple steps (BayesC π and GBLUP) were calculated either with (GEBV) or without (DGV) such index. Table 4 shows accuracy and bias of DGV/GEBV of studied traits and methodologies. The predictions of GEBV using GBLUP were less accurate than predictions of DGV for most analyzed traits, except for FCR. This probably means that the contribution of parent average is more effective for prediction accuracy of less heritability traits (FCR, $h^2=0.11$). Nonetheless, the bias of GEBV predictions were so much higher, which means that probably this value is underestimated (Neves et al., 2014). The accuracies of GEBV obtained using BayesC π were higher than those for DGV, mostly for the low heritabilities traits (RFI, $h^2=0.17$ and FCR, $h^2=0.11$). Using BayesC π , predictions of GEBV for ADG and DMI were equally accurate to that using single step methodology. However, the bias showed that the predictions using BayesC π of low heritability traits were biased. On the other hand, the estimate of GEBV for traits with high heritability (ADG- $h^2=0.39$ and DMI- $h^2=0.43$) were equally or only a bit more accurate than predictions of DGV. This result differs to those found by Lourenco et al. (2014), which reported greater accuracies for PA in a study using a small genotyped dairy population. However, according to Bjima (2012) accuracy of PA is most strongly reduced by selection. So, once 88% of studied population has undergone selection the accuracy and bias of prediction using an index with PA could probably be affected by selection.

Table 4. Accuracies of DGV/GEBV by studied traits and methodologies by RANDOM cross-validation layout and regression coefficient between adjusted phenotype and DGV/GEBV (between parentheses).

Traits	GBLUP		BayesC π		ssGBLUP
	GEBV	DGV	GEBV	DGV	GEBV
RFI	0.29(1.62)	0.36(0.90)	0.40(2.13)	0.35(1.60)	0.45(1.16)
FCR	0.32(2.92)	0.23(0.78)	0.43(3.82)	0.23(3.10)	0.30(0.99)
ADG	0.44(1.12)	0.46(0.83)	0.46(1.13)	0.46(1.09)	0.47(0.68)
DMI	0.45(1.04)	0.48(0.83)	0.49(1.11)	0.48(1.05)	0.49(0.75)

GBLUP= genomic blup; BayesC π = Bayesian C π methodology; ssGBLUP=single step GBLUP; RFI= residual feed intake; FCR= feed conversion ratio; ADG= average daily gain; and DMI= Dry Matter Intake.

In general, the regression coefficient was close to 1, except for the low heritability traits using DGV from BayesC π which in most analysis was over 1, meaning that predictions were underestimated. Similar results were reported by Neves et al. (2014), where BayesC and Bayesian Lasso provided the most underestimated predictions comparing to GBLUP. It is expected the decrease of bias of prediction with a larger number of genotyped and recorded animals. Previous results with data from this same population but with smaller number of genotyped animals showed higher bias of prediction, especially for the low heritability traits (SILVA et al., 2013).

Among the studied cross-validation designs, the RANDOM provided the most accurate genomic prediction, ranging from 0.23 to 0.49 (Table 5). This probably happened because the RANDOM design had the highest proportion of additive relationships between training and testing over 0.25 (Table 3). Also, in RANDOM design about 2.14% of relationship coefficients between animals on training and testing subsets are between 0.25 and 0.50 (Table 3). The relationship within each fold in RANDOM design was weak. According to Pszczola et al. (2012) higher accuracies are obtained when relationships between animals in the training population are weak and the relationship between the training and validation populations is high. In both subsets (training and testing) animals from different generations were used, which allows validating the model on close relatives and/or validating in animals from the same generation and the same herd. Comparing different cross-validation layouts in a dairy cattle population, Pérez-Cabal et al. (2012) also found the highest accuracies in the RANDOM design and concluded that the number of close relatives on training and testing subsets of cross-validation

influences accuracy even in high or low heritability traits. According to Pryce et al. (2012) and Chen et al. (2013), the ability to predict genomic breeding values within and between population/breed depends on the strength of relationships between all pair-wise combinations of individuals; as higher is the level of genomic relationship among individuals, more accurately genomic breeding values can be predicted in that population.

Table 5. Heritability ($\pm$ standard error), average accuracy ($\pm$ standard error) on BLUP, ssGBLUP, GBLUP (DGV), and BayesC π (DGV) for all studied traits using different cross-validation layout (RANDOM, UNREL and YOUNG).

Traits	h^2	Method	RANDOM	UNREL	YOUNG
RFI	0.17 $\pm$ 0.07	BLUP	0.23 (0.07)	0.10 (0.08)	0.24
		ssGBLUP	0.45 (0.06)	0.29 (0.10)	0.22
		GBLUP	0.36 (0.11)	0.22 (0.08)	0.09
		BayesC π	0.35 (0.10)	0.22 (0.08)	0.06
FCR	0.11 $\pm$ 0.06	BLUP	0.29 (0.08)	0.32 (0.06)	0.30
		ssGBLUP	0.30 (0.05)	0.29 (0.02)	0.31
		GBLUP	0.23 (0.05)	0.10 (0.04)	0.14
		BayesC π	0.23 (0.04)	0.08 (0.05)	0.17
ADG	0.39 $\pm$ 0.08	BLUP	0.45 (0.09)	0.24 (0.01)	0.58
		ssGBLUP	0.47 (0.09)	0.23 (0.03)	0.47
		GBLUP	0.46 (0.10)	0.18 (0.03)	0.54
		BayesC π	0.46 (0.10)	0.17 (0.02)	0.49
DMI	0.43 $\pm$ 0.08	BLUP	0.45 (0.08)	0.27(0.04)	0.51
		ssGBLUP	0.49 (0.06)	0.35 (0.02)	0.48
		GBLUP	0.48 (0.07)	0.32 (0.02)	0.45
		BayesC π	0.48 (0.07)	0.31 (0.01)	0.47

h^2 = heritability; RFI= Residual Feed Intake; FCR= Feed conversation; ADG= Average daily gain; and DMI= Dry Matter Intake.

On the other hand, the general mean accuracy of genomic predictions for young animals (YOUNG design) was intermediate to those for UNREL and RANDOM. The results found by Saatchi et al. (2011) also support results in this study, which agrees that accuracies of genomic prediction on young animals were intermediate to the accuracies obtained from unrelated populations and random clustering for most traits.

The predictions obtained for young animals (YOUNG design) for ADG and DMI were higher or the same as those obtained by RANDOM design. Comparing to RANDOM, the model apparently loses power of predicting GEBV for young animals.

This happened mainly because there was information for animals in the next generations on training and testing subsets in RANDOM which account for more accurate predictions. This result agrees to those obtained for Saatchi et al. (2010) and Habier et al. (2010), which concluded that the number of generations separating training and validation subsets also influences accuracy, with lower accuracies occurring when the relationship is more distant. Also, RANDOM and YOUNG designs had very similar number of animals and also similar relationship between training and testing subsets (Table 3). Considering that RANDOM design is an average of four repetitions with high standard deviation, the value of accuracy for YOUNG design (which had no repetition), in this case, could probably be considered as other repetition of RANDOM.

Indeed, the main reason to study YOUNG cross-validation design is because of the industry interest in predicting the performance for future generations. So, even for a small population, accurately genomic prediction can be achieved for younger animals, especially for high heritability traits (Table 5). Still, even for low heritability traits accuracies as high as 0.31 were obtained.

It is reasonable to assume that the number of animals in testing population can affect the accuracy of prediction (VANRADEN et al., 2009; CALUS, 2010; DAETWYLER et al., 2010). Usually, for traits with a large number of phenotypic information available, such as milk yield and growth traits, accuracies of genomic prediction of 0.8 are currently achievable. The accuracy of genomic prediction of feed efficiency was around 0.30 in beef and dairy cattle studies (BOLOORMA et al., 2013). Much larger reference populations need to be assembled to improve this accuracy. Comparing multistep procedures for feed efficiency traits, Boloorma et al. (2013) reported that traits with a large number of recorded and genotyped animals and with high heritability provide greatest accuracy of GEBV.

The UNREL layout was designed to have the highest relationship within subsets and small relationship between them (Table 3). Over 95% of all relationship coefficients between animals in training and testing subsets were lesser than 0.10, which means that a strong proportion of animals in training were less related to those in testing. On average, genomic predictions obtained in this design were the least accurate, ranging from 0.08 to 0.34 (Table 5). According to Pérez-Cabal et al. (2012),

the number of close relatives on training and testing populations can also affect the accuracy of prediction. In our study, the average accuracy for UNREL was 0.29 using ssGBLUP for RFI, which was not extremely low. This shows how accurate the prediction would be for a population less related to that where the prediction equation was obtained.

Using ssGBLUP for evaluation of small genotyped populations provides the most accurate predictions and should be considered as an option to simplify the evaluations especially for low heritability traits. On average, this method appeared to be suitable for genomic evaluation of a small genotyped population, and it allows using similar model as in traditional evaluations (LOURENCO et al., 2014).

5. CONCLUSIONS

The single-step genomic BLUP seems to be more suitable for obtaining genomic predictions for feed efficiency traits on a small population of genotyped animals.

As more the cross-validation subsets are related more accurately genomic breeding values can be predicted.

The prediction of DGV or GEBV obtained using Bayesian methodology can be biased, especially for low heritability traits.

6. REFERENCES

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

BIJMA, P. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. **Journal of Animal Breeding and Genetics**. 129, 345-358, 2012.

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, p. 7: 3088-3104, 2013.

CALUS, M. P. L. Genomic breeding value prediction: methods and procedures. **Animal**, v. 42, p. 157-164, 2010.

CHEN, L.; SCHENKEL, F.; VINSKY, M.; CREWS JR, D. H.; LI, C. Accuracy of predicting values for residual feed intake in Angus and Charolais beef cattle. **Journal of Animal Science**, v. 91, p. 4669-4678, 2013.

CROWLEY, J. J.; EVANS, R. D.; MC HUGH, N.; PABIOU, T.; KENNY, D. A.; MCGEE, M.; CREWS JR., D. H.; BERRY, D. P. Genetic associations between feed efficiency measured in a performance test station and performance of growing cattle in commercial beef herds. **Journal of Animal Science**, v. 89, p. 3382-3393, 2011.

DAETWYLER, H. D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J. A. The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, p. 1021-1031, 2010.

DIOLORENZO, N.; LAMB, G. C. Environmental and Economic Benefits of Selecting Beef Cattle for Feed Efficiency. UF/IFAS Extension, AN276 Gainesville- FL, 2012.

DING, C AND HE, X. K-means Clustering via Principal Component Analysis. Proc. of Int'l Conf. Machine Learning (ICML 2004): 225–232.

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

GUO, G.; LUND, M. S.; ZHANG, Y.; SU, G. Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. **Journal of Animal Breeding and Genetics**, v. 127, p. 423-432, 2010.

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

HABIER, D.; TETENS, J.; SEEFRIED, F.; LICHTNER, P.; THALLER, G. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics Selection Evolution**, v. 425, 2010.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The Impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, p. 2389-2397, 2007.

LEGARRA, A.; RICARD, A.; FILANGI, O. 2010. GS3–Genomic selection, Gibbs sampling, Gauss Seidel and BayesCπ. Available at: <<http://snp.toulouse.inra.fr/~alegarra/>>. Accessed on: 4 August 2015.

LETTRE, G. Recent progress in the study of the genetics of height. **Human Genetics**, v. 129, p. 465-472, 2011.

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

LOURENCO, D. A.; TSURUTA, S.; FRAGOMENI, B. O.; MASUDA, Y.; AGUILAR, I.; LEGARRA, A.; BERTRAND, J. K.; AMEN, T. S.; WANG, L.; MOSER, D. W.; MISZTAL, I. Genetic evaluation using single-step genomic BLUP in American Angus. **Journal of Animal Science**, v. 93, p. 2653-2662, 2015.

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

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. H. BLUPF90 and related programs (BGF90). In: World Congress Genetics Application Livestock Production, 7th., 2002, Montpellier. **Proceedings...** Montpellier, 2002. p. 28.

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

MOSER, G.; KHATKAR, M. S.; HAYES, B. J.; RAADSMA, H. W. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. **Genetics Selection Evolution**, v. 42, n. 37, 2010.

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

ONOGI, A.; KOMATSU, T.; SHOJI, N.; SIMIZU, K.; KUROGI, K.; YASUMORI, T.; TOGASHI, K.; IWATA, H. Genomic prediction in Japanese Black cattle: application of a single-step approach to beef cattle. **Journal of Animal Science**, v. 95, p. 1931-1938, 2014.

PÉREZ-CABAL, M. A.; VAZQUEZ, A. I.; GIANOLA, D.; ROSA, G. J. M.; WIEGEL, K. A. Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. **Frontiers in Genetics**, v. 3, 2012.

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

PRYCE, J. E.; ARIAS, J.; BOWMAN, P. J.; DAVIS, S. R.; MACDONALD, K. A.; WAGHORN, G. C.; WALES, W. J. WILLIAMS, Y. J.; SPELMAN, R. J.; HAYES, B. J. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. **Journal of Dairy Science**, v. 95, p. 2108-2119, 2012.

PSZCZOLA, M.; STRABEL, T.; MULDER, H. A.; CALUS, M. P. L. Reliability of direct genomic values for animals with different relationships within and to the reference population. **Journal of Dairy Science**, v. 95, p. 389-400, 2012.

RESENDE, M. D. V.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira**, n.56, p. 63-77, 2008.

SAATCHI, M.; MIRAEI-ASHTIANI, S. R.; NEJATI-JAVAREMI, A.; MORADI-SHAHREBABA, M.; MEHRABANI-YEGANEH, H. The impact of information quantity and strength of relationship between training set and validation set on accuracy of genomic estimated breeding values. **African Journal of Biotechnology**, v. 9, p. 438-442, 2010.

SAATCHI, M.; MCCLURE, M.; MCKAY, S. D.; ROLF, M. M.; KIM, J. W.; DECKER, J. E.; TAXIS, T. M.; CHAPPLE, R. H.; RAMEY, H. R.; NORTHCUTT, S. L.; BAUCK, S.; WOODWARD, B.; DEKKERS, J. C. M.; 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, 2011.

SAATCHI, M.; WARD, J.; GARRICK, D. J. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. *J ANIM SCI* 2013, 91:1538-1551. **Journal of Animal Science**, v. 91, p.1538-1551, 2013.

SARGOLZAEI, M.; IWAISAKI, H.; COLLEAU, J. J. CFC: A tool for monitoring genetic diversity. In: World Congress on Genetics Applied To Livestock Production, 8th., 2006, Belo Horizonte. **Proceedings...** Belo Horizonte, 2006. p. 27-28.

SILVA, R. M. O.; TAKADA, L.; BRANCO, R. H.; MERCADANTE, M. E.; CARVALHEIRO, R.; ALBUQUERQUE, L. G. Habilidade de predição genômica para características de consumo e eficiência alimentar em bovinos Nelore. In: Simpósio Brasileiro de Melhoramento Animal, X., 2013, Uberaba. **Proceedings...** Uberaba, 2013.

VANRADEN, P. M.; WRIGHT, J. R.; COOPER, T. A. Adjustment of selection index coefficients and polygenic variance to improve regressions and reliability of genomic evaluations. **Journal of Dairy Science**, v. 95, p. 520, 2012.

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, p.16-24, 2009.

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

ZHANG, Z.; ZHANG, Q.; DING, X. D. Advances in genomic selection in domestic animals. **Chinese Science Bulletin**, v. 56, p. 2655-2663, 2011.

CHAPTER 3 - Genome-wide association study for carcass traits in an experimental Nelore cattle population

ABSTRACT -The purpose of this study was to identify genomic regions associated with carcass traits in a Nelore cattle population. The data set contained 2,306 ultrasound records for *longissimus* muscle area (LMA), 1,832 for backfat thickness (BF), and 1,830 for rump fat thickness (RF). After genomic data quality control, there were available 437,197 SNPs for 721 animals for LMA, 669 for BF and 718 for RF. SNP solutions were estimated by genome-wide association using a single-step genomic BLUP approach (ssGWAS). Variances were calculated for windows of 50 consecutive SNPs. The regions that accounted for more than 0.5% of additive genetic variance were used to search for candidate genes (CG). The results indicated a total of 12, 18, and 15 different windows explained more than 0.5% of the genetic variance for LMA, BF, and RF, respectively. CG associated to expression of steroid hormone biosynthesis and adipose tissue were found in these regions. Many genomic regions were identified associated with the studied carcass traits, which confirms the polygenic nature of them. The results found in this study should help to better understand the genetic and physiologic mechanism associated to LMA, BF, and RF in Zebu animals. Many regions with known genes that have their functions already described were confirmed associate to *longissimus* muscle area, backfat thickness, and rump fat thickness. Nevertheless, the proportion of variance explained by windows of markers was too low to recommend the candidate genes found in this study as selection criterions.

Keywords: *longissimus* muscle area, single nucleotide polymorphisms, single-step gwas, subcutaneous fat

1. INTRODUCTION

The beef cattle production in tropical and subtropical regions is predominantly based on *Bos indicus* (Zebu) breeds and their crosses with *Bos taurus*. Even if Zebu breeds have adaptive advantages to tropical conditions over Taurine breeds, but they have some productivity and quality carcass limitations. The Zebu meat is not well accepted in most demanding markets mainly because their production has been focused on quantitative level without standardization. Thus, due to international consumer demand of better meat quality, the (sub) tropical countries have focused on obtaining phenotypic information and they have done genetic evaluations for carcass traits in its Zebu breeding programs.

The *longissimus* muscle area (LMA) is an indicator of carcass composition and it is related to carcass muscularity (BOGGS; MERKEL, 1990), carcass weight, fat and muscle traits in steers (BERGEN et al., 2005). The subcutaneous fat covering the *longissimus* muscle is an efficient indicator of finishing carcass (HEDRICK, 1983). In this sense, Zebu breeds have disadvantage over Taurine breeds (*Bos taurus*), since the proportion of fat and intramuscular fat percentage in Zebu animals carcasses is lower than *Bos taurus* (CUNDIFF, 2004). Still, favorable genetic correlations between subcutaneous fat and reproductive traits have been reported, indicating that high subcutaneous fat deposition could denote early finishing and result in animals more sexually precocious (CAETANO et al., 2013). The production of fat cover is an important trait in the meat industry especially to protect the carcass after slaughter. The conventional processing of cattle carcasses refrigeration after slaughter may result in tougher meat, thus, an adequate quality carcass must have enough fat covering to guarantee its preservation and desirable quality for consume (CUNDIFF et al., 1993).

Some recent breakthroughs about the use of DNA information has been applied for the improvement of quantitative traits, and may be especially helpful for traits that are hard or expensive to measure and because of that are not routinely recorded. In this sense, genome-wide association analysis (GWAS) has been applied to find associations between genome regions and economical important traits. Recently, studies of livestock breeding have focused on identifying regions in the

genome that are associated with carcass traits (KIM et al., 2011; LU et al., 2013). The success of genetic selection using major SNPs is dependent on the proportion of additive genetic variance explained by each marker or region of the genome (FRAGOMENI et al., 2014).

A common used method for GWAS is based on testing one marker at time as fixed effect (HIRSCHHORN; DALY, 2005). However, other GWAS methods have been reported. An alternative to simplify traditional GWAS consists to integrate all genotypes and phenotypic information available (from genotyped and ungenotyped animals) in one-step procedure (single-step GWAS) that allows the use of any model, and all relationships simultaneously (WANG et al., 2012). According to Fernando and Garrick (2013), the use of Bayesian approach for GWAS has the principal advantage the power of detecting associations. Comparing two iteratively weighted single-step genomic BLUP procedures, a single-marker model, and BayesB, Wang et al. (2014) concluded that the association is strongly dependent on methodologies and details of implementations. The BayesB excessively shrinks regions towards zero while overestimating the amount of genetic variation attributed to other SNP effects.

Considering that the most of genomic studies for carcass traits has been done on Taurine breeds in temperate regions (MUJIBI et al., 2011; PRYCE et al., 2012; ELZO et al., 2012; CHEN et al., 2013; PRYCE et al., 2014) there is a need to study those traits in Zebu breeds under tropical conditions to unveil the genetic architecture of carcass traits in these breeds.

2. OBJECTIVE

The purpose of this study was to identify associations between chromosomal regions to *longissimus* muscle area (LMA), subcutaneous rump fat thickness (RF), and subcutaneous backfat thickness (BF) in an experimental Nelore cattle population aiming to better understand the genetic architecture of those traits.

3. MATERIAL AND METHODS

3.1 Data

The analyzed Nelore cattle data set was provided by the APTA Beef Cattle Center - Institute of Animal Science (IZ), Sertãozinho, SP, Brazil. This farm has three experimental herds: selection herd (NeS) it is a closed herd which has been selected for yearling weight since 1978; traditional herd (NeT) submitted to the same selection criterion as NeS but, eventually, receives animals from other herds; and control herd (NeC): selected for yearling weight mean. The estimated annual genetic gain is 0.73% of the phenotypic average in NeS and NeT. Figure 1 shows the principal components analysis made based on genomic relationship matrix with the sub-structure of analyzed population. The animals showed in blue are from NeC, in red are from NeS, and animals in green are from NeT.

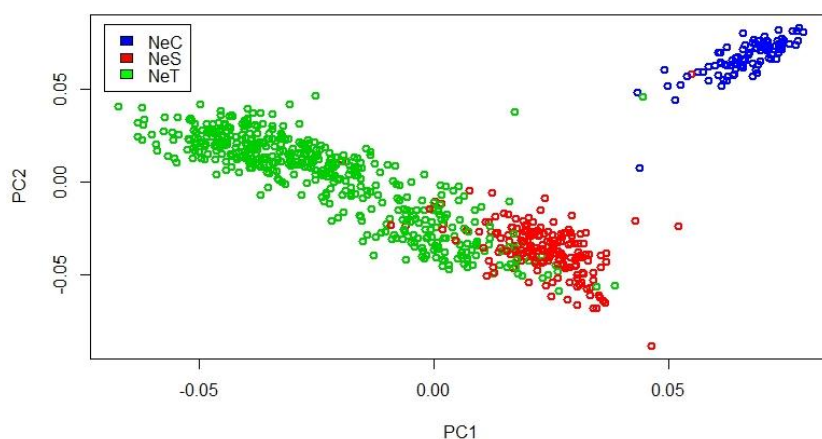


Figure 1. Distribution of animals by herd, provided by principal component analysis using genomic relationship matrix.

The data set contained pedigree information on 9,529 animals, of which had 2,306 ultrasound records for *longissimus* muscle area (LMA), 1,832 for backfat thickness (BF), and 1,830 rump fat thickness (RF), which were born from 1996 to 2013. Table 1 shows the descriptive statistics for the studied traits. The phenotypes were obtained at 12 and 18 months of age for male and female, respectively, mainly because the management used in that farm differs by sex.

The animals were kept on pasture until seven months of age, when they were weaned. After this period, males were submitted to a performance test on feedlot. Females remained on pasture, except for those born in 2004, 2005, 2008, 2009, 2010 and 2011 that were also submitted to performance test on feedlot after weaning.

Table 1. Descriptive statistics for studied carcass traits.

Traits	Np		Ngp		Mean	SD
	Male	Female	Male	Female		
LMA (cm ²)	1,384	922	471	250	47.94	7.06
BF(mm)	1002	830	437	232	1.87	1.16
RF(mm)	916	914	469	249	4.85	1.98

LMA = *longissimus* muscle area; BF = backfat thickness; RF = rump fat thickness; Np= number of animals with phenotypic records; Ngp= number of genotyped animals with phenotypic records; SD= Standard deviation.

The studied traits were measured by ultrasound in males and females between 1995 and 2013. To obtain the ultrasound measurements, the PIE MEDICAL - Aquila equipment with probe of 7 inches and 3,5MHz was used. Then, the images were analyzed using the Echo Image Viewer 1.0 (Pie Medical Equipament B.V., 1996) software. The probe was perpendicularly positioned between the 12^a and 13^a ribs on the left side to obtain the LMA and BF phenotypes. To obtain the RF the probe was positioned at intersection of *Gluteus Medius* and *Biceps Femoris*, which are located between ilium and ischium.

3.2 Genotyping procedure

The high-density SNP chip (BovineHD *BeadChip* - Illumina, 777k) was used for genotyping the animals. SNP markers with minor allele frequency (MAF) and call rate less than 2 and 98%, respectively, were deleted. Also, samples with a call rate less than 90% were not considered in analyses. After genomic data quality control, there were available 437,197 SNPs for 761 animals; however, only 721 had phenotypes for LMA, 669 for BF, and 718 for RF (Table 1).

3.3 (Co) variance component estimation

The (co)variance components and genetic parameters were estimated by Bayesian inference (GIANOLA; FERNANDO, 1986), considering a linear single-trait animal model. Direct additive genetic and residual effects were included as random effects. Analyses were performed using GIBBS2f90 (MISZTAL et al., 2002; AGUILAR et al., 2010). The variance of **a** and **e** are:

$$var = \begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{H}\sigma_a^2 & 0 \\ 0 & \mathbf{I}\sigma_e^2 \end{bmatrix}$$

where σ_a^2 and σ_e^2 are total genetic additive and residual variances, respectively, and **H** is a matrix that combines pedigree and genomic relationships, and its inverse consists on the integration of additive and genomic relationship matrices, **A** and **G**, respectively (AGUILAR et al., 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{A}^{-1}$ is the inverse of relationship matrix based on pedigree information, $\mathbf{G}^{-1}$ is the inverse of genomic relationship matrix, which was constructed as described by VanRaden (2008), and $\mathbf{A}_{22}^{-1}$ is the inverse of pedigree-based relationship matrix for genotyped animals.

The general model can be represented as follows:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e},$$

where **Y** is the vector of phenotypic observations, **X** is a incidence matrix of phenotypes and fixed effects, **b** is the vector of fixed effects, that included sex and contemporary groups (CG: sex, year of birth), and age of ultrasound measurement as covariable, **Z** is a incidence matrix that relates animals to phenotypes, **a** is the vector of direct additive genetic effect, and **e** is a vector of residual effects. Also, the linear effect of two principle components calculated based on genomic relationship matrix (**G**) were considered as covariable to correct for sub-structure of population as suggested by Price et al. (2006).

Assumptions were:

$$E[\mathbf{y}] = \mathbf{Xb}, \text{ var}[\mathbf{y}] = \mathbf{Z}\Sigma\mathbf{Z}' + \mathbf{R}$$

with $\Sigma = \text{var}(a) = H\sigma_a^2$ and $R = I\sigma_r^2$ in the single-trait model, where $H\sigma_a^2$ is the additive genetic variance and $I\sigma_r^2$ the residual variance, H is the numerator relationship matrix among animals and I is the appropriate identity matrix. An inverted qui-square distribution was used for the prior values of the direct and residual genetic variances.

The *a posteriori* conditional distributions of β , a , and e effects were sampled from a multivariate normal distribution. The analysis consisted of a single chain of 300,000 cycles with a "burn-in" of 100,000 cycles, taking a sample every 10 iterations. Thus, 20,000 samples were used to obtain the parameters. Chain convergence was assessed by visual examination. The *a posteriori* estimates were obtained retrospectively using the POSTGIBBSF90 program (MISZTAL et al., 2002).

3.4 Single Step Genome Wide Association (ssGWAS)

The analyses were performed using the single-step GWAS methodology (WANG et al., 2012) considering the same linear animal model used to estimate the (co) variance components described above. The animal effects were decomposed in genotyped (a_g) and ungenotyped (a_n) animals as describe by Wang et al., (2012), with the animal effect of genotyped animal:

$$a_g = Zu,$$

where Z is a matrix that relates genotypes of each locus and u is a vector of marker effects. The variance of animal effects was assumed as:

$$\text{var}(a_g) = \text{var}(Zu) = ZDZ'\sigma_u^2 = G^*\sigma_a^2,$$

where D is a diagonal matrix of weights for variances of markers ($D=I$ for GBLUP), σ_u^2 is the genetic additive variance captured by each SNP marker when no weights are present and G^* is the weighted genomic relationship matrix.

Thus, the SNP effects were obtained following equation, as described by Wang et al. (2012):

$$\hat{u} = \lambda DZ'G^*^{-1}\hat{a}_g = DZ'[ZDZ']^{-1}\hat{a}_g$$

where λ is a variance ratio or a normalizing constant. According to VanRaden et al. (2009),

$$\lambda = \frac{\sigma_u^2}{\sigma_a^2} = \frac{1}{\sum_{i=1}^M 2pi(1-pi)},$$

where M is the number of SNPs and p_i is the allele frequency of the second allele of the i^{th} SNP.

The following iterative process described by Wang et al. (2012) was used considering D to estimate the SNP effects:

1. $D=I$;
2. To calculate GEBVs for all animals in data set using ssGBLUP;
3. To calculate the SNP effect: $\hat{u} = \lambda DZ'G^*^{-1}\hat{a}_g$;
4. To calculate the variance of each SNP: $d_i = \hat{u}_i^2 2p_i(1 - p_i)$, where i is the i -th marker;
5. To normalize the values of SNPs to keep constant the additive genetic variance;
6. To calculate the G matrix;
7. Exit, or loop to step 2.

The effects of markers were obtained by 2 iterations from step 2 to 6 as showed by Wang et al. (2012). The percentage of genetic variance explained by i -th region was calculated as described by WANG et al. (2014):

$$\frac{Var(a_i)}{\sigma_a^2} \times 100\% = \frac{Var(\sum_{j=1}^{50} Z_j \hat{u}_j)}{\sigma_a^2} \times 100\%$$

where a_i is genetic value of the i -th region that consists of contiguous 50 consecutive SNPs, σ_a^2 is the total genetic variance, Z_j is vector of SNP content of the j -th SNP for all individuals, and $\hat{u}_j$ is marker effect of the j -th within the i -th region.

Analyses were performed using BLUPF90 family software (MISZTAL et al., 2002) modified to include genomic information (AGUILAR et al., 2010). The results were presented by the proportion of variance explained by each window of 50 SNPs with average of 280Kb.

3.5 Search for Associated Genes

The chromosome segments that explained more than 0.5% of additive genetic variance were selected to explore and determine possible quantitative trait loci. The Map Viewer of bovine genome was used for identification of genes, available at "National Center for Biotechnology Information" (NCBI - <http://www.ncbi.nlm.nih.gov>)

in UMD3.1 version bovine genome and Ensembl Genome Browser (<http://www.ensembl.org/index.html>). In these databanks genes were identified in those chromosome regions that explained more than 0.5% of additive genetic variance of studied traits. The classification of genes for biological function, identification of metabolic pathways and enrichment of genes was performed on the website "The Database for Annotation, Visualization and Integrated Discovery (DAVID) v. 6.7" (<http://david.abcc.ncifcrf.gov/>), GeneCards (<http://www.genecards.org/>), and UniProt (<http://uniprot.org>).

4. RESULTS AND DISCUSSION

The posterior heritabilities estimated in this study are in Table 2. The heritabilities' confidence interval of all studied traits is very low, which can indicate a good adjustment of **H** matrix for the structure of analyzed population.

Table 2. Descriptive statistics analysis for studied carcass traits.

Traits	σ_a^2	σ_e^2	h^2	95% Confidence Interval	
				CL	CU
LMA	17.94	19.80	0.47±0.05	0.44	0.51
BF	0.21	0.56	0.28±0.05	0.24	0.31
RF	0.80	1.75	0.31±0.05	0.28	0.35

LMA = *longissimus* muscle area; BF = backfat thickness; RF = rump fat thickness; σ_a^2 =additive genetic variance; σ_e^2 = residual variance; h^2 = heritability; CL, 95%= lower limit for 95% confidence intervals; CU, 95%= upper limit for 95% confidence intervals

Since, the heritability estimates for studied traits show that a great part of total phenotypic variance of these traits is due to genes effects, it is important to know whether there is some major genes. The known genes found in the regions that accounted for more than 0.5% of additive genetic variance were used to search for candidate genes (CG) which are presented in Tables 3-5 according to the studied trait. The results indicated a total of 12, 18 and 15 different windows with known genes explaining more than 0.5% of the genetic variance for LMA, RF, and BF, respectively. Also, many uncharacterized genes (LOC) were found in the regions associated to all studied traits. Thus, this information about those uncharacterized genes might be helpful for posterior studies. However, as mentioned by Fragomeni et al. (2014), GWAS results should be carefully interpreted avoiding to determine an association as a causative effect since many QTL have been described for many traits but just a few of them have been validated by others studies.

Among the regions associated to LMA, 28 known genes were found in those windows that explained more than 0.5% of additive genetic variance (Figure 2 and Table 3). In addition, 15 uncharacterized regions were found in the regions associated to LMA. The ALKBH3 gene (BTA15) codifies an intrinsic DNA repair protein that suppresses transcription associated DNA damage at highly expressed genes (LIEFKE et al., 2015). Even though this gene has not been associated for those traits in a cattle population, Nay et al. (2012) reported that deleting it made mice's fibroblasts more susceptible to death. Thus, it is reasonable to assume that this gene may be acting on the number of muscle fiber and consequently providing bigger or smaller *longissimus* muscle area.

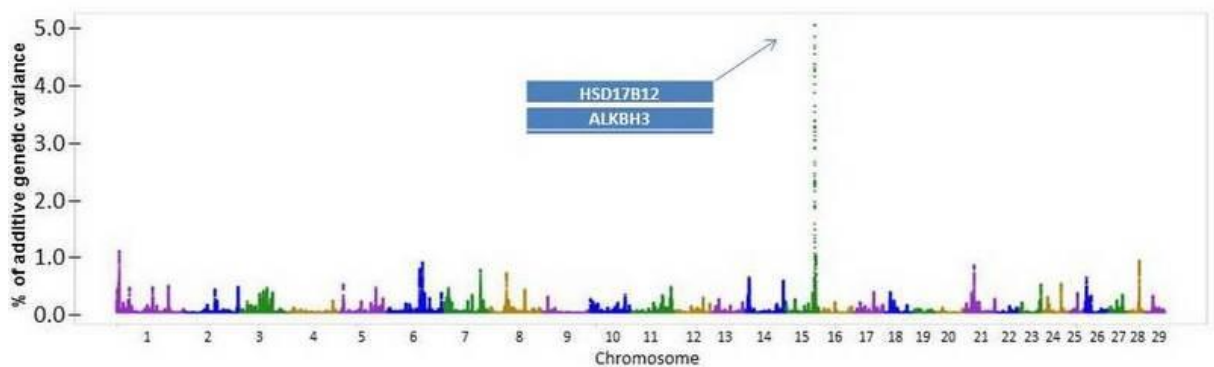


Figure 2. Manhattan plot of additive genetic variance explained (axis y) by windows of 50 adjacent SNPs distributed by chromosomes for *longissimus* muscle area in Nelore.

The results of enrichment analysis for LMA showed that the first cluster is related to processes that decrease the rate of cell death and any process that stops, prevents or reduces the rate of cell death by apoptotic process (Appendix A). Also, the second cluster was for DNA repair, mechanisms that minimize acute damage to the cell's overall integrity and response to DNA damage stimulus. Both processes are dependent to each other, since fail in DNA repair increases the rate of cell death.

Located in the *Bos taurus* chromosome 15 (BTA15), HSD17B12 gene is a member of the hydroxysteroid dehydrogenase superfamily, involved in the metabolism of steroids, retinoids, bile and fatty acids which apparently can be related in metabolic pathways involved in tumor (VISUS et al., 2011). Considering that tumor is the results of uncontrolled cells multiplication, it could be related to DNA repair.

Table 3. Identification of genes based on additive genetic variance explained by windows of 50 adjacent SNPs for *longissimus* muscle area.

Chromosome	Position (bp)	Genes*	Var (%)
BTA1	5646552 - 5785986	GRIK1 LOC101907301	1.06
BTA1	4012541 - 4421333	KRTAP7-1 LOC101906373 LOC101907866 LOC101907950 LOC785105	0.51
BTA6	69231499 - 69368947	CWH43 DCUN1D4 LOC101905687	0.75
BTA7	89201585 - 89448769	LOC104968974 RASA1 CCNH	0.74
BTA8	31762825 - 31969601	LOC782470 LOC104969317	0.68
BTA14	80670924 - 80817593	RALYL	0.55
BTA15	74119235 - 74308370	LOC104974311 API5	0.61
BTA15	74571599-74924022	HSD17B12 ALKBH3 C15H11orf96	4.93
BTA15	76538448 - 76894000	CHST1 LOC104974325 LOC104974324 SLC35C1 CRY2 MAPK8IP1 C15H11orf94 PEX16 GYLTL1B PHF21A	0.98
BTA21	21279610 - 21595190	RLBP1 RHCG LOC104975343 TICRR KIF7 PLIN1 PEX11A WDR93 LOC104975344	0.82
BTA24	48873432 - 49085371	CTIF	0.50
BTA28	33644873 - 33873100	LOC101907374 DLG5	0.89

*Official gene symbol (assembly UMD_3.1, annotation release 103); Var: Additive genetic variance explained by the 50 adjacent SNP windows.

For the analysis of enrichment of genes for BF the clusters obtained by DAVID software are related to ATP- binding, transcription regulation, protein amino acid phosphorylation (Appendix B). Similarly to what happened for RF, many genes that are related to lipids and fat expression were found in windows that explained more than 0.5% of additive genetic variance of BF (Figure 3). Among those genes, the HTR2B gene, located on BTA2 (Table 4) has been also associated to fat deposition in humans by Sohle et al., 2012. This gene encodes one of the several different receptors for 5-hydroxytryptamine (serotonin). The serotonin hormone functions as a neurotransmitter, a hormone, and a mitogen. Two CG genes were found in the region that explained the greatest part of genetic variance (BTA29): ALDH3B1 and CHKA. The ALDH3B1 gene that participates on human lipids metabolism (KITAMURA et al., 2013) is associated to diabetes in humans (JEFF et al., 2014). The CHKA gene is also associated to lipid metabolism. Gabás-Rivera et al. (2013) reported that its expression increased in mice liver which had been fed with high fatty acids levels.

Table 4. Identification of genes based on additive genetic variance explained by windows of 50 SNPs for backfat thickness.

Chromosome	Position (bp)	Genes*	Var (%)
BTA1	27032026 - 27189437	ARHGAP31 TMEM39A POGLUT1 TIMMDC1 LOC104970865 CD80	0.64
BTA2	99181462 - 99535059	LOC101906862	0.51
BTA2	119720331 - 119790546	HTR2B PSMD1 ARMC9	1.35
BTA7	104518902 - 104847810	LOC786544 GIN1 LOC101905525 PPIP5K2 LOC101905593 C7H5orf30 LOC786544	0.70
BTA9	40405496 - 40651230	METTL24 CDC40 WASF1 LOC104969538	0.51

Chromosome	Position (bp)	Genes*	Var (%)
BTA10	41293230 - 41672195	LOC104973133	0.54
BTA11	26550217 - 26891283	SLC3A1 LOC101906743 PREPL CAMKMT	0.52
BTA11	70031657 - 70122506	LOC101905929 LOC104968430	0.66
BTA14	21074886 - 21224382	PRKDC MCM4 LOC100139903	1.62
BTA14	22714764 - 22909901	PCMTD1 LOC104974020 ST18	0.70
BTA14	24376195 - 24543370	XKR4	0.61
BTA14	24874608 - 25102663	LYN RPS20 MOS PLAG1 CHCHD7	1.89
BTA14	25203669 - 25492467	PENK LOC101907667	0.62
BTA16	60839844 - 61165403	LOC100299281 LOC101902436 LOC104972944 MIR2285X RASAL2 LOC101902850	0.74
BTA21	6054709 - 6291692	LOC104975304 LOC100301305 LOC101907131 ASB7 LINS CERS3	0.66
BTA29	46115170 - 46387810	DOC2G NUDT8 TBX10 LOC508879 UNC93B1 ALDH3B1 NDUFS8 TCIRG1 CHKA SUV420H1 LOC104976291	2.59

*Official gene symbol (assembly UMD_3.1, annotation release 103); Var: Additive genetic variance explained by the 50 SNP windows.

The XKR4 gene located on BTA14 was identified into a region associated to BF in this study, also, this gene was previously reported to be associated with rump fat thickness in a recent genome-wide association study in Belmont Red and Santa

Gertrudis breed by Porto Neto et al. (2012). These authors found three SNPs within XKR4 gene significantly ($P < 0.001$) associated with subcutaneous rump fat thickness, which only rs41724387 explained 5.9% of additive genetic variance. This gene has also been reported as a candidate gene due to associations with residual feed intake, average daily feed intake and average daily gain (BOLORMMA et al., 2011; LINDHOLM-PERRY et al., 2012). Also, Bastion et al. (2014) suggested that the XKR4 gene participates in the regulation of prolactin secretion in cattle.

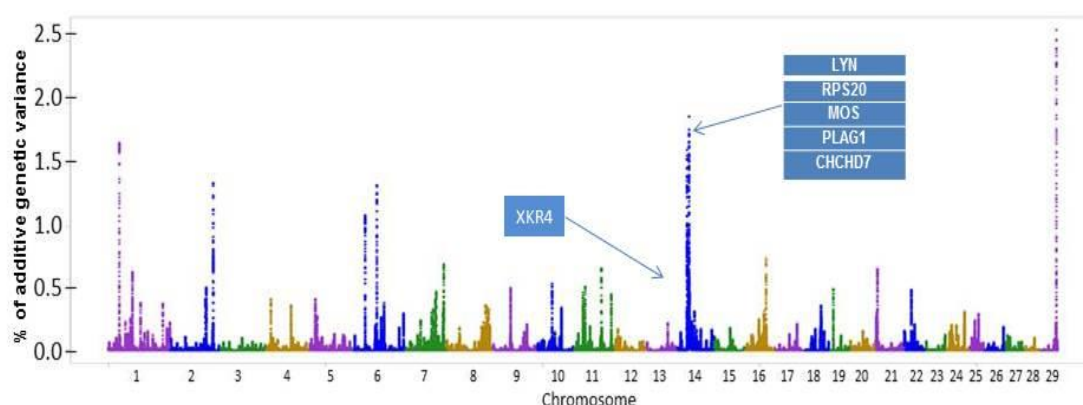


Figure 3. Manhattan plot of additive genetic variance explained (axis y) by windows of 50 adjacent SNPs distributed by chromosomes for backfat thickness in Nelore.

The PLAG1, CHCHD7, MOS, RPS20, LYN, RDHE2 (SDR16C5) and PENK genes have been previously reported associated to several important traits in bovine by many independent studies (KARIM et al., 2011; LITTLEJOHN et al., 2012; NIHSHIMURA et al., 2012). This region was associated with both BF and RF (Table 4 and Table 5). Many studies have reported association between PLAG1 gene with fat in beef cattle (FORTES et al., 2013a), beyond some others carcass traits, including Nelore breed (MAGALHÃES, 2015) and some others breeds (NISHIMURA et al., 2012; LEE et al., 2013; HOSHIBA et al., 2013; SHARMA et al., 2014; SAATCHI et al., 2014). Furthermore, this gene has been associated to expression of feed intake, performance and reproductive traits in many others bovines breeds (LITTLEJOHN et al., 2011; RILEY et al., 2013; FORTES et al., 2013a; FORTES et al., 2013b; SAATCHI et al., 2014; UTSUNOMIYA et al., 2014).

It is important to mention that the PENK gene, that was associated to RF in a swine population (JIAO et al., 2014), is located pretty close to the PLAG1 gene in

bovine. Thus, it is hard to say whether these genes has independent function or the window that PENK is located, in this case, is in linkage disequilibrium with a PLAG1 gene window, once they are juxtaposed and with possible physical disequilibrium.

Some other genes that were not associated to BF (except for the window that PLAG1 gene is placed) were found in the regions that explained part of additive genetic variance of rump fat thickness. These results agree to those reported by Li et al. (2012) that concluded that the place of fat deposition is under epigenetic effect. This means that the genes that act depositing lipids in several regions may or may not be the same because depending on the region that the genes are located they might or not be methylated.

For the analysis of enrichment of genes for RF the clusters obtained by DAVID software are related to ATP binding, response to hormone stimulus and intracellular signaling cascade (Appendix A). As a polygenic trait, many regions in bovine genome have some influence on expression of RF (Figure 2). In those regions were found 20 known genes (located in BTA2, BTA5, BTA6, BTA9, BTA13, BTA14, BTA15, BTA19 and BTA20 chromosomes) which some of them have been reported to be associated to phenotypes of production of lipids (Table 4). According to Saez et al. (2007), CAPN5 gene (BTA15) is associated to blood cholesterol level in humans. In this same region is located the MYO7A gene which has been reported being associated with abdominal fat in chicken. Also, the LUM gene (BTA5) was related with omental fat deposition in humans (OLIVA et al., 2013). The BAI3 gene, placed in a region of BTA9 acts on the fusion of myoblast and it is expressed on extracellular matrix (HAMOUND et al., 2014) that participates in the formation of adipose tissue.

Some others genes had their expression identified in skeletal muscle and performance capability on adipose tissue, since that tissue are juxtaposed and have many genes with common acting. The RPS6KB1 gene (BTA19) is sub expressed in skeletal muscle of calves which were experimental fed with high levels of protein (WANG et al., 2014). The DCN gene (BTA5) operates in *intramyocellular* space co-acting with myostatin gene (ALBRECHT et al., 2011) and also, it is in extracellular matrix with highly potential to contribute for fat deposition (DOUGLAS et al., 2006).

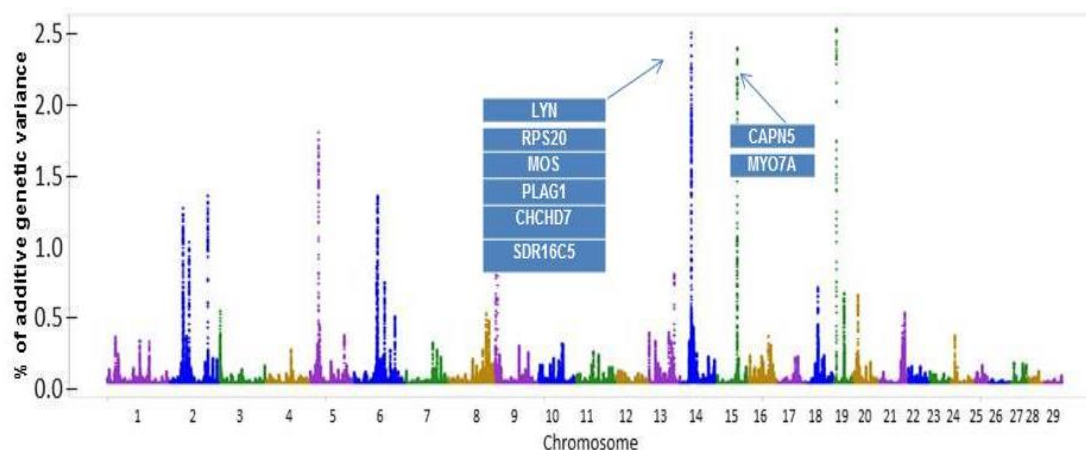


Figure 4. Manhattan plot of additive genetic variance explained (axis y) by windows of 50 adjacent SNPs distributed by chromosomes for rump fat thickness in Nelore.

Table 5. Identification of genes based on additive genetic variance explained by windows of 50 SNPs for rump fat thickness.

Chromosome	Position (Bp)	Genes*	Var (%)
BTA2	33484646 - 33713607	KCNH7	1.19
BTA2	106280299 - 106622973	TRNAC-ACA	1.27
BTA2	34215251- 34433259	IFIH1 FAP GCG	0.65
BTA2	50272472 - 50597681	LOC510454 LOC100295230	0.96
BTA5	20993029 - 21117110	KERA TRNAC-ACA LUM DCN	1.70
BTA6	53086533 - 53397963	LOC100139828 LOC788223	0.79
BTA6	54794939 - 54953177	LOC104968862 LOC104968863	1.22
BTA6	55345470 - 55570440	LOC100296974 LOC100296505 LOC104968886	1.27
BTA9	8440022 - 8597025	BAI3	1.03
BTA13	69726309 - 69962829	LOC104973877	0.74
BTA14	24877166 - 25105265	LYN RPS20 MOS PLAG1 CHCHD7 SDR16C5	2.38
BTA14	25203669 - 25492467	PENK LOC101907667	1.93
BTA15	57005358 - 57277927	LOC104974276 ACER3	1.00

Chromosome	Position (bp)	Genes*	Var (%)
BTA15	57005358 - 57277927	LOC786726 LOC104974277 B3GNT6 CAPN5	1.00
BTA15	57296895 - 57435696	CAPN5 MYO7A LOC786996	2.28
BTA19	10837090 - 11303460	CLTC PTRH2 VMP1 LOC104974991 MIR21 LOC104968960 TUBD1 RPS6KB1 RNFT1 LOC101907104	2.41
BTA19	31689778 - 31937809	MYOCD LOC104975039 ARHGAP44	0.58
BTA19	32305077 - 32647787	HS3ST3A1 LOC104975040	0.61
BTA20	16231525 - 16430411	LOC104975227	0.60

*Official gene symbol (assembly UMD_3.1, annotation release 103); Var: Additive genetic variance explained by the 50 SNP windows.

Considering that *longissimus* muscle area, backfat thickness and rump fat thickness are governed by variants with small effects, the results found in this study confirm the polygenic effect of these traits, since many regions close to important QTL were found to be associated with studied phenotypes. Also, many uncharacterized genes were identified in those regions. The genes identified in this study have been already reported to be associated with these traits. Thus, some of those genes could probably be considered as candidate genes for genetic selection and future investigations are needed to better characterize them and their respective functions in the bovine.

The results found should help to better understand the genetic and physiologic mechanism that regulates the muscle tissue deposition and subcutaneous fat cover expression of Zebu animals. The analyzed data set contained information from proved animals which have progenies participating in many breeding programs in several regions of the country. Thus, the identification of those genes should contribute for animal breeding programs of Zebu breeds using carcass traits as selection criteria.

5. CONCLUSION

Many regions with known genes that have their functions already described were confirmed associate to *longissimus* muscle area, backfat thickness, and rump fat thickness. Nevertheless, the proportion of variance explained by windows of markers was too low to recommend the candidate genes found in this study as selection criterions. In addition, studies focusing on identifying the role of many uncharacterized genes associated to studied traits should be required to better understanding the genetic architecture of those traits.

6. ACKNOWLEDGMENTS

The costs associated to genotyping of animals and purchase of the GrowSafe System was supported by the thematic project "Genomic Tools for the genetic improvement of economically important traits directly in Nelore cattle" (FAPESP grant number 2009/16118-5).

Center APTA Beef Cattle - Institute of Animal Science (IZ), São Carlos, SP, Brazil for provide the dataset.

7. REFERENCES

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

ALBRECHT, E.; LIU, X.; YANG, X.; ZHAO, R.; JONAS, L.; MAAK, S. Colocalization of myostatin and decorin in bovine skeletal muscle. **Archiv Tierzucht**, v. 54, n. 2, p. 147-156, 2011.

BASTIN, B.C.; HOUSER, A.; BAGLEY, C. P.; ELY, K. M.; PAYTON, R. R.; SAXTON, A. M.; SCHRICK, F. N.; WALLER, J. C.; KOJIMA, C. J. A polymorphism in XKR4 is significantly associated with serum prolactin concentrations in beef cows grazing tall fescue. **Animal Genetics**, v. 45, n. 3, 2014.

BERGEN, R.; MILLER, S. P.; WILTON, J. W. Genetic correlations among indicator traits for carcass composition measured in yearling beef bulls and finished feedlot steers. **Canadian Journal of Animal Science**, v. 85, n. 4, p. 463-473, 2005.

BOLORMAA, S.; HAYES, B. J.; SAVIN, K.; BARENDSE, W.; ARTHUR, P. F.; HERD, R. M.; GODDARD, M. E. Genome-wide association studies for feedlot and growth traits in cattle. **Journal of Animal Science**, v. 89, p. 1684-1697, 2011.

BOGGS, D. L.; MERKEL, A. R. **Live animal carcass evaluation and selection manual**. 3 ed. Dubuque, Iowa, Kendall/Hunt Publishing Co., 1990. 211p.

CAETANO, S. L.; SAVEGNAGO, R. P.; BOLIGON, A. A.; RAMOS, S. B.; CHUD, T. C. S.; LÔBO, R. B.; MUNARI, D. P. Estimates of genetic parameters for carcass, growth and reproductive traits in Nelore cattle. **Livestock Science**, v. 155, n. 1, p. 1-7, 2013.

CHEN, L.; SCHENKEL, F.; VINSKY, M.; CREWS, D. H.; LI, C. Accuracy of predicting genomic breeding values for residual feed intake in Angus and Charolais beef cattle. **Journal of Animal Science**, v. 91, p. 4669-4678, 2013.

CUNDIFF, L. V.; KOCH, R. M.; GREGORY, K. E.; CROUSE, J. D.; DIKEMAN, M. E. Characteristics of Diverse Breeds in Cycle IV of the Cattle Germplasm Evaluation Program Breeds and genetics. University of Nebraska - Lincoln. 2004.

DOUGLAS, T.; HEINEMANN, S.; BIERBAUM, S.; SCHARNWEBER, D.; WORCH, H. Fibrillogenesis of Collagen Types I, II, and III with Small Leucine-Rich Proteoglycans Decorin and Biglycan. **Biomacromolecules**, v. 7, p. 2388-2393, 2006.

ELZO, M. A.; LAMB, G. C.; JOHNSON, D. D.; THOMAS, M. G.; MISZTAL, I.; RAE, D. O.; MARTINEZ, C. A.; WASDIN, J. G.; DRIVER, J. D. Genomic-polygenic evaluation of Angus-Brahman multibreed cattle for feed efficiency and postweaning growth using the Illumina 3K chip. **Journal of Animal Science**, v. 90, p. 2488-2497, 2012.

FERNANDO, R. L.; GARRICK, D. Bayesian methods applied to GWAS. **Methods in Molecular Biology**, v. 1019, p. 237-274, 2013.

FORTES, M. R. S.; REVERTER, A.; KELLY, M.; MCCULLOCH, R.; LEHNERT, S. A. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. **Andrology**, v. 1, p. 644–650, 2013a.

FORTES, M. R. S.; KEMPER, K.; SASAZAKI, S.; REVERTER, A.; PRYCE, J. E.; BARENDSE, W.; BUNCH, R.; MC CULLOCH, R.; HARRISON, B.; BOLORMAA, S.; ZHANG, Y. D.; HAWKEN, R. J.; GODDARD, M. E. and LEHNERT, S. A. Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. **International Foundation for Animal Genetics**, v. 44, p. 636–647, 2013b.

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

GABÁS-RIVERA, C.; MARTÍNEZ-BEAMONTE, R.; RÍOS, J. L.; NAVARRO, M. A.; SURRA, J.C.; ARNAL, C.; RODRÍGUEZ-YOLDI, M. J.; OSADA, J. Dietary oleanolic acid mediates circadian clock gene expression in liver independently of diet and animal model but requires apolipoprotein A1. **The Journal of Nutritional Biochemistry**, v. 24, p. 2100-2109, 2013.

HAMOUND, N.; TRANA, V.; CROTEAUA, L.; KANIAA, A.; CÔTÉ, J. G-protein coupled receptor BAI3 promotes myoblast fusion in vertebrates. **PNAS**, v. 111, p. 3745-3750, 2014.

HEDRICK, H.B. Methods of estimating live animal and carcass composition. **Journal of Animal Science**, Champaign. v.57, n.5, p.1316-26, 1983.
HIRSCHHORN, J. N.; DALY, M. J. Genome-wide association studies for common diseases and complex traits. **Nature Reviews Genetics**, v. 6, p. 95-108, 2005.

HOSHIBA, H.; SETOGUCHI, K.; WATANABE, T.; KINOSHITA, A.; MIZOSHITA, K.; SUGIMOTO, Y.; TAKASUGA, A. Comparison of the effects explained by variations in the bovine PLAG1 and NCAPG genes on daily body weight gain, linear skeletal measurements and carcass traits in Japanese Black steers from a progeny testing program. **Animal Science Journal**, v. 84, p. 529-534, 2013.

JEFF, J. M.; ARMSTRONG, L. L.; RITCHIE, M. D.; DENNY, J. C.; KHO, A. N.; BASFORD, M. A.; WOLF, W. A.; PACHECO, J. A.; LI, R.; CHISHOLM, R. L.; RODEN, D. M.; HAYES, M. G; CRAWFORD, D. C. ADMIXTURE Mapping and Subsequent Fine-Mapping Suggests a Biologically Relevant and Novel Association on Chromosome 11 for Type 2 Diabetes in African Americans. **PLoS One**, v. 9, 2014.

JIAO, S.; MALTECCA, C.; GRAY, K. A.; CASSADY, J. P. Feed intake, average daily gain, feed efficiency, and real-time ultrasound traits in Duroc pigs: II. Genome-wide association. **Journal of Animal Science**, v. 92, p. 2846-2860, 2014.

KARIM, L.; TAKEDA, H.; LIN, L.; DRUET, T.; ARIAS, J.A.; BAURAIN, D.; CAMBISANO, N.; DAVIS, S.R.; FARNIR, F.; GRISART, B.; HARRIS, B.L.; KEEHAN, M.D.; LITTLEJOHN, M.D.; SPELMAN, R.J.; GEORGES, M.; COPPIETERS, W. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. **Nature Genetics**, v. 43, p. 405-413, p. 2011.

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, p. 361-365, 2011.

KITAMURA, T.; NAGANUMA, T.; ABE, K.; NAKAHARA, K.; OHNO, Y.; KIHARA, A. Substrate specificity, plasma membrane localization, and lipid modification of the aldehyde dehydrogenase ALDH3B1. **Biochimica et Biophysica Acta**, v. 1831, p. 1395-1401, 2013.

LEE, S. H.; CHOI, B. H.; LIM, D.; GONDRO, C.; CHO, Y. M.; DANG, C. G.; SHARMA, A.; JANG, G. W.; LEE, K. T.; YOON, D.; LEE, H. K.; YEON, S. H.; YANG, B. S.; KANG, H. S.; HONG, S. K. Genome-wide association study identifies major loci for carcass weight on BTA14 in Hanwoo (Korean cattle). **PLoS One**, v. 8, 2013.

LI, M.; WU, H.; WANG, T.; XIA, Y.; JIN, L.; JIANG, A.; ZHU, L.; CHEN, L.; LI, R.; LI, X. Co-methylated Genes in Different Adipose Depots of Pig are Associated with Metabolic, Inflammatory and Immune Processes. **International Journal of Biological Sciences**, v. 8, p. 831-837, 2012.

LIEFKE, R.; WINDHOF-JAIDHAUSER, I. M.; GAEDCKE, J.; SALINAS-RIESTER, G.; WU, F.; GHADIMI, M.; DANGO, S. The oxidative demethylase ALKBH3 marks hyperactive gene promoters in human cancer cells. **Genome Medicine**, v. 7, 2015.

LINDHOLM-PERRY, A. K.; KUEHN, L. A.; SMITH, T. P.; FERRELL, C. L.; JENKINS, T. G.; FREETLY, H. C.; SNELLING, W. M. A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle. **Animal Genetics**, v. 43, 216-219, 2012.

LITTLEJOHN, M.; GRALA, T.; SANDERS, K.; WALKER, C.; WAGHORN, G.; MACDONALD, K.; COPPIETERS, W.; GEORGES, M.; SPELMAN, R.; HILLERTON, E.; DAVIS, S.; SNELL, R. Genetic variation in PLAG1 associates with early life body weight and peripubertal weight and growth in *Bos taurus*. **Animal Genetics**, v. 43, p. 591-594, 2012.

LU, D.; SARGOLZAEI, M.; KELLY, M.; VOORT, G. V.; WANG, Z.; MANDELL, I.; MOORE, S.; PLASTOW, G. AND MILLER, S. P. Genome-wide association analyses for carcass quality in crossbred beef cattle. **BMC Genetics**, v. 14, n. 80, 2013.

MAGALHÃES, A. F. B. **Utilização de Informações Genômicas para o Melhoramento Genético de características da carne em bovinos da Raça Nelore**. 2015. 59f. Tese (Doutorado em Zootecnia) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal, 2015.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. H. BLUPF90 and related programs (BGF90). In: World Congress Genetics Application Livestock Production, 7th., 2002, Montpellier, p. 28.

MUJIBI, F. D. N.; NKRUMAH, J. D.; DURUNNA, O. N.; STOTHARD, P.; MAH, J.; WANG, Z.; BASARAB, J.; PLASTOW, G.; CREWS, D. H.; MOORE, S. S. Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. **Journal of Animal Science**, v. 89, n. 11, p. 3353-3361, 2011.

NAY, S. L.; LEE, D. H.; BATES, S. E.; O'CONNOR, T. R. Alkbh2 protects against lethality and mutation in primary mouse embryonic fibroblasts. **DNA Repair**, v. 11, p. 502-510, 2012.

NISHIMURA, S.; WATANABE, T.; MIZOSHITA, N.; TATSUDA, K.; FUJITA, T.; WATANABE, N.; SUGIMOTO, Y.; TAKASUGA, A. Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. **BMC Genetics**, v. 13, 2012.

OLIVA, K.; BARKER, G.; RICE, G. E.; BAILEY, M. J.; LAPPAS, M. 2D-DIGE to identify proteins associated with gestational diabetes in omental adipose tissue. **Journal of Endocrinology**, v. 218, p. 165-178, 2013.

PORTO NETO, L. R.; BUNCH, R. J.; HARRISON, B. E. and BARENDSE, W. Variation in the XKR4 gene was significantly associated with subcutaneous rump fat thickness in indicine and composite cattle. **Animal Genetics**, v. 43, p. 785-789, 2012.

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

PRYCE, J. E.; ARIAS, J.; BOWMAN, P. J.; DAVIS, S. R.; MACDONALD, K. A.; WAGHORN, G. C.; WALES, W. J. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. **Journal of Dairy Science**, v. 95, p. 2108-2119, 2012.

PRYCE, J.E.; WALES, W. J. HAAS, Y., VEERKAMP, R. F. AND HAYES B. J. Genomic selection for feed efficiency in dairy cattle. **Animal**, v. 8, n. 1, p. 1-10, 2014.

RILEY, D. G.; WELSH JR, T. H.; GILL, C. L.; HULSMAN, L. L.; HERRING, A. D.; RIGG, P. K.; SAWYER, J. E.; SANDERS, J. O. Whole genome association of SNP with new born calf cannon bone length. **Livestock Science**, v. 155, p. 186-196, 2013.

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

SÁEZ, M. E.; MARTÍNEZ-LARRAD, M.T.; RAMÍREZ-LORCA, R.; GONZÁLEZ-SÁNCHEZ, J. L.; ZABENA, C.; MARTINEZ-CALATRAVA, M. J.; GONZÁLEZ, A.; MORÓN, F. J.; RUIZ, A.; SERRANO-RÍOS, M. CALPAIN-5 gene variants are associated with diastolic blood pressure and cholesterol levels. **BMC Medical Genetics**, v. 8, 2007.

SHARMA, A.; DANG, C. G.; KIM, K. S.; KIM, J. J.; LEE, H. K.; KIM, H. C.; YEON, S. H.; KANG, H. S.; LEE, S. H. Validation of genetic polymorphisms on BTA14 associated with carcass trait in a commercial Hanwoo. **Animal Genetics**, v. 45, p. 863-867, 2014.

SOHLE, J.; MACHUY, N.; SMAILBEGOVIC, E.; HOLTZMANN, U.; GRONNIGER, E.; WENCK, H.; STAB, F.; WINNEFELD, M. Identification of New Genes Involved in Human Adipogenesis and Fat Storage. **PLoS One**, v. 7, 2012.

UTSUNOMIYA, Y. T.; DO CARMO, A. S.; CARVALHEIRO, R.; NEVES, H. H.; MATOS, M. C.; ZAVAREZ, L. B.; PÉREZ O'BRIEN, A. M.; SÖLKNER, J.; MCEWAN, J. C.; COLE, J. B.; VAN TASSELL, C. P.; SCHENKEL, F. S.; DA SILVA, M. V.; PORTO NETO, L. R.; SONSTEGARD, T. S.; GARCIA, J. F. 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, 2013.

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

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

VISUS, C.; ITO, D.; DHIR, R.; SZCZEPANSKI, M. J.; CHANG, Y.J.; LATIMER, J. J.; GRANT, S. G.; DELEO, A. B. Identification of Hydroxysteroid (17 β) dehydrogenase type 12 (HSD17B12) as a CD8+ T-cell-defined human tumor antigen of human carcinomas. **Cancer Immunology Immunotherapy**, v. 60, n. 7, p. 919-929, 2011.

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

WANG, P.; DRACKLEY, J. K.; STAMEY-LANIER, J. A.; KEISLER, D.; LOOR, J. J. Effects of level of nutrient intake and age on mammalian target of rapamycin, insulin, and insulin-like growth factor-1 gene network expression in skeletal muscle of young Holstein calves. **Journal of Dairy Science**, v. 97, p. 383-391, 2014.

APPENDIX

Appendix A

Table 1A. Gene enrichment clustering for *longissimus* muscle area.

Annotation Cluster 1		Enrichment Score: 1.13			
Category	Term	Count	%	PValue	FDR
SMART	SM00326:SH3	3	1.31	0.01	9.69
UP_SEQ_FEATURE	domain:SH3	3	1.31	0.02	18.83
SP_PIR_KEYWORDS	sh3 domain	3	1.31	0.03	24.38
INTERPRO	IPR001452:Src homology-3 domain	3	1.31	0.03	23.58
GOTERM_BP_FAT	GO:0043066~negative regulation of apoptosis	3	1.31	0.09	70.10
GOTERM_BP_FAT	GO:0043069~negative regulation of programmed cell death	3	1.31	0.09	71.01
GOTERM_BP_FAT	GO:0060548~negative regulation of cell death	3	1.31	0.09	71.19
GOTERM_BP_FAT	GO:0042981~regulation of apoptosis	4	1.75	0.10	75.11
GOTERM_BP_FAT	GO:0043067~regulation of programmed cell death	4	1.75	0.10	75.94
GOTERM_BP_FAT	GO:0010941~regulation of cell death	4	1.75	0.10	76.24
SP_PIR_KEYWORDS	cytoplasm	6	2.62	0.37	99.25
GOTERM_BP_FAT	GO:0007242~intracellular signaling cascade	3	1.31	0.54	100.00
Annotation Cluster 2		Enrichment Score: 0.93			
Category	Term	Count	%	PValue	FDR
GOTERM_BP_FAT	GO:0033554~cellular response to stress	4	1.75	0.04	44.01
GOTERM_BP_FAT	GO:0006281~DNA repair	3	1.31	0.06	55.54
GOTERM_BP_FAT	GO:0006974~response to DNA damage stimulus	3	1.31	0.10	73.44
GOTERM_BP_FAT	GO:0006259~DNA metabolic process	3	1.31	0.16	89.64
SP_PIR_KEYWORDS	nucleus	6	2.62	0.61	100.00
Annotation Cluster 3		Enrichment Score: 0.44			
Category	Term	Count	%	PValue	FDR
GOTERM_BP_FAT	GO:0019725~cellular homeostasis	3	1.31	0.14	85.98
UP_SEQ_FEATURE	transmembrane region	9	3.93	0.22	93.21
SP_PIR_KEYWORDS	transmembrane	9	3.93	0.22	93.38
GOTERM_BP_FAT	GO:0042592~homeostatic process	3	1.31	0.28	98.82
SP_PIR_KEYWORDS	membrane topological	10	4.37	0.32	98.32
UP_SEQ_FEATURE	domain:Cytoplasmic	6	2.62	0.38	99.48
GOTERM_CC_FAT	GO:0016021~integral to membrane	9	3.93	0.49	99.91
GOTERM_CC_FAT	GO:0031224~intrinsic to membrane	9	3.93	0.54	99.97
UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	5	2.18	0.76	100.00
SP_PIR_KEYWORDS	glycoprotein	5	2.18	0.79	100.00
Annotation Cluster 4		Enrichment Score: 0.21			

Category	Term	Count	%	PValue	FDR
GOTERM_CC_FA T	GO:0005654~nucleoplasm	3	1.31	0.36	99.04
SP_PIR_KEYWOR DS	nucleus	6	2.62	0.61	100.00
GOTERM_CC_FA T	GO:0031981~nuclear lumen	3	1.31	0.62	100.00
GOTERM_CC_FA T	GO:0070013~intracellular organelle lumen	3	1.31	0.74	100.00
GOTERM_CC_FA T	GO:0043233~organelle lumen	3	1.31	0.75	100.00
GOTERM_CC_FA T	GO:0031974~membrane-enclosed lumen	3	1.31	0.76	100.00
Annotation Cluster 5		Enrichment Score: 0.17			
Category	Term	Count	%	PValue	FDR
GOTERM_BP_FAT	GO:0051252~regulation of RNA metabolic process	4	1.75	0.48	99.98
SP_PIR_KEYWOR DS	nucleus	6	2.62	0.61	100.00
SP_PIR_KEYWOR DS	transcription regulation	3	1.31	0.71	100.00
SP_PIR_KEYWOR DS	Transcription	3	1.31	0.73	100.00
GOTERM_BP_FAT	GO:0006355~regulation of transcription, DNA-dependent	3	1.31	0.73	100.00
GOTERM_BP_FAT	GO:0045449~regulation of transcription	4	1.75	0.74	100.00
GOTERM_BP_FAT	GO:0006350~transcription	3	1.31	0.82	100.00
Annotation Cluster 6		Enrichment Score: 0.07			
Category	Term	Count	%	PValue	FDR
SP_PIR_KEYWOR DS	cell membrane	3	1.31	0.76	100.00
GOTERM_CC_FA T	GO:0044459~plasma membrane part	3	1.31	0.84	100.00
GOTERM_CC_FA T	GO:0005886~plasma membrane	4	1.75	0.94	100.00

Table 2A. Gene enrichment clustering for backfat thickness.

Annotation Cluster 1		Enrichment Score: 0.90			
Category	Term	Count	%	PValue	FDR
SP_PIR_KEYWORDS	phosphotransferase	3	0.83	0.02	18.47
SP_PIR_KEYWORDS	transferase	7	1.93	0.03	24.82
SP_PIR_KEYWORDS	kinase	5	1.38	0.03	28.14
SP_PIR_KEYWORDS	atp-binding	6	1.66	0.06	50.13
GOTERM_MF_FAT	GO:0005524~ATP binding	6	1.66	0.10	67.07
GOTERM_MF_FAT	GO:0032559~adenyl	6	1.66	0.10	68.56
GOTERM_MF_FAT	ribonucleotide binding	6	1.66	0.12	74.74
GOTERM_MF_FAT	GO:0030554~adenyl	6	1.66	0.12	74.74
GOTERM_MF_FAT	nucleotide binding	6	1.66	0.12	75.74
GOTERM_MF_FAT	GO:0001883~purine	6	1.66	0.12	75.74
GOTERM_MF_FAT	nucleoside binding	6	1.66	0.12	76.49
GOTERM_MF_FAT	GO:0001882~nucleoside	6	1.66	0.12	76.49
GOTERM_MF_FAT	binding	6	1.66	0.12	76.49
GOTERM_MF_FAT	GO:0016310~phosphorylati	6	1.66	0.12	76.49
GOTERM_BP_FAT	on	4	1.10	0.13	84.82
SP_PIR_KEYWORDS	nucleotide-binding	6	1.66	0.13	79.45
GOTERM_MF_FAT	GO:0032553~ribonucleotid	6	1.66	0.19	90.20
GOTERM_MF_FAT	e binding	6	1.66	0.19	90.20
GOTERM_MF_FAT	GO:0032555~purine	6	1.66	0.19	90.20
GOTERM_MF_FAT	ribonucleotide binding	6	1.66	0.19	90.20
GOTERM_BP_FAT	GO:0006793~phosphorus	4	1.10	0.19	94.53
GOTERM_BP_FAT	metabolic process	4	1.10	0.19	94.53
GOTERM_BP_FAT	GO:0006796~phosphate	4	1.10	0.19	94.53
GOTERM_BP_FAT	metabolic process	4	1.10	0.19	94.53
GOTERM_MF_FAT	GO:0017076~purine	6	1.66	0.21	92.93
GOTERM_MF_FAT	nucleotide binding	6	1.66	0.21	92.93
UP_SEQ_FEATURE	nucleotide phosphate-	4	1.10	0.24	96.10
UP_SEQ_FEATURE	binding region:ATP	4	1.10	0.24	96.10
GOTERM_MF_FAT	GO:0004672~protein	3	0.83	0.27	96.69
UP_SEQ_FEATURE	kinase activity	3	0.83	0.27	97.66
GOTERM_BP_FAT	binding site:ATP	3	0.83	0.31	99.36
GOTERM_BP_FAT	GO:0006468~protein amino	3	0.83	0.31	99.36
GOTERM_MF_FAT	acid phosphorylation	3	0.83	0.31	99.36
GOTERM_MF_FAT	GO:0000166~nucleotide	6	1.66	0.32	98.61
GOTERM_MF_FAT	binding	6	1.66	0.32	98.61

Annotation Cluster 2		Enrichment Score: 0.45			
Category	Term	Count	%	PValue	FDR
GOTERM_MF_FAT	GO:0003677~DNA binding	6	1.66	0.19	89.57
SP_PIR_KEYWORDS	transcription regulation	5	1.38	0.26	96.35
SP_PIR_KEYWORDS	Transcription	5	1.38	0.35	99.08
GOTERM_BP_FAT	GO:0006350~transcription	5	1.38	0.38	99.85
SP_PIR_KEYWORDS	dna-binding	4	1.10	0.42	99.77
SP_PIR_KEYWORDS	nucleus	8	2.21	0.46	99.90
GOTERM_BP_FAT	GO:0045449~regulation of	5	1.38	0.57	100.00
GOTERM_BP_FAT	transcription	5	1.38	0.57	100.00

Annotation Cluster 3		Enrichment Score: 0.04			
Category	Term	Count	%	PValue	FDR
SP_PIR_KEYWORDS	metal-binding	4	1.10	0.85	100.00
GOTERM_MF_FAT	GO:0046914~transition	4	1.10	0.86	100.00
GOTERM_MF_FAT	metal ion binding	4	1.10	0.86	100.00
GOTERM_MF_FAT	GO:0043169~cation	5	1.38	0.94	100.00
GOTERM_MF_FAT	binding	5	1.38	0.94	100.00

GOTERM_MF_FAT	GO:0043167~ion binding	5	1.38	0.94	100.00
GOTERM_MF_FAT	GO:0046872~metal ion binding	4	1.10	0.98	100.00
Annotation Cluster 4		Enrichment Score: 0.01			
Category	Term	Count	%	PValue	FDR
SP_PIR_KEYWORDS	glycoprotein	4	1.10	0.96	100.00
UP_SEQ_FEATURE	transmembrane region	5	1.38	0.97	100.00
UP_SEQ_FEATURE	glycosylation site:N-linked (GlcNAc...)	4	1.10	0.97	100.00
SP_PIR_KEYWORDS	membrane	6	1.66	0.98	100.00
SP_PIR_KEYWORDS	transmembrane	5	1.38	0.99	100.00
GOTERM_CC_FAT	GO:0016021~integral to membrane	5	1.38	1.00	100.00
GOTERM_CC_FAT	GO:0031224~intrinsic to membrane	5	1.38	1.00	100.00

Table 3A. Gene enrichment clustering for rump fat thickness.

Annotation Cluster 1		Enrichment Score: 1.29			
Category	Term	Count	%	PValue	FDR
SP_PIR_KEYWORDS	proteoglycan	3	0.98	0.00	2.56
UP_SEQ_FEATURE	repeat:LRR 12	3	0.98	0.01	7.20
UP_SEQ_FEATURE	repeat:LRR 11	3	0.98	0.01	8.91
	IPR000372:Leucine-rich repeat, cysteine-rich flanking region, N-terminal	3	0.98	0.01	9.13
INTERPRO		3	0.98	0.01	8.07
SMART	SM00013:LRRNT	3	0.98	0.01	11.45
UP_SEQ_FEATURE	repeat:LRR 10	3	0.98	0.01	14.79
UP_SEQ_FEATURE	repeat:LRR 9	3	0.98	0.02	17.57
UP_SEQ_FEATURE	repeat:LRR 8	3	0.98	0.02	19.42
	compositionally biased region:Cys-rich	3	0.98	0.02	23.49
UP_SEQ_FEATURE	repeat:LRR 7	3	0.98	0.03	30.86
UP_SEQ_FEATURE	repeat:LRR 6	3	0.98	0.04	35.64
UP_SEQ_FEATURE	repeat:LRR 5	3	0.98	0.04	33.60
INTERPRO	IPR001611:Leucine-rich repeat	3	0.98	0.04	36.90
SP_PIR_KEYWORDS	extracellular matrix	3	0.98	0.04	40.41
UP_SEQ_FEATURE	repeat:LRR 4	3	0.98	0.06	50.75
UP_SEQ_FEATURE	repeat:LRR 3	3	0.98	0.06	54.72
UP_SEQ_FEATURE	repeat:LRR 1	3	0.98	0.06	54.92
UP_SEQ_FEATURE	repeat:LRR 2	3	0.98	0.07	54.29
SP_PIR_KEYWORDS	leucine-rich repeat	3	0.98	0.07	61.75
	GO:0005578~proteinaceous extracellular matrix	3	0.98	0.08	66.71
GOTERM_CC_FAT		3	0.98	0.09	86.00
GOTERM_CC_FAT	GO:0031012~extracellular matrix	7	2.30	0.15	88.37
UP_SEQ_FEATURE	disulfide bond	7	2.30	0.17	88.86
SP_PIR_KEYWORDS	disulfide bond	5	1.64	0.17	99.26
SP_PIR_KEYWORDS	Secreted	6	1.97	0.42	99.82
GOTERM_CC_FAT	GO:0005576~extracellular region	3	0.98	0.42	99.81
SP_PIR_KEYWORDS	signal	6	1.97	0.42	99.86
	GO:0044421~extracellular region part	3	0.98	0.42	99.86
GOTERM_CC_FAT		6	1.97	0.42	100.00
UP_SEQ_FEATURE	signal peptide	3	0.98	0.56	100.00
	GO:0007186~G-protein coupled receptor protein signaling pathway	3	0.98	0.56	100.00
GOTERM_BP_FAT		4	1.31	0.60	100.00
	GO:0007166~cell surface receptor linked signal transduction	4	1.31	0.60	100.00
GOTERM_BP_FAT		4	1.31	0.60	100.00

Annotation Cluster 2		Enrichment Score: 0.98			
Category	Term	Count	%	PValue	FDR
	GO:0050953~sensory perception of light stimulus	3	0.98	0.05	49.31
GOTERM_BP_FAT		3	0.98	0.05	49.31
GOTERM_BP_FAT	GO:0007601~visual perception	5	1.64	0.06	55.39
GOTERM_BP_FAT	GO:0050890~cognition	5	1.64	0.13	85.21
	GO:0050877~neurological system process	4	1.31	0.14	88.13
GOTERM_BP_FAT		5	1.64	0.17	88.86
GOTERM_BP_FAT	GO:0007600~sensory perception	5	1.64	0.35	99.26
SP_PIR_KEYWORDS	Secreted	5	1.64	0.35	99.26
GOTERM_CC_FAT	GO:0005576~extracellular region	5	1.64	0.35	99.26

Annotation Cluster 3		Enrichment Score: 0.69			
Category	Term	Count	%	PValue	FDR

GOTERM_BP_FAT	GO:0009991~response to extracellular stimulus	3	0.98	0.05	50.48
GOTERM_BP_FAT	GO:0008284~positive regulation of cell proliferation	3	0.98	0.14	88.69
GOTERM_BP_FAT	GO:0042127~regulation of cell proliferation	3	0.98	0.37	99.84
SP_PIR_KEYWORDS	nucleus	6	1.97	0.68	100.00
Annotation Cluster 4		Enrichment Score: 0.68			
Category	Term	Count	%	PValue	FDR
SP_PIR_KEYWORDS	phosphotransferase	3	0.98	0.03	27.99
SP_PIR_KEYWORDS	ATP	3	0.98	0.04	35.78
UP_SEQ_FEATURE	active site:Proton acceptor	4	1.31	0.05	47.86
SP_PIR_KEYWORDS	nucleotide-binding	6	1.97	0.06	52.76
SP_PIR_KEYWORDS	atp-binding	5	1.64	0.09	66.39
GOTERM_CC_FAT	GO:0045202~synapse	3	0.98	0.10	68.57
SP_PIR_KEYWORDS	transferase	5	1.64	0.10	71.77
UP_SEQ_FEATURE	domain:Protein kinase	3	0.98	0.12	79.48
	nucleotide phosphate-binding region:ATP	4	1.31	0.13	80.63
	IPR017441:Protein kinase, ATP binding site	3	0.98	0.14	79.95
INTERPRO	IPR000719:Protein kinase, core binding site:ATP	3	0.98	0.15	82.41
UP_SEQ_FEATURE	GO:0032555~purine ribonucleotide binding	6	1.97	0.17	87.03
GOTERM_MF_FAT	GO:0032553~ribonucleotide binding	6	1.97	0.17	87.03
GOTERM_MF_FAT	GO:0017076~purine nucleotide binding	6	1.97	0.19	90.55
GOTERM_MF_FAT	GO:0005524~ATP binding	5	1.64	0.21	93.06
	GO:0032559~adenyl ribonucleotide binding	5	1.64	0.22	93.72
GOTERM_MF_FAT	kinase	3	0.98	0.22	94.81
SP_PIR_KEYWORDS	GO:0030554~adenyl nucleotide binding	5	1.64	0.25	95.89
GOTERM_MF_FAT	GO:0001883~purine nucleoside binding	5	1.64	0.26	96.40
GOTERM_MF_FAT	GO:0004672~protein kinase activity	3	0.98	0.26	96.48
GOTERM_MF_FAT	GO:0001882~nucleoside binding	5	1.64	0.26	96.61
GOTERM_MF_FAT	GO:0000166~nucleotide binding	6	1.97	0.29	97.96
	GO:0006468~protein amino acid phosphorylation	3	0.98	0.30	99.25
GOTERM_BP_FAT	GO:0005829~cytosol	4	1.31	0.32	98.70
GOTERM_BP_FAT	GO:0016310~phosphorylation	3	0.98	0.38	99.86
GOTERM_CC_FAT	GO:0044430~cytoskeletal part	3	0.98	0.42	99.80
	GO:0006793~phosphorus metabolic process	3	0.98	0.48	99.99
GOTERM_BP_FAT	GO:0006796~phosphate metabolic process	3	0.98	0.48	99.99
GOTERM_BP_FAT	GO:0043228~non-membrane-bounded organelle	5	1.64	0.56	99.99
GOTERM_CC_FAT	GO:0043232~intracellular non-membrane-bounded organelle	5	1.64	0.56	99.99
GOTERM_CC_FAT	GO:0005856~cytoskeleton	3	0.98	0.62	100.00
SP_PIR_KEYWORDS	cytoplasm	5	1.64	0.65	100.00
GOTERM_CC_FAT	GO:0044459~plasma membrane	4	1.31	0.66	100.00

SP_PIR_KEYWORDS	part nucleus	6	1.97	0.68	100.00
-----------------	-----------------	---	------	------	--------

Annotation Cluster 5		Enrichment Score: 0.56			
Category	Term	Count	%	PValue	FDR
GOTERM_BP_FAT	GO:0009725~response to hormone stimulus	3	0.98	0.12	82.89
GOTERM_BP_FAT	GO:0009719~response to endogenous stimulus	3	0.98	0.14	87.73
GOTERM_BP_FAT	GO:0010033~response to organic substance	3	0.98	0.33	99.62
GOTERM_CC_FAT	GO:0000267~cell fraction	3	0.98	0.49	99.95
GOTERM_BP_FAT	GO:0007242~intracellular signaling cascade	3	0.98	0.62	100.00

Annotation Cluster 6		Enrichment Score: 0.37			
Category	Term	Count	%	PValue	FDR
GOTERM_CC_FAT	GO:0031090~organelle membrane	4	1.31	0.22	94.08
GOTERM_BP_FAT	GO:0016192~vesicle-mediated transport	3	0.98	0.24	97.84
GOTERM_CC_FAT	GO:0005829~cytosol	4	1.31	0.32	98.70
GOTERM_CC_FAT	GO:0044459~plasma membrane part	4	1.31	0.66	100.00
GOTERM_CC_FAT	GO:0005886~plasma membrane	6	1.97	0.70	100.00
SP_PIR_KEYWORDS	phosphoprotein	9	2.95	0.79	100.00

Annotation Cluster 7		Enrichment Score: 0.37			
Category	Term	Count	%	PValue	FDR
SP_PIR_KEYWORDS	glycoprotein	9	2.95	0.18	90.07
	glycosylation site:N-linked (GlcNAc...)	8	2.62	0.29	98.17
UP_SEQ_FEATURE	membrane	10	3.28	0.43	99.85
SP_PIR_KEYWORDS	transmembrane region	8	2.62	0.47	99.95
UP_SEQ_FEATURE	transmembrane	8	2.62	0.48	99.95
GOTERM_CC_FAT	GO:0016021~integral to membrane	9	2.95	0.56	99.99
GOTERM_CC_FAT	GO:0031224~intrinsic to membrane	9	2.95	0.61	100.00
UP_SEQ_FEATURE	topological domain:Cytoplasmic	5	1.64	0.67	100.00

Annotation Cluster 8		Enrichment Score: 0.03			
Category	Term	Count	%	PValue	FDR
SP_PIR_KEYWORDS	zinc	3	0.98	0.79	100.00
GOTERM_MF_FAT	GO:0008270~zinc ion binding	3	0.98	0.91	100.00
SP_PIR_KEYWORDS	metal-binding	3	0.98	0.92	100.00
GOTERM_MF_FAT	GO:0046914~transition metal ion binding	3	0.98	0.96	100.00
GOTERM_MF_FAT	GO:0046872~metal ion binding	4	1.31	0.98	100.00
GOTERM_MF_FAT	GO:0043169~cation binding	4	1.31	0.98	100.00
GOTERM_MF_FAT	GO:0043167~ion binding	4	1.31	0.99	100.00