UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

A GENOMIC ASSOCIATION AND PREDICTION OF PRINCIPAL COMPONENTS OF GROWTH TRAITS AND VISUAL SCORES IN NELLORE CATTLE

Giovana Vargas

Zootecnista

2018

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

A GENOMIC ASSOCIATION AND PREDICTION OF PRINCIPAL COMPONENTS OF GROWTH TRAITS AND VISUAL SCORES IN NELLORE CATTLE

Giovana Vargas

Orientador: **Dr. Roberto Carvalheiro**

Coorientadores: **Prof. Dr. Danísio Prado Munari**

Dr. Haroldo Henrique de Rezende Neves

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

2018

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: A GENOMIC ASSOCIATION AND PREDICTION OF PRINCIPAL COMPONENTS OF GROWTH TRAITS AND VISUAL SCORES IN NELLORE CATTLE

AUTORA: GIOVANA VARGAS
ORIENTADOR: ROBERTO CARVALHEIRO
COORIENTADOR: DANISIO PRADO MUNARI
COORIENTADOR: HAROLDO HENRIQUE DE REZENDE NEVES

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

Prof. Dr. DANISIO PRADO MUNARI
Departamento de Ciências Exatas / FCAV / Unesp - Jaboticabal

Prof. Dr. MARCOS ELI BUZANSKAS
Departamento de Zootecnia / Universidade Federal da Paraíba - Areia/PB

Profa. Dra. LUCIA GALVÃO DE ALBUQUERQUE
Departamento de Zootecnia / FCAV / Unesp - Jaboticabal

Pós-doutorando RODRIGO PELICIONI SAVEGNAGO
Instituto de Zootecnia / Sertãozinho/SP

Pós-doutorando GERARDO ALVES FERNANDES JÚNIOR
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Jaboticabal, 18 de dezembro de 2018

DADOS CURRICULARES DO AUTOR

Giovana Vargas – Nasceu em Jaboticabal, São Paulo em 27 de Janeiro de 1988, filha de Márcia Fátima Paula Rodrigues Vargas e Cleyton Vargas. Iniciou sua graduação em Zootecnia na Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal (FCAV/Unesp) em março de 2008. Durante a graduação (2008 – 2012) participou de projetos científicos, foi bolsista CNPq, monitora na disciplina Processamento de Dados e professora voluntária no Cursinho Ativo da FCAV/Unesp. Em Março de 2013, iniciou seu mestrado no programa de pós-graduação em Genética e Melhoramento Animal da FCAV/Unesp, onde foi bolsista CAPES e FAPESP. Durante o mestrado dedicou-se a um estudo quantitativo e de associação genômica ampla da característica aprumo em bovinos da raça Nelore sob orientação do pesquisador Dr. Roberto Carvalheiro, recebendo o título de mestre em Fevereiro de 2015. No mesmo ano iniciou o Doutorado no programa de pós-graduação em Genética e Melhoramento Animal da FCAV/Unesp sob orientação do pesquisador Dr. Roberto Carvalheiro, sendo bolsista CAPES e FAPESP. Durante Julho/2017 à Junho/2018 obteve a Bolsa Estágio de Pesquisa no Exterior (BEPE) para a realização do estágio sanduíche, desenvolvendo parte do projeto na Universidade de Guelph, Canadá, sob orientação do prof. Dr. Flávio Schramm Schenkel.

"A persistência é o caminho do êxito."

Charles Chaplin

Dedico

Aos meus pais, **Márcia** e **Cleyton** por me proporcionarem a educação que
tenho hoje, pelo apoio, dedicação e ensinamentos.

AGRADECIMENTOS

A Deus, por sempre me guiar pelo caminho certo, por me iluminar nos momentos difíceis e não me deixar desanimar diante das dificuldades.

Aos meus pais Márcia e Cleyton e minhas irmãs Bárbara e Larissa, pelo apoio, incentivo, paciência, por acreditarem no meu potencial como pessoa e profissional e por serem exemplos de força e coragem.

Ao meu companheiro de vida Guilherme, por sempre estar ao meu lado, me guiando e apoiando, me ensinando a crescer e me melhorar a cada dia.

Aos meus amigos Baltasar, Daiane, Grazyella, Henrique, Laiza e Thaise, pelo companheirismo, convivência, apoio, festinhas de aniversário e cumplicidade. Não é por acaso que certas pessoas cruzam a nossa vida, a escolha de mantê-las por perto só cabe a nós. Vocês meus amigos, eu quero levar para a vida toda.

Ao meu orientador Dr. Roberto Carvalheiro, por dedicar seu tempo e paciência em me orientar, pela atenção, pelos ensinamentos, pelas oportunidades que me foram dadas durante toda a pós-graduação, por ser ético e responsável e por sempre se preocupar com meu crescimento profissional.

Aos meus coorientadores Prof. Dr. Danísio Prado Munari e Dr. Haroldo Henrique de Rezende Neves, pela paciência, atenção, suporte, ensinamentos, amizade e contribuição.

Ao Prof. Dr. Flávio Schramm Schenkel, por ter me orientado durante meu estágio sanduíche em Guelph, pelo conhecimento adquirido, pela disposição e atenção que me foi dada durante todo o período de estágio.

As bancas de qualificação e defesa por todas as contribuições e sugestões feitas ao trabalho.

A Unesp e Universidade de Guelph pelas oportunidades de desenvolver e aprimorar este trabalho e a todas as pessoas que de alguma forma contribuíram para a preparação do mesmo.

SUMMARY

A GENOMIC ASSOCIATION AND PREDICTION OF PRINCIPAL COMPONENTS OF GROWTH TRAITS AND VISUAL SCORES IN NELLORE CATTLE

ABSTRACT – Principal component analysis (PCA) is a multivariate statistical technique that allows evaluating relationships among different traits in order to eliminate the redundancy resulting from their correlations. In animal breeding, PCA has been used to explore possible biological interpretations associated with the principal components (PCs) that can lead to the characterization of distinguished animal's biotype. The objectives of the present study were: i) to evaluate relationships among growth, visual scores, and reproductive traits by performing a PCA; ii) to identify genomic regions associated with PCs by performing a genome-wide association study (GWAS) on the main PCs; and iii) to evaluate the prediction ability of genomic breeding values (GEBVs) obtained for the PCs. Phenotypic data from 355,524 Nellore animals provided by the Alliance Nellore database, were used in this investigation. A total of 3,382 Nellore animals were genotyped using the Illumina® BovineHD chip (HD, ~777,000 SNPs) and 137 animals were genotyped using the GeneSeek Genomic Profiler Bovine HD chip (~76,000 SNPs). The GGP-HD genotypes were imputed to the HD genotypes. After genomic data quality control, 471,880 SNPs from 3,519 animals were available. The PCA was applied on the additive genetic (co)variance matrix (A_T) obtained using multi-trait analysis. For GWAS, SNP effects were estimated using the weighted single-step GBLUP and the BayesC methods. The genes identified within the top-10 ranking windows that explained the highest proportion of variance were used for further functional analyses. For the genomic prediction study, the GEBVs were predicted using three distinguish response variables: EBV of the original traits, EBV of the PCs, and EBV of a selection index used by some Nellore cattle commercial breeding programs. The genomic predictive ability was calculated by correlation between GEBVs and response variables. The first three PCs explained 87.11% of the total additive genetic variance for the traits. The first component contrasted the animals according to the growth rate, the second component contrasted the animals with early or late biotype, while the third component differentiate weaning and yearling traits. GWAS detected potential genomic regions associated with growth, carcass traits, conformation, and fatty acid composition traits that may be affecting the PCs. These findings are of relevance to the biological understanding of the PCs and their associated biotypes in Nellore cattle. Genomic predictions with moderate accuracies were obtained for the nine original traits, PCs and selection index, indicating the possibility of using PCA for implementing genomic selection for Nellore cattle.

Keywords: beef cattle, GEBV, SNP effects, wssGBLUP

ASSOCIAÇÃO E PREDIÇÃO GENÔMICA DE COMPONENTES PRINCIPAIS DE CARACTERÍSTICAS DE CRESCIMENTO E ESCORES VISUAIS EM BOVINOS DA RAÇA NELORE

RESUMO – A análise de componentes principais (ACP) é uma técnica da estatística multivariada usada para avaliar as relações entre diferentes características a fim de eliminar a redundância resultante de suas correlações. No melhoramento genético animal, a ACP tem sido usada para explorar possíveis interpretações biológicas associadas aos componentes principais (CPs) que podem levar a caracterização de diferentes biotipos de animais. Os objetivos do presente estudo foram: i) avaliar as relações entre características de crescimento, escore visual e reprodutiva, por meio de ACP; ii) identificar, por meio de estudo de associação genômica ampla (GWAS), regiões genômicas que diferenciam os animais quanto aos diferentes componentes; e iii) avaliar a habilidade de predição de valores genéticos genômicos (GEBVs) obtidos para os CPs. Foram utilizados dados fenotípicos de 355.524 animais da raça Nelore provenientes da base de dados Aliança Nelore. Destes, foram genotipados 3.382 animais em painel Illumina® BovineHD (HD, ~777.000 SNPs) e 137 animais em painel GeneSeek Genomic Profiler Bovine HD (~76.000 SNPs). Os animais genotipados com o painel GGP-HD tiveram seus genótipos imputados para o painel mais denso (HD). Após o controle de qualidade, 3.519 animais com informações genotípicas de 471.880 SNPs permaneceram nas análises. A ACP foi realizada utilizando-se a matriz de (co)variância genética aditiva ($\mathbf{A_T}$) obtida a partir de análise multi-característica. As estimativas dos efeitos dos SNPs foram obtidas utilizando-se as metodologias weighted single-step GBLUP e BayesC. Os genes identificados para as top 10 janelas que explicaram a maior proporção da variância foram usados para realizar as análises funcionais. Os GEBVs foram preditos utilizando como variáveis respostas: EBV das características originais, EBV dos CPs e EBV de um índice de seleção utilizado por programas de melhoramento da raça Nelore. A habilidade de predição da seleção genômica foi calculada pela correlação entre os GEBVs e as variáveis respostas. Os três primeiros CPs explicaram 87,11% da variância genética aditiva total das características. O primeiro CP contrastou animais de acordo com a taxa de crescimento, o segundo CP diferenciou os animais em biotipos tardios e precoces, e o terceiro CP contrastou características mensuradas ao desmame e ao sobreano. Foram identificadas possíveis regiões genômicas associadas a características de crescimento, carcaça, conformação e ácidos graxos, que sugerem possível associação com os CPs. As regiões identificadas ajudam na interpretação biológica dos CPs e seus respectivos biotipos em bovinos Nelore. No estudo de predição genômica, acurácias de magnitude moderada foram obtidas para as nove características estudadas, para os CPs e índice de seleção, indicando que a ACP poderia ser utilizada para a seleção de bovinos Nelore.

Palavras-chave: bovinos de corte, efeitos dos SNPs, GEBV, wssGBLUP

CHAPTER 1 – General considerations

1.1 INTRODUCTION

Brazil ranks as the top beef exporter and the second commercial beef producer in the world (USDA, 2018). The majority of the Brazilian herds is composed of Zebu cattle (*Bos taurus indicus*), especially the Nellore breed (ABCZ, www.abcz.com.br). This is primarily the result of adaptation of these animals to harsh environments and tropical pasture conditions. The main raising system of beef cattle occurs on pasture and sometimes animals are finished at feedlots to accelerate weight gain and carcass finishing. Due to the satisfactory economic performance of Nellore cattle as a result of genetic selection processes on increasing the production efficiency, this breed is of great importance to Brazilian agribusiness and the beef cattle industry.

In beef cattle, growth trait measurements at different ages have been commonly used by breeding programs as selection criteria in an attempt to improve general animal growth (Boligon et al., 2010; Araújo et al., 2014). Considering that animal weight or weight gain are associated with skeletal size, body shape and fat content, visual scores have been included as selection criteria in Zebu beef cattle breeding programs to select animals that meet industry requirements (Bertipaglia et al., 2012; Everling et al., 2014). The use of visual scores is an advantageous alternative of measurement without either stressing out the animals or increasing the recording costs for performing genetic evaluation (Barrozo et al., 2015). In addition, the inclusion of visual scores in selection programs is a suitable alternative to improve carcass quality, muscle mass distribution and finishing precocity.

In Brazil, visual score evaluation systems are mainly based on recording animals for conformation, finishing precocity and muscling traits at weaning and yearling. Conformation score estimate the amount of meat produced in the animal carcass and the general body's growth (Weber et al., 2009). Finishing precocity score estimate the animal's capacity to accumulate fat deposition without reaching

high body weight, while muscling score represents the degree of muscle mass present in the carcass. Over the recent years, several studies have reported moderate to high heritability estimates and favorable genetic correlations between body weight and weight gain at different ages and visual scores (Toral et al., 2011; Barrozo et al., 2015; de Lacerda et al., 2017), suggesting that there is sufficient additive genetic variation for selection and genetic progress in these traits. According to Cardoso et al. (2004), high genetic correlations between visual scores of conformation, finishing precocity and muscularity are expected, because conformation comprises aspects of muscle mass, size, and finishing precocity.

When working with multiple traits, joint analysis may not be feasible to large data sets, resulting in increased computational demand. In addition, traits that are highly correlated may be redundant and therefore, will not contribute to genetic evaluation (Roso and Fries, 1995). Principal component analysis (PCA) has been used in animal breeding attempting to elucidate the structural relationships among different traits, eliminate redundant information and reduce the size of the direct additive genetic (co)variance matrix of multiple trait models (Savegnago et al., 2011; Buzanskas et al., 2013; Boligon et al., 2016).

The PCA allows unraveling potential biological associations that are usually not observed in the original data. Roso and Fries (1995) showed the possibility of selecting Polled Hereford animals for distinct size or body volumes based on the first principal component (PC). Boligon et al. (2016) used PCA for genetic evaluation of growth and reproductive traits in Nellore cattle based on the (co)variance matrix among estimated breeding values (EBV) and concluded that the first three PCs contrasted animals into different biotypes (set of desirable phenotypes), allowing particular biological interpretation of each PC.

Predicting selection response is a key step for planning optimum genetic improvement strategies. Selection methods capable to explore efficiently the available genetic material for identification of superior genotypes is fundamental to maximize the genetic gain of economically important traits. The definition of an aggregate genotype, commonly called the breeding objective on which selection is practiced, is a result of the combination of certain traits, in which the selection process will be applied simultaneously, allowing identify superior genotypes. In this

sense, when the PCs with favorable biological meaning are in agreement to desired breeding objectives, they would represent the aggregate genotype of individuals, which could be used directly in selection processes. According to Smith (1936) and Hazel (1943) the aggregate genotype can be specified as:

$$H = a' g$$

where **H** (total merit) is the aggregate genotype, **a** is a vector of economic values and **g** is a vector of additive genetic effects. The vector **a** provides the importance of changes in the genetic levels of traits and is defined as the net economic effects of changes in each trait in the vector **g** for a unit change in the additive genetic effects of the trait. The **H** can be used directly in selection process when estimates of the EBVs are available for all traits in the breeding objective, and those estimates account for genetic correlations among traits.

Traditionally, animals were selected for breeding purpose using information on phenotypes and pedigrees. However, the increasing availability of genome-wide dense molecular markers [e.g., single nucleotide polymorphisms (SNPs)] have permitted livestock breeding programs explore new pathways for obtaining additional genetic gain of individuals based on linkage disequilibrium (LD) between markers and genes (Meuwissen et al., 2001). Several genome-wide association studies (GWAS) in livestock have led to the conclusion that the effect of individual quantitative trait loci (QTL) on economic important traits are likely small and, therefore, a large number of QTL are needed to explain their genetic variation. For instance, the identification of genetic variants for growth, reproductive and carcass traits in beef cattle (Martínez et al., 2017; Silva et al., 2017; Hay and Roberts, 2018) allowed breeders to investigate the genetic basis affecting these traits.

Genomic prediction by means of dense markers molecular panels were firstly applied to animal breeding, especially in dairy cattle (Hayes et al., 2009), and later tested on plants (Jannink et al., 2010). Genomic prediction is focused on modeling the association between genome-wide SNP markers and phenotypes to predict breeding values of selection candidates. Genomic prediction relies on the expectation that all QTL are likely to be in LD with at least one marker. The success of genomic prediction is measured by its accuracy, i.e. how reliable a future phenotype of target individuals can be predicted, and depends on several

parameters, such as sample size of the reference population, accuracy of the pseudo-phenotypes, level of LD, and the degree of relationship between training and validation populations (Bolormaa et al., 2013; Boddhireddy et al., 2014).

1.2 OBJECTIVES

1.2.1 General objectives

To perform a principal component analysis (PCA) using an eigen-decomposition of the additive genetic (co)variance matrix for growth, visual score and reproductive traits, to perform a genome-wide association study on the main principal components (PC) and to investigate the feasibility of genomic predictions of the PCs.

1.2.2 Specific objectives

- ➢ To compare different approaches applied to PCA using: an eigen-decomposition of either the additive genetic (co)variance matrix (A_T) or the (co)variance matrix of the EBVs;
- ➢ To perform GWAS using EBVs from PCs as pseudo-phenotypes aiming to identify genomic regions associated with the main PCs;
- ➢ To compare the genetic progress achieved from selection based on PCs and from current selection indexes used by Nellore breeding programs;
- ➢ To evaluate the accuracy of genomic predictions for the original traits, the three main PCs and a selection index used by some Nellore breeding programs.

1.3 LITERATURE REVIEW

1.3.1 Principal Component Analysis (PCA)

Principal component analysis (PCA) introduced by Pearson (1901) and developed independently by Hotelling (1933), is one of the most traditional techniques used in multivariate data analysis. In animal breeding, PCA has assisted quantitative geneticists by using orthogonal transformation to reduce the dimensionality of originally correlated variables into a smaller set of non-correlated and independent variables named principal components (PCs) explaining as much as possible of the variability in the original variables (Hair et al., 2009). The PCs are obtained by diagonalization of positive semi-definite symmetric matrices and represents linear combinations of a set of the original variables that reflect trends and patterns of covariation in the data.

Considering that the original variables X_1, X_2, ..., X_p leads to a data matrix $\mathbf{X}$ (n x p) in which:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

with means μ_1, μ_2, ..., μ_p and variances σ_1^2, σ_2^2, ..., σ_p^2, respectively, these variables are not independent and, therefore, have (co)variance between the ith and k^{th} variable defined by σ_{ik}, for $i \neq k = 1, 2, ..., p$. Thus, p variables can be expressed in vector form as: $X = [X_1, X_2, ..., X_p]'$, with vector of means $\mu = [\mu_1, \mu_2, ... \mu_p]'$ and (co)variance matrix $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1p}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix}$$

From $\mathbf{\Sigma}$, the eigenvalue and eigenvector pairs can be calculated (λ_1, e_1), (λ_2, e_2), ... (λ_p, e_p), where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ are associated to $\mathbf{\Sigma}$ and, therefore, the ith PC is defined as:

$$Z_i = e_i'X = e_{i1}X_1 + e_{i2}X_2 + ... + e_{ip}X_p$$

where $\mathbf{Z_i}$ is a latent variable, i.e. is not measured from the experiment or sample survey. The aim is to determine $\mathbf{Z_i}$ from the p variables contained in $\mathbf{X}$. The main idea is to project the original coordinate points in a plane by maximizing the distance between them. This is the equivalent to maximize the variability of $\mathbf{Z_i}$. The variance of $\mathbf{Z_i}$ can be obtained as:

$$Var(Z_i) = Var(e_i^{'}X) = e_i^{'}Var(X)e_i = e_i^{'}\Sigma e_i$$

where $i = 1, \ldots, p$.

Using the spectral decomposition of $\mathbf{\Sigma}$, given by $\mathbf{\Sigma = P\Lambda P'}$, where $\mathbf{P}$ is the matrix composed of eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues, then, $tr(\mathbf{\Sigma}) = tr(\mathbf{P\Lambda P^{'}}) = tr(\mathbf{\Lambda P^{'}}P) = tr(\mathbf{\Lambda I}) = tr(\mathbf{\Lambda}) = \sum_{i=1}^{p} \lambda_i$ and

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_k \end{bmatrix}$$

The $tr(\mathbf{\Sigma})$ is given by the sum of the elements of the diagonal:

$$tr(\Sigma) = \sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i$$

Thus, the total variability contained in the original variables is equal to total variability contained in the PCs. The contribution of each PC ($\mathbf{Z_i}$) is expressed in percentage, and the proportion of the variance explained by the kth PC can be explained as:

$$PC_k = \frac{Var(Z_i)}{\sum_{i=1}^{p} Var(Z_i)} * 100$$

By the proportion of variance explained by the PCs it is possible to determine the number of components to be considered. According to Meyer (2007), when the original variables are highly correlated the first few PCs are responsible for explaining the greatest original data variation and those with smallest contribution on the variance can be excluded without notably altering the accuracy of the estimates.

The PCA technique has been incorporated into genetic evaluations in poultry (Savegnago et al., 2011; Venturini et al., 2013), beef cattle (Buzanskas et al., 2013; Boligon et al., 2016) and buffaloes (Agudelo-Gómez et al., 2015) in an attempt to reduce the size of the direct additive genetic (co)variance matrix in multiple trait models, as well as to explore the genetic relationship between estimated breeding

values (EBVs) for important economic traits. When applied on the (co)variance matrix of the EBVs, PCA may reveal biological interpretation of the main PCs that was not previously identified in the original data. Boligon et al. (2016) showed the possibility of selecting Nellore cattle on the bases of the first three PCs obtained from EBVs predicted for 9 different traits using single-trait analyses. According to the authors, these components could represent 3 different potential selection criteria for improving desirable biotypes.

Despite being widely applied on the (co)variance matrix of the EBVs, PCA can also be applied on the additive genetic (co)variance matrix ($\mathbf{A_T}$). In this type of analysis, the PCs are obtained by performing an eigen-decomposition of the $\mathbf{A_T}$ matrix. The use of $\mathbf{A_T}$ matrix allows access the additive genetic variation within traits and the corresponding (co)variance among traits, which is not possible when performing PCA based on the (co)variance matrix of the EBVs obtained from single trait analysis. When performing the eigen-decomposition of the $\mathbf{A_T}$ matrix, it is possible to contrast the individuals in relation to the additive genetic effect of the studied set of traits (Roso and Fries, 1995).

1.3.2 Genome-wide association study

Genotypic information coupled with measurements for traits of interest allow the conduction of genome-wide association studies (GWAS) to uncover genomic regions potentially associated with economically important traits. By performing GWAS, the knowledge on biological expression of the traits can be achieved. For this purpose, several statistical methods have been used. Differently from the traditional method based on testing one marker at a time and only using genotyped animals with known phenotype (or pseudo-phenotypes), the single-step GBLUP (ssGBLUP) proposed by Legarra et al. (2009) combines all pedigree, phenotypic and genotypic information, including phenotypic information on non-genotyped individuals, in a single step.

Usually when performing ssGBLUP, SNP effects are assumed to follow a normal distribution with a common variance and the same weight for SNP variance. However, alternative genomic evaluation methods have been proposed to allow differences between variance of SNP effects. For instance, Bayesian methods, such as BayesA, BayesB, and BayesC (Meuwissen et al., 2001; Kizilkaya et al., 2010; Habier et al., 2011), are used to assign to SNP effect a prior distribution, and then sample from the posterior distribution via Markov Chain Monte Carlo (MCMC) (Meuwissen et al., 2001; Habier et al., 2011; Gianola et al., 2013). The main difference between these methods is in the definition of a priori distribution of the SNP effects included in the genomic model. Previous studies have shown that the application of Bayesian methods was better than GBLUP approaches in simulation studies using few QTL with large effects and many QTL with small effects (Meuwissen and Mike, 2004; Lund et al., 2009; Guo et al., 2010). However, for traits presenting polygenic nature, which is the majority of the traits of interest in livestock, GBLUP approaches has been proven to be better than Bayesian methods (Cole et al., 2009; Su et al., 2010; Forni et al., 2011; Wang et al., 2014). In addition, because the application of Bayesian methods requires that animals must be phenotyped and genotyped, phenotypes from non-genotyped animals cannot be included in the analysis.

In an attempt to modify the ssGBLUP to obtain SNP weights, Wang et al. (2012) proposed the weighted single-step GBLUP (wssGBLUP). Over the years, many studies have performed GWAS using the wssGBLUP as alternative method to identify genes and genomic regions associated with traits of interest (Magalhães et al., 2016; Melo et al., 2017; Vargas et al., 2018). Melo et al. (2016), in a simulation study using the wssGBLUP method, reported that the inclusion of all available phenotypic records even from non-genotyped animals can contribute to improve QTL detection for complex traits.

When performing GWAS, many studies have used different strategies for defining the window size that has led to identify genomic regions associated with important traits. These strategies include non-overlapping fixed length genomic windows [e.g., Irano et al., 2016; Melo et al., 2016] and sliding windows [e.g., Dikmen et al., 2013; Fernandes Júnior et al., 2016b; Valente et al., 2016; Saowaphak et al.,

2017]. However, the use of arbitrarily window sizes in GWAS analyses, either by a fixed number of SNP markers or by fixed physical length in terms of base pairs, could lead to splitting LD blocks into separate windows as well as significant haplotype blocks affecting the traits of interest. Therefore, the use of haplotype blocks as windows in such analyses could be a better alternative to identify genetic variants and biological mechanisms underlying these traits. For this purpose, after performing GWAS and estimating SNP effects, the proportion of variance explained by each haplotype block is estimated by the sum of the variance explained by all SNPs within haplotype block. In this way, alleles linked along the chromosomes and that would likely be transmitted together across generations would not be split when estimating variance explaining by certain genomic regions.

Principal component analysis has been incorporated into GWAS by extracting PCs from related traits, thus, allowing obtain essential and non-redundant information to identify potentially candidate variants. Lee et al. (2014) using records of Duroc pig breed, identified genomic regions associated with the main PCs from 14 meat quality traits that contributed to improve the knowledge on their biological expression. Macciota et al. (2017) reported the possibility of identifying some putative candidate genes associated with PCs from indicator variables of tolerance to heat stress from milk production data in dairy cattle. In animal breeding field, these findings are of great relevance, as they suggest that PCA could contribute to GWAS to detect important genomic regions potentially associated with PCs from a set of targeted traits and thus, improve the knowledge on their genetic expression.

1.3.3 Genomic prediction

Genomic prediction has been widely used to estimate the genetic merit of all genotyped individuals based on dense markers across the entire genome (Meuwissen et al., 2001). The principle of genomic prediction is the use of both genotypic and phenotypic data available in reference population to obtain predictive equation to calculate genomic estimated breeding values (GEBVs). This information

is then applied to a second population, consisting of selection candidates, for which genotype information is available but not necessarily phenotype information.

An important advantage of genomic prediction compared to traditional selection process, is that the prediction accuracy on selection candidates is higher (Hayes et al., 2009). In addition, the inclusion of genomic information allows a substantial reduction of generation interval, especially for species with long generation interval, making it possible to anticipate selection decisions (Meuwissen et al., 2001; Schaeffer, 2006).

The accuracy of genomic predictions is important to the application of genomic selection in animal breeding and depends on several parameters, such as the size of the reference population (Andonov et al., 2017), the heritability estimate of the trait (Viana et al., 2017), the density of the SNP panel and level of LD, the degree of relationship between training and validation populations (Meuwissen et al., 2001), and the statistical methods implemented (Bolormaa et al., 2013; Neves et al., 2014). For instance, several studies have shown that genomic predictions are more accurate if the genomic relationship between the reference and validation population is higher (Lee et al., 2008; Legarra et al., 2008; Clark et al., 2012). According to Habier et al. (2013), the co-segregation of QTL predicted from markers genotypes with a pedigree is a different source of information important to understand when designing reference populations for breeding programs, as may influence the accuracy of genomic predictions.

In general, studies have reported accuracies of genomic predictions in beef cattle lower than in dairy cattle. For instance, the accuracy of genomic prediction in dairy cattle exceeds 0.7 for production, fertility, somatic cell count, and other traits (Wiggans et al., 2011; Lund et al., 2011), while accuracies in the range 0.2 to 0.7 are reported for growth, visual score, reproductive and carcass traits in beef cattle (Neves et al., 2014, Fernandes Júnior et al., 2016b). This may be explained by the fact that in dairy cattle reference populations are of higher quality and target populations may be more closely related to the reference populations than in beef cattle.

1.4 REFERENCES

Agudelo-Gómez D, Pineda-Sierra S, Cerón-Muñoz MF (2015) Genetic evaluation of dual-purpose Buffaloes (Bubalus bubalis) in colombia using principal component analysis. **Plos One** 10:e0137148.

Andonov S, Lourenco DAL, Fragomeni BO, Masuda Y, Pocrnic I, Tsuruta S, Misztal I (2017) Accuracy of breeding values in small genotyped populations using different sources of external information-A simulation study. **Journal of Dairy Science** 100:395-401.

Araújo CV, Lôbo RB, Figueiredo LGG, Mousquer CJ, Laureano MMM, Bittencourt TCBSC, Araújo SI (2014) Estimates of genetic parameters of growth traits of Nellore cattle in the Midwest region of Brazil. **Revista Brasileira de Saúde e Produção Animal** 15:846-853.

Barrozo D, Mattar M, Tholon P, de Oliveira JA, Scalez DCB, Queiroz SA (2015) Genetic and environmental effects on scores of conformation, precocity and muscling in long yearling Nellore cattle. **Semina: Ciências Agrárias** 36:4447-4456.

Bertipaglia TS, Carreño LOD, Machado CHC, Andrighetto C, Fonseca R (2012) Estimates of genetic parameters for visual scores and their correlation with production and reproductive traits in Brahman cattle. **Revista Brasileira de Zootecnia** 41:1407–1411.

Boligon AA, Albuquerque LG, Mercadante MEZ, Lôbo RB (2010) Study of relations among age at first calving, average weight gains and weights from weaning to maturity in Nellore cattle. **Revista Brasileira de Zootecnia** 39:746–751.

Boligon AA, Vicente IS, Vaz RZ, Campos GS, Souza FRP, Carvalheiro R, Albuquerque LG (2016) Principal component analysis of breeding values for growth and reproductive traits and genetic association with adult size in beef cattle. **Journal of. Animal Science** 94:5014–5022.

Bolormaa S, Pryce JE, et al. (2013) Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality in Bos taurus, Bos indicus, and composite beef cattle. **Journal of Animal Science** 91:3088–3104.

Boddhireddy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, DeNise S (2014) Genomic predictions in Angus cattle: comparisons of sample size, response variables, and clustering methods for cross-validation. **Journal of Animal Science** 92:485–497.

Buzanskas ME, Savegnago RP, Grossi DA, Venturini GC, Queiroz SA, Silva LO, Júnior RA, Munari DP, Alencar MM (2013) Genetic parameter estimates and principal component analysis of breeding values of reproduction and growth traits in female Canchim cattle. **Reproduction, Fertility and Development** 25:775–781.

Cardoso FF, Cardellino RA, Campos LT (2004) Componentes de (co)variância e parâmetros genéticos de caracteres pós-desmama em bovinos da raça Angus. **Revista Brasileira de Zootecnia** 33:313-319.

Clark SA, Hickey JM, Daetwyler HD, van der Werf JH (2012) The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. **Genetics Selection Evolution** 44:4.

Cole JB, VanRaden PM, O'Connell JR, Van Tassell CP, Sonstegard TS, Schnabel RD, Taylor JF, Wiggans GR (2009) Distribution and location of genetic effects for dairy traits. **Journal of Dairy Science** 92:2931-46.

de Lacerda VV, Campos GS, Silveira DD, Roso VM, Santana Junior ML, Souza FRP, Boligon AA (2017) Genetic associations between mature size and condition score of Nelore cows, and weight, subjective scores and carcass traits as yearlings. **Animal Production Science**

Dikmen S, Cole JB, Null DJ, Hansen PJ (2013) Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle. **Plos One** 8:e69202.

Everling DM, Bresolin T, Rorato PRN, Araujo RO, Boligon AA, Weber T, Dornelles MA, Campos LT (2014) Finishing precocity visual score and genetic associations with growth traits in Angus beef cattle. **Genetics and Molecular Research** 13:7757–7765.

Fernandes Júnior GA, Rosa GJM, Valente BD, Carvalheiro R, Baldi F, Garcia DA (2016a) Genomic prediction of breeding values for carcass traits in Nellore cattle. **Genetics Selection Evolution** 48:7.

Fernandes Júnior GA (2016b) Genome scan for postmortem carcass traits in Nellore cattle. **Journal of Animal Science** 94:4087–4095.

Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution** 43:1.

Gianola D (2013) Priors in whole-genome regression: the Bayesian alphabet returns. **Genetics** 194:573–596.

Guo G, Lund MS, Zhang Y, Su G (2010) Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. **Journal of Animal Breeding and Genetics** 127:423-32.

Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the bayesian alphabet for genomic selection. **BMC Bioinformatics** 12:186.

Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. **Genetics** 194:597–607.

Hair JF, Black WC, Babin BJ, Anderson RE (2009) **Multivariate data analysis**. Prentice Hall, Upper Saddle River, p. 761.

Hazel LN (1943) The genetic basis for constructing selection indexes. **Genetics** 28:476–490.

Hay EH and Roberts A (2018) Genome-wide association study for carcass traits in a composite beef cattle breed. **Livestock Science** 213:35-43.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. **Journal of Dairy Science** 92:433–443.

Hotelling H (1933) Analysis of a Complex of Statistical Variables Into Principal Components. **Journal of Educational Psychology** 24:417-441and 498-520.

Irano N, De Camargo GM, et al. (2016) Genome-Wide Association Study for Indicator Traits of Sexual Precocity in Nellore Cattle. **Plos One** 11:e0159502.

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. **Briefings in functional genomics** 9:166–177.

Kizilkaya K, Fernando RL, Garrick DJ (2010) Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. **Journal of Animal Science** 88:544-51.

Lee S, van der Werf J, Hayes B, Goddard M, Visscher P (2008) Predicting unobserved phenotypes for complex traits from whole-genome SNP data. **PLOS Genetics** 4:e1000231.

Lee T, Shin D-H, Cho S, Kang HS, Kim SH, Lee H-K, Kim H, Seo K-S (2014) Genome-wide Association Study of Integrated Meat Quality-related Traits of the Duroc Pig Breed. **Asian-Australasian Journal of Animal Sciences** 27:303–309.

Legarra A, Robert-Granie C, Manfredi E, Elsen JM (2008) Performance of genomic selection in mice. **Genetics** 180:611–8.

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

Lund MS, Sahana G, de Koning DJ, Su G, Carlborg O (2009) Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. **BMC Proceeding** 1:S1.

Lund MS, de Roos APW, de Vries AG, Druet T, Ducrocq V, Fritz S, Guillaume F, Guldbrandtsen B, Liu Z, Reents R, Schrooten C, Seefried F, Su G (2011) A common reference population from four European Holstein populations increases reliability of genomic predictions. **Genetics Selection Evolution** 43:43.

Macciotta NPP, Biffani S, Bernabucci U, Lacetera N, Vitali A, Ajmone-Marsan P, Nardone A (2017) Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. **Journal of Dairy Science** 100:4683-4697.

Magalhães AFB, de Camargo GMF, et al. (2016) Genome-Wide Association Study of Meat Quality Traits in Nellore Cattle. **Plos One** 11:e0157845.

Martínez R, Bejarano D, Gómez Y, Dasoneville R, Jiménez A, Even G, Sölkner J, Mészáros G (2017) Genome-wide association study for birth, weaning and yearling weight in Colombian Brahman cattle. **Genetics and Molecular Biology** 40:453–459.

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

Melo TP, Garcia Neto BF, Fortes MRS, Albuquerque LG, Carvalheiro R (2017) Agreement among GWAS results from different statistical methods as a strategy to increase the power of QTL detection. In: 22nd CONFERENCE OF THE ASSOCIATION FOR THE ADVANCEMENT OF ANIMAL BREEDING AND GENETICS. **Proceedings…** Queensland, p. 425-428

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. **Genetics** 157:1819–29.

Meuwissen TH and Goddard ME (2004) Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. **Genetics Selection Evolution** 36:261-79.

Meyer K (2007) Multivariate analyses of carcass traits for Angus cattle fitting reduced rank and factor analytic models. **Journal of Animal Breeding and Genetics** 124:50–63.

Neves HHR, Carvalheiro R, et al. (2014) Accuracy of genomic predictions in Bos indicus (Nellore) cattle. **Genetics Selection Evolution** 46:17.

Pearson K (1901) On lines and planes of closest fit to systems of points in space. **Philosophical Magazine** 2:559-572.

Roso VM and Fries LA (1995) Componentes principais em bovinos da raça Polled Hereford à desmama e sobreano. **Revista Brasileira de Zootecnia** 24:728–735.

Saowaphak P, Duangjinda M, Plaengkaeo S, Suwannasing R, Boonkum W (2017) Genetic correlation and genome-wide association study (GWAS) of the length of productive life, days open, and 305-days milk yield in crossbred Holstein dairy cattle. **Genetics and Molecular Research** 16:gmr16029091.

Savegnago RP, Caetano SL, Ramos SB, Nascimento GB, Schmidt GS, Ledur MC, Munari DP (2011) Estimates of genetic parameters, and cluster and principal components analysis of breeding values related to egg production traits in a White Leghorn population. **Poultry Science** 90:2174–2188.

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

Silva RMO, Stafuzza NB, Fragomeni BO, et al. (2017) Genome-wide association study for carcass traits in an experimental Nelore cattle population. **Plos One** 12:e0169860.

Smith HF (1936) A discriminant function for plant selection. Annals of Eugenics, London, 7:240-250.

Su G, Guldbrandtsen B, Gregersen VR, Lund MS (2010) Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. **Journal of Dairy Science** 93:1175-83.

Toral FLB, Roso VM, Araújo CV, Filho JCR (2011) Genetic parameters and response to selection for post-weaning weight gain, visual scores and carcass traits in Hereford and Hereford ✕ Nellore cattle. **Livestock Science** 137:231–237.

United States Department of Agriculture—USDA (2018) **Livestock and Products Annual**. Foreign Agricultural Service, 10p.

Valente TS, Baldi F, Sant'anna AC, Albuquerque LG, da Costa MJRP (2016) Genome-wide association study between single nucleotide polymorphisms and flight speed in Nellore cattle. **Plos One** 11:e0156956.

Vargas G, Neves HHR, Camargo GMF, Cardoso V, Munari DP, Carvalheiro R (2018) Genome-wide association study and functional analysis of feet and leg conformation traits in Nellore cattle. **Journal of Animal Science** 96:1617–1627.

Venturini GC, Savegnago RP, Nunes BN, Ledur MC, Schmidt GS, El Faro L, Munari DP (2013) Genetic parameters and principal component analysis for egg production from White Leghorn hens. **Poultry Science** 92:2283–2289.

Viana JMS, Piepho H-P, Silva FF (2017) Quantitative genetics theory for genomic selection and efficiency of genotypic value prediction in open-pollinated populations. **Scientia Agricola** 74:41–50.

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

Wang H, Misztal I, Aguilar I, Legarra A, Fernando RL, Vitezica Z, Okimoto R, Wing T, Hawken R, Muir WM (2014) Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics** 5:134.

Weber T, Rorato PRN, Lopes JS, Comin JG, Dornelles MA, De Araújo RO (2009) Parâmetros genéticos e tendências genéticas e fenotípicas para características produtivas e de conformação na fase pré-desmama em uma população da raça Aberdeen Angus. **Revista Brasileira de Zootecnia** 38:832–842.

Wiggans GR, Cooper TA, Vanraden PM, Cole JB (2011) Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. Journal of Dairy Science 94:6188–6193.

CHAPTER 2 - UNRAVELLING BIOLOGICAL BIOTYPES FOR GROWTH, VISUAL SCORE AND REPRODUCTIVE TRAITS IN NELLORE CATTLE VIA PRINCIPAL COMPONENT ANALYSIS[1]

ABSTRACT – Principal component analysis (PCA) is used to summarize important information from multivariate data in sets of new variables named principal components (PCs). In animal breeding, these new composite variables can be used to study the associations among multiple traits using the magnitude and direction of the PCA coefficients (in the eigenvectors) for each trait. Phenotypic data from 355,524 Nellore animals were used to estimate genetic parameters and explore the relationship among growth (weaning and post-weaning weight gain), visual score (weaning and yearling conformation, finishing precocity and muscling) and reproductive (scrotal circumference) traits using PCA. Genetic parameters were estimated by multi-trait analysis using a mixed linear animal model. The eigen-decomposition of the additive genetic (co)variance matrix (A_T matrix) obtained using multi-trait analysis were used to calculate the PCs. In addition, PCA using the (co)variance matrix of the breeding values (EBVs) from single- and multi-trait analyses were investigated for comparison purposes. The direct heritability estimates for the weaning and yearling traits ranged from 0.17 (birth-to-weaning weight gain and conformation) to 0.21 (finishing precocity) and from 0.18 (weaning-to-yearling weight gain) to 0.46 (scrotal circumference), respectively. Genetic correlations estimated among all analyzed traits were positive (favorable) ranging from 0.15 (conformation at weaning and scrotal circumference) to 0.96 (finishing precocity and muscling at weaning). The first three PCs from multi-trait analysis using the eigen-decomposition of the A_T matrix, explained 87.11% of the total additive genetic variance for the traits. The first PC (PC1) had negative and relatively similar coefficients for all traits, the second PC (PC2) contrasted the animals with early or late biotype, and the third PC (PC3) characterized a contrast between weaning and yearling traits. Our findings suggest that the PCA could be explored in breeding programs to select Nellore cattle to tailor selection towards specific PC, targeting, for instance, faster growth and precocious biotype.

Keywords: beef cattle, Bos taurus indicus, eigen-decomposition, genetic correlation, principal components

2.1 INTRODUCTION

Currently, Brazil has the second largest commercial beef cattle herd and is among the top beef exporters in the world (Brazilian Association of Meat Exporters, ABIEC, 2017). This is partially due to the prevalence of the Nellore breed (*Bos taurus indicus*, ~80% of the total beef cattle herd), which is well adapted to harsh environments and performs well under tropical conditions. Compared to taurine cattle breeds, Nellore is heat tolerant and parasite resistant, and has good maternal ability and high metabolic and reproductive efficiency under tropical pasture conditions (Silva et al., 2010). Well-structured breeding schemes and large genetic variability (Magnabosco et al., 2014; Buzanskas et al., 2017) have contributed substantially to the increase of production efficiency of the Nellore breed through genetic selection.

Successful livestock breeding programs start with the definition of breeding objectives, which requires a balance between indicator traits of economic efficiency, animal health and welfare and final product quality to meet the consumers' needs. Breeding objectives are a combination of economically important traits recorded in the production system and that will constitute the selection criteria (Smith, 1983). Thus, the genetic variation and relationship among these various traits are of great importance, as highly genetic correlated traits will add little extra information to the analyses. For instance, in Nellore cattle, there is remarkable variability in breeding objectives and different biotypes (set of desirable traits/phenotypes) are observed (Cardoso et al., 2003; Carvalheiro and Cavalcanti, 2008).

When working with a set of correlated traits, principal component analysis (PCA) is one of the most popular multivariate techniques to perform dimension reduction for these traits and therefore, facilitate simultaneous genetic selection for all traits of interest. In brief, PCA aims to summarize a set of originally-correlated variables into a smaller set of uncorrelated variables (principal components, PCs), but still retaining most of the original variability (Hair et al., 2009). This set of

uncorrelated variables is expected to contain most of the genetic variation for the traits of interest and is often associated with just the first two or three PCs, allowing a better description of the studied population with a relatively small number of parameters (Kirkpatrick and Meyer, 2004). An alternative PCA approach introduced by Meyer and Kirkpatrick (2005) has been used to estimate genetic PCs directly through a reparameterisation of the usual linear mixed model. These authors showed that reduced rank estimation can substantially reduce computational demand of multivariate analyses and improve convergence rates. In animal breeding, PCA have been used to investigate genetic relationships among traits and to unravel potential biological associations among traits, usually not observed in the original data (Savegnago et al., 2011; Vohra et al., 2015; Boligon et al., 2016).

Usually, when performing PCA, PCs are calculated based on the (co)variance matrix of the estimated breeding values (EBVs) obtained for each trait (e.g., Savegnago et al., 2011; Buzanskas et al., 2013). For instance, Boligon et al. (2016) used PCA to investigate the variability and the relationship among EBVs for growth and reproductive traits in Nellore cattle from single-trait analyses and showed that the first three PCs were sufficient to explain the majority of the variability among the EBVs. The authors observed that the first PC contrasted mainly animals with high or low growth. The second PC differentiate the animals according to their maturation rate (weaning versus yearling performance), while the third PC contrasted the animals with early or late biotype. Nonetheless, when performing PCA applied on the (co)variance matrix of the EBVs obtained using single-trait analyses, the covariance among traits are not considered and, therefore, the EBVs are estimated not taking into account the genetic relationships between traits. In an effort to overcome this limitation, an alternative would be to estimate the PCs from the eigen-decomposition of the additive genetic (co)variance matrix ($\mathbf{A_T}$). The reason to use the $\mathbf{A_T}$ matrix instead of the (co)variance matrix of the EBVs is because the latter is an expedite approximation of $\mathbf{A_T}$. The $\mathbf{A_T}$ matrix corresponds to the additive genetic relationship matrix among traits and summarizes the additive genetic variation within the set of traits and the (co)variance among them. The eigen-decomposition of the $\mathbf{A_T}$ matrix can be used to capture the distribution (orientation) for different genetic effects in

multivariate space, contrasting the individuals according to the additive genetic effect of the studied traits (Roso and Fries, 1995).

The objectives of the present study were: (1) to estimate genetic parameters for growth (weaning and post-weaning weight gain), visual score (weaning and yearling conformation, precocity and muscling) and reproductive (scrotal circumference) traits in Nellore cattle; (2) to compare three different approaches applied to PCA using: an eigen-decomposition of either the $\mathbf{A_T}$ matrix or the (co)variance matrix of the EBVs and unravel potential biological interpretation for principal components; and (3) to investigate the feasibility of using the PCs to genetically select for a specific breeding objective based on desirable biotypes for Nellore cattle.

2.2 MATERIAL AND METHODS

2.2.1 Dataset description

Phenotypic and pedigree information from 355,524 Nellore animals born between 1990 and 2010 were obtained from Alliance Nellore database (www.gensys.com.br). These animals were raised in pasture production systems under tropical conditions in 246 farms located in Brazil and Paraguay. The number of observations, mean or median (for score traits), standard deviation (SD) and number of sires, dams, and contemporary groups (CG) for each trait are presented in Table 1. In this study we analyzed the following traits: birth-to-weaning weight gain (BWG), and visual scores of conformation (WC), finishing precocity (WP) and muscling (WM) at weaning (about 205 days of age), weaning-to-yearling weight gain (WYG), and visual scores of conformation (YC), finishing precocity (YP) and, muscling (YM) at yearling and, scrotal circumference (SC) at yearling (about 550 days of age).

Table 1. Description of the dataset used in the analyses, including number of observations (N), mean or median, standard deviation (SD) and number of sires, dams, and contemporary group (CG) in Nellore cattle

Traits[a]	N	Mean	SD	Sires	Dams	CG
BWG	334,123	147.20	29.86	3773	178,295	5127
WC	334,123	3[b]	-	3773	178,295	5127
WP	334,123	3[b]	-	3773	178,295	5127
WM	334,123	3[b]	-	3773	178,295	5127
WYG	141,131	92.92	34.84	2802	94,090	2639
YC	165,287	3[b]	-	3033	111,341	2203
YP	165,287	3[b]	-	3033	111,341	2203
YM	165,287	3[b]	-	3033	111,341	2203
SC	58,912	26.42	3.55	2298	47,668	1133

[a]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference.
[b]Median.

Conformation, finishing precocity and muscling were evaluated based on the CPMU (conformation, precocity, muscling and navel) method, which was developed by Gensys (gensys.com.br) and has been used by several beef cattle breeding programs in Brazil. According to this method, scores are assigned for each animal by trained technicians on the basis of the variation within the CG. Visual scores ranged from one to five, in which five represents the best expression of the trait and one represents the worst. The evaluation of conformation score is based on the amount of meat on the animal carcass, mainly influenced by size and muscularity. Finishing precocity score is related to the capacity of the animal to store subcutaneous fat precociously. For this trait, the evaluation of the proportion between ribs and legs is also performed. The muscling score is a measure that reflects muscle development of the animal body. Scrotal circumference was measured (in centimeters) at yearling and is a sexual precocity indicator in Nellore breeding programs. This set of traits is

used in selection indexes of some Nellore breeding programs. For instance, a selection index that is used to select animals at weaning is composed of the following traits (relative importance in %): days to reach 160 kg from birth to weaning (60%), conformation (8%), finishing precocity (16%) and muscling (16%) at weaning. Another selection index used to rank yearling animals is composed by of the following traits (relative importance in %): days to reach 160 kg from birth to weaning (23%), conformation (4%), finishing precocity (8%) and muscling (8%) at weaning, days to reach 240 kg post-weaning (23%), conformation (4%), finishing precocity (8%) and muscling (8%) at yearling, and scrotal circunference (14%).

2.2.2 Definition of contemporary groups and covariates

The fixed effects considered in the formation of the contemporary groups (CG) for each trait are presented in Table 2. Birth seasons were defined as (1) January to March, (2) April to June, (3) July to September, and (4) October to December. Records of weight gains and scrotal circumference exceeding 3.5 standard deviations (SD) from the overall mean of the CG were removed. For all traits, CGs with fewer than 10 records were also removed. The model included the covariate 'age of animal at recording' as linear effect for all traits and 'age of dam at calving' as linear and quadratic effects for the weaning traits.

Table 2. Fixed effects included in the definition of contemporary groups for birth to weaning weight gain (BWG), conformation, precocity, and muscling at weaning (WC, WP, and WM, respectively), weaning to yearling weight gain (WYG), conformation, precocity, and muscling at yearling (YC, YP, and YM, respectively) and scrotal circumference (SC) in Nellore cattle.

Fixed effect	Traits								
	BWG	WC	WP	WM	WYG	YC	YP	YM	SC
Farm at birth	x	x	x	x	x	x	x	x	x
Birth year	x	x	x	x	x	x	x	x	x
Birth season	x	x	x	x	x	x	x	x	x

Sex	x	x	x	x	x	x	x	x	
Weaning management group	x	x	x	x	x				x
Weaning date	x	x	x	x	x				x
Yearling management group						x	x	x	x	x
Yearling date						x	x	x	x	x

2.2.3 Estimates of genetic parameters

Both single- and multiple-trait model were fitted using Wombat software (Meyer, 2007) to estimate variance components and EBVs. For weaning traits, the statistical model included the fixed effect of CG and covariates previously described as well as random direct additive genetic, maternal genetic, maternal permanent environmental and residual effects. In matrix notation the equation model for each weaning trait was:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z_a a} + \mathbf{Z_m m} + \mathbf{Wp} + \mathbf{e}$$

where: $\mathbf{y}$ is the vector of observations for each trait, $\mathbf{\beta}$ is a vector of systematic effects (CG and covariates), $\mathbf{a}$ is a vector of random direct additive genetic effects, $\mathbf{m}$ is a vector of random maternal genetic effects, $\mathbf{p}$ is a vector of random maternal permanent environmental effects, and $\mathbf{e}$ is a vector of random residual effects. $\mathbf{X}$, $\mathbf{Z_a}$, $\mathbf{Z_m}$, and $\mathbf{W}$ are the incidence matrices relating elements in $\mathbf{\beta}$, $\mathbf{a}$, $\mathbf{m}$, and $\mathbf{p}$ to $\mathbf{y}$, respectively. For yearling traits the model did not include the effects $\mathbf{m}$ and $\mathbf{p}$. The assumptions of the multi-trait model were:

$$\text{Var}\begin{pmatrix} \mathbf{a} \\ \mathbf{m} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} \mathbf{G_a} \otimes \mathbf{A} & \mathbf{G_{am}} \otimes \mathbf{A} & 0 & 0 \\ \mathbf{G_{ma}} \otimes \mathbf{A} & \mathbf{G_m} \otimes \mathbf{A} & 0 & 0 \\ 0 & 0 & \mathbf{G_p} \otimes \mathbf{I_c} & 0 \\ 0 & 0 & 0 & \mathbf{R} \otimes \mathbf{I_n} \end{pmatrix}$$

$$\text{E}\begin{pmatrix} \mathbf{a} \\ \mathbf{m} \\ \mathbf{p} \\ \mathbf{e} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and $A_T = \begin{pmatrix} G_a & G_{am} \\ G_{ma} & G_m \end{pmatrix}$

where: G_a is a variance-covariance matrix of random direct genetic effects within and across traits, G_m is a variance-covariance matrix of random maternal genetic effects within and across traits, G_{am} and G_{ma} are matrices of genetic (co)variances between direct and maternal effects within and across traits, G_p is a variance-(co)variance matrix of random maternal permanent environmental effects within and across traits, R is a variance-covariance matrix of random residual effects within and across traits, A is the additive genetic relationship matrix, I_c is an identity matrix of order equal to the number of dams, I_n is an identity matrix of order equal to the number of animals with phenotypic records, $\otimes$ is the Kronecker product operator, and 0 is a vector of zeros.

Principal components were calculated from the eigen-decomposition of the A_T matrix using the eigen function implemented in the R software (R Development Core Team, 2017). The PC eigenvalue is associated with the variance of all nine traits included in the PC. Each eigenvalue has a corresponding unitary vector (weight) named eigenvector (Rencher, 2002), which represent the strength and direction of the variance of each variable within the PC. Thus, in practice, PCs are a combination of traits that potentially have a biological meaning. We used the Kaiser criterion (Kaiser, 1960) to identify the PCs that explained the largest proportion of the total genetic variation of the traits. According to the Kaiser criterion, only the PCs with eigenvalues above the unit should be considered. The first PC (PC1) explains the largest percentage of the genetic variation of genetic variance of the traits, while the second one (PC2) explains the second largest percentage, and so on.

In order to compare the PC approach used in this study to the method used by Boligon et al. (2016), in which genetic parameters and breeding values from single-trait analyses were obtained using a model including the same effects as considered in the present study, PCA was also applied to the (co)variance matrix of the EBVs obtained from single- and multi-trait analyses. The PCA were carried out using the *prcomp* function implemented in the R software (R Development Core Team, 2017). The EBVs were standardized to zero mean and variance equal to one to avoid scale

effects and to facilitate comparison between traits. In addition, in order to compare the relative ranking of animals based on EBVs, Spearman rank correlation coefficients between the EBVs obtained using the PC approaches were calculated for the first three PCs.

2.3 RESULTS

2.3.1 Estimation of genetic parameters

We have observed direct heritability estimates for weaning traits (BWG, WC, WP and WM) ranging from 0.17 (BWG and WC) to 0.21 (WP) and for yearling traits (WYG, YC, YP, YM and SC) from 0.18 (WYG) to 0.46 (SC), indicating that all these traits are under moderate genetic control (Table 3). The contribution of maternal effects to the phenotypic variance for BWG (0.15) was higher compared to the estimates for conformation, precocity and muscling (range: 0.07 to 0.08). The proportion of phenotypic variance due to maternal permanent environmental effects was almost the same for BWG and all visual scores at weaning (range: 0.12 to 0.13).

Table 3. Variance estimates, heritability and proportion of phenotypic variance due to maternal permanent environmental effects from multi-trait analysis for growth, visual score and reproductive traits in Nellore cattle.

Traits[a]	Variance components[b]					Heritability[b]		
	σ_a^2	σ_m^2	σ_{am}	σ_{pe}^2	σ_e^2	h_a^2	h_m^2	c^2
BWG	57.70	50.21	9.67	44.18	181.62	0.17 (0.01)	0.15 (0.01)	0.13
WC	0.17	0.08	0.01	0.13	0.62	0.17 (0.01)	0.08 (0.01)	0.13
WP	0.25	0.08	0.01	0.14	0.69	0.21 (0.01)	0.07 (0.01)	0.12
WM	0.24	0.09	0.01	0.16	0.71	0.20 (0.01)	0.07 (0.01)	0.13
WYG	57.71	-	-	-	252.56	0.18 (0.01)	-	-
YC	0.33	-	-	-	0.65	0.34 (0.01)	-	-
YP	0.39	-	-	-	0.74	0.34 (0.01)	-	-

YM	0.38	-	-	-	0.75	0.34 (0.01)	-	-
SC	3.23	-	-	-	3.83	0.46 (0.02)	-	-

[a]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference

[b]σ_a^2: direct additive genetic variance, σ_m^2: maternal genetic variance, σ_{am}: covariance between direct and maternal effects, σ_{pe}^2: maternal permanent environmental variance, σ_e^2: residual variance, h_a^2: direct genetic heritability and standard error, h_m^2: maternal genetic heritability and standard error, c^2: proportion of phenotypic variance due to maternal permanent environmental effects

Genetic correlations ranged from 0.39 (WC and WP) to 0.96 (WP and WM) and from 0.25 (YC and YP, YC and YM) to 0.94 (YP and YM) for weaning and yearling traits, respectively (Table 4). The genetic correlations between weaning and yearling traits were of low to high magnitude (0.15 to 0.86). Phenotypic correlations ranged from 0.60 (WC and WP) to 0.82 (WP and WM) and from 0.23 (WYG and SC) to 0.78 (YP and YM) for weaning traits and yearling traits, respectively. The phenotypic correlation between weaning and yearling traits were of null to moderate magnitude (−0.009 to 0.56).

Table 4. Estimates of genetic correlations (above the diagonal) and phenotypic correlations (below the diagonal) between the studied traits of Nellore cattle.

Traits[a]	BWG	WC	WP	WM	WYG	YC	YP	YM	SC
BWG		0.88 (0.01)	0.52 (0.02)	0.56 (0.02)	0.25 (0.03)	0.85 (0.009)	0.43 (0.02)	0.46 (0.02)	0.39 (0.03)
WC	0.76 (0.01)		0.39 (0.02)	0.42 (0.02)	0.19 (0.03)	0.85 (0.01)	0.21 (0.03)	0.22 (0.03)	0.15 (0.03)
WP	0.66 (0.01)	0.60 (0.01)		0.96 (0.01)	0.22 (0.03)	0.42 (0.02)	0.86 (0.01)	0.83 (0.01)	0.46 (0.03)
WM	0.67 (0.01)	0.63 (0.01)	0.82 (0.01)		0.18 (0.03)	0.45 (0.02)	0.79 (0.01)	0.84(0.01)	0.42 (0.03)
WYG	-0.07 (0.01)	-0.009 (0.01)	-0.04 (0.01)	-0.05 (0.01)		0.57 (0.02)	0.53 (0.02)	0.49 (0.02)	0.29 (0.03)
YC	0.56 (0.01)	0.49 (0.01)	0.31 (0.01)	0.31 (0.01)	0.43 (0.01)		0.25 (0.02)	0.25 (0.02)	0.29 (0.02)
YP	0.30 (0.01)	0.19 (0.01)	0.43 (0.01)	0.39 (0.01)	0.43 (0.01)	0.45 (0.01)		0.94 (0.01)	0.43 (0.02)
YM	0.31 (0.01)	0.21 (0.01)	0.41 (0.01)	0.42 (0.01)	0.42 (0.01)	0.47 (0.01)	0.78 (0.01)		0.43 (0.02)
SC	0.33 (0.01)	0.22 (0.01)	0.26 (0.01)	0.25 (0.01)	0.23 (0.01)	0.31 (0.01)	0.30 (0.01)	0.29 (0.01)	

[a]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference
Standard errors of estimated correlations are presented within brackets

2.3.2 Principal component analysis

Table 5 presents the eigenvalues, proportion and cumulative sum of the explained variance over the nine components for the three PC approaches. The first three PCs were chosen using the eigen-decomposition of the $\mathbf{A_T}$ matrix (which explained 87.11% of the total additive genetic variance for the traits) and using the PCA applied on the (co)variance matrix of the EBVs obtained using single-trait analysis (which explained 79.33% of the total variance of the EBVs). For the PC approach applied on the (co) variance matrix of the EBVs obtained using multi-trait analysis, the first two PCs would be chosen according to the Kaiser's criterion (explaining 79.67% of the total variance of the EBVs). However the third component was also included for purposes of comparison.

Table 6 shows the eigenvector coefficients for the first three PCs. From the nine original dimensions using the eigen-decomposition of the $\mathbf{A_T}$ matrix, PC1 explained 55.11% of the total variation and had negative and moderate coefficients for all traits. The PC2 accounted for 20.78% of the total genetic variation and had contrasting coefficients for weight gain and conformation to finishing precocity and muscling, both at weaning and yearling, and scrotal circumference at yearling. The third PC explained 11.22% of the total genetic variation, suggesting that this PC differentiates the animals according to their maturation rate (weaning versus yearling performance).

Table 5. Principal components (PC), eigenvalues (λ_i), proportion and cumulative sum of the explained variance over the nine components for three approaches for principal component analysis (PCA).

PC[a]	PCA applied to the $\mathbf{A_T}$ matrix obtained using multi-trait analysis			PCA applied to the (co)variance matrix of the EBVs from single-trait analyses			PCA applied to the (co)variance matrix of the EBVs from multi-trait analysis		
	λ_i	Proportion (%)	Cumulative sum (%)	λ_i	Proportion (%)	Cumulative sum (%)	λ_i	Proportion (%)	Cumulative sum (%)
PC1	4.96	55.11	55.11	4.56	50.67	50.67	5.45	60.56	60.56
PC2	1.87	20.78	75.89	1.55	17.22	67.89	1.72	19.11	79.67
PC3	1.01	11.22	87.11	1.03	11.44	79.33	0.73	8.11	87.78
PC4	0.76	8.44	95.55	0.76	8.44	87.77	0.52	5.78	93.56
PC5	0.17	1.89	97.44	0.50	5.55	93.32	0.38	4.22	97.78
PC6	0.15	1.67	99.11	0.22	2.44	95.76	0.09	1.00	98.78
PC7	0.05	0.56	99.67	0.16	1.78	97.54	0.06	0.67	99.45
PC8	0.03	0.33	100	0.14	1.56	99.1	0.03	0.33	99.78
PC9	0.00	0.00	100	0.08	0.90	100	0.02	0.22	100

[a]PC*n*: *n*th principal component

Table 6. Eigenvectors for the first three principal components (PCs) obtained using three approaches for principal component analysis (PCA).

PC[a]	PCA applied to the $\mathbf{A_T}$ matrix obtained using multi-trait analysis			PCA applied to the (co)variance matrix of the EBVs from single-trait analyses			PCA applied to the (co)variance matrix of the EBVs from multi-trait analysis		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
BWG[b]	-0.36	0.39	-0.13	0.35	-0.42	0.13	0.32	-0.42	0.07
WC	-0.29	0.51	-0.19	0.32	-0.46	0.29	0.19	-0.50	0.51
WP	-0.39	-0.26	-0.26	0.39	-0.17	-0.36	0.38	0.26	0.22
WM	-0.39	-0.22	-0.33	0.39	-0.23	-0.34	0.39	0.20	0.32
WYG	-0.22	0.01	0.84	0.24	0.39	0.43	0.29	-0.11	-0.72
YC	-0.30	0.52	0.15	0.31	0.10	0.59	0.26	-0.56	-0.19
YP	-0.37	-0.32	0.12	0.36	0.38	-0.13	0.39	0.25	-0.05
YM	-0.38	-0.30	0.06	0.37	0.36	-0.14	0.39	0.24	0.01
SC	-0.24	-0.10	0.13	0.21	0.29	-0.26	0.31	0.11	-0.13

[a]PCn: nth principal component
[b]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference
Standard errors of estimated correlations are presented within brackets

For the PCA applied on the (co)variance matrix of the EBVs using single-trait analyses, PC1 explained 50.67% of the total variation and had positive and moderate coefficients for all traits. The PC2 explained 17.22% of total genetic variation and had the same direction of the coefficients obtained for PC3 using the eigen decomposition of the A_T matrix, allowing to contrast weaning and yearling traits. The third PC explained 11.44% of the total genetic variation, presented similar direction obtained for PC2 using the A_T matrix and, in general, differentiate the animals for weight gain and conformation in contrast to finishing precocity and muscling, at both at ages (weaning and yearling), and scrotal circumference at yearling. For the PCA applied on the (co)variance matrix of the EBVs using multi-trait analyses, PC1 explained 60.56% of the total variation and had positive and moderate coefficients for all traits. The PC2 and PC3 explained 19.11% and 8.11% of the total variation, respectively, and had similar direction for the coefficients obtained from the PC approach using the A_T matrix.

Spearman rank correlations for the first three PCs ranged from 0.61 to 0.99 and the largest correlations were obtained for PC1 (0.93 to 0.99). For PC2 and PC3, the correlations were calculated considering the correspondence between these PCs obtained from the different approaches. In other words, the PC2 obtained based on the eigen-decomposition of the A_T matrix and from multi-trait analysis corresponds to the PC3 obtained from single-trait analyses and vice-versa. For PC2, Spearman rank correlation obtained between the EBVs using the A_T matrix and from multi-trait analyses was higher (0.98) compared to the correlation obtained between the EBVs from single-trait analyses and the A_T matrix (0.71) and from single- and multi-trait analysis (0.76). Similar results were observed for the correlations obtained for PC3, where the highest correlation were obtained between EBVs using the A_T matrix and from multi-trait analyses (0.90), versus the correlations obtained between EBVs from single-trait analyses and the A_T matrix (0.61) and from single- and multi-trait analysis (0.62).

2.4 DISCUSSION

The direct heritability estimates obtained for visual scores at both ages (weaning and yearling) were close to those described in the literature for tropical beef cattle, with estimates ranging from 0.19 to 0.54 (Regatieri et al., 2011; Boligon et al., 2016; Gordo et al., 2016). Overall, as expected, the direct heritability estimates for growth traits and visual scores suggest that these traits are under moderate genetic control and should respond to selection. Furthermore, the response for traits measured at yearling is expected to be higher compared to that for traits measured at weaning. For weaning visual scores, the contribution of the maternal genetic effect to phenotypic variance ranged from 7 to 8%. These findings are in agreement with Boligon et al. (2011b and 2016) and Koury Filho et al. (2010), who reported heritability estimates ranging from 0.04 to 0.07 in Nellore cattle. The maternal influence is related to any contribution on offspring phenotypes that is attributed to its mother (Willham, 1972), and models that do not account for maternal effects could yield higher estimates of additive direct genetic variance (Meyer, 1992). From this study, significant genetic changes for weaning visual scores are not expected from the maternal genetic effect, since most of the total additive genetic variance is due to direct genetic effects.

Genetic correlations estimated between the nine traits were positive (favorable) and ranged from 0.15 (WC and SC) to 0.96 (WP and WM). The genetic correlations between WYG and visual scores at yearling were of moderate magnitude (0.49 to 0.57) and followed the same trend as the correlations between BWG and visual scores at weaning (0.52 to 0.88). The genetic correlations estimated between visual scores at different ages were of moderate to high magnitude, suggesting that genetic gains in visual scores from weaning to yearling might be obtained by the correlated responses of these traits. According to Koury Filho et al. (2010), both stages (weaning and yearling) are of interest for genetically evaluating the animals since the evaluation at weaning does not involve pre-selection and, at yearling, the morphological traits can better express the direct genetic merit of an animal, since there is greater trait variability among individuals.

The genetic correlations between the same trait evaluated in both periods were higher for finishing precocity and muscling, indicating that these scores are mainly controlled by a similar group of genes. Similar results were also reported by Cardoso et al. (2004), Bertipaglia et al. (2012) and Boligon et al. (2016), who estimated genetic correlation between the two periods ranging from 0.71 to 0.95 for finishing precocity and muscling in Angus, Brahman, and Nellore cattle, respectively. This finding was expected since precocious animals present faster and greater muscle mass development (Koury Filho et al., 2009). Thus, the identification of the best genotypes for muscularity is also associated with the best genetic make-up for finishing precocity. The genetic correlations between conformation and finishing precocity were higher at weaning (0.39) than at yearling (0.25), indicating that selection for body size does not necessarily result in precocious animals. Some Nellore animals present considerable height and length, without a high depth of ribs, resulting in higher conformation scores and lower finishing precocity scores. Thus, larger animals tend to reach the finishing stage at an older age. The same was observed for the genetic correlation between conformation and muscling, which was 0.42 at weaning and 0.25 at yearling, indicating that selection for larger animals at yearling does not result to substantial changes in muscle mass development.

In general, the genetic relationship among traits is related to the number of PCs that explain most of the total variance in the data. In the present study, three PCs explained a large proportion of the total additive genetic variance for growth and visual score traits at weaning and yearling. Buzanskas et al. (2013) performed a PCA using standardized genetic values predicted for weight at 420 days of age, age at first and second calving, and calving interval in Canchim beef cattle and reported that only PC1 met the Kaiser criterion (> 1) and explained 48.51% of the total variance. Boligon et al. (2016) reported that the first three PCs comprised 79.06% of the total estimated breeding value variation of growth and reproductive traits in Nellore cattle.

The additive genetic variance explained by the first three PCs using the A_T matrix was higher than that explained using the (co)variance matrix of the EBVs from single-trait analyses (87.11% vs 79.33%, respectively). However, the proportion of variance explained by the first three PCs in using A_T matrix was smaller when compared with the PCA applied on the (co)variance matrix of the EBVs from multi-

trait analysis (87.11% vs 87.78%, respectively). The PCA applied to the (co) variance matrix of the EBVs from single-trait analyses presented the smallest additive genetic variance explained by the first three PCs. For the three PC approaches, the magnitude between the coefficients (in the eigenvectors) obtained for the first PC was similar. However, the differences between the coefficients for the other PCs were more substantial.

The magnitude of the eigenvector (either positive or negative) indicates the importance of the corresponding trait on the PC, thus, the larger coefficient would indicate greater discriminatory power. For the three PC approaches evaluated, the first PC showed a moderate coefficient for the nine original traits, suggesting a weighted average of the original traits. The common sign (either negative or positive) for all coefficients indicates that these variables have the same direction of variation. Thus, it would follow that animals with high PC estimates would tend to have, on average, higher performance compared to animals with low values for the same variables, especially contrasting high or low growth rate animals. These findings are in agreement with Boligon et al. (2016), who reported positive and similar coefficients for the first PC. Similar results were found by Roso and Fries (1995), who performed a PCA to investigate the relationships between adjusted weight and visual score traits in Polled Hereford. The authors observed that the first PC presented positive and similar coefficients for all traits, contrasting animals of distinct size or body volumes.

The second PC obtained using the **A$_T$** matrix and the (co)variance matrix of the EBVs from multi-trait analysis, and the third PC obtained using the (co)variance matrix of the EBVs from single-trait analyses, show contrasting animals with early (better for finishing precocity, muscling and scrotal circumference) and late (better for weight gain and conformation) maturity biotypes at weaning and yearling. Conformation scores are attributed to each animal according to the visual assessment of carcass weight, by evaluating the length and body depth as well as muscle development. Finishing precocity and muscling are related to the capacity of an animal to achieve a minimum degree of carcass finishing at a relatively low weight and muscle mass development as a whole, respectively. Tall and thin animals, with a shallow rib depth are considered to be late, thus receiving low scores for finishing

precocity. While animals with deeper ribs, a better degree of finishing and fat deposition, mainly at the base of the tail and groin are earlier maturity animals (Boligon et al., 2011b). Biologically, larger animals (higher conformation scores) tend to present delayed muscle mass development and carcass finishing. This might be explained by considering the growth curve definition, in which the skeleton develops first, followed by the muscle mass and, finally, adipose tissue (Boggs and Merkel, 1993). Thus, larger animals require more time to complete skeletal development and, therefore, will reach the finishing precocity at an older age. Ideally, a combination of growth precocity (conformation and weight gain) with finishing precocity is desirable to yield animals with a balanced biotype in a short time period, which would be more profitable to beef cattle producers. The difference between the order of the PCs probably is due to the fact the (co)variance of EBVs from univariate analyses did not take into account the genetic correlation between traits (when calculating the EBVs) contrarily to when using A_T or (co)variance matrix of the EBVs from multi-trait analysis.

The third PC obtained using the A_T matrix and the (co)variance matrix of the EBVs from multi-trait analysis, and the second PC obtained using the (co)variance matrix of the EBVs from single-trait analyses, differentiate weaning traits and yearling traits, contrasting animals according to their maturation rate (weaning versus yearling performance), being possibly associated with genetic potential of the animals to growth at different ages. In beef cattle, the growth during the pre-weaning period is important, because is the highest growth rate of the animal's life and factors such as maternal influence (mainly regarding the milk production and maternal ability), year of birth and age of dam (Teixeira and Albuquerque, 2003) are considered essential to the expression of the animal's genetic potential. In contrast, the post-weaning period represents the environment condition in which the animals are raised without direct influence of maternal effects (Cardoso et al., 2004).

In general, the rank correlations between EBVs for the PC1 were high using the different PC approaches (0.93 to 0.99), indicating that practically the same animals would be selected based on EBVs for PC1 predicted based on A_T matrix or the (co)variance matrix of the EBVs. Rank correlations between the PCs obtained using the A_T matrix and the (co)variance matrix of the EBVs from multi-trait analysis

(0.99) were higher when compared to the other pair of correlations (0.93 and 0.94). However, the correlations based on the PC2 and PC3 obtained using the (co)variance matrix of the EBVs from single-trait analysis and from the $\mathbf{A_T}$ matrix or the (co)variance matrix of the EBVs from multi-trait analysis were smaller (0.61 to 0.76), indicating substantial re-ranking between animals derived from multi-trait PC analyses and from single-trait analysis.

Our findings suggest that PCA can be used in Nellore commercial breeding programs aiming to select for specific biotypes (breeding objectives). For selection purposes, we recommend the use of the first three PCs from the eigen-decomposition of the additive genetic (co) variance matrix to identify the top genetic merit individuals out of the pool of breeding candidates.

2.5 CONCLUSIONS

The growth and visual score traits measured at weaning and yearling in Nellore cattle are under moderate genetic control. Principal component analysis showed that the first three principal components are sufficient to explain most of the genetic variance among weaning and yearling growth and score traits, and scrotal circumference in Nellore cattle. For the principal component approaches investigated, the eigen-decomposition of the additive genetic (co)variance matrix and of the (co)variance matrix of the estimated breeding values, allowed for similar biological interpretation of the genetic PCs. Our results suggest that the principal component analysis could be explored in breeding programs to select Nellore cattle to tailor selection towards specific PC, targeting, for instance, faster growth and precocious biotype.

2.6 REFERENCES

Associação Brasileira das Indústrias Exportadoras de Carnes (ABIEC) (2017) Disponível em: <http://www.abiec.com.br/download/estatisticas-ago17.pdf>. Acesso em: 10 out. 2017.

Bertipaglia TS, Carreño LOD, Machado CHC, Andrighetto C, Fonseca R (2012) Estimates of genetic parameters for visual scores and their correlation with production and reproductive traits in Brahman cattle. **Revista Brasileira de Zootecnia** 41:1407-1411.

Boggs DL, Merkel RA (1993) Live Animals Carcass Evaluation and Selection Manual. Kendall/Hunt Pub. Co., Metro Washington Park Zoo (USA) p. 221.

Boligon AA, Mercadante MEZ, Albuquerque LG (2011b) Genetic associations of conformation, finishing precocity and muscling visual scores with mature weight in Nelore cattle. **Livestock Science** 135:238-243.

Boligon AA, Vicente IS, Vaz RZ, Campos GS, Souza FRP, Carvalheiro R, Albuquerque LG (2016) Principal component analysis of breeding values for growth and reproductive traits and genetic association with adult size in beef cattle. **Journal of Animal Science** 94:5014-5022.

Buzanskas ME, Savegnago RP, Grossi DA, Venturini GC, Queiroz SA, Silva LOC, Torres Jr RAA, Munari DP, Alencar MM (2013) Genetic parameter estimates and principal component analysis of breeding values of reproduction and growth traits in female Canchim cattle. **Reproduction, Fertility and Development** 25:775-781.

Buzanskas ME, Piresa PS, Chuda TCS, Bernardesa PA, Rola LD, Savegnago RP, Lôbo RB, Munari DP (2017) Parameter estimates for reproductive and carcass traits in Nelore beef cattle. **Theriogenology** 92:204-209.

Cardoso V, Roso VM, Severo JLP, Queiroz SA, Fries LA (2003) Formando lotes uniformes de reprodutores múltiplos e usando-os em acasalamentos dirigidos, em populações Nelore. **Revista Brasileira de Zootecnia** 32:834-842.

Cardoso FF, Cardellino RA, Campos LT (2004) Componentes de (co)variância e parâmetros genéticos de caracteres pós-desmama em bovinos da raça Angus. **Revista Brasileira de Zootecnia** 33:313-319.

Carvalheiro R and Cavalcanti JR (2008) Does the biotype of the animals from your herd match your breeding objective? In: VII SIMPÓSIO BRASILEIRO DE MELHORAMENTO ANIMAL. **Proceedings...** São Carlos.

Gordo DGM, Espigolan R, et al (2016) Genetic parameter estimates for carcass traits and visual scores including or not genomic information. **Journal of Animal Science** 94:1821-1826.

Hair JF, Black WC, Babin BJ, Anderson RE (2009) **Multivariate data analysis**. Prentice Hall, Upper Saddle River, p. 761.

Kaiser HF (1960) The application of electronic computers to factor analysis. **Educional and Psychological Measurement** 20:141-151.

Kirkpatrick M and Meyer K (2004) Direct estimation of genetic principal components: simplified analysis of complex phenotypes. **Genetics** 168:2295-2306.

Koury Filho W, Albuquerque LG, Alencar MM, Forni S, Silva JAV, Lôbo RB (2009) Estimativas de herdabilidade e correlações para escores visuais, peso e altura ao sobreano em rebanhos da raça Nelore. **Revista Brasileira de Zootecnia** 38:2362-2367.

Koury Filho W, Albuquerque LG, Forni S, Silva JAV, Yokoo MJ, Alencar MM (2010) Genetic parameters estimates for visual scores and their association with body weight in beef cattle. **Revista Brasileira de Zootecnia** 39:1015-1022.

Magnabosco CU, Lopes FB, Miyagi ES, Lôbo RB, Sainz RD (2014) Multivariate approach of inter-relationships among growth, consumption and carcass traits in Nellore cattle. **Revista Ciência Agronômica** 45:168-176.

Meyer K (1992) Variance components due to direct and maternal effects for growth traits of Australian beef cattle. **Livestock Production Science** 31:179-204.

Meyer K and Kirkpatrick M (2005) Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. **Genetics Selection Evolution** 37:1-30.

Meyer K (2007) Multivariate analyses of carcass traits for Angus cattle fitting reduced rank and factor analytic models. **Journal of Animal Breeding and Genetics** 124:50-63.

R Development Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. R Core Team.

Rencher AC (2002) Methods of Multivariate Analysis. Wiley-Interscience, New York, NY, USA.

Regatieri IC, Boligon AA, Albuquerque LG (2011) Genetic analysis of visual scores and their relationships to mature female weight in Nellore breed. **Revista Brasileira de Zootecnia** 40:100-105.

Roso VM and Fries LA (1995) Componentes principais em bovinos da raça Polled Hereford à desmama e sobreano. **Revista Brasileira de Zootecnia** 24:728-735.

Savegnago RP, Caetano SL, Ramos SB, Nascimento GB, Schmidt GS, Ledur MC, Munari DP (2011) Estimates of genetic parameters, and cluster and principal components analysis of breeding values related to egg production traits in a White Leghorn population. **Poultry Science** 90:2174-2188.

Silva AM, Alencar MM, Regitano LCA, Oliveira MCS (2010) Infestação natural de fêmeas bovinas de corte por ectoparasitas na Região Sudeste do Brasil. **Revista Brasileira de Zootecnia** 39:1477-1482.

Smith C (1983) Effects of changes in economic weight on the efficiency of index selection. **Journal of Animal Science** 56:1057-1064.

Teixeira RA and Albuquerque LG (2003) Efeitos ambientais que afetam o ganho de peso pré-desmama em animais Angus, Hereford, Nelore e mestiços Angus-Nelore e Hereford-Nelore. **Revista Brasileira de Zootecnia** 32:887-890.

Vohra V, Niranjan SK, Mishra AK, Jamuna V, Chopra A, Sharma N, Jeong DK (2015) Phenotypic characterization and multivariate analysis to explain body conformation in Lesser Known Buffalo (Bubalus bubalis) from North India. **Asian Australasian Journal of Animal Science** 28:311-317.

Willham RL (1972) The role of maternal effects in animal breeding. 3. Biometrical aspects of maternal effects in animals. **Journal of Animal Science** 35:1288-1293.

CHAPTER 3 - GENOMIC REGIONS ASSOCIATED WITH PRINCIPAL COMPONENTS FOR GROWTH, VISUAL SCORE AND REPRODUCTIVE TRAITS IN NELLORE CATTLE[1]

ABSTRACT – The use of principal component (PC) analysis allows exploring the most relevant relationships using a reduced number of variables that explain the majority of the data variation. The search for candidate genes underlying the expression of PCs for different traits is a useful tool to better understand biological mechanisms associated with the traits of interest. The aim of this study was to identify genomic regions associated with PCs for growth (weaning and post-weaning weight gains), visual score (conformation, finishing precocity and muscling) and reproductive (indicated by scrotal circumference) traits in Nellore cattle by performing a genome-wide association study (GWAS) on the main PCs. Phenotypic and pedigree data from 355,524 animals and genotypes from 3,519 animals were used in this investigation. The estimated breeding values (EBV) were obtained from a multi-trait analysis using a mixed linear animal model. The eigen-decomposition of the additive genetic (co)variance matrix among traits ($\mathbf{A_T}$) was used to calculate the EBVs for the main PCs, which explained 87% of the variation in $\mathbf{A_T}$. The PCs were used as pseudo-phenotypes in the GWAS analyses. The SNP effects were estimated using the weighted single-step GBLUP and the BayesC method. The top-10 ranking windows that explained the highest proportion of variance were identified for further functional analyses. The most important genomic regions were identified on BTA7 and BTA24 for PC1, BTA8 for PC2, and BTA3 and BTA10 for PC3. The functional analyses contributed to unravel biological interpretation of PCs by identifying genes potentially associated with growth, carcass traits, conformation, and fatty acid composition traits. These findings are of relevance to the biological understanding of the PCs and their associated biotypes in Nellore cattle, potentially allowing for genetic selection for more specific breeding goals, such as animals with faster growth and precocious biotype.

Keywords: BayesC; eigen-decomposition, haplotype block window; QTL; weighted single step GBLUP

3.1 INTRODUCTION

The identification and selection of superior animals for breeding is accomplished through the recording of several economically important traits. The knowledge of the importance and contribution of these traits to the breeding goals, as well as their genetic relationships, is of relevance for livestock industry, as it will determine the adequate choice of selection criteria to guarantee the success of breeding programs. When working with several correlated traits, principal component analysis (PCA) is a multivariate technique that can be used for dimension reduction, keeping most of the original trait information. By performing PCA on (co)variance matrix of economically important traits, important biological interpretations could be revealed by evaluating the magnitude and direction of the coefficients in the eigenvectors of each principal components (PC) (Roso and Fries, 1995).

Vargas et al. (2018) used PCA to investigate the additive genetic (co)variance matrix (A_T) among growth, visual score and reproductive traits in Nellore cattle and showed that the first three components contrasted animals into different biotypes with discernible biological meaning. Moreover, PCA has also been used in GWAS for further investigate the genetic background of the few leading PCs. Macciotta et al. (2017) used PCA to describe the overall level and the slope of response of milk production traits across increasing levels of temperature-humidity index and were able to identify some putative candidate genes associated with the main PCs. For livestock breeding programs, these findings are of great relevance, as they suggest that GWAS could uncover genomic regions potentially associated with PCs from a set of targeted traits and improve the knowledge on their biological expression.

The objectives of the present study were: 1) to perform GWAS using estimated breeding values (EBVs) from PCs for growth, visual score, and yearling reproductive traits as pseudo-phenotypes aiming to identify genomic regions associated with main PCs; and 2) to detect potential candidate genes, and the biological mechanisms underlying these components.

3.2 MATERIALS AND METHODS

3.2.1 Phenotypic and pedigree data

Data for growth, visual score and yearling reproductive traits from 355,524 Nellore animals from Alliance Nellore database (www.gensys.com.br) were recorded between 1990 and 2010 and used for this study. The animals were raised under tropical pasture systems in 246 farms located in Brazil and in Paraguay. The number of observations, mean and standard deviation, coefficient of variation, number of sires, dam and contemporary groups for each trait are presented in Table 1. The following traits were analyzed: birth-to-weaning weight gain (BWG) and visual scores of conformation (WC), finishing precocity (WP) and muscling (WM) at weaning (about 205 days of age), weaning-to-yearling weight gain (WYG) and visual scores of conformation (YC), finishing precocity (YP) and muscling (YM), and scrotal circumference (SC) at yearling (about 550 days of age). Conformation, precocity, and muscling traits were evaluated by trained technicians, who assigned scores (from 1 to 5) on the basis of the variation within the contemporary groups. The higher the score, the more notable is the expression of the trait.

Table 1. Number of observations (N), mean and standard deviation (SD), coefficient of variation (CV), number of sires, dam and contemporary groups (CG) for birth-to-weaning weight gain (BWG), conformation, precocity, and muscling at weaning (WC, WP, and WM, respectively), weaning to yearling weight gain (WYG), conformation, precocity, and muscling at yearling (YC, YP, and YM, respectively) and scrotal circumference (SC) in Nellore cattle.

Traits	N	Mean±SD	CV (%)	Sires	Dams	CG
BWG (Kg)	334,123	147.20±29.86	20.28	3,773	178,295	5,127
WC (1 to 5)	334,123	3.07±1.07	34.85	3,773	178,295	5,127
WP (1 to 5)	334,123	3.17±1.11	35.01	3,773	178,295	5,127
WM (1 to 5)	334,123	3.03±1.13	37.29	3,773	178,295	5,127
WYG (Kg)	141,131	92.92±34.84	37.49	2,802	94,090	2,639
YC (1 to 5)	165,287	3.11±1.05	33.76	3,033	111,341	2,203

YP (1 to 5)	165,287	3.16±1.11	35.13	3,033	111,341	2,203
YM (1 to 5)	165,287	2.96±1.12	37.84	3,033	111,341	2,203
SC (cm)	58,912	26.42±3.55	13.43	2,298	47,668	1,133

The contemporary group (CG) for all traits included farm, year and season of birth, and sex (except for SC). The following effects were also added to CG: management group at weaning and weaning date for BWG, WC, WP, and WM; management group at yearling and yearling recording date for YC, YP, and YM; weaning and yearling management group, and weaning and yearling date for WYG and SC. Birth season was defined as 1= January to March, 2= April to June, 3= July to September, and 4= October to December. Records of the weight gains and SC exceeding 3.5 standard deviations (SD) from the overall mean of the CG were removed. For all traits, CG with fewer than 10 records were also removed from further analyses. The model included the covariate 'age of animal at recording' as linear effect for all traits and 'age of dam at calving' as linear and quadratic effects for the weaning traits.

3.2.2 Genotypic data and quality control

A total of 3,382 Nellore animals were genotyped using the Illumina® BovineHD chip (HD; ~777,000 SNPs; Illumina, Inc., San Diego, CA, USA) and 137 animals were genotyped using the GeneSeek Genomic Profiler Bovine HD chip (GGP-HD; ~76,000 SNPs; GeneSeek, Lincoln, NE, USA). The FImpute v2.2 software (Sargolzaei et al., 2014) was used for imputation of genotypes from the GGP-HD chip to the HD chip. The genotyping quality control (QC) filtered out markers located on non-autosomal regions that mapped to the same position, deviated from Hardy-Weinberg equilibrium (HWE) test ($P<10^{-5}$), with GenCall (GC) score lower than 0.15, SNP call rate lower than 0.95 and minor allele frequency (MAF) less than 0.02. All

samples presented call rate higher than 0.90 and were used in the GWAS. The remaining number of SNPs and samples after QC were 471,880 and 3,519, respectively.

3.2.3 Principal component analysis

A single multi-trait animal model, for the nine traits, was applied to estimate (co)variance components and EBVs using Wombat software (Meyer, 2007). For weaning traits, the statistical model included the fixed effects of CG and covariates (age of animal at recording and age of dam at calving), random direct additive genetic, maternal genetic, maternal permanent environmental and residual effects. In matrix notation the equation model for each weaning trait was:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_m\mathbf{m} + \mathbf{W}\mathbf{p} + \mathbf{e}$$

where: $\mathbf{y}$ is the vector of observations for each trait, $\boldsymbol{\beta}$ is a vector of systematic effects (CG and covariates), $\mathbf{a}$ is a vector of random direct additive genetic effects, $\mathbf{m}$ is a vector of random maternal genetic effects, $\mathbf{p}$ is a vector of random maternal permanent environmental effects, and $\mathbf{e}$ is a vector of random residual effects. $\mathbf{X}$, $\mathbf{Z}_a$, $\mathbf{Z}_m$, and $\mathbf{W}$ are the incidence matrices relating elements in $\boldsymbol{\beta}$, $\mathbf{a}$, $\mathbf{m}$, and $\mathbf{p}$ to $\mathbf{y}$, respectively. For yearling traits the model did not include maternal genetic and permanent environmental effects.

The PCs were calculated from the eigen-decomposition of the additive genetic (co)variance matrix among traits ($\mathbf{A_T}$) using the *eigen* function implemented in the R software (R Development Core Team, 2017). The Kaiser criterion (1960) was used to select the PCs that explained the largest proportion of the total genetic variation of the traits. This criterion takes into account only PCs with eigenvalue above the unit. The EBVs of the corresponding PCs (EBV_{PC_i}) previously selected according to Kaiser's criterion were obtained as: $EBV_{PC_{ij}} = e_{i1}*EBV_{j1} + e_{i2}*EBV_{j2} + \ldots + e_{i9}*EBV_{j9}$, where e_{i1} is the coefficient of the eigenvector of the ith PC for the first trait (BWG), and EBV_{j1} is the EBV for the jth animal for the first trait (BWG), e_{i2} is the coefficient of

the eigenvector of the ith PC for the second trait (WC), and EBV_{j2} is the EBV for the jth animal for the second trait (WC), and so on.

3.2.4 Genome-wide association analyses

The following statistical methods were used to estimate SNP effects: (i) the weighted single-step method (wssGBLUP) proposed by Wang et al. (2012) and (ii) the BayesC method (Habier et al., 2011). The wssGBLUP model can be described as:

$$\mathbf{y}^* = \mathbf{\mu} + \mathbf{Z_a a} + \mathbf{e}$$

where $\mathbf{y}^*$ is the vector of pseudo-phenotypes (EBVs of the PCs in this study); $\mathbf{\mu}$ is a vector of the overall mean; $\mathbf{Z_a}$ is an incidence matrix that relates animals to pseudo-phenotypes; $\mathbf{a}$ is the vector of direct additive genetic effects and e is the vector of random residuals. It was assumed that a~N(0,$\mathbf{H}\sigma_a^2$) and e~N(0,$\mathbf{R}\sigma_e^2$), where $\mathbf{H}$ is the relationship matrix based on genomic and pedigree information, σ_a^2 is the additive genetic variance, $\mathbf{R}$ is an diagonal matrix, whose elements account for the differences in the reliabilities of the observations in $\mathbf{y}$ due to differences in the amount of available information on offspring to estimate EBVs, and σ_e^2 is the residual variance. A proxy for the reliabilities of the pseudo-observations ($EBV_{PC_{ij}}$) was obtained as $SE(EBV_{PC_{ij}}) = \sqrt{e_{i1}^2 * SE_{j1}^2 + e_{i2}^2 * SE_{j2}^2 + \ldots + e_{i9}^2 * SE_{j9}^2}$, where $SE(EBV_{PC_{ij}})$ is the approximated SE of the $EBV_{PC_{ij}}$ for the ith PC of the jth animal, e_{i1}^2 is the square of the coefficient of the eigenvector for the ith PC for the first trait (BWG), and SE_{j1}^2 is the square of the standard error of the EBV of the jth animal for the first trait (BWG), e_{i2}^2 is the square of the coefficient of the eigenvector for the ith PC for the second trait (WC), and SE_{j2}^2 is the square of the standard error of the EBV of the jth animal for the second trait (WC), and so on. The inverse of $SE(EBV_{PC_{ij}})$ was used in the diagonal of the $\mathbf{R}$ matrix. This proxy for the reliability of the pseudo-observations assumed no estimation error covariance between EBVs and that the coefficients of the

eigenvector for each PC were constants. The inverse of **H** matrix can be defined as (Aguilar et al., 2010):

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

where **A** is the numerator relationship matrix based on pedigree for all animals; **A₂₂** is the numerator relationship matrix based on pedigree for genotyped animals only; and **G** is the genomic relationship matrix for genotyped animals.

The solutions of SNP effect estimates (**û**) were obtained as a function of the breeding values using the formula: $\hat{\mathbf{u}} = \mathbf{D}\mathbf{Z}'[\mathbf{Z}\mathbf{D}\mathbf{Z}']^{-1}\hat{\mathbf{a}}_{\mathbf{g}}$, where: **D** is a diagonal matrix with weights for SNPs, **Z** is an incidence matrix of genotypes for each locus, **Z'** is the transpose of **Z** matrix, and **âg** is the vector of predicted breeding values (EBVs of PCs) of genotyped animals. The **û** vector and the **D** matrix were iteratively recomputed over two iterations using the following algorithm (Wang et al., 2012):

1. In the first iteration t = 0 and $\mathbf{D}_t = \mathbf{I}$, where t is the iteration number, $\mathbf{D}_t$ is matrix **D** at iteration t, and **I** is an identity matrix.

2. Calculate the SNP effects at iteration $t(\hat{u}_{(t)})$.

3. Recalculate the diagonal elements of **D** as: $d^{*}_{i(t+1)} = \hat{u}^2_{i(t)} 2p_i(1-p_i)$ for all SNPs, where p_i is the allele frequency of the reference allele of the ith marker and i is the ith SNP.

4. Normalize $\mathbf{D}_{(t+1)} = (tr(\mathbf{D}_{(0)})/tr(\mathbf{D}^{*}_{(t+1)}))\mathbf{D}^{*}_{(t+1)}$.

5. t = t +1.

6. Exit, or loop to step 2.

The wssGBLUP analyses were performed using the BLUPF90 family programs (Misztal, 2017).

The BayesC method consisted of fitting a mixture model for SNP effects using the following model:

$$\mathbf{y}^{*} = \mathbf{1}\mu + \sum_{i=1}^{n}(\mathbf{g}_i\mathbf{b}_i\boldsymbol{\delta}_i) + \mathbf{e}$$

where $\mathbf{y}^{*}$, μ and **e** are the same as described above, **1** is a vector of ones, **gᵢ** is the vector containing the genotypes of the animals for the ith SNP, **bᵢ** is the allele substitution effect of the ith SNP, and **δᵢ** is an indicator variable (0, 1) sampled from a binomial distribution with parameters **n** and π, where **n** is the number of SNPs and π

is the fraction of SNPs not included in the model (equal to 1 if the ith SNP has a non-zero effect on the trait and 0 otherwise). Prior beta distribution with parameters $\alpha=10^8$ and $\beta=10^{10}$ were assumed for π so that, in practice, π was almost fixed to be 0.99 (Legarra et al., 2017). The BayesC was performed using the Markov Chain Monte Carlo (MCMC) algorithm implemented in the software GS3 (Legarra et al., 2017), running a single chain with 550,000 iterations, a burn-in period of 50,000 and a thin interval of 50. Compared to the Bayes Cπ, the BayesC method can achieve better results under specific situations, such as a trait with low heritability, low number of records and many QTLs affecting the trait of interest, leading to a not reliable estimation of π (Van den Berg et al., 2013).

3.2.5 Defining QTL regions

The results from the GWAS performed with wssGBLUP and BayesC were used to identify genomic windows associated with the main PCs. Quantitative trait loci regions of interest were defined based on haplotype block windows. Haploview v4.2 (Barrett et al., 2005) software, which uses by default the haplotype block definition proposed by Gabriel et al. (2002), was used to identify the haplotype blocks for the 29 autosomal chromosomes. The haplotype block definition was based on D' and its 95% confidence interval, and each SNP-pair comparison was called "strong LD", "inconclusive" or "strong recombination". A haplotype block was then defined as follows: let $C = \langle g_1, \ldots, g_n \rangle$ be a chromosome of n SNPs, $S = \langle s_i, \ldots, s_j \rangle$ a region of adjacent SNPs in C, l the number of strong LD SNP pairs in S, and r the number of strong evidence of historical recombination SNP pairs in S. Then, S is a haplotype block if (a) the two outermost SNPs, s_i and s_j, are in strong LD, and (b) there is at least a proportion d of informative pairs that are in strong LD, i.e., $l / (l + r) \geq d$.

A total of 75,615 haplotype blocks were obtained (Figure 1). The largest haplotype block consisted of 109 SNPs, while the smallest one contained two SNPs. The average size of the blocks ($\pm$ SD) was 4.82 SNPs ($\pm$ 0.30). The distribution of haplotype blocks varied depending on the chromosome length. The largest and

smallest groups of haplotype blocks were observed on BTA1 and BTA25, respectively. A total of 102,915 SNPs did not cluster in haplotype blocks, but were still kept in the analyses. Interestingly, some of these markers were in the top-10 ranking genomic regions that explained the highest proportion of variance estimated for the PCs. In order to verify if these markers were located at low confidence genomic regions, the findings were compared to the results obtained by Utsunomiya et al. (2016), who developed an approach to identify misassembled segments in the bovine reference genome assembly (UMD3.1.1).
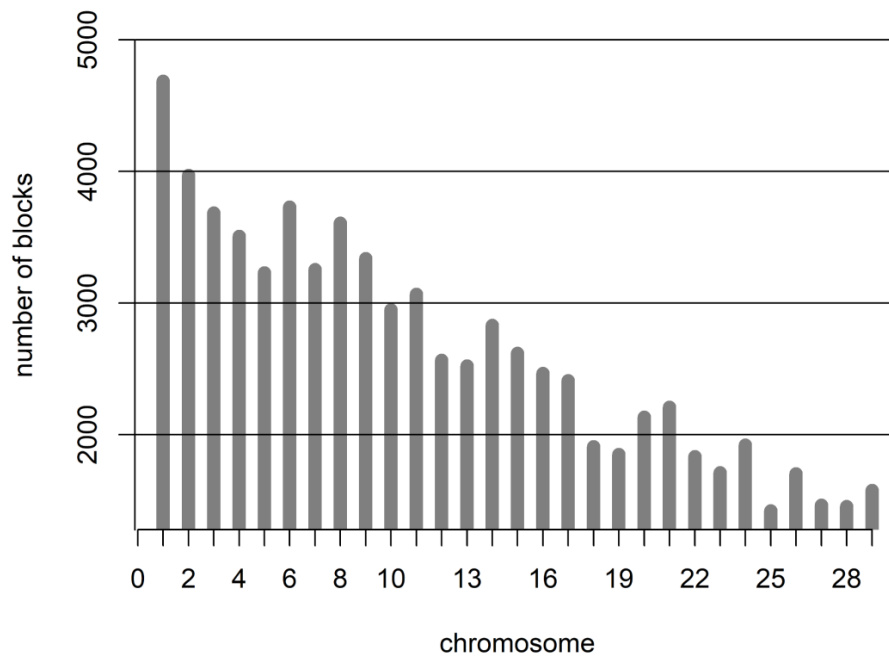
Figure 1. Number of haplotype windows per chromosome.

For the PCs, the top-10 ranking windows that explained the highest proportion of genetic variance were identified. These regions were used to identify QTLs based on the starting and ending coordinates of each window by consulting the Animal QTLdb database (Hu et al., 2013).

3.2.6 Functional annotation analyses

The NCBI Genome Data Viewer for the bovine genome was used for identification of the genes, using the UMD3.1.1 assembly as the reference map (https://www.ncbi.nlm.nih.gov/genome/gdv/?org=bos-taurus). Fasta sequences of the genes located within the top-10 windows were downloaded from the ENSEMBL Biomart Martview application (http://www.ensembl.org) and then uploaded into Blast2GO (Conesa et al., 2005). The BLAST results were then mapped to GO terms to obtain the GO annotation. All unique sequences were aligned against the reference sequences in the NCBI database using the BLASTp algorithm from Blast2GO software (Götz et al., 2008). Metabolic pathways associated with significant sequences were identified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000; Kanehisa et al., 2016).

3.3 RESULTS

3.3.1 Principal component analysis

According to Kaiser's criterion (1960), the first three PCs were selected from multi-trait analysis. These three main PCs explained 87% of the total additive genetic variance of the traits included in this study (Table 2). From the nine original dimensions, PC1 showed negative and moderate coefficients for all traits. The common negative signal for all coefficients suggests that these traits present the same direction of variation in this PC. The PC2 had contrasting animals with early (better for finishing precocity, muscling and scrotal circumference) and late (better for weight gain and conformation) maturity biotype both at weaning and yearling. The PC3 differentiated weaning and yearling traits, contrasting animals according to their maturation rate (weaning versus yearling performance).

Table 2. Eigenvectors, eigenvalues (λ) and proportion (Prop) of the explained variance over the first three principal components (PC1 to PC3) for growth, visual score, and yearling reproductive traits in Nellore cattle.

PCs	BWG	WC	WP	WM	WYG	YC	YP	YM	SC	λ	Prop (%)
PC1	-0.36	-0.29	-0.39	-0.39	-0.22	-0.30	-0.37	-0.38	-0.24	4.96	55.11
PC2	0.39	0.51	-0.26	-0.22	0.01	0.52	-0.32	-0.30	-0.10	1.87	20.78
PC3	-0.13	-0.19	-0.26	-0.33	0.84	0.15	0.12	0.06	0.13	1.01	11.22

BWG: birth to weaning gain; WC: weaning conformation; WP: weaning precocity; WM: weaning muscling; WYG: weaning to yearling gain; YC: yearling conformation; YP: yearling precocity; YM: yearling muscling; SC: scrotal circumference.

3.3.2 Genome-wide association studies

The proportions of variance explained by the top-10 windows associated with the PCs are listed in Table 3. Some of the genomic regions obtained for the main PCs using wssGBLUP analyses were the same as obtained using BayesC analyses (highlighted in bold). For PC1, the common genomic region is located on BTA21 (59,058,057 bp). For PC2, these genomic regions are located on BTA8 (59,138,856 bp to 59,143,141 bp and 87,504,583 bp to 87,505,342 bp) and on BTA17 (1,180,289 bp). For PC3, common genomic regions were identified on BTA10 (99,643,238 bp and 99,650,200 bp) and BTA20 (11,603,778 bp). The highest proportion of variance observed using wssGBLUP analyses were found on BTA7, BTA8 and BTA10 for PC1, PC2 and PC3, respectively, and on BTA24, BTA8 and BTA3 using BayesC analyses. The sum of the proportion of variances explained by the top-10 windows for these three PCs were 3.59%, 5.19% and 4.61%, respectively, using wssGBLUP analyses, and 5.81%, 11.11%, 6.39% using BayesC analyses.

Table 3. Top-10 windows explaining the highest proportion of genetic variance for the principal components (PC1 to PC3) obtained by wssGBLUP and BayesC analyses.

PCs	wssGBLUP			BayesC		
	Chr	Region (bp)	Pvar	Chr	Region (bp)	Pvar
PC1	7	99,833,568*	0.41	24	29,182,953-29,191,540	1.92
	21	**59,058,057***	0.40	9	10,562,531-10,569,900	0.81
	2	44,257,626*	0.39	15	15,087,965*	0.66
	24	29,204,164-29,209,985	0.38	26	3,803,966-3,821,201	0.48
	26	9,009,427*	0.35	10	67,991,764-68,000,131	0.48
	2	19,678,063-19,687,209	0.35	21	63,891,592-63,930,408	0.33
	12	14,292,205*	0.34	3	119,764,973-119,775,433	0.33
	18	33,880,245*	0.34	2	19,275,577-19,283,611	0.28
	4	36,147,273-36,155,792	0.32	19	43,023,638-43,042,919	0.27
	4	96,646,598-96,653,736	0.31	**21**	**59,058,057***	0.25
PC2	**8**	**59,138,856-59,143,141**	0.89	**8**	**59,138,856-59,143,141**	5.86
	17	**1,180,289***	0.88	17	73,599,986*	0.92
	12	23,593,713-23,594,586	0.66	11	74,478,406-74,487,412	0.72
	3	58,098,646-58,110,679	0.60	20	39,648,515-39,665,731	0.59
	11	75,813,087*	0.43	**8**	**87,504,583-87,505,342**	0.58
	8	40,990,076-41,005,284	0.39	18	20,489,314-20,505,923	0.57
	11	36,967,411-36,973,700	0.36	14	46,985,993-47,022,598	0.53
	8	**87,504,583-87,505,342**	0.34	**17**	**1,180,289***	0.46
	18	16,266,983-16,267,862	0.33	8	102,445,910*	0.45
	19	35,672,589-35,676,953	0.31	2	61,284,557-61,294,243	0.42
PC3	**10**	**99,643,238-99,650,200**	0.89	3	14,369,634-14,405,644	1.89
	20	**11,603,778***	0.57	**20**	**11,603,778***	1.05
	18	3,844,284*	0.43	16	60,336,209-60,356,570	0.53
	24	36,333,389*	0.43	11	2,048,759-2,057,847	0.50
	2	108,529,237-108,546,737	0.43	16	16,079,646-16,083,267	0.47
	3	114,937,690-	0.40	14	19,933,581-19,943,530	0.41

	114,944,278				
2	53,059,581-53,060,713	0.39	24	59,846,358[*]	0.41
2	131,580,148-131,605,519	0.37	**10**	**99,643,238-99,650,200**	0.39
19	52,665,363-52,691,522	0.36	18	27,366,008-27,412,578	0.39
8	4,830,615[*]	0.34	8	85,649,054[*]	0.35

Chr: chromosome; Region (bp): starting and ending coordinates of the haplotype block; Pvar: % genetic variance explained by the SNPs within the window; [*]Single SNPs that did not cluster in haplotype blocks. The common genomic windows obtained for the main PCs using wssGBLUP and BayesC analyses are highlighted in bold.

For the three PCs, several genomic regions were previously reported by different authors as containing QTLs for growth (Kim et al., 2003; Rolf et al., 2012), carcass (McClure et al., 2010; Baeza et al., 2011), conformation (Pryce et al., 2011; Berkowicz et al., 2012), and fatty acid composition (Peters et al., 2012; Cesar et al., 2014) traits for beef and dairy cattle (Table 4).

Table 4. Previously published QTLs regions for growth, carcass, conformation and fatty acid composition traits for beef and dairy cattle which harbor the top-10 windows (haplotype blocks) explaining the highest proportion of genetic variance for the principal components (PC1 to PC3) obtained by wssGBLUP and BayesC analyses.

PCs	wssGBLUP			BayesC		
PC1	QTL Trait	Chr	Region (bp)	QTL Trait	Chr	Region (bp)
	Body weight (birth)	2	18,250,280-49,471,250	Body weight (birth)	2	18,250,280-49,471,250
	Body weight (weaning)	2	43,430,199-49,295,365	Stature	2	17,953,475-22,876,474
	Stature	2	17,953,475-22,876,474	Body weight (weaning and yearling)	9	8,888,721-18,194,890
	Longissimus muscle area	4	28,268,762-38,039,020	Body weight (mature)	10	67,861,579-72,420,258
	Marbling score	4	28,268,762-38,039,020	Carcass weight	10	65,419,585-72,420,258
	Fat thickness at the 12th rib	14	10,961,016-14,785,148	Body weight (birth)	15	14,694,158-18,600,230
	Body weight (weaning)	18	22,926,481-36,500,992	Body weight (weaning)	15	10,762,815-18,600,230
	Body weight (mature)	18	33,011,652-36,500,992	Intramuscular fat	19	37,554,862-46,978,329
	Fat acid content	18	28,994,493-	Body weight (weaning)	24	27,303,121-

			46,084,260			29,303,627
	Body weight (weaning)	24	27,303,121-29,303,627	Carcass weight	24	26,571,082-30,004,563
PC2	Average daily gain	3	55,119,878-68,212,933	Body weight (slaughter)	2	48,134,573-62,205,631
	Body weight (birth)	3	56,210,430-58,174,550	Carcass weight	2	59,275,414-65,118,834
	Body weight (birth)	8	29,543,811-42,343,887	Body weight (birth)	8	48,094,361-64,725,844
	Body weight (weaning)	8	58,434,111-59,768,227	Body weight (weaning)	8	58,434,111-59,768,227
	Fatty acid content	11	75,039,655-75,976,999	Body weight (weaning and mature)	11	70,068,881-80,930,488
	Marbling score	11	62,118,964-82,382,031	Marbling score	11	62,118,964-82,382,031
	Body weight (yearling), Height	12	21,356,707-36,550,507	Carcass weight	14	46,763,361-47,157,787
	Body weight (weaning)	19	21,579,122-37,460,465	Body weight (weaning)	17	70,975,160-74,702,234
	Marbling score	19	31,254,823-37,460,465	Marbling score	20	29,876,187-45,415,528

PC3	Body weight (birth)	2	49,295,365-85,095,065	Body weight (slaughter), Carcass weight	3	14,298,989-14,500,187
	Marbling score	2	49,295,365-62,205,631	Fat thickness at the 12th rib	3	14,300,994-23,367,147
	Fat thickness at the 12th rib	8	4,473,073-17,294,027	Body weight	8	85,304,450-86,538,180
	Body weight (birth)	19	49,937,105-59,464,484	Average daily gain, Body weight	14	19,715,680-25,062,335
	Marbling score	19	49,942,658-59,447,271	Body weight (weaning)	16	12,204,152-19,401,272
	Body weight (yearling and mature)	20	7,696,273-15,713,179	Average daily gain	18	27,034,490-29,073,969
	Body weight (weaning)	24	30,004,563-41,590,595	Body weight (yearling and mature)	20	7,696,273-15,713,179

Chr = chromosome; Region (bp) = starting and ending coordinates for the QTL.

Table 5 presents the annotated genes within the top-10 windows associated with the first three PCs using wssGBLUP and BayesC analyses. For PC1, ten annotated genes were found within nine genomic windows identified in this study. The PC2 showed the lowest number of annotated genes, i.e. six, which were within five genomic windows. For PC3, a total of eleven genes were identified within nine genomic windows.

Table 5. Annotated genes within the top 10 windows associated with the first three principal components (PC1 to PC3) obtained by wssGBLUP and BayesC analyses.

PCs	wssGBLUP			BayesC		
	Chr	Region (bp)	Genes	Chr	Region (bp)	Genes
PC1	2	44,257,626	*CACNB4*	3	119,764,973-119,775,433	*NDUFA10*
	2	19,678,063-19,687,209	*NFE2L2, LOC107133476*	10	67,991,764-68,000,131	*FBXO34*
	4	96,646,598-96,653,736	*PLXNA4*	19	43,023,638-43,042,919	*STAT5A*
	24	29,204,164-29,209,985	*CDH2*	24	29,182,953-29,191,540	*CDH2*
	26	9,009,427	*SGMS1*	26	3,803,966-3,821,201	*LOC104976760*
PC2	8	87,504,583-87,505,342	*ROR2*	8	87,504,583-87,505,342	*ROR2*
	11	75,813,087	*LOC101907994*	11	74,478,406-74,487,412	*PTRHD1, CENPO*
				14	46,985,993-47,022,598	*NOV*
				17	73,599,986	*SPECC1L*
PC3	2	131,580,148-131,605,519	*HSPG2, LDLRAD2*	3	14,369,634-14,405,644	*MEF2D, LOC100848553*
	8	4,830,615	*GALNTL6*	8	85,649,054	*IPPK*
	19	52,665,363-	*NPTX1*	11	2,048,759-	*KCNIP3*

Chr	Region (bp)	Genes	Chr	Region (bp)	Genes
	52,691,522			2,057,847	
24	36,333,389	*LOC786055*	16	16,079,646-16,083,267	*BRINP3*
			16	60,336,209-60,356,570	*LOC104974482*

Chr: chromosome; Region (bp): starting and ending coordinates; Genes: ENSEMBL symbol of annotated genes using the *Bos taurus* UMD3.1 assembly.

Gene Ontology enrichment analyses were performed for all candidate genes to investigate whether the loci associated with the three PCs corresponded to genes involved in known pathways. For PC1, most of the biological processes were involved in cellular (17.9%), metabolic (15.4%) and cellular component organization or biogenesis (10.3%) processes. For PC2, the genes were mainly enriched into cellular processes (28.6%), whereas for PC3, the biological processes were involved in cellular (28.6%), multicellular organismal (21.4%), metabolic (14.3%) and developmental processes (14.3%). For PC3, annotation using KEGG allowed the identification of genes involved in the "mucin type O-glycan biosynthesis" pathway.

3.4 DISCUSSION

In this study, the large number (75,615) of haplotypes blocks identified with an average size of about 5 SNPs confirms the feasibility of the Bovine HD chip to capture LD in the Nellore population. This finding suggests that the use of haplotype blocks to define the window size in association studies could be a better alternative to identify genetic variants and biological mechanisms underlying the traits of interest than using fixed length haplotypes.

Overlapping windows were identified by the top-10 ranking genomic regions obtained using wssGBLUP and BayesC analyses. These windows are located on BTA21 for PC1, BTA8 and BTA17 for PC2 and on BTA10 and BTA20 for PC3. The identification of common genomic regions strengthens the potential of the candidate genes identified as possible associated with the PCs of the traits of interest. The

proportions of genetic variance explained by the top-10 ranking genomic regions indicate the polygenic nature of the PCs, i.e., a large number of genomic regions explained a relatively small proportion of the total additive genetic variance of each PC.

For those windows harboring genomic regions previously described in the literature, it was observed a trend regarding the QTLs for each PC. For PC1, in general, the QTLs previously identified were associated with growth traits, such as body weight at birth, weaning, yearling and slaughter (Kim et al., 2003), fatty acid composition traits (Peters et al., 2012; Cesar et al., 2014) and carcass traits (McClure et al., 2010; Baeza et al., 2011). For PC2, most of the QTLs were associated with body weight at birth and weaning and carcass traits, including marbling score and subcutaneous fat (Kim et al., 2003; McClure et al., 2010; Baeza et al., 2011). For PC3, the QTLs previously identified were associated with the same traits described for PC1, but also for conformation traits, such as height at yearling and maturity, and rump width (McClure et al., 2010). The existence of previously reported QTL covering areas surrounding the windows detected in this study provides more support for the detected associations.

For the first PC, two genomic windows located on BTA24 (29,204,164 bp - 29,209,985 bp and 29,182,953 bp - 29,191,540 bp) have a common gene, named cadherin 2 gene (*CDH2*). Cadherins are single chain transmembrane glycoproteins that are essential for tissue development, regulation of cell proliferation, differentiation, and survival (Larue et al., 1996; Gumbiner, 2005). The *CDH2* has been described as regulator of postnatal skeletal growth and bone mass maintenance in mice, developing several functions in the osteogenic lineage (Benedetto et al., 2010). The genetic ablation of this protein results in skeletal growth defects and impaired bone formation in humans and mice (Marie et al., 2014). The region found on BTA 2 (19,678,063 bp -19,687,209 bp) harbors the *NFE2L2* gene also frequently referred as *NRF2*, which develop a critical role in adipocyte differentiation (Seo and Lee, 2013). Adipocyte cells are connective-tissue cells specialized in synthesis and contain large globules of fat and, therefore, are essential for maintaining the animal's energy balance (Gregoire et al., 1998). Previous studies have reported that the *NRF2* is involved in regulating whole body weight, obesity and

hepatic lipid homeostasis in mice (Shin et al., 2009; Huang et al., 2010; Yu et al., 2011).

Another gene found for PC1 was the *STAT5A* located on BTA19 (43,023,638 bp – 43,042,919 bp). The *STAT5A* is a member of the universal transcription factors family (*STAT*) that plays important roles in the cellular response to several hormones and cytokines, and, therefore, is essential to cell proliferation and growth (Lin and Leonard, 2000; Rochman et al., 2009). Interestingly, the *STAT5* gene is the main mediator of growth hormone (GH) action on target tissues, such as hepatocytes, where its role is related to modulating cellular metabolism and the production of IGF-1, an important regulator of postnatal body growth (Hennighausen and Robinson, 2008). Previous studies have demonstrated that disruption of both *STAT5A* and *STAT5B* resulted in a more pronounced reduction of body growth in mice and, therefore, both genes are necessary in the skeletal muscle for normal growth (Teglund et al., 1998; Klover and Hennighausen, 2007). Selvaggi et al. (2015), studying the *STAT5A* polymorphism and its influence on growth traits in cattle, observed that the most frequent genotype in the studied population showed an initial faster growth, resulting in higher body weight at 90 and 270 days of age, whereas the less frequent genotype exhibit a faster growth in the post-weaning period, which determined higher body weight at 450 days of age.

The general functions described above for *CDH2*, *NFE2L2* and *STAT5A* corroborate the biological interpretation for PC1, which showed a moderate coefficient for all the nine original traits, mainly discriminating high or low growth rate animals. Growth is related to an increase mass (weight) per unit time and involves changes in form and composition resulting from differential growth of the different parts of the body. In beef cattle, growth is evaluated in major tissues of the carcass, i.e., muscle, fat and bone, combined with adequate proportions of these three major tissues. Thus, it was expected that for PC1 the genomic regions obtained would be associated with growth and body composition traits.

Regarding the PC2, the present study identified a genomic region on BTA20 (39,648,515 bp - 39,665,731 bp) that had been previously associated with fatty acid composition traits in beef cattle (Saatchi et al., 2013). Some of the genes found for this PC have been mainly associated with bone and muscle formation and skeletal

development (*ROR2* and *NOV*). The common genomic region identified among methods located on BTA8 (87,504,583 bp to 87,505,342 bp) harbors the receptor tyrosine kinase *ROR2*. In mammals, *ROR2* plays major roles in skeletal development and mutations within this gene are responsible for short stature, limb bone shortening, and segmental defects of the spine (Afzal et al., 2000; van Bokhoven et al., 2000). The *ROR2* is expressed in human's osteoblasts, a group of cells responsible for the formation of new bone in the skeleton, being essential to promote differentiation at early and late stages of osteoblastogenesis (Liu et al., 2007).

The *NOV* (nephroblastoma overexpressed) gene identified on BTA14 (46,985,993 bp - 47,022,598 bp) belongs to the CNN family of proteins that are known for having growth-regulatory functions. This gene is expressed by osteoblasts in a variety of tissues, including bone, cartilage and muscle, and its transcription is regulated by transforming growth factor (TGF)-beta and bone morphogenetic proteins (BMP) (Canalis, 2007). A previous study has demonstrated that mutation of *NOV* gene results in abnormal skeletal development and muscle atrophy in mice (Heath et al., 2008). These findings contribute to better understand the biological interpretation of this second PC, which contrasts animals with early maturity biotype (animals with higher capacity to achieve a minimum degree of carcass finishing at moderate live weight and with adequate muscle mass development), and late maturity biotype (tall and thin animals with less fat deposition and higher live weight) both at weaning and yearling ages.

The functions described for the candidate genes found for PC2 stand out by their associations with general tissue development. In livestock, during typical growth development, the animal's body goes through a process of mass increase and body shape changes, which differs depending on the priorities of tissues development. The post-natal bone growth occurs early in animal's life and remains constant during development, while the muscle tissue has its greatest development after birth and is the main component of the weight gain (Owens et al., 1995). At this moment, the genes related to muscle and skeletal formation, such as the annotated genes described for this PC, are essential for a normal growth development. Finally, the accumulation of adipose tissue occurs when the muscle development decrease. Thus, biologically, larger animals require more time to complete skeletal development

and, therefore, will present delayed muscle mass development and carcass finishing. For beef cattle producers, a desirable biotype is represented by an animal with precocious growth (weight gain) and finishing, resulting in a more profitable system.

For PC3, a specific window on BTA3 (14,369,634 bp – 14,405,644 bp) was reported by Connor et al. (2017) in a genome-wide copy number variant analysis for production traits, including residual feed intake and dry matter intake in Holstein cattle. In another GWAS using Brahman cattle, Martínez et al. (2017) also detected suggestive SNPs on BTA16 (16,079,646 bp to 16,083,267 bp) associated with birth weight. The window found on BTA16 (60,336,209 bp – 60,356,570 bp) harbors a gene (*LOC104974482*) that was recently associated with significant copy number variation regions (CNVRs) for fatty acids in intramuscular fat of *longissimus thoracis* muscle of Nellore cattle (Lemos et al., 2018).

The genomic window located on BTA2 (131,580,148 bp - 131,605,519 bp) harbors a member of the low-density lipoprotein receptor family, the *LDLRAD2* gene, which has its ubiquitous expression in fat tissues (Hussain et al., 1999). The *LDLR* is the major cholesterol-carrying lipoprotein of plasma, developing important roles in the regulation of cholesterol homeostasis in mammalian cells (Defesche, 2004). Members of this family have been previously associated with regulating body weight and glucose metabolism in mice (Liu et al., 2012). The genomic window located on BTA24 (36,333,389 bp) contains a variant form of the second most abundant protein in the cell, the *LOC786055* also known as *EEF1A1* (Lee et al., 1992). Its major expression in specific tissues, such as skeletal muscle, occurs early in animal's life and declines during development, such that in adult skeletal muscle, there is an almost complete loss of *EEF1A1* expression (Khalyfa et al., 2001). The *MEF2D* gene found on BTA3 (14,369,634 bp – 14,405,644 bp) belongs to the MEF2 family of transcription factors (Estrella et al., 2015). Members of this family are involved in control of muscle and skeletal cell development through their effects on cell differentiation (Pon and Marra, 2016).

The overlap of the identified genomic regions for production traits with previously reported QTL in the literature, as well as the general functions identified for the candidate genes, supports the biological interpretation for PC3, which discriminates weaning versus yearling traits, contrasting animals according to their

maturation rate (weaning versus yearling performance). This difference is possibly associated with the genetic potential of the animals to growth at different ages and environments: a) a pre-weaning environmental, where there is maternal influence (mainly due to the cow's milk production and maternal ability), and b) a post-weaning environment, where the animals are raised without the direct influence of maternal effects, and usually under a challenging condition due the dry season which affects the quality of pasture.

The GO enrichment analysis performed in this study for the three main PCs, revealed that different annotated genes are enriched in a large number of GO terms. For PC1, the GO terms identified are mainly involved in biological processes directly associated with the general organism development. For instance, the GO term "regulation of multicellular organism growth" (GO:0040014) is the process responsible to modulate the frequency, rate or extent of growth of the body. Another notable biological process for PC1 was GO:0006629 term "lipid metabolic process", which is defined by any chemical reactions and pathways involving lipid in general, such as fatty acids, glycolipids and sterols (http://www.ensembl.org).

In agreement with the general functions described for the annotated genes found for PC2, two GO terms - "skeletal system development" (GO:0001501) and "limb development" (GO:0060173) - were identified in this study. The GO term "BMP signaling pathway" (GO:0030509) is represented by part of the transforming growth factor-β (TGF-β) superfamily of proteins, the bone morphogenetic proteins (BMPs), which plays roles in bone and cartilage formation (Wang et al., 2014).

Biological processes enriched for the gene list obtained for PC3 were associated with the development of muscle, skeletal and vital organs. The GO term "osteoblast differentiation" (GO:0001649) describes the process responsible for the formation of bone. For this component, annotation using KEEG allowed the identification of the "mucin type O-glycan biosynthesis" pathway. Previous studies reported that mutations in genes directly involved in this biological pathway are associated with growth retardation phenotypes and increased bone density in mouse (Ichikawa et al., 2009; Duncan et al., 2011). In general, the identified biological pathways allowed for specific interpretation across the three main PCs and are in agreement with the biological functions previous described for each component.

3.5 CONCLUSIONS

The results from the genome-wide association studies showed that the genetic merit of the animals for each of the three main principal components for growth, visual score and reproductive traits in Nellore cattle is affected by many loci with small effects. Important chromosomal genomic regions associated with the main principal components were also identified. Some of these genomic regions overlap with previously reported regions associated with growth, carcass traits, conformation, and fatty acid composition traits in Nellore and other cattle breeds. Novel candidate regions for principal components were detected and some of them have suggestive important biological functions. The functional processes, pathways, and regulatory mechanisms identified in this study contributed to a better biological interpretation of the principal components and its associated biotypes, potentially allowing for genetic selection for more specific breeding goals, such as animals with faster growth and precocious biotype.

3.6 REFERENCES

Afzal AR, Rajab A, et al. (2000) Recessive Robinow syndrome, allelic to dominant brachydactyly type B, is caused by mutation of ROR2. **Nature Genetics** 25:419-422.

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science** 93:743-752.

Baeza MC, Corva PM, et al. (2011) Genetic markers of body composition and carcass quality in grazing Brangus steers. **Genetics and Molecular Research** 10:3146-56.

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

Benedetto AD, Watkins M, Grimston S, Salazar V, Donsante C, Mbalaviele G, Radice GL, Civitelli R (2010) N-cadherin and cadherin 11 modulate postnatal bone growth and osteoblast differentiation by distinct mechanisms. **Journal of Cell Science** 123:2640-2648.

Berkowicz EW, Magee DA, Berry DP, Sikora KM, Howard DJ, Mullen MP, Evans RD, Spillane C, MacHugh DE (2012) Single nucleotide polymorphisms in the imprinted bovine insulin-like growth factor 2 receptor gene (IGF2R) are associated with body size traits in Irish Holstein-Friesian cattle. **Animal Genetics** 43:81-87.

Boligon AA, Vicente IS, Vaz RZ, Campos GS, Souza FRP, Carvalheiro R, Albuquerque LG (2016) Principal component analysis of breeding values for growth and reproductive traits and genetic association with adult size in beef cattle. **Journal of Animal Science** 94:5014-5022.

Buzanskas ME, Savegnago RP, Grossi DA, Venturini GC, Queiroz SA, Silva LOC, Torres Júnior RAA, Munari DP, Alencar MM (2013) Genetic parameter estimates and principal component analysis of breeding values of reproduction and growth traits in female Canchim cattle. **Reproduction, Fertility and Development** 25:775-781.

Canalis E (2007) Nephroblastoma overexpressed (Nov) is a novel bone morphogenetic protein antagonist. **Annals of the New York Academy Sciences** 1116:50-58.

Cesar AS, Regitano LC, et al. (2014) Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. **BMC Genetics** 15:39.

Conesa A1, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics** 21:3674-3676.

Connor EE, Zhou Y, Wiggans GR, Lu Y, Tempelman RJ, Schroeder SG, Chen H, Liu G (2017) Genome-wide copy number variant analysis in Holstein cattle reveals variants associated with 10 production traits including residual feed intake and dry matter intake. In: ADSA ANNUAL MEETING. Pittsburgh, Pennsylvania, United States.

Defesche JC (2004) Low-density lipoprotein receptor--its structure, function, and mutations. **Seminars in Vascular Medicine** 4:5-11.

Duncan EL, Danoy P, et al. (2011) Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. **Plos Genetics** 7:e1001372.

Estrella NL, Desjardins CA, Nocco SE, Clark AL, Maksimenko Y, Naya FJ (2015) MEF2 Transcription factors regulate distinct gene programs in mammalian skeletal muscle differentiation. **Journal of Biological Chemistry** 290:1256-1268.

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

Götz S1, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. **Nucleic Acids Research** 36:3420-35.

Gregoire FM, Smas CM, Sul HS (1998) Understanding adipocyte differentiation. **Physiology Reviews** 78(3).

Gumbiner BM (2005) Regulation of cadherin-mediated adhesion in morphogenesis. **Nature Reviews Molecular Cell Biology** 6:622-634.

Habier D, Fernando R, Kizilkaya K, Garrick DJ (2011) Extension of the Bayesian alphabet for genomic 179 selection. **BMC Bioinformatics** 186:186-197.

Heath E, Tahri D, Andermarcher E, Schofield P, Fleming S, Boulter CA (2008) Abnormal skeletal and cardiac development, cardiomyopathy, muscle atrophy and cataracts in mice with a targeted disruption of the Nov (Ccn3) gene. **BMC Development Biology** 8:18.

Hennighausen L and Robinson GW (2008) Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. **Genes and Development** 15:711-21.

Hu Z-L, Park CA, Wu X-L, Reecy JM (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research** 41:871-879.

Huang J, Tabbi-Anneni I, Gunda V, Wang L (2010) Transcription factor Nrf2 regulates SHP and lipogenic gene expression in hepatic lipid metabolism. **American Journal of Physiology** 299:G1211-G1221.

Hussain MM, Strickland DK, Bakillah A (1999) The mammalian low-density lipoprotein receptor family. **Annual Reviews of Nutrition** 19:141-72.

Ichikawa S, Sorenson AH, Austin AM, Mackenzie DS, Fritz TA, Moh A, Hui SL, Econs MJ (2009) Ablation of the Galnt3 gene leads to low-circulating intact fibroblast growth factor 23 (Fgf23) concentrations and hyperphosphatemia despite increased Fgf23 expression. **Endocrinology** 150:2543-2550.

Kaiser HF (1960) The application of electronic computers to factor analysis. **Educional and Psychological Measurement** 20:141-151.

Kanehisa M and Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. **Nucleic Acids Research** 28:27-30.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2016) KEGG: new perspectives on genomes, pathways, diseases and drugs. **Nucleic Acids Research** 45:D353-D361.

Khalyfa A, Bourbeau D, Chen E, Petroulakis E, Pan J, Xu S, Wang E (2001) Characterization of elongation factor-1A (eEF1A-1) and eEF1A-2/S1 protein expression in normal and wasted mice. **Journal of Biology Chemistry** 276:22915-22922.

Kim JJ, Farnir F, Savell J, Taylor JF (2003) Detection of quantitative trait loci for growth and beef carcass fatness traits in a cross between Bos taurus (Angus) and Bos indicus (Brahman) cattle. **Journal of Animal Science** 81:1933-1942.

Klover P and Hennighausen L (2007) Postnatal body growth is dependent on the transcription factors signal transducers and activators of transcription 5a/b in muscle: a role for autocrine/paracrine insulin-like growth factor I. **Endocrinology** 148:1489-97.Larue L, Antos C, Butz S, Huber O, Delmas V, Dominis M, Kemler R (1996) A role for cadherins in tissue formation. **Development** 122:3185-3194.

Lee S, Francoeur AM, Liu S, Wang E (1992) Tissue-specific expression in mammalian brain, heart, and muscle of S1, a member of the elongation factor-1 alpha gene family. **Journal of Biological Chemistry** 267:24064-24068.

Legarra A, Ricard A, Filangi O. GS3. User Manual. France; 2014. Available from: <http://snp.toulouse.inra.fr/~alegarra/manualgs3_last.pdf>. Accessed 20 Dec. 2017.

Lemos MVA, Peripolli E, et al. (2018) Association study between copy number variation and beef fatty acid profile of Nellore cattle. **Journal of Applied Genetics** 59:203-223.

Lin JX and Leonard WJ (2000) The role of Stat5a and Stat5b in signaling by IL-2 family cytokines. **Oncogene** 19:2566-2576.

Liu W, Singh R, Choi CS, Lee HY, Keramati AR, Samuel VT, Lifton RP, Shulman GI, Mani A (2012) Low density lipoprotein (LDL) receptor-related protein 6 (LRP6) regulates body fat and glucose homeostasis by modulating nutrient sensing pathways and mitochondrial energy expenditure. **Journal of Biological Chemistry** 287:7213-23.

Liu Y, Bhat RA, Seestaller-Wehr LM, Fukayama S, Mangine A, Moran RA, Komm BS, Bodine PVN, Billiard J (2007) The orphan receptor tyrosine kinase Ror2 promotes osteoblast differentiation and enhances ex vivo bone formation. **Molecular Endocrinology** 21:376-387.

Macciotta NPP, Biffani S, Bernabucci U, Lacetera N, Vitali A, Ajmone-Marsan P, Nardone A (2017) Derivation and genome-wide association study of a principal component-based measure of heat tolerance in dairy cattle. **Journal of Dairy Science** 100:4683-4697.

Marie PJ, Haÿ E, Modrowski D, Revollo L, Mbalaviele G, Civitelli R (2014) Cadherin-Mediated Cell-Cell Adhesion and Signaling in the Skeleton. **Calcified Tissue International** 94:46-54.

Martínez R, Bejarano D, Gómez Y, Dasoneville R, Jiménez A, Even G, Sölkner J, Mészáros G (2017) Genome-wide association study for birth, weaning and yearling weight in Colombian Brahman cattle. **Genetics and Molecular Biology** 40:453-459.

McClure MC, Morsci NS, et al. (2010) A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. **Anim Genetics** 41:597-607.

Meyer K (2007) Multivariate analyses of carcass traits for Angus cattle fitting reduced rank and factor analytic models. **Journal of Animal Breeding and Genetics** 124:50-64.

Misztal I. BLUPF90—a flexible mixed model program in Fortran 90. 2012. Available: <http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90.pdf>. Accessed 17 Dec 2017.

Owens FN, Gill DR, Secrist DS, Coleman SW (1995) Review of some aspects of growth and development of feedlot cattle. **Journal of Animal Science** 73:3152.

Peters SO, Kizilkaya K, Garrick DJ, Fernando RL, Reecy JM, Weaber RL, Silver GA, Thomas MG (2012) Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. **Journal of Animal Science** 90:3398-3409.

Pon JR and Marra MA (2016) MEF2 transcription factors: developmental regulators and emerging cancer genes. **Oncotarget** 7:2297-2312.

Pryce JE, Hayes BJ, Bolormaa S, Goddard ME (2011) Polymorphic regions affecting human height also control stature in cattle. **Genetics** 187:981-984.

R Development Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. R Core Team.

Rochman Y, Spolski R, Leonard WJ (2009) New insights into the regulation of T cells by gamma(c) family cytokines. **Nature Reviews Immunology** 9:480-490.

Rolf MM, Taylor JF, Schnabel RD, McKay SD, McClure MC, Northcutt SL, Kerley, Weaber RL (2012) Genome-wide association analysis for feed efficiency in Angus cattle. **Animal Genetics** 43:367-374.

Roso VM and Fries LA (1995) Componentes principais em bovinos da raça Polled Hereford à desmama e sobreano. **Revista Brasileira de Zootecnia** 24:728-735.

Saatchi M, Garrick DJ, Tait Jr RG, Mayes MS, Drewnoski M, Schoonmaker J, Diaz C, Beitz DC, Reecy J (2013) Genome-wide association and prediction of direct genomic breeding values for composition of fatty acids in Angus beef cattle. **BMC Genomics** 14:730.

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

Savegnago RP, Caetano SL, Ramos SB, Nascimento GB, Schmidt GS, Ledur, M.C., Munari DP (2011) Estimates of genetic parameters, and cluster and principal components analysis of breeding values related to egg production traits in a White Leghorn population. **Poultry Science** 90:2174-2188.

Selvaggi M, D'Alessandro AG, Dario C (2015) Bovine STAT5A gene polymorphism and its influence on growth traits in Podolica breed. **Animal Production Science** 56(7).

Seo HA and Lee IK (2013) The Role of Nrf2: Adipocyte Differentiation, Obesity, and Insulin Resistance. **Oxidative Medicine Cellular Longevity** 2013:184598.

Shin S, Wakabayashi J, Yates MS, Wakabayashi N, Dolan PM, Aja S, Liby KT, Sporn MB, Yamamoto M, Kensler TW (2009) Role of Nrf2 in prevention of high-fat diet-induced obesity by synthetic triterpenoid CDDO-imidazolide. **European Journal of Pharmacology** 620:138-44.

Teglund S, McKay C, Schuetz E, van Deursen JM, Stravopodis D, Wang D, Brown M, Bodner S, Grosveld G, Ihle JN (1998) Stat5a and Stat5b proteins have essential and nonessential. **Cell** 93:841-50.

Utsunomiya ATH, Santos DJA, et al. (2016) Revealing misassembled segments in the bovine reference genome by high resolution linkage disequilibrium scan. **BMC Genomics** 17:705.

Van Bokhoven H, Celli J, et al. (2000) Mutation of the gene encoding the ROR2 tyrosine kinase causes autosomal recessive Robinow syndrome. **Nature Genetics** 25:423-426.

Van Den Berg I, Fritz S, Boichard D (2013) QTL fine mapping with Bayes C (π): a simulation study. **Genetics Selection Evolution** 45:19.

Vargas G, Schenkel FS, Brito LF, Neves HHR, Munari DP, Boligon A, Carvalheiro R (2018) Unravelling biological biotypes for growth, visual score and reproductive traits in Nellore cattle via principal component analysis. **Livestock Science** 207:37-43.

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

Wang RN, Green J, et al. (2014) Bone Morphogenetic Protein (BMP) signaling in development and human diseases. **Genes & Diseases** 1:87-105.

Yu Z, Shao W, Chiang Y, Foltz W, Zhang Z, Ling W, Fantus IG, Jin T (2011) Oltipraz upregulates the nuclear factor (erythroid-derived 2)-like 2 [corrected](NRF2) antioxidant system and prevents insulin resistance and obesity induced by a high-fat diet in C57BL/6J mice. **Diabetologia** 54:922-34.

CHAPTER 4 - PREDICTION OF RESPONSE TO SELECTION AND GENOMIC ACCURACY OF PRINCIPAL COMPONENTS FOR GROWTH, VISUAL SCORE AND REPRODUCTIVE TRAITS IN NELLORE CATTLE

ABSTRACT – Selection index is an efficient tool to combine various traits into a single value based on their biological and economic importance. Alternatively, an index derived from various traits can be calculated based on Principal Component Analysis (PCA) by using the eigenvalue of the Principal Component (PC) and the eigenvector of traits in each PC. This study was carried out to: 1) compare the genetic progress that would be achieved from selection based on PCs derived from nine growth, visual score and reproductive traits or using a current selection index (Final Index, FI) implemented in some Nellore cattle commercial breeding programs and a Harmonic Index (a selection index based on PCA); and, 2) investigate the performance of genomic predictions (accuracy and inflation) of these nine traits, the first three PCs, and the selection index. For this, phenotypes from 355,524 animals and genotypes from 3,519 animals (containing 471,880 Single Nucleotide Polymorphisms, SNPs) were used. The estimated breeding values (EBVs) were obtained from a multi-trait analysis using a mixed linear animal model. The eigen-decomposition of the additive genetic (co)variance matrix ($\mathbf{A_T}$) was used to calculate the EBVs for the first three PCs. These EBVs were used as pseudo-phenotypes in the subsequent genomic analyses. Genomic predictions using the EBVs of all nine original traits and EBVs of selection indexes were also performed for comparison. Prediction accuracy was measured as the Pearson's correlation between pseudo-phenotypes and Genomic Estimated Breeding Values (GEBVs). The Harmonic Index (HI) yielded higher genetic gains for the studied traits when compared to PCs and FI and, therefore, could be used as alternative for increasing expected response to selection in Nellore cattle. Our findings indicated that genomic selection for growth, visual score and reproductive traits in Nellore cattle is feasible since moderate to high genomic predictions were obtained.

Keywords: eigen-decomposition, selection index, single step GBLUP

4.1 INTRODUCTION

Commercial breeding programs have focused on selecting animals for economically important traits that contribute to the production efficiency and long-term sustainability of the beef production systems (Moreira et al., 2015; Lopes et al., 2016; Fernandes et al., 2018). Over the years, selection has been applied to several traits simultaneously, as selection based on a single trait is not sufficient to improve the overall genetic merit of a population. When important traits differ in terms of variability, heritability, economic importance, and phenotype and genetic correlations, Selection Index has been proved to be an effective breeding approach for animal selection (Hazel, 1943; Falconer and Mackay, 1996).

The theory of selection indexes is based on the fact that each individual has an overall genetic merit value that corresponds to the sum of breeding values associated with economically important traits weighted according to their relative economic importance (Smith, 1936; Hazel, 1943). Therefore, traits that have larger impact on profit or production goals consequently have greater economic weights associated with them. The strategy for properly designing selection indexes are: well-estimated phenotypic and genetic (co)variance matrices and proper definition of the relative economic values of each trait part of the overall breeding goals. However, assessing the relative economic value for each trait can be challenging, which can considerably affect the selection index efficiency (Smith, 1983).

The use of an alternative index that takes into account the correlation between the combined variables and the breeding value of the animal and the desired response to selection of each trait was proposed as an option to overcome the unavailability of accurate economic values for all important traits (Pesek and Baker, 1969; Rouvier, 1969). Based on this approach, traits are weighted on the basis of the importance as determined by the researcher or breeder, in order to obtain a desired genetic gain without directly considering their relative economic importance (Pesek and Baker, 1969; Rouvier, 1969; Yamada et al., 1975). Alternatively, the index based on principal component analysis (PCA) has been proposed as an advantageous approach (Buzanskas et al., 2013; Boligon et al., 2016) that uses a well-known

methodology to construct linear combinations between traits by using the eigenvalue of the principal component (PC) and the eigenvector of traits in each PC. The index based on PCA does not take into account the economic importance of each trait, however, the PCs can show particular biological interpretations which could be used to select animals according to desirable breeding objectives.

These potential biological associations between traits were previously demonstrated when PCA was applied on the (co)variance matrix among traits or the (co)variance matrix of the estimated breeding values (EBVs) obtained for each trait in beef cattle studies (Buzanskas et al., 2013; Boligon et al., 2016; Vargas et al., 2018). For instance, Roso and Fries (1995) demonstrated the possibility of selecting Hereford cattle for distinct size or body volumes based on the first PC.

Alternatively to traditional methods of selection that use information on phenotypes and pedigrees to predict breeding values, genomic selection allows selecting candidates based on prediction of genomic estimated breeding values (GEBV) obtained from their estimated marker effects. The existence of linkage disequilibrium (LD) between single nucleotide polymorphisms (SNPs) and quantitative trait loci (QTL) is an essential assumption when using markers to accurately predict the genetic merit of animals (Meuwissen et al., 2001). However, additional sources of genetic information also have influence on prediction accuracy of genomic selection, such as co-segregation and relationships that are implicitly captured by SNP genotypes (Habier et al., 2007; 2013). In order to evaluate the application of genomic selection based on PCs, it is needed to assess the accuracy of GEBVs, which can determine the successful of implementing genomic selection for such traits or indexes and, therefore, optimize genetic response in breeding programs.

The design of selection indexes based on PCs has been suggested to be a useful alternative for improving animal performance (Agudelo-Gómez et al., 2016; Boligon et al., 2016). However, to our knowledge, there are no studies investigating the efficiency of this approach compared to pre-existing selection indexes. Thus, the objectives of this study were: 1) to compare the genetic progress achieved by selecting breeding candidates based on PCs or on two current selection indexes used by some Nellore cattle commercial breeding programs; and 2) investigate the

performance of genomic predictions (accuracy and inflation) of these nine traits, the first three PCs, and the selection indexes.

4.2 MATERIALS AND METHODS

4.2.1 Phenotypic and contemporary groups

Phenotypic records for growth, visual score and yearling reproductive (scrotal circumference) traits from 355,524 Nellore animals from the Alliance Nellore dataset (www.gensys.com.br), collected between 1990 and 2010, were used for this study. The animals were raised under tropical pasture systems in 246 farms located in Brazil and in Paraguay. The evaluated traits included birth-to-weaning weight gain (BWG) and visual scores of conformation (WC), finishing precocity (WP) and muscling (WM) at weaning (about 205 days of age), weaning-to-yearling weight gain (WYG) and visual scores of conformation (YC), finishing precocity (YP) and muscling (YM), and scrotal circumference (SC) at yearling (about 550 days of age) as an indicator of reproductive performance. Conformation, finishing precocity, and muscling traits were based on recorded visual scores assigned by well-trained technicians in a discrete ordered scale (from 1 to 5) on the basis of the variation within the contemporary groups. The higher the score, the more notable was the expression of the trait. Heritability estimates for the weaning and yearling traits range from 0.17 (BWG and WC) to 0.21 (WP) and from 0.18 (WYG) to 0.46 (SC), respectively (Vargas et al., 2018).

Contemporary groups (CGs) for all traits included farm, year and season of birth, and sex. The following effects were also added to the CGs: management group at weaning and weaning date for BWG, WC, WP, and WM; management group at yearling and yearling date for YC, YP, and YM; weaning and yearling management group, and weaning and yearling date for WYG and SC. Birth season was defined as 1 = January to March, 2 = April to June, 3 = July to September, and 4 = October to

December. Records of the weight gains and SC deviating 3.5 standard deviations (SD) from the overall mean of the CG were removed. For all traits, CG with less than 10 records were also removed from further analyses. Furthermore, the models included the covariate 'age of animal at recording' as linear effect for all traits and 'age of dam at calving' as linear and quadratic effects for the weaning traits.

4.2.2 Genotypic data and quality control

Genotypic data were available for 3,519 Nellore animals. Genotypes were generated with the Illumina® BovineHD SNP chip (HD; ~777,000 SNPs; Illumina, Inc., San Diego, CA, USA) and the GeneSeek Genomic Profiler Bovine HD SNP chip (GGP-HD; ~76,000 SNPs; GeneSeek, Lincoln, NE, USA). The FImpute v2.2 software (Sargolzaei et al., 2014) was used for imputation of genotypes from the GGP-HD chip to the HD chip. Samples with call rate lower than 90% were removed from the analysis. Non-autosomal and unmapped SNP were discarded. Markers deviating from Hardy-Weinberg equilibrium test (P-value$<10^{-5}$), with GenCall (GC) score lower than 0.15, SNP call rate lower than 0.95 and minor allele frequency (MAF) less than 0.02 were removed. The remaining number of SNPs and samples after QC were 471,880 and 3,519, respectively.

4.2.3 Principal component analysis

Multi-trait analysis was performed using Wombat software (Meyer, 2007) to estimate (co)variance components and EBVs. For weaning traits, the animal model included the fixed effect of CG and covariates (age of animal at recording and age of dam at calving), random direct additive genetic, maternal genetic, maternal permanent environmental and residual effects. The statistical model for each weaning trait was as follows:

$$y = X\beta + Z_a a + Z_m m + Wp + e$$

where: y is the vector of observations for each trait, β is a vector of systematic effects (CG and covariates), a is a vector of random direct additive genetic effects, m is a vector of random maternal genetic effects, p is a vector of random maternal permanent environmental effects, and e is a vector of random residual effects. X, Z_a, Z_m, and W are the incidence matrices relating elements in β, a, m, and p to y, respectively. Maternal genetic and permanent environmental effects were not included in the models for yearling traits.

The PCs were calculated from the eigen-decomposition of the additive genetic (co)variance matrix among traits (A_T) using the *eigen* function implemented in the R software (R Development Core Team, 2018) as performed by Vargas et al. (2018). The Kaiser criterion (Kaiser, 1960) was used to select the PCs that explained the largest proportion of the total genetic variation of the traits. This criterion takes into account only PCs with eigenvalue above the unit. The PCs were used for calculating indexes to evaluate animals for the studied traits.

The EBVs of the corresponding PCs (EBV_{PC_i}) previously selected according to Kaiser's criterion were obtained as:

$$EBV_{PC_{ij}} = ev_{i1} * EBV_{j1} + ev_{i2} * EBV_{j2} + \ldots + ev_{i9} * EBV_{j9} \qquad (1)$$

where ev_{i1} is the coefficient of the eigenvector of the ith PC for the first trait (BWG), and EBV_{j1} is the EBV for the jth animal for the first trait (BWG), ev_{i2} is the coefficient of the eigenvector of the ith PC for the second trait (WC), and EBV_{j2} is the EBV for the jth animal for the second trait (WC), and so on.

4.2.4 Predictions of the genetic gain

In practical terms, the equation defined in (1) can be compared to the aggregate genotype (H) of an individual, in which the economic weight for each trait is represented by the eigenvector's coefficient obtained from PCA, as described above, and the genetic value (a_i) is represented by the EBVs. Thus, the vector of

response to selection for each trait in H is directly obtained from the following equation:

$$\Delta G = [\Delta G_1 + \Delta G_2 + \ldots + \Delta G_q] = i\frac{(b'G)}{\sqrt{b'P'b}}$$

where **i** is the selection intensity, **b** is the weighting factors column vector, **G** is the genetic (co)variance matrix and **P** is the phenotypic (co)variance matrix. The ΔG represents the amount of increase in performance that is expected to be achieved per generation of selection. In this study, estimates of ΔG were standardized in terms of genetic standard deviation of the corresponding trait.

Response to selection for a Harmonic Index (HI), an index based on PC1 and PC2, were also calculated:

$$\Delta G_{HI} = \sqrt{1'GE_{ij}D_{ij}E'_{ij}G'}$$

where **1** is a (9x1) summing vector of ones, **E** is a matrix containing the eigenvectors as columns and **D** is a diagonal matrix with the ith and j^{th} eigenvalues.

In order to compare the genetic progress expected to be achieved with selecting breeding candidates based on PCs, estimates of genetic gains for a similar set of traits that compose a selection index (Final Index) currently used by some Nellore cattle breeding programs were also calculated (Table 1).

Table 1. Weight factors (in %) for each trait used to compose the Final Index (FI) currently used for selection in Nellore cattle.

Index	Trait[2]								
	BWG	WC	WP	WM	WYG	YC	YP	YM	SC
FI	23	4	8	8	23	4	8	8	14

[1]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference

4.2.5 Genomic predictions

From the multi-trait analysis, three alternative response variables for genomic predictions were used, EBVs of the original traits (EBV_T), EBVs of the PCs (EBV_{PC}), and EBVs of the selection indexes for Nellore cattle (EBV_{IND}). Genomic predictions were performed using single-step GBLUP method (ssGBLUP), under the following model:

$$\mathbf{y}^* = \mathbf{\mu} + \mathbf{W_a}\mathbf{a} + \mathbf{e}$$

where $\mathbf{y}^*$ is the vector of pseudo-phenotypes (EBV_T, EBV_{PC} or EBV_{IND}); $\mathbf{\mu}$ is a vector of the overall mean; $\mathbf{W_a}$ is an incidence matrix that relates animals to pseudo-phenotypes; $\mathbf{a}$ is the vector of direct additive genetic effects and $\mathbf{e}$ is the vector of random residuals. It was assumed that $a \sim N(0, \mathbf{H}\sigma_a^2)$ and $e \sim N(0, \mathbf{R}\sigma_e^2)$, where $\mathbf{H}$ is the relationship matrix based on genomic and pedigree information, σ_a^2 is the additive genetic variance, $\mathbf{R}$ is an diagonal matrix, whose elements account for the differences in the reliabilities of the observations in $\mathbf{y}$ due to differences in the amount of available information on offspring to estimate EBVs, and σ_e^2 is the residual variance. A proxy for the reliabilities of the pseudo-observations ($EBV_{PC_{ij}}$ or EBV_{IND}) was obtained as $SE(EBV_{PC_{ij}}) = \sqrt{ev_{i1}^2 * SE_{j1}^2 + ev_{i2}^2 * SE_{j2}^2 + \ldots + ev_{i9}^2 * SE_{j9}^2}$, where $SE(EBV_{PC_{ij}})$ is the approximated SE of the $EBV_{PC_{ij}}$ for the ith PC of the ith animal, ev_{i1}^2 is the square of the coefficient of the eigenvector for the ith PC for the first trait (BWG), and SE_{J1}^2 is the square of the standard error of the EBV of the jth animal for the first trait (BWG). Analyses were carried out using the BLUPF90 family programs (Misztal et al., 2016). The inverse of $SE(EBV_{PC_{ij}})$ was used in the diagonal of the $\mathbf{R}$ matrix. This proxy for the reliability of the pseudo-observations assumed no estimation error covariance between EBVs and that the coefficients of the eigenvector for each PC were constant.

4.2.6 Validation scheme

A forward prediction validation scheme was implemented to assess the feasibility of using genomic predictions for the PCs in Nellore cattle. The datasets were split into training and validation populations. The training population consisted of animals born between 1993 and 2008, whereas the validation set consisted of animals born in 2009 and 2010. This validation scheme ensured that information of own performance and/or progeny records of the validation animals did not contribute to the estimation of the training population EBVs, thus avoiding overlapping between training and validation sets (Amer and Banos, 2010). Prediction accuracy was measured as the Pearson's correlation between pseudo-phenotypes and GEBVs. In order to measure the degree of inflation/deflation of GEBVs, the regression coefficient (b1) of EBV_T ($b1_{EBV_T,GEBV}$), EBV_{PC} ($b1_{EBV_{PC},GEBV}$) and EBV_{FI} ($b1_{EBV_{FI},GEBV}$) was evaluated.

4.3 RESULTS

Estimates of genetic parameters based on the full phenotypic datasets as well as the results obtained from PCA were reported by Vargas et al. (2018). According to the authors, the first three PCs (PC1 to PC3) attained the Kaiser criterion (Kaiser, 1960) and explained 87.11% of the total additive genetic variance for the traits. The first component (PC1) was characterized by similar coefficients (eigenvectors) for all traits, indicating that all of them are similarly important for the animal's performance. The second component (PC2) distinguishes animals with contrasting performance for weight gain and conformation in relation to finishing precocity, muscling and scrotal circumference (both at weaning and yearling), mainly contrasting precocious versus late finishing animals. The third component (PC3) contrasted between weaning and yearling traits (Table 2).

Table 2. Eigenvectors of economically important traits in Nellore cattle, reduced to the first three principal components (PC1 to PC3).

Trait[1]	PC1	PC2	PC3
BWG	0.36	-0.39	-0.13
WC	0.29	-0.51	-0.19
WP	0.39	0.26	-0.26
WM	0.39	0.22	-0.33
WYG	0.22	-0.01	0.84
YC	0.30	-0.52	0.15
YP	0.37	0.32	0.12
YM	0.38	0.30	0.06
SC	0.24	0.10	0.13

[1]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference

It was observed in the two-dimensional graph of PC1 and PC2 (Figure 1) and the linear correlation between the original variables with PCs (Table 3) that the PC1 was more closely associated with WP, WM, YC, and YP, with linear correlations ranging from 0.83 to 0.87. Considering the PC2, BWG, WC and YC presented greater discriminatory power when compared to the other traits, and they were negatively correlated with this component. For PC3, the largest linear correlation was obtained for WYG, with a correlation of 0.84.
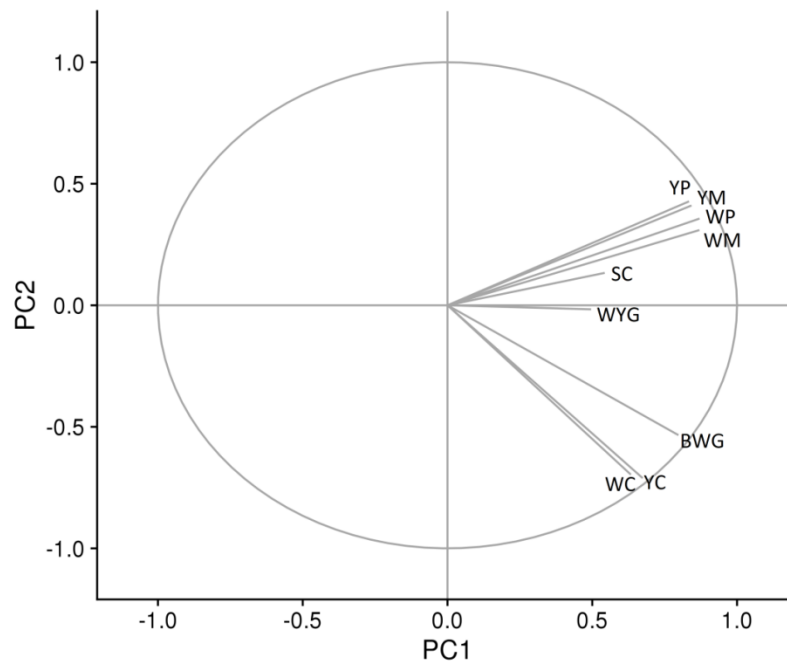
Figure 1. Two-dimensional principal component's plot of principal components 1 (PC1) and 2 (PC2).

Table 3. Correlation coefficients between original variables and the first three principal components (PC1 to PC3)

Traits[1]	PC1	PC2	PC3
BWG	0.79	-0.53	-0.14
WC	0.63	-0.69	-0.20
WP	0.87	0.36	-0.26
WM	0.87	0.31	-0.32
WYG	0.49	-0.02	0.84
YC	0.83	-0.71	0.14
YP	0.84	0.43	0.12
YM	0.67	0.41	0.06
SC	0.54	0.13	0.14

[1]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference

Table 4 summarizes the structure of training and validation datasets as well as the average accuracy of EBV$_{PC}$ for each PC. The size of the training and validation

sets was the same for PCs, selection indexes and individual traits. Accuracy of EBV_{PC} for the training set ranged from 0.43 (PC3) to 0.67 (PC2) and for the validation set ranged from 0.41 (PC3) to 0.66 (PC2). For WI and FI, accuracies of EBV_{IND} for the training set were equal to 0.53 and 0.68, and for the validation set were equal to 0.54 and 0.71, respectively. For the nine original traits, average accuracies of EBVs were equal to 0.60, 0.75, 0.57, 0.58, 0.43, 0.58, 0.57, 0.57 and 0.45 for BWG, WC, WP, WM, WYG, YC, YP, YM and SC, respectively.

Table 4. Structure of training and validation datasets for the first three PCs (PC1 to PC3).

PC[1]	Training population				Validation population		
	Birth year	Phenotype	Phenotype and genotype	Acc[2]	Birth year	N[3]	Acc[2]
PC1	1990 - 2008	317,954	2,500	0.63	2009 - 2010	238	0.62
PC2	1990 - 2008	317,954	2,500	0.67	2009 - 2010	238	0.66
PC3	1990 - 2008	317,954	2,500	0.43	2009 - 2010	238	0.41

[1]PCn: n[th] principal component
[2]Acc: proxy of the accuracy for EBV_{PC} in the training and validation population
[3]N: sample size

Genetic gains obtained per generation for the traits studied under different selection indexes are presented in Table 5. The genetic gains were assessed under a standardized scenario (i=1.76, which means that about 10% of individuals are kept for breeding after each selection round). For PC1, expected genetic gains ranged from 0.68 (WYG) to 1.21 (YP). For this PC, genetic gains obtained for weaning and yearling traits were slightly similar to those obtained for IAlliance. The PC2 showed negative gains for BWG (-0.65), WC (-0.82), WYG (-0.03), and YC (-0.85) and positive (favorable) genetic gains for the other traits. For PC3, expected genetic gains obtained for weaning traits were negative and ranged from -0.28 (WM) to -0.10 (BWG), while positive genetic gains were obtained for yearling traits [range: 0.03 (YM) to 0.81 (WYG)]. Positive genetic gains were obtained for all traits when using HI (range: 1.69 to 2.20). For this index the highest genetic gains were obtained for finishing precocity and muscling both at weaning and yearling. Genetic gains obtained for HI were higher when compared to the PC and FI indexes.

Table 5. Genetic gains obtained by generation for growth, visual score and scrotal circumference traits based on principal components (PCs) and current selection indexes considering 10% selection intensity (i=1.76).

PC[1]	Traits[2]								
	BWG	WC	WP	WM	WYG	YC	YP	YM	SC
PC1	1.09	0.87	1.21	1.20	0.68	0.99	1.12	1.14	0.73
PC2	-0.65	-0.82	0.36	0.31	-0.03	-0.85	0.52	0.51	0.14
PC3	-0.10	-0.17	-0.23	-0.28	0.81	0.09	0.08	0.03	0.13
HI	1.91	1.71	2.20	2.16	1.88	1.82	2.15	2.11	1.69
Selection Index[3]	Traits[2]								
	BWG	WC	WP	WM	WYG	YC	YP	YM	SC
FI	1.07	1.04	1.10	0.98	1.12	1.39	1.43	1.45	1.43

[1]PCn: n[th] principal component, HI: Harmonic Index
[2]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference
[3]FI: Final Index

The genomic prediction accuracies for the PCs ranged from 0.48 (PC2) to 0.62 (PC1), and for FI, accuracies was equal to 0.64 (Table 6). For the nine original traits, the predictive accuracies ranged from 0.49 (WC) to 0.70 (SC). The regression coefficients of EBV$_T$ on GEBV ranged from 0.67 (WYG) to 1.17 (WM) and were, in general, higher than those estimated for PC1 (0.74), PC2 (0.58), PC3 (0.51) and FI (0.83).

Table 6. Prediction accuracies measured by Pearson's correlation between estimated breeding values and genomic breeding values [r(EBV,GEBV)] and estimates of regression coefficients (b1) for the nine traits.

PC[1]	r(EBV$_{PC}$,GEBV)	b1
PC1	0.62	0.74
PC2	0.48	0.58
PC3	0.60	0.51
Selection Index[2]	r(EBV$_{FI}$,GEBV)	b1
FI	0.64	0.83
Trait[3]	r(EBV$_T$,GEBV)	b1
BWG	0.66	1.28
WC	0.49	1.01
WP	0.69	1.05
WM	0.67	1.17
WYG	0.67	0.67
YC	0.63	1.12
YP	0.68	1.02
YM	0.68	1.00
SC	0.70	0.80

[1]PCn: nth principal component
[2]FI: Final Index
[3]BWG: birth-to-weaning weight gain, WC: conformation at weaning, WP: precocity at weaning, WM: muscling at weaning, WYG: weaning-to-yearling weight gain, YC: conformation at yearling, YP: precocity at yearling, YM: muscling at yearling, SC: scrotal circumference

4.4 DISCUSSION

The PCA allowed visualizing genetic associations among the studied traits by summarizing the number of variables, using only the first three PCs instead of the nine traits, with a minimum loss of information. By definition, the correlation between PCs is zero, which means that the variation explained in PC1 is independent of that explained in PC2 and so on, implying that selection of individuals based on any PC

should not result in correlated response with other PC. Based on the results, PCA allows selecting animals according to general growth and distinguished maturity rates. In this sense, when the PCs with favorable biological meaning are in agreement to desired breeding objectives, they would represent selection indexes, which could be used directly in selection processes.

Overall, the PC1 showed higher and balanced genetic progress for the studied traits when compared to PC2 and PC3. Thus, the selection to increase general growth in Nellore cattle should result in positive response when using PC1 as criteria for selection. This result could be explained by the favorable genetic correlations between the nine original traits (range: 0.15 to 0.86) and the moderate coefficients (eigenvectors) obtained for this PC. Similar genetic gains were obtained for the nine traits when using PC1 and FI, indicating that comparable response to selection could be achieved when using either one. Despite FI assigns higher weight factors for growth traits, such as BWG and WYG, when compared to visual scores and SC, the genetic correlations between these traits have influence on their genetic gain estimates.

From the eigenvectors obtained for PC2 it is possible to select animals for early biotype (better for finishing precocity and muscling) at weaning and yearling, i.e. animals that are able to achieve a minimum degree of carcass finishing at a relatively low weight and suitable muscle mass development in a shorter time and, thus, are prepared for slaughtering earlier. The expected genetic gains estimated for this PC are in accordance with this finding, in which negative values were obtained for BWG (-0.65), WC (-0.82), WYG (-0.03), and YC (-0.85) and positive (favorable) genetic gains were obtained for finishing precocity and muscling at different ages, and scrotal circumference. The genetic gain obtained for WYG (-0.03) was probably due to the lower magnitude of the eigenvector obtained for this trait (0.01).

For PC3, negative genetic gains were obtained for weaning traits (range: -0.28 to -0.10) and positive (favorable) genetic gains were obtained for yearling traits (range: 0.03 to 0.81), indicating that this PC index could be used for selecting animals according to weaning and yearling performance. This finding was expected because PC3 is composed of negative eigenvectors for weaning traits and positive eigenvectors for yearling traits, differentiating the animals according to both ages.

This PC was responsible for presenting the lowest genetic gains for conformation, finishing precocity, and muscling at yearling when compared to the other PCs and indexes, which is possibly associated to the lower magnitude of the eigenvectors obtained for these traits.

In this study, PC1 and PC2 were combined in order to create a Harmonic Index (HI) that could be used to select animals more profitable for particular production systems, focusing on select animals with adequate growth rate and precocious biotype. Including PC1 along with PC2 had a meaningful impact on the direction and amount of genetic gains. For this new index, genetic gains of higher magnitude were obtained when compared to the other indexes, indicating that superior response to selection would be expected when using HI as criteria for selection. For this index, no difference in terms of direction were observed between traits as showed by PC2, which contrasted weigh gain and conformation, and finishing precocity, muscling and scrotal circumference both at weaning and yearling. However, higher response to selection would be expected for finishing precocity and muscling when selecting animals based on this index.

Previous studies have suggested using PCA as selection indexes attempting select superior individuals according to specific breeding objectives. For these purposes, the eigenvectors of each trait would represent the weights, i.e. the importance of changes in the genetic levels of traits and thus creating linear combinations of original traits that will lead to greater genetic progress in the desired direction. In this study, substantial increase in genetic gains would be achieved for the traits using HI when compared to FI, indicating that PCA approach could be implemented for selecting individuals.

In this study, we compared the accuracies of GEBVs across nine important economic traits, three PCs, and a selection index indicated for Nellore cattle by using their respective EBVs as response variables. Based on the forward-prediction validation, the accuracies predicted for the original studied traits ranged from 0.49 (WC) to 0.70 (SC) and were on average similar to the estimates obtained for PC1 (0.62), PC2 (0.48), PC3 (0.60) and IAlliance (0.64). The perceived disadvantage associated with EBV_{PC} accuracies when used as response variables can possibly be associated to the low reliabilities used as weighting factors in genomic analysis.

Once the EBV$_{PC}$ was calculated taking into account linear combinations of EBVs of each trait and their corresponding eigenvectors, the approximated standard errors represent an approximation of the reliabilities of EBV$_{PC}$ and, thus, was implemented as an alternative strategy of weighting in genomic analysis.

The validation accuracies obtained for all nine traits were on average higher than those previously reported by Neves et al. (2014), when using a smaller training set. The authors reported accuracy estimates ranging from 0.38 (BWG) to 0.68 (YP). Several factors can lead to these differences, such as the size of the training population, accuracy of the pseudo-phenotypes, level of LD, and the degree of relationship between training and validation populations (Bolormaa et al., 2013; Boddhireddy et al., 2014).

Regression coefficients close to 1 indicates that genomic predictions are on the same scale as the pseudo-phenotypes, i.e. neither inflated nor deflated. Regression coefficients lower than 1 indicate that GEBVs are overestimated and, thus, present higher variability than expected, while regression coefficients higher than 1 indicate that GEBVs present lower variability than expected (Wiggans et al., 2011). Unbiased predictions allows accurately ranking the animals for a fair comparison across generations (Patry and Ducrocq, 2009; Aguilar et al., 2010). In general, the regression coefficients for the nine original traits were close to 1, indicating that predictions were not biased. The regression coefficients for FI were equal to 0.83 and the PCs regression coefficients ranged from 0.51 to 0.74. Differences in the scale of genomic predictions were also observed by Neves et al. (2014). According to the authors, predictions of genomic breeding values obtained for growth, visual score and reproductive traits using GBLUP method tended to be slightly inflated.

4.5 CONCLUSIONS

The first principal component and the Final Index (selection index for Nellore cattle) yielded to similar genetic gains for the nine original traits and, therefore, there

is no advantage in using principal component approach for increasing expected response to selection in Nellore cattle. However, the Harmonic Index (a selection index based on combining principal components) would lead to higher genetic gain for the studied traits when compared to the other indexes and, therefore, could be used as an alternative to select individuals. We demonstrated the possibility of improving growth, visual score and reproductive traits, principal components and selection index in Nellore cattle since moderate to high genomic predictions can be achieved. These results indicate the possibility of using PCA for implementing genomic selection for Nellore cattle.

4.6 REFERENCES

Agudelo-Gómez DA, Agudelo-Trujillo JH, Cerón-Muñoz MF (2016) Selection index for meat and milk traits of buffaloes in Colombia. **Livestock Science** 191: 6-11.

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science** 93:743-752.

Amer PR and Banos G (2010) Implications of avoiding overlap between training and testing data sets when evaluating genomic predictions of genetic merit. **Journal of Dairy Science** 93:3320-3330.

Boddhireddy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, DeNise S (2014) Genomic predictions in Angus cattle: comparisons of sample size, response variables, and clustering methods for cross-validation. **Journal of Animal Science** 92:485–497.

Boligon AA, Vicente IS, Vaz RZ, Campos GS, Souza FRP, Carvalheiro R, Albuquerque LG (2016) Principal component analysis of breeding values for growth and reproductive traits and genetic association with adult size in beef cattle. **Journal of Animal Science** 94:5014-5022.

Bolormaa S, Pryce JE, et al. (2013) Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality in *Bos taurus*, *Bos indicus*, and composite beef cattle. **Journal of Animal Science** 91:3088–3104.

Buzanskas ME, Savegnago RP, Grossi DA, Venturini GC, Queiroz SA, Silva LOC, Torres Jr. RAA, Munari DP, Alencar MM (2013) Genetic parameter estimates and principal component analysis of breeding values of reproduction and growth traits in female Canchim cattle. **Reproduction, Fertility and Development** 25:775-781.

Falconer DS and Mackay TFC (1996) Introduction to quantitative genetics. 4.ed. Edinburgh : Longman Group Limited, 464p.

Fernandes GM, Savegnago RP, El Faro L, Mosaquatro Roso V, de Paz CCP. (2018) Economic values and selection index in different Angus-Nellore cross-bred production systems. **Journal of Animal Breeding and Genetics** 135:208-220.

Hazel LN (1943) The genetic basis for constructing selection indexes. **Genetics** 28:476–490.

Habier D, Fernando R, Dekkers J (2007) The impact of genetic relationship information on genome-assisted breeding values. **Genetics** 177:2389–97.

Habier D, Fernando R, Garrick D (2013) Genomic BLUP decoded: a look into the black box of genomic prediction. **Genetics** 194:597–607.

Kaiser HF (1960) The application of electronic computers to factor analysis. **Educional and Psychological Measurement** 20:141-151.

Koury Filho W, Albuquerque LG, Forni S, Silva JAV, Yokoo MJ, Alencar MM (2010) Genetic parameters estimates for visual scores and their association with body weight in beef cattle. **Revista Brasileira de Zootecnia** 39:1015-1022.

Lopes FB, da Silva MC, Magnabosco CU, Goncalves Narciso M, Sainz RD (2016) Selection indices and multivariate analysis show similar results in the evaluation of growth and carcass traits in beef cattle. **Plos One** 11:e0147180.

Meyer K (2007) Multivariate analyses of carcass traits for Angus cattle fitting reduced rank and factor analytic models. **Journal Animal Breeding Genetics** 124:50-63.

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

Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, et al., 2016. Manual for BLUPF90 Family of Programs. University of Georgia, Athens, GA.

Moreira HL, Buzanskas ME, Munari DP, Canova EB, Lôbo RB, de Paz CCP (2015) Reproductive traits selection in Nelore beef cattle. **Ciência e Agrotecnologia** 39(4).

Neves HHR, Carvalheiro R, et al. (2014) Accuracy of genomic predictions in Bos indicus (Nellore) cattle. **Genetics Selection Evolution** 46:17.

Patry C and Ducrocq V (2011) Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. **Journal of Dairy Science** 94:1011–1020.

Pesek J and Baker RJ (1969) Desired improvement in relation to selection indices. **Canadian Journal of Plant Science** 49:803–804.

R Development Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. R Core Team.

Roso VM and Fries LA (1995) Componentes principais em bovinos da raça Polled Hereford à desmama e sobreano. **Revista Brasileira de Zootecnia** 24:728-735.

Rouvier R (1969). Ponderation des valeurs genotypiques dans la selection par indexe sur plusieurs caractdres. **Biometrics** 25:295–307.

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

Smith HF (1936) A discriminant function for plant selection. Annals of Eugenics, London, 7:240-250.

Smith C (1983) Effects of changes in economic weights on the efficiency of index selection. **Journal of Animal Science** 56:1057-1064.

Vargas G, Schenkel FS, Brito LF, Neves HHR, Munari DP, Boligon AA, Carvalheiro R (2018) Unravelling biological biotypes for growth, visual score and reproductive traits in Nellore cattle via principal component analysis. **Livestock Science** 217:37-43.

Wiggans GR, Cooper TA, Vanraden PM, Cole JB (2011) Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. Journal of Dairy Science 94:6188–6193.

Yamada Y, Yokouchi K, Nishida A (1975). Selection index when genetic gains of individual traits are of primary concern. **The Japanese Journal of Genetics** 50:33-41.

CHAPTER 5 – FINAL CONSIDERATIONS

In the present study results related to PCA for growth, visual scores, and reproductive traits in Nellore cattle, as well as its application in genomic association and prediction studies have been shown. Three PCA approaches were investigated for obtaining the PCs. From these analyses, genetic associations among the studied traits were unraveled, allowing the characterization of specific biotypes that were suggested to be used as breeding objectives for selection purposes. Although the PCA using the (co)variance matrix of EBVs obtained from multi-trait analysis resulted in less PCs explaining the greatest proportion of the variance of EBVs when compared to the eigen-decomposition of the additive genetic (co)variance matrix ($\mathbf{A_T}$) or the PCA using the (co)variance matrix of EBVs obtained from single-trait analysis, the use of the first three PCs from the eigen-decomposition of $\mathbf{A_T}$ matrix seems to be the indicated approach to select Nellore cattle according to specific biotypes. Contrary to the (co)variance matrix of EBVs from single-trait analyses, the $\mathbf{A_T}$ matrix accounts for the genetic relationship among traits, allowing to contrast the individuals according to the additive genetic effect of the studied traits.

Our study showed important results from the genome-wide association studies (GWAS) using the main PCs for growth, visual score and reproductive traits in Nellore cattle. As most traits of interest in livestock, these PCs have presented a polygenic nature, which means that they are influenced by many loci with small effects. The identification of important chromosomal genomic regions associated with the PCs and their biological functions contributed to the further understanding of the genetic control acting on these components and their associated biotypes. These findings may contribute to the inclusion of these components on selection process and provide support for future functional genome studies. In addition, the observation of QTL previous reported in the literature covering areas surrounding the genes herein identified provides more evidence for these associations. Future studies targeting the genomic regions identified in this study could provide further knowledge to uncover the genetic architecture underlying the PCs.

The use of selection index based on PCs analysis was previously suggested as an attractive option because it simplifies the problem to a series of non-correlated new traits that can be useful as a simple selection strategy. The choice of selection criteria may vary according to the producer's decision and profile. The PC1 and FI indexes lead to similar response to selection when compared to the other indexes. However, it is important to emphasize that the main goal of using PCA for economically important traits is to capture the distribution for different genetic effects that allow contrasting individuals according to distinguished biotypes, while the selection indexes are developed according to the relative importance of the traits that meets a general production system. In this sense, the PCA would still represent an alternative approach that could be used by breeding programs to select individuals according to specific PCs that agree with the breeding objectives. In addition, when combining PC1 and PC2 to create a Harmonic Index, genetic gains of higher magnitude can be obtained.

Because the great relevance of Nellore cattle in beef production systems in Brazil, breeding programs have focused on using genomic selection to accelerate genetic improvement of economically important traits. Several studies have proven the feasibility of implementing genomic selection for growth, carcass, and meat quality traits in this breed (Neves et al., 2014, Fernandes Júnior et al., 2016, Magalhães et al., 2018). In this study, genomic predictions with moderate accuracies were obtained for growth, visual score and reproductive traits, PCs and a selection index in Nellore cattle. In general, these findings were slightly higher when compared to the reported in the literature for a similar set of traits. These differences may be due to the size of the training population, accuracy of the pseudo-phenotypes, level of LD, and the degree of relationship between training and validation populations (Bolormaa et al., 2013; Boddhireddy et al., 2014).

5.1 REFERENCES

Bolormaa S, Pryce JE, et al. (2013) Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality in Bos taurus, Bos indicus, and composite beef cattle. **Journal of Animal Science** 91:3088–3104.

Boddhireddy P, Kelly MJ, Northcutt S, Prayaga KC, Rumph J, DeNise S (2014) Genomic predictions in Angus cattle: comparisons of sample size, response variables, and clustering methods for cross-validation. **Journal of Animal Science** 92:485–497.

Fernandes Júnior GA, Rosa GJM, Valente BD, Carvalheiro R, Baldi F, Garcia DA (2016) Genomic prediction of breeding values for carcass traits in Nellore cattle. **Genetics Selection Evolution** 48:7.

Magalhães AFB, Fonseca LFS, Cardoso DF, Fernandes Júnior GA, Chardulo LAL, Albuquerque LG (2018) Genomic selection for meat quality traits In Nellore cattle. In: 55ª REUNIÃO ANNUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA. **Resumos...** Goiânia-GO.

Neves HHR, Carvalheiro R, et al. (2014) Accuracy of genomic predictions in Bos indicus (Nellore) cattle. **Genetics Selection Evolution** 46:17.