SÃO PAULO STATE UNIVERSITY
SCHOOL OF AGRICULTURAL AND VETERINARIAN SCIENCES
CAMPUS OF JABOTICABAL

ESTIMATION OF GENOTYPE-ENVIRONMENT INTERACTION USING GENOMIC REACTION NORM AND ANALYSIS OF GENE NETWORK FOR REPRODUCTIVE TRAITS IN NELLORE CATTLE

Lúcio Flávio Macêdo Mota
Animal Scientist

2019

SÃO PAULO STATE UNIVERSITY
SCHOOL OF AGRICULTURAL AND VETERINARIAN SCIENCES
CAMPUS OF JABOTICABAL

ESTIMATION OF GENOTYPE-ENVIRONMENT INTERACTION USING GENOMIC REACTION NORM AND ANALYSIS OF GENE NETWORK FOR REPRODUCTIVE TRAITS IN NELLORE CATTLE

Lúcio Flávio Macêdo Mota

Advisor: Prof. Dr. Lucia Galvão de Albuquerque

Thesis presented to the São Paulo State University - School of Agricultural and Veterinarian Sciences, Campus of Jaboticabal in partial fulfillment of requirements for the degree of Ph.D. in Genetics and Animal Breeding.

2019

M917e

Mota, Lúcio Flávio Macêdo
Estimation of genotype-environment interaction using genomic reaction norm and analysis of gene network for reproductive traits in Nellore cattle / Lúcio Flávio Macêdo Mota. -- Jaboticabal, 2019
87 p. : il., tabs.

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

1. gene function. 2. genomic prediction. 3. genotype-environment interaction. 4. Nellore cattle. 5. sexual precocity. I. Título.

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

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: ESTIMATION OF GENOTYPE-ENVIRONMENT INTERACTION USING GENOMIC REACTION NORM AND ANALYSIS OF GENE NETWORK FOR REPRODUCTIVE TRAITS IN NELLORE CATTLE

AUTOR: LÚCIO FLÁVIO MACÊDO MOTA
ORIENTADORA: LUCIA GALVÃO DE ALBUQUERQUE

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

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

Dr. MÁRIO LUIZ SANTANA JÚNIOR (Videoconferência)
Universidade Federal de Rondonópolis / Rondonópolis/MT

Prof. Dr. FERNANDO SEBASTIAN BALDI REY
Departamento de Zootecnia / FCAV / Unesp - Jaboticabal

Pós-doutoranda ANA FABRÍCIA BRAGA MAGALHÃES
Instituto de Zootecnia / Sertãozinho/SP

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

Jaboticabal, 26 de fevereiro de 2019

AUTHOR'S CURRICULAR DATA

Lucio Flavio Macedo Mota – born on July 06th of 1988 in Aracaju – SE, son of Luiz Fernando da Mota and Antonia da Conceição Macedo. Started the university study in Animal Science at Federal University of Alagoas – *Campus* Arapiraca, in August 2007 and concluded in July 2012. During the undergraduate, I was a teaching assistant in chemistry and biochemistry for two years and granted with a scientific/technological initiation scholarship from the “Fundação de Amparo a Pesquisa do Estado de Alagoas – FAPEAL” and presented him undergraduate thesis entitled “Influencia do sexo em medidas e índices zoométrico em bovinos Nellore jovens”. Started the master's degree in Animal Science in September 2012 and obtained the master’s degree in 2014 with the dissertation entitled “Expressão do gene leptina, proteômica e modelos para estimação do CAR em animais da raça Nellore” with Prof. Dr. Cristina Moreira Bonafé as advisor. Started the Ph.D. course in Genetics and Animal Breeding at School of Agricultural and Veterinarian Sciences - São Paulo State University, Campus of Jaboticabal in March 2015. Was granted with a scholarship from "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES" and later was granted with scholarship from “Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP”. During the doctorate’s course, held an internship at the University of Wisconsin – Madison, under the supervision of Prof. Dr. Guilherme Jordão de Magalhães Rosa, with a scholarship from FAPESP and a project entitled “A search for pleiotropic effects on reproductive traits in Nellore cattle considering genotype by environment interaction”.

“When you run so fast to get somewhere, you miss half the fun of getting there.

When you worry and hurry through your day, it's like an unopened gift thrown away.

Life isn't a race, so take it slower, hear the music before your song is over.”

David L. Weatherford

I dedicate this thesis to my parents and relatives, which near or far, were always present, always believing in me and encouraging me to go further, for giving me support for the realization of my dreams and you made this achievement possible. Thank you for trusting and believing in me. I could not have done it without your support!

You are the greatest reason for my existence and example of love, trust, encouragement, and dedication.

Dedication

ACKNOWLEDGMENTS

Firstly, I would like to thank God for life and bringing peace and comfort, even in the most difficult moments, enabling the realization of a great dream.

I am thankful to my family and principally my mother Antonia da Conceição Macedo, for always being by my side and which have always supported me in the construction of this dream. I am grateful to all people who helped shape my personality and to be a better person, thank you for believing in me.

My appreciation goes to my advisor, Dr. Lucia Galvão de Albuquerque. I am extremely grateful for her meaningful assistance, tireless guidance, and patience. Thank you for the extensive professional and personal advice, for believing in my potential and in my work. Always concerned with the proper development of the work and with professional and personal advice.

I am grateful for the support from my friends Alexandre Cominotte, Ana Cristina, Ana Fabrícia, Anderson, Andrés, André Mauric, André Vieira, Ândrea, Bianca, Bresolin, Burna Salata, Camila, Cherley, Daiane, Daniel Jordan, Dani Beraldo, Diércles, Diego, Diogo, Elisa, Espigolan, Fabielli, Gabi Bonfá, Gerardo, Giovana, Gordo, Henrique, Ivan, Laiza, Larissa, Lucas, Malane, Medeiros, Marisa, Natália, Tonussi, Rafael Espigolan, Samuel, Thaise and William. To all the other friends for friendship and supporting me. I wish to thank all friends and colleagues I have met during my Ph.D. course. To my friends, thank you for your thoughts, well-wishes/prayers and being there whenever I needed a friend.

I would like to express my deep gratitude to Professor Dr. Guilherme Jordão de Magalhães Rosa, for all support, attention, suggestions to improve this work, and for spending his time transmitting me valuable knowledge.

I am most grateful to the collaborators for lending me their expertise and intuition to my scientific and technical problems: Dr. Lucia Galvão, Dr. Roberto Carvalheiro and Dr. Guilherme Rosa.

I wish to thank all the friends that I have made in Madison, especially Anna, Arthur, Fernando, Paulo, Renata, Thiago Passafaro and Vera, with whom I have shared good moments and big excitement. Their presence was very important in a process that is often felt as tremendously solitaire. Thank you for given your time, energy, and expertise. I am richer for it.

I am also very grateful to my friends that I met during the internship period: Andrea Bellingeri, Angelina, Alisson Quick, Beth Lett, Carol, Di Liang, Hasan Khatib, Isabela, João Dórea, Jorge Barrientos, Karen Nielsen, Manfei Li, Nicole Lee, Nora Bello, Paulina, Rafael Caputo, Thiago Barros, Toshihiro Okamura and all friends from animal sciences department.

I would like to say thank the members of my PhD dissertation committee, for the valuable suggestions that contributed to the improvement of the work.

This journey would not have been possible without the support of my family, professors and mentors, and friends especially to Aldrin (in memoriam) and Lagrotta (in memoriam). To my family, thank you for encouraging me in all of my pursuits and inspiring me to follow my dreams. I am especially grateful to my parents, who have supported me. I always knew that you believed in me and wanted the best for me. Thank you for teaching me that my job in life was to learn, to be happy, and to know and understand myself only then, could I know and understand others.

This research was financially supported by FAPESP # 2009/16118-5 and CNPq # 559631/2009-0.

I am thankful for the scholarships from São Paulo Research Foundation - FAPESP #2015/25356-8 and #2017/02291-3. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

SUMMARY

ESTIMAÇÃO DA INTERAÇÃO GENÓTIPO-AMBIENTE UTILIZANDO NORMA DE REAÇÃO GENÔMICA E ANÁLISE DA REDE GÊNICA PARA CARACTERÍSTICAS REPRODUTIVAS DE BOVINOS DA RAÇA NELORE

Resumo - A interação genótipo-ambiente (GxE) pode ser uma importante fonte de variação em características reprodutivas com um efeito notável no início da puberdade animal. Desta forma, os objetivos do presente estudo foram: i) Avaliar a interação GxE em características indicadores de precocidade sexual em animais da raça Nelore em diferentes condições ambientais (EC) e ii) identificar regiões genômicas e vias biológicas associadas a características indicadores de precocidade sexual e verificar se seus efeitos mudam de acordo com os níveis de EC. Informações fenotípicas para idade ao primeiro parto (AFC), ocorrência de prenhez precoce (HP), reconcepção de novilhas (HR) e perímetro escrotal (SC), foram coletados em 128.994, 85.339, 90.831 e 151.053 animais, respectivamente. Destes, 1800 novilhas, 3050 touros jovens e 800 touros foram genotipados com BovineHD BeadChip. Um modelo de norma de reação foi usado para estimar a resposta do animal às mudanças nas condições ambientais. Para avaliar a capacidade preditiva, foram utilizados os esquemas de validação em animais jovens e em ambiente específico. Para varredura genômica ampla os efeitos dos marcadores SNP para as características reprodutivas foram estimados em três níveis de EC Baixo (EC = -3.0), Médio (EC = 0.0) e Alto (EC = 3.0) usando uma transformação linear dos valores genômicos genéticos. As regiões pleiotrópicas associadas com características reprodutivas (AFC, SC, HP e HR) em três EC foram identificadas utilizando a combinação estatística dos resultados de GWAS uni-característica e consideradas significativas quando $-log_{10}(p\text{-valor}) > 6.0$. A inclusão de informação genômica combinada com pedigree para avaliar a interação GxE, proporcionou estimativas dos componentes de variância e parâmetros genéticos mais acurados quando comparado com apenas a informação de pedigree. A interação GxE apresentou um efeito importante na prenhez precoce e perímetro escrotal levando a sensibilidade do valor genético genômico. Além disso, uma forte interação marcador SNP-ambiente foi observada com mudanças no efeito e variância explicada nos diferentes níveis ambientais. O esquema de validação em animais mais jovens apresentou maior capacidade preditiva para SC e HP em comparação com a validação específica do ambiente. A utilização da combinação estatística dos resultados de GWAS uni-característica identificou um total de 56 marcadores SNP com efeito pleiotrópico para características reprodutivas em diferentes níveis de EC. Tais marcadores foram distribuídos nos *Bos taurus autosome* (BTA) 5, 7, 10, 14, 16 e 29, e apresentaram mudanças em seu efeito de acordo com as mudanças nas condições ambientais. Um total de 45 genes foram identificados dentro de 200kb dos marcadores significativos e apresentam um efeito importante em processos biológicos associado a deposição de gordura, crescimento embrionário, diminuição dos níveis hormonais e diminuição da fertilidade em machos e fêmeas (principalmente na diminuição do número de oócitos e espermatogênese). Em resumo, este estudo indicou o efeito da interação GxE no valor genético de características reprodutivas em bovinos da raça Nelore. Além disso, as regiões genômicas identificadas apresenteram um papel fundamental nos mecanismos

biológicos que codificam as principais vias reprodutivas e as mudanças nas condições ambientais apresentam efeitos importantes em sua expressão. Estes resultados contribuem para uma melhor compreensão de diferenças no desempenho reprodutivo de animais Nelore criados em diferentes condições ambientais, podendo auxiliar o melhoramento genético de características reprodutivas em bovinos Nelore.

Palavras-chave: bovinos Nelore, função gênica, interação genótipo-ambiente, precocidade sexual, predição genômica

ESTIMATION OF GENOTYPE-ENVIRONMENT INTERACTION USING GENOMIC REACTION NORM AND ANALYSIS OF GENE NETWORK FOR REPRODUCTIVE TRAITS IN NELLORE CATTLE

Abstract – Genotype-environment (GxE) interactions could be an important source of variation in reproductive traits with a striking effect on the onset of animal puberty. Thus, the objectives of the present study were: i) to assess the GxE interaction in Nellore sexual precocity indicator traits under different environmental conditions (EC) and ii) to identify, genomic regions and biological pathways associated to Nellore sexual precocity indicator traits and to investigate whether their effects changes according to EC levels. Phenotypic records for age at first calving (AFC), heifer early pregnancy (HP), heifer rebreeding (HR) and scrotal circumference (SC) were collected on 128,994; 85,339; 90,831 and 151,053 animals, respectively. From those, 1800 heifers, 3050 young bulls, and 800 sires were genotyped with BovineHD BeadChip. A reaction norm model was used to estimate the animal's response to environmental conditions changes. To assess the predictive ability the *younger scheme* and *environment-specific scheme* were used. For genome-wide scan, the SNP effects for reproductive traits were estimated in three EC levels: Low (EC = -3.0), Medium (EC = 0.0) and High (EC = 3.0) using a linear transformation of the genomic breeding values. The pleiotropic regions associated to reproductive traits (AFC, SC, HP and HR) in three EC levels, were identified using the statistical combination of the single-trait GWAS results and considered significant when $-log_{10}(p\text{-}value) > 6.0$. The inclusion of genomic information combined with pedigree, lead to variance components and genetic parameter estimates more accurate than that considering pedigree information. The GxE interaction had an important effect on heifer early pregnancy and scrotal circumference leading to genomic breeding value sensitivity in different environmental condition. In addition, a strong SNP marker-environment interaction was observed for Nellore sexual precocity indicator traits with changes in their effect and explained variance in different EC levels. The *younger scheme* leads to a higher predictive ability for SC and EP compared to *environment-specific scheme*. Using the statistical combination of single-trait GWAS were identified, a total of 56 SNP markers with pleiotropic effect for reproductive traits in different EC levels. Those SNP markers were distributed on *Bos Taurus* autosome (BTA) 5, 7, 10, 14, 16, and 29 and showed changes in their effect according to environmental condition changes. A total of 45 genes were mapped within 200 kb and show a striking effect in biological process associated to fat deposition, embryonic growth, decreasing in hormonal levels and decreasing female and male fertility (mainly on decreased oocyte number and spermatogenesis). In summary, this study pointed to the effect of GxE interaction on breeding value of reproductive traits in Nellore cattle. In addition, genomic regions identified show a key role in biological mechanisms that encode the major reproductive pathways, and that the environmental conditions changes have important effects on its expression. These results contribute to a better understanding of differences in reproductive

performance of Nellore cattle raised under different environmental conditions, might aid the genetic improvement of reproductive traits in Nellore cattle.

Keywords: gene function, genomic prediction, genotype-environment interaction, Nellore cattle, sexual precocity

Chapter 1 – General Considerations

1.1 Introduction

Brazilian beef cattle animals are raised and selected on pasture conditions in a wide range of environments, with variations in nutritional levels and heat stress. Differences in these environmental conditions have been pointed out as an important factor to phenotypic variation (Burrow, 2012). Usually, traits as age at first calving, heifer early pregnancy and scrotal circumference are used to identify and select precocious animals for reproduction. However, Nellore heifers are raised on pasture during their growth period with seasonal variations in forage quality and quantity. These environmental factors are partially responsible for the high age at puberty in Brazilian herds. In general, genetic-quantitative studies have shown that these traits are affected by environmental conditions leading to genotype-environment (GxE) interaction (Chiaia et al., 2015; Santana Jr et al., 2017; Santana Jr et al., 2018a; Santana Jr et al., 2018b).

In beef cattle breeding programs, generally, animals are selected under improved environmental conditions, which allow high production. However, due to the presence of GxE interactions, sires genetic rank can change when their offspring are raised under poor nutritional level, affecting genetic gain (Cardoso and Tempelman, 2012; Rauw and Gomez-Raya, 2015). Nevertheless, to mitigate these effects, breeding programs should target animals more adapted to different environments (Kolmodin et al., 2003; Kolmodin and Bijma, 2004).

When GxE interactions occur, the animal's selection could be driven to identifying the genotype with better performance in different environmental conditions increasing their robustness to environmental variation (Mulder and Bijma, 2005). This selection strategy might lead to better genetic gains due to the similarity of environmental conditions between seedstock herd and production systems (Nirea and Meuwissen, 2017). Moreover, this selection design, based on their resilience, has been proposed to reduce GxE interaction effects on genetic gain, using the robustness trait in breeding programs (Knap, 2005; Hermesch et al., 2015; Colditz and Hine, 2016).

The effect of animal sensitivity to environmental conditions changes have been assessed through reaction norm models, due to considering that the trait could change continuously across the environment (Kolmodin et al., 2002; Calus et al.,

2002; Kolmodin and Bijma, 2004; Su et al., 2006; Cardoso and Tempelman, 2012; Santana et al., 2017; Grenier et al., 2016). Hence, animal sensitivity is expressed as a function of the reaction norm slope (de Jong, 1995). In this sense, reaction norm is fitted similarly to random regression model (Schaeffer, 2004), using as explanatory variable, environmental gradients. This approach provides intercept and slope estimates for each animal, where the estimates for the slope could be used to select animals more robustness (Knap, 2005; Knap and Su, 2008).

The development of modern genotyping and sequencing technologies, e.g. single nucleotide polymorphism (SNP), have been applied to animal selection enabling higher accuracy of selection (Meuwissen et al., 2001). Thus, GxE interaction evaluation trough genomic information in reaction norm models represents a promising alternative to evaluate animal sensitivity to environmental changes (Lillehammer et al., 2007). Indeed, genomic reaction norms have played an important role to knowledge the genetic variability of animals with performance environment-dependent (Lillehammer et al., 2009; Silva et al., 2014; Mulder, 2016).

The analysis of the GxE interaction using genomic reaction norm models (RNM) represents a promising alternative to evaluate characteristics strongly influenced by environmental factors, i.e. age at first calving, heifer early pregnancy, scrotal circumference, and heifer rebreeding. The inclusion of genomic information in RNM shows advantages for the estimation of genetic parameters and prediction of breeding value, increasing prediction accuracy, mainly, for fertility traits (Zhang et al., 2019). Besides, it permits to identify changes in genomic regions affecting economically relevant traits in different environments. The knowledge of genomic regions involved with GxE interaction for reproductive traits in Nellore cattle can aid the development of breeding programs in harsh conditions.

1.2 Literature review

1.2.1 Structure of beef cattle programs

Beef cattle breeding programs are hierarchically organized with animal selection performed in seedstock herds, to subsequently widespread the genetic material through artificial insemination across the different commercial herds (Simm, 1998; Figure 1). Beef cattle breeding programs are structurally organized to select animals in seedstock herd for high production levels in more favorable conditions,

i.e. better nutrition levels and management (Figure 1), supporting the expression of their genetic potential (Rauw and Gomez-Raya, 2015). In contrast, these animals, with genetic superiority in seedstock herds, will produce offspring raised in commercial herds that show a wide range of environmental conditions defined through temperature, seasonally poor nutritional levels, health status, as well as other stressor factors. Hence, when sires are selected for increasing performance under high nutrition level and their offspring are raised in harsh conditions the differences between sires and/or the sires rank can change (Burrow, 2012).

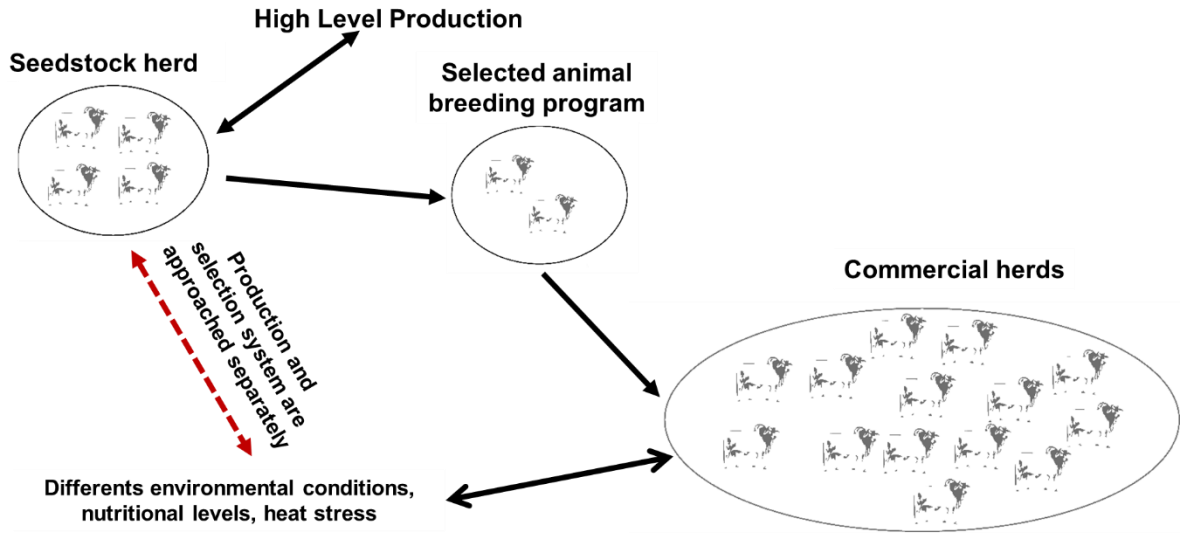


Figure 1. Hierarchical structure of beef cattle breeding programs with animal selection in seedstock herd through better nutrition, and management, to subsequently disseminate in commercial herds. Adapted from Simm (1998) and Klopčič (2009).

1.2.2 Genotype – environment interaction (GxE)

The GxE interaction occurs when the performance of genotypes, exhibit different response to environmental condition changes (Falconer and Mackay, 1996). The GxE could show different kinds of effect on trait expression, so that different genotypes can have different response when exposed to harsh condition (Figure 2). When the genotypes show differences in estimated breeding value (EBV) across environmental conditions, but without changes in ranking the environmental effect lead to a scaling effect of EBV (Figure 2 A). However, when genotypes exhibit different levels of response according to environmental condition, often the re-ranking of genotypes could occur (Figure 2 B). The scaling, without re-ranking, effect

caused by GxE interaction on EBV value show less importance, because the best genotype in determined environment would still in other environments.

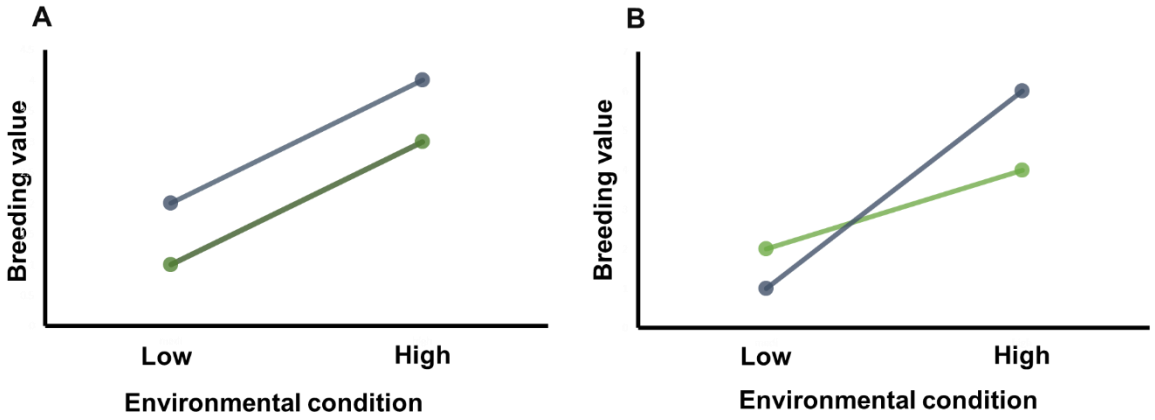


Figure 2. Environment effect on genotype variation, the breeding value can vary depending on the genotype and environment, and often a complex interaction between genotype and environment. (A) Example of environment effect leading to scaling effect on breeding value of genotype and (B) re-ranking of breeding value in response to environmental change.

The GxE interaction leading to animal re-rank of EBV (Figure 2 B), of the trait evaluated in different environmental conditions can be considered as different traits (Kolmodin et al., 2002) and the magnitude of GxE interaction effect could be assessed by the genetic correlation of trait in different environmental conditions (Falconer and Mackay, 1996; Kolmodin et al., 2002). Robertson (1959), suggested that correlation for a trait evaluated in different environments lower than 0.80 would indicate an important re-ranking and would justify the separate selection scheme for breeding progams (Mulder and Bijma 2005; Mulder et al., 2006). Hence, the selection of animal for determinate trait in a determined environmental condition may not be favorable for offspring raised in a different environment.

1.2.3 Modeling GxE interactions

To evaluate the GxE interaction effect and to obtain breeding values (EBV) in different environmental conditions for beef cattle, two models are frequently used: multitrait and reaction norm models.

1.2.3.1 Multi-trait model for GxE interactions

The multi-trait model is useful when environmental conditions can be described by few categories, e.g., when system of production (pasture x feedlot), level of nutrition (high x low), system of production (nucleus x production or feedlot and pasture progeny test) or performance test in different countries (Interbull - http://www.interbull.org), are evaluated.

In multi-trait models for GxE interaction assumes that the same phenotypic information, measured in different enviromental conditions, are different traits (Falconer and Latyszewski 1952; Malosetti et al., 2013). Thus, the (co)variance components can be estimated for a trait in each environmental condition and the animal re-ranking across environmental conditions is quantified by the genetic correlation (Robertson, 1959).

The multitrait model approach to evaluate the GxE interaction for three environments may be described as:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

where the y_1, y_2 and y_3 represents the phenotypic information for same trait (*y*) recorded in different environmental conditions (*1* to *3*), *b* represents the vector of fixed effect for trait in each environmental conditions, *a* represents the vector of random animal effect for each environment (*1* to *3*), e_1, e_2 and e_3 represents the residual effect for environments and *X* and *Z* are the incidence matrix of fixed effect and random effect of animal, respectively.

The random effects are considered to follows the multivariate normal distribution (MVN): $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \sim MVN(0, A \otimes G)$ and $\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \sim MVN(0, I \otimes R)$, where *A* is the additive genetic relationship matrix and *G* is the additive genetic (co)variances matrix as follow: $G = \begin{bmatrix} \sigma^2_{aE1} & \sigma_{aE2,1} & \sigma_{aE3,1} \\ \sigma_{aE1,2} & \sigma^2_{aE2} & \sigma_{aE3,2} \\ \sigma_{aE1,3} & \sigma_{aE2,3} & \sigma^2_{aE3} \end{bmatrix}$, σ^2_{aE1}, σ^2_{aE2} and σ^2_{aE3} represents the additive genetic variances on environment E_1, E_2 and E_3 and σ_{aE1}, σ_{aE2} and σ_{aE3} are the genetic covariance between environments. The residual (co)variance matrix denoted

by $R = \begin{bmatrix} \sigma^2_{eE1} & 0 & 0 \\ 0 & \sigma^2_{eE2} & 0 \\ 0 & 0 & \sigma^2_{eE3} \end{bmatrix}$, where σ^2_{eE1}, σ^2_{eE2} and σ^2_{eE3} represents the residual variances on environmentE_1, $E_2$and E_3 and residual covariance between environments (E) is zero ($\emptyset$) because the animal are raised in only one environment, *I* is the identity matrix and $\otimes$ is the Kronecker product.

The multitrait models enable to estimate the genetic correlation between the same traits under different environmental conditions. Genetic correlations are dependent on environmental conditions, with an important implication on genetic gain and tradeoffs (Sgrò and Hoffmann, 2004). Thus the genetic correlation of trait across environments is defined as $r_{aE1,aE2} = {\sigma_{aE1,aE2}} \Big/ {\sqrt{\sigma^2_{aE1} * \sigma^2_{aE2}}}$ and estimated through genetic parameters estimated by the multitrait models.

Raidan et al. (2016) evaluated differences in breeding values using multi-trait models for growth and reproductive traits with animals raised on pastures and feedlots. They reported the occurrence of GxE interaction and concluded that a higher response to selection was obtained evaluating and selecting animals on pasture instead of feedlots. Nirea and Meuwissen (2017) assessed the GxE interaction using multi-trait models in pigs and an important effect of GxE interaction was observed by the low genetic correlations among nucleus and production systems.

Robertson (1959) suggested that genetic correlations below 0.80, indicate an important biological effect of GxE interaction in a trait. In various studies using multi-trait models, genetic correlations below 0.80 have been associated to reductions in selection response (Fikse et al., 2003; Ruiz-Sánchez et al., 2007; Tsuruta et al., 2015; Raidan et al., 2016).

1.2.3.2 Reaction norm model for GxE interaction

The multi-trait models used to assess the GxE interaction show limitations when many environmental conditions are evaluated, leading to problems in convergence of models. Hence, the genetic parameters estimation could be biased. In this context, the reaction norm (RN) model may be useful, particularly in situations in which the environmental conditions can be described as continuous variable (Kolmodin et al., 2002; Schaeffer, 2004). The GxE interaction is fitted considering

that the animal performance could change continuously across the environment (Schaeffer, 2004).

The RN models assume that phenotypic value is express as a polynomial function associated with the environmental condition, in which the polynomial coefficients show genetic influence (de Jong, 1995). Assuming a linear effect, the animal sensitivity can be expressed in function of the regression slope of the animal performance (e.g. reproductive traits) on known environmental conditions (de Jong, 1995). Hence, genetic variations across environmental conditions resulting from variations on slope coefficient.

The animals' response to environmental conditions (EC) changes can be describes by RN model as follows:

$$y_{ij} = XB + \sum_{f=0}^{1} \omega_f \Phi_f(EC_j) + \sum_{f=0}^{1} \alpha_{if} \Phi_f(EC_j) + e_{ij}$$

where, y_{ij} is the vector for the phenotypic information of animal *i* recorded in the environment *j*; $X\beta$ is the fixed effects, ω_f are the fixed regression coefficients on Φ_f; Φ_f are the Legendre polynomials corresponding to *EC* level (EC_j); α_{if} are the random regression for additive effects of intercept and slope for animal *i* on the EC_j and e_{ij} is a random residual.

The α_{if} is assumed to follow a normal distribution $\alpha \sim N(0, A \otimes K_{ab})$, where *A* is the relationship matrix and $K_{ab} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$ is the additive genetic (co)variance matrix for the reaction norm coefficients, where σ_a^2, σ_b^2 and σ_{ab} genetic variance for intercept and slope, and genetic covariance between intercept and slope, respectively. In case of heteroscedastic residual variance between environments assume $e \sim N(0, IR)$ where I is the identity matrix and $R = \begin{bmatrix} \sigma_{eE1}^2 & 0 \\ 0 & \sigma_{eEp}^2 \end{bmatrix}$ is the residual variances in the *E1* and *Ep* environment, respectively.

The covariance components in different environmental conditions are obtained by covariance functions (Kirkpatrick et al., 1990; Schaeffer, 2004): $V_{AE} = \alpha_{if} K_{ab} \alpha_{if}'$, where V_{AE} represents the genetic covariance matrix across environmental conditions and α_{if} represents the polynomial coefficients for each level of the environments and α_{if}' transposed of α_{if}. The genetic correlation

between environments can be estimates though covariance functions $r_{En,Ep}$= Cov(A_{En},A_{Ep})/$\sqrt{V_{AEn} * V_{AEp}}$. Thus, the breeding value (EBV) of animals in a determined environment also can be estimated in different environmental conditions as a function: $EBV = \alpha_{if}\ \Phi_f{}'$, were α_{if} the genetic values estimated for reaction norm coefficients intercept (a) and the slope (b) and Φ_f represents the transposed polynomial coefficients for each level of the environments.

The EBV variation for a determined animal in reaction norm model is expressed as functions relating EBV of animal to environmental variables (Figure 3). The reaction norm for an animal is measured by combination of two distinct components denominated intercept and slope (Figure 3 A).

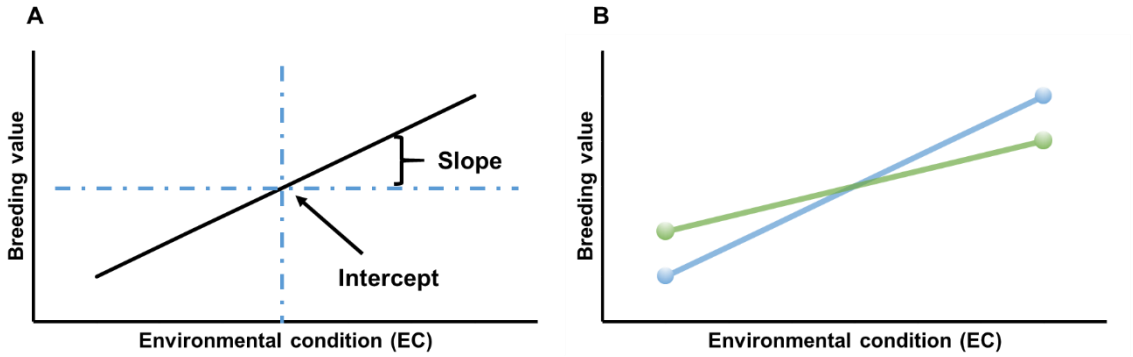


Figure 3. Graphical representation of reaction norm for breeding value (EBV) to environmental changes. Linear reaction norm model is characterized by the intercept (represents the expected average EBV of animal) and the slope (represents the animal EBV for environmental sensitivity). [A] represents the variation of EBV as function across environmental condition (EC) and *a* represents the intercept and *b* represents the slope of reaction norm. [B] Representation of animals that show differences in the parameters of the reaction norm (intercept and slope).

The advantages of GxE interaction assessed by reaction norm model, is that EBV can be estimated across the continuous environments, but also their environmental sensitivity through the slope estimation (Fuller et al., 2005). Thus, slope estimation determines the intensity EBV sensitivity for the animal to environmental changes (Figure 3 B). Moreover, the (co)variance estimates for intercept and slope can be used to evaluate the genetic variation and genetic correlations of a trait across environmental conditions applying covariance functions (Kirkpatrick et al., 1990). Thus, the level of GxE interaction can be also measured by the ratio of genetic variance of the slope (σ^2_{a1}) and intercept (σ^2_{a0}) (Kolmodin and

Bijma 2004) and by their genetic correlation ($r_g = \sigma_{a0,a1}/\sqrt{\sigma_{a0}^2\sigma_{a1}^2}$) (Kolmodin et al., 2002).

The GxE interaction in dairy and beef cattle have been evaluated by RNM in a wide range of environmental factors and animals re-ranking has been found when the environments show a large difference, e.g. different temperature index or production level (Kolmodin et al., 2002; Schaeffer, 2004; Satana et al., 2018a; Santana et al., 2018b). In dairy cattle, RNM has been frequently used because the sires have their offspring raised in a wide range of herd environments. Using RNM considering GxE interaction, Kolmodin et al. (2003), and Kolmodin, and Bijima (2004) occurs that, in restrictive environments, the animals were re-ranked, indicating different response to selection across the environments.

In beef cattle, the animal sensitivity have been identified for reproductive (Chiaia et al., 2015; Santana et al., 2017; Araujo Neto et al., 2018; Santana Jr et al., 2018a; Santana Jr et al., 2018b) and growth traits (Pegolo et al., 2011; Cardoso and Tempelman, 2012; Chiaia et al. 2015; Oliveira et al., 2016, Oliveira et al., 2018). These authors identified an important effect of GxE interaction and low correlation estimates between extreme environments, leading to a significant re-ranking of animals in different environments. Santana et al. (2017) observed the evidences of GxE interaction on non-return rate in Brazilian Holstein cattle as effect of heat stress. They concluded that the genetic evaluation and animal selection should take into account the GXE interaction effects caused by heat stress. Araujo Neto et al. (2018), evaluating the effect of GxE interaction on age at first calving and production traits in Nellore cattle, observed different responses in breeding value of sires to changes in the environmental gradient. The authors concluded that it is important to consider the GxE effect in breeding program schemes in order to maximize the genetic gain.

1.2.4 Genotype x environment interaction in genomic era

Genotype-environmental interaction has been an important source of phenotypic variation in beef cattle and have been assessed at the phenotypic level (Rauw and Gomez-Raya, 2015; Mulder, 2016). These differences in the phenotypic plasticity could be explained by the fact that some alleles might have different expressions according to the environments. Currently, the development of genomic tools has provided a substantial gain in knowledge about the genetic mechanisms

underlying complex traits and accuracy of selection for selection candidates (Meuwissen et al., 2001). In this sense, combining reaction norm (RN) models and genomic information, can lead to a better estimation of GxE interaction estimate compared to conventional selection methods using pedigree information (Silva et al., 2014). In addition, provide new opportunities to identify and to explore the genetic factors and molecular variants that affect the animal sensitivity (Des Marais et al., 2013; Mota et al., 2017).

1.2.4.1 Genomic selection exploring the GxE interaction

Genomic selection (GS) as proposed by Meuwissen et al. (2001) permits to use the genome-wide SNP markers information to predict the breeding values of animals, allowing the more accurate predictions of breeding value (EBV), improving the animal selection by breeding programs. Combining genomic information in model that assess the GxE interaction, e.g. RN models, high accuracy for selection have been achieved compared to pedigree information (Silva et al. 2014; Mulder, 2016). The increase in accuracy of breeding value prediction could result of genomic information to overcome the problem of few sire having offspring in various environments (Hayes et al., 2016). This occur because the breeding value for animals can be calculated based on genomic information, requiring that the genotyped animal have been evaluated in different environmental conditions (Nirea and Meuwissen, 2017).

The genomic models considering GxE interaction have increased prediction accuracy of phenotypic response for traits such as total number born in pigs (Silva et al., 2014), tick resistance in Hereford and Braford beef cattle (Mota et al., 2016), feed efficiency traits in pigs (Nirea and Meuwissen, 2017), yearling weight in Nellore cattle (Oliveira et al., 2018). The inclusion of GxE interactions into genomic prediction models, can aid breeding programs to select animals adapted to different environmental conditions and it generates higher accuracy of animal selection.

The GxE interaction shows an important effect on prediction accuracies with the lowest estimates when only a subset of environments is used in the reference population (Nirea and Meuwissen, 2017). These reduction on predictive ability of model occurs due to GxE interaction effect on re-ranking of animals, mainly when observe lower genetic correlations estimates between environments [$r < 0.80$]

(Bohlouli et al., 2017). Nirea and Meuwissen (2017) concluded that using a reference population with animal from different environments achieves more accuracy of prediction than only environments. Thus, the reference population can be oriented towards a customized approach by identifying the best designs that account the environmental sensitivity to improve the genetic gain.

1.2.4.2 Genome-wide scan exploring the GxE interaction

The occurrence of GxE interaction in target traits of breeding programs presents potential challenges for the selection, caused by differences in the performance of genotypes. Mainly, by the fact that the presence of GxE interaction can lead to changes of genetic variants effects among environments and their association with the trait might be significant in one environment but not significant in another environment (Lillehammer et al., 2009). Genome-wide scan to the mapping genomic regions combined with RN models permits to explore the understanding of different mechanisms involving the GxE interaction and their relative importance in animal sensitivity (Streit et al., 2012; van Gestel and Weissing, 2016). Recently, Des Marais et al. (2013) have provided insight about the most important genomic regions associated with GxE interaction and their control appears to involve different genetic regions affecting major pathways for phenotypic expression.

The genome-wide association study (GWAS) under different environments, on dairy cattle (Lillehammer et al., 2007; Hayes et al., 2009; Lillehammer et al., 2009), beef cattle (Mota et al., 2017) and swine (Silva et al. 2014), have identified several genomic regions associated to an important effect on phenotype expression in different environmental conditions. Remarkably, animal adaptation to harsh conditions is modulated by diverse genomic factors, related to major metabolic pathways (Pfau and Russo, 2015). These factors also have been associated with shared effects that can range from antagonistic effects to, what is most commonly observed, differential sensitivity of alleles across environments (Pigliucci, 2005). In addition, the knowledge of genomic regions involved with GxE interactions could be important to aid the development of breeding programs in harsh conditions and with climate change.

Recently, the single-trait GWAS results have been largely used in the statistical combination aiming to uncover important genomic regions among

correlated traits (Bolormaa et al., 2014 and Melo et al., 2018). The major advantage of combination statistical of single-trait GWAS is the increase of statistical power to detect genomic variants associated with correlated traits, i.e. pleiotropic effect (Bolormaa et al., 2014). In this sense, the statistical combination of GWAS results in different environmental conditions can be used to identify the existence of possible pleiotropic effects for different traits evaluated across environments. Thus, the identification of pleiotropic genomic regions among traits in different environmental conditions can contribute to a better understanding of physiological mechanisms involved in GxE interaction.

1.2.5 Influence of environmental conditions on reproductive traits in beef cattle

In the context of sexual precocity and reproductive traits selection in beef cattle, the GxE interaction is an important effect (Chiaia et al., 2015; Santana Jr at al., 2018b, Araujo Neto et al., 2018). Several studies have focusing on the effects of production systems on reproductive traits and have found that their differences are responsible for differences on reproduction indexes (Barth et al., 2008; Samadi et al., 2014; Nepomuceno et al., 2017; Ferraz Jr et al., 2018).

Production systems are a key factor on heifers and young bulls' sexual precocity and heifer rebreeding (Barth et al., 2008; Nepomuceno et al., 2017; Ferraz Jr et al., 2018). Ferraz Jr et al. (2018) observed that an improvement in the production system, i.e. nutritional factors, increases sexual precocity in Nellore heifers. In young bulls, the nutritional factors can reduce the age at puberty and promote earlier onset of spermatogenesis (Barth et al., 2008). These authors concluded that post-weaning body weight and average daily gain are strongly associated to age at puberty.

Differences in nutrition levels affect sexual precocity and reproductive performance in cattle by differences in metabolic status, through changes in glucose, insulin and reproductive hormones (Robinson et al., 2006; Barth et al., 2008; Garnsworthy et al., 2008; D'Occhio et al., 2019). Perhaps these factors are involved directly with specific metabolites with a striking effect on oocyte and spermatogenesis development, ovulation, embryo growth and survival and pregnancy rate (Robinson et al., 2006; Ashworth et al., 2009). These suggest that genes regulate the energy expenditure, implying that cattle fertility is an energy-

dependent process. Zhang et al. (2019), evaluating fertility traits in Danish Holstein cattle using genomic reaction norm models, observed that growth, immunity, and reproduction related genes affect directly fertility traits.

Differences in production systems can give rise to changes in gene expression (Kantanen et al., 2015). The understanding of this mechanism could be better explored by dense SNP-markers technology to be used to search for adaptation patterns. Thus, from a GxE interaction genomic point of view, animal adaptation to environmental conditions or nutritional factors are associated with structural and functional changes in genomics regions (Lv et al., 2014) and can be estimated from genomic reaction norms (Zhang et al., 2019).

1.3 Objective

1.3.1 General objective

The main objectives of this study were the understanding of the genetic and genomic basis of the animal sensitivity for Nellore sexual precocity indicator traits in different environmental condition and the biological pathways behind genotype-environment interaction.

1.3.2 Specific objectives

• To verify the occurrence of GxE interaction in Nellore sexual precocity indicator traits by RNM using pedigree and combining pedigree and genomic relationship aiming to increase accuracy of prediction applying different scheme of training and validation. In addition, the effect of GxE interaction on SNP markers in different environmental conditions was also evaluated.

• To carry carried out to identifying major pleiotropic genomic regions associated to age at first calving (AFC), heifer early pregnancy (HP), heifer rebreeding (HR), and scrotal circumference (SC) traits, in Nellore cattle raised in different environmental conditions (EC).

1.4 References

Araujo Neto FR de, Pegolo NT, Aspilcueta-Borquis RR, Pessoa MC, Bonifácio A, Lobo RB, de Oliveira HN (2018) Study of the effect of genotype–environment interaction on age at first calving and production traits in Nellore cattle using

multi-trait reaction norms and Bayesian inference. **Animal Science Journal** 89:939–945.

Ashworth CJ, Toma LM, Hunter MG (2009) Nutritional effects on oocyte and embryo development in mammals: implications for reproductive efficiency and environmental sustainability. **Philosophical Transactions of the Royal Society B: Biological Sciences** 364:3351–3361.

Barth AD, Brito LFC, Kastelic JP (2008) The effect of nutrition on sexual development of bulls. **Theriogenology** 70:485–494.

Bohlouli M, Alijani S, Javaremi AN, König S, Yin T (2017) Genomic prediction by considering genotype × environment interaction using different genomic architectures. **Annals of Animal Science** 17:683–701.

Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, Tier B, Savin K, Hayes BJ, Goddard ME 2014. A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. **PLOS Genetics** 10:e1004198.

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

Burrow HM (2012) Importance of adaptation and genotype × environment interactions in tropical beef breeding systems. **Animal** 6:729–740.

Calus MPL, Groen AF and de Jong G (2002) Genotype × Environment Interaction for Protein Yield in Dutch Dairy Cattle as Quantified by Different Models. **Journal of Dairy Science** 85:3115–3123.

Cardoso FF and Tempelman RJ (2012) Linear reaction norm models for genetic merit prediction of Angus cattle under genotype by environment interaction. **Journal of Animal Science** 90:2130–2141.

Chiaia HLJ, de Lemos MVA, Venturini GC, Aboujaoude C, Berton MP, Feitosa FB, Carvalheiro R, Albuquerque LG, de Oliveira HN, Baldi F (2015) Genotype × environment interaction for age at first calving, scrotal circumference, and yearling weight in Nellore cattle using reaction norms in multitrait random regression models. **Journal of Animal Science** 93:1503–1510.

Colditz IG, Hine BC (2016) Resilience in farm animals: Biology, management, breeding and implications for animal welfare. **Animal Production Science** 56:1961–1983.

de Jong G (1995) Phenotypic plasticity as a product of selection in a variable environment. **American Naturalist** 145:493–512.

Des Marais DL, Hernandez KM, Juenger TE (2013) Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. **Annual Review of Ecology, Evolution, and Systematics** 44:5–29.

Falconer DS and Latyszewski M (1952) The environment in relation to selection for size in mice. **Journal of Genetics** 51:67–80.

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. **Pearson**, Burnt Mill, England p. 480.

Ferraz Jr MVC, Pires A V, Santos MH, Silva RG, Oliveira GB, Polizel DM, Biehl M V, Sartori R, Nogueira GP (2018) A combination of nutrition and genetics is able to reduce age at puberty in Nelore heifers to below 18 months. animal 12:569–574.

Fikse WF, Rekaya R, Weigel KA (2003) Genotype × Environment Interaction for Milk Production in Guernsey Cattle. **Journal of Dairy Science** 86:1821–1827.

Fuller T, Sarkar S, Crews D (2005) The use of norms of reaction to analyze genotypic and environmental influences on behavior in mice and rats. **Neuroscience & Biobehavioral Reviews** 29:445–456.

Grenier S, Barre P, Litrico I (2016) Phenotypic Plasticity and Selection: Nonexclusive Mechanisms of Adaptation. **Scientifica** 2016:1–9.

Hayes BJ, Bowman PJ, Chamberlain AJ, Savin K, van Tassell CP, Sonstegard TS, Goddard ME (2009) A Validated Genome Wide Association Study to Breed Cattle Adapted to an Environment Altered by Climate Change. **Plos One** 4:e6676.

Hayes BJ, Daetwyler HD, Goddard ME (2016) Models for Genome × Environment Interaction: Examples in Livestock. **Crop Science** 56:2251–2259.

Hermesch S, Li L, Doeschl-Wilson AB, Gilbert H (2015) Selection for productivity and robustness traits in pigs. **Animal Production Science** 55:1437–1447.

Kantanen J, Løvendahl P, Strandberg E, Eythorsdottir E, Li M-H, Kettunen-Præbel A, Berg P, Meuwissen T (2015) Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries. **Frontiers in genetics** 6:1–9.

Klopčič M, Reents R, Philipson J, Kuipers A (2009) Breeding for robustness in cattle. **Wageningen Academic Publishers**, The Netherlands p 289.

Kirkpatrick M, Lofsvold D, Bulmer M (1990) Analysis of the inheritance, selection and evolution of growth trajectories. **Genetics** 124: 979–993..

Knap PW (2005) Breeding robust pigs. **Australian Journal of Experimental Agriculture** 45:763–773.

Knap PW, Su G (2008) Genotype by environment interaction for litter size in pigs as quantified by reaction norms analysis. **Animal** 2:1742–1747.

Kolmodin R, Strandberg E, Madsen P, Jensen J, Jorjani H (2002) Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. Acta **Agriculturae Scandinavica** 52:11–24.

Kolmodin R, Strandberg E, Jorjani H, Danell B (2003) Selection in the presence of a genotype by environment interaction: response in environmental sensitivity. **Animal Science** 76:375–385.

Kolmodin R, Bijma P (2004) Response to mass selection when the genotype by environment interaction is modelled as a linear reaction norm. **Genetics Selection Evolution** 36:435–454.

Lillehammer M, Árnyasi M, Lien S, Olsen HG, Sehested E, Ødegård J, Meuwissen THE (2007) A Genome Scan for Quantitative Trait Locus by Environment Interactions for Production Traits. **Journal of Dairy Science** 90:3482–3489.

Lillehammer M, Hayes BJ, Meuwissen THE, Goddard ME (2009) Gene by environment interactions for production traits in Australian dairy cattle. **Journal of Dairy Science** 92:4008–4017.

Lv FH, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, Joost S, Li MH, Marsan PA (2014) Adaptations to climate-mediated selective pressures in sheep. **Molecular Biology and Evolution** 31:3324–3343.

Malosetti M, Ribaut J-M, van Eeuwijk FA (2013) The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. **Frontiers in Physiology** 4:1–17.

Melo TP, Fortes MRS, Bresolin T, Mota LFM, Albuquerque LG, Carvalheiro R (2018) Multitrait meta-analysis identified genomic regions associated with sexual precocity in tropical beef cattle. **Journal of Animal Science** 96:4087–4099.

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

Mota RR, Silva FF, Lopes PS, Tempelman RJ, Sollero BP, Aguilar I, Cardoso FF (2017) Analyses of reaction norms reveal new chromosome regions associated with tick resistance in cattle. **Animal** 12:205-214.

Mulder HA, Bijma P (2005) Effects of genotype × environment interaction on genetic gain in breeding programs. **Journal of Animal Science** 83:49–61.

Mulder HA, Veerkamp RF, Ducro BJ, van Arendonk JAM, Bijma P (2006) Optimization of Dairy Cattle Breeding Programs for Different Environments with Genotype by Environment Interaction. **Journal of Dairy Science** 89:1740–1752

Mulder HA (2016) Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. **Frontiers in Genetics** 7:1–11.

Nepomuceno DD, Pires A V., Ferraz MVC, Biehl M V., Gonçalves JRS, Moreira EM, Day ML (2017) Effect of pre-partum dam supplementation, creep-feeding and post-weaning feedlot on age at puberty in Nellore heifers. **Livestock Science** 195:58–62.

Nirea KG, Meuwissen THE (2017) Improving production efficiency in the presence of genotype by environment interactions in pig genomic selection breeding programmes. **Journal of Animal Breeding and Genetics** 134:119–128.

Oliveira DP, Lourenco DAL, Tsuruta S, Misztal I, Santos DJA, de Araújo Neto FR, Aspilcueta-Borquis RR, Baldi F, Carvalheiro R, de Camargo GMF, Albuquerque LG and Tonhati H 2018. Reaction norm for yearling weight in beef cattle using single-step genomic evaluation. **Journal of animal science** 96, 27–34.

Oliveira MM, Santana ML and Cardoso FF 2016. Multiple-breed reaction norm animal model accounting for robustness and heteroskedastic in a Nelore–Angus crossed population. **animal** 10, 1093–1100.

Pegolo NT, Albuquerque LG, Lobo RB and de Oliveira HN 2011. Effects of sex and age on genotype x environment interaction for beef cattle body weight studied using reaction norm models. **Journal of Animal Science** 89, 3410–3425.

Pfau ML, Russo SJ (2015) Peripheral and central mechanisms of stress resilience. **Neurobiology of Stress** 1:66–79.

Pigliucci M (2005) Evolution of phenotypic plasticity: where are we going now? **Trends in Ecology & Evolution** 20:481–486.

Raidan FSS, Santos DCC, Moraes MM, Araújo AEM, Ventura HT, Bergmann JAG, Turra EM, Toral FLB (2016) Selection of performance-tested young bulls and

indirect responses in commercial beef cattle herds on pasture and in feedlots. **Genetics, selection, evolution: GSE** 48:1-11.

Rauw WM, Gomez-Raya L (2015) Genotype by environment interaction and breeding for robustness in livestock. **Frontiers in Genetics** 6:1–15.

Robinson JJ, Ashworth CJ, Rooke JA, Mitchell LM, McEvoy TG (2006) Nutrition and fertility in ruminant livestock. **Animal Feed Science and Technology** 126:259–276.

Robertson A (1959) The sampling variance of the genetic correla- tion coefficient. **Biometrics** 15:469–485.

Ruiz-Sánchez R, Blake RW, Castro-Gámez HMA, Sánchez F, Montaldo HH, Castillo-Juárez H (2007) Changes in the Association Between Milk Yield and Age at First Calving in Holstein Cows with Herd Environment Level for Milk Yield. **Journal of Dairy Science** 90:4830–4834.

Samadi F, Blache D, Martin GB, D'Occhio MJ (2014) Nutrition, metabolic profiles and puberty in Brahman (Bos indicus) beef heifers. **Animal Reproduction Science** 146:134–142.

Santana ML, Bignardi AB, Stefani G, El Faro L (2017) Genetic component of sensitivity to heat stress for nonreturn rate of Brazilian Holstein cattle. **Theriogenology** 98:101–107.

Santana Jr ML, Eler JP, Bignardi AB, Boligon AA, Ferraz JBS (2018a) Genetic correlation between growth and reproductive performance of beef females depends on environment. **Animal Production Science** 58:1201–1209.

Santana Jr ML, Eler JP, Oliveira GA, Bignardi AB, Pereira RJ, Ferraz JBS (2018b) Genetic variation in Nelore heifer pregnancy due to heat stress during the breeding season. **Livestock Science** 218:101–107.

Schaeffer LR (2004) Application of random regression models in animal breeding. **Livestock Production Science** 86:35–45.

Sgrò CM, Hoffmann AA (2004) Genetic correlations, tradeoffs and environmental variation. **Heredity** 93:241–248.

Silva FF, Mulder HA, Knol EF, Lopes MS, Guimaraes SEF, Lopes PS, Mathur PK, Viana JMS, Bastiaansen JWM (2014) Sire evaluation for total number born in pigs using a genomic reaction norms approach. **Journal of Animal Science** 92:3825–3834.

Simm G (1998) Genetic improvement of cattle and sheep. CABI Publishing p.446.

Streit M, Reinhardt F, Thaller G, Bennewitz J (2012). Reaction norms and genotype-by-environment interaction in the German Holstein dairy cattle. **Journal of Animinal Breeding and Genetics** 129:380–389.

Su G, Madsen P, Lund MS, Sorensen D, Korsgaard IR, Jensen J (2006) Bayesian analysis of the linear reaction norm model with unknown covariates. **Journal of Animal Science** 84:1651–1657.

Tsuruta S, Lourenco DAL, Misztal I, Lawlor TJ (2015) Genotype by environment interactions on culling rates and 305-day milk yield of Holstein cows in 3 US regions. Journal of dairy science 98:5796–805.

van Gestel J, Weissing FJ (2016) Regulatory mechanisms link phenotypic plasticity to evolvability. **Scientific reports** 6:24524.

Chapter 2 – Genomic reaction norm models exploiting genotype x environment interaction effects on sexual precocity indicator traits in Nellore cattle

Abstract – Brazilian beef cattle are raised, predominantly, on pasture conditions in a wide range of environments. In this scenario, genotype by environment (GxE) interactions are an important source of phenotypic variation in the reproductive traits. Hence, the evaluation of GxE interaction for heifer's early pregnancy (HP) and scrotal circumference (SC) traits in Nellore cattle, belonging to three breeding programs, was carried out to unravel the animal sensitivity to the environmental condition (EC). The dataset consisted of 85,874 records for HP and 151,553 records for SC, from which 1,800 heifers and 3,343 young bulls were genotyped with the BovineHD BeadChip. Genotypic information of 826 sires was also used in the analyses. Reaction norm models, using pedigree information (RNM_A) or pedigree and genomic information (RNM_H), were used to infer upon GxE interactions. Two validation schemes were used to assess the predictive ability using as training populations: a) younger scheme and b) environment-specific scheme: low EC (-3.0 and -1.5) and high EC (1.5 and 3.0). The inclusion of the H matrix in RNM increased the genetic variance of intercept and slope in 18.55% and 23.00% on average, respectively, and provided genetic parameter estimates more accurate than the analysis considering pedigree only. The same trend was observed for heritability estimates, which were of 0.28 to 0.56 for SC and of 0.26 to 0.49 for HP, using RNM_H, and of 0.26 to 0.52 for SC and of 0.22 to 0.45 for HP, using RNM_A. The lowest correlation observed between unfavorable (-3.0) and favorable (3.0) EC levels were of 0.30 for EP and -0.12 for SC indicating the presence of GxE interaction. The GxE interaction effect implied in differences on animals' genetic merit and re-ranking of animals on different environmental conditions. Strong SNP marker-environment interaction was detected for Nellore sexual precocity indicator traits with changes in effect and variance across EC levels. The RNM_H captured GxE interaction effects better than RNM_A, and improved the predictive ability around 14.04% for SC and 21.31% for HP. The younger scheme increased the overall predictive ability for SC and HP compared to environment-specific scheme. The results suggest that inclusion of genomic information combined with pedigree to assess the GxE interaction, lead to variance components and genetic parameter estimates more accurate.

Key-words: accuracy; beef cattle breeding; genomic prediction; random regression

2.1 Introduction

In beef cattle breeding programs, sexual precocity indicator traits have been used as selection criteria, aiming to increase the herd efficiency. Usually, traits as heifer early pregnancy (EP) and scrotal circumference (SC) are used to identify and select animals sexually precocious. However, sexual precocity in cattle is affected by many environmental factors, e.g., heat stress and nutrition, which play a major role on the onset of puberty (Robinson et al., 2006). Differences in the environments have been pointed out as a main issue for sexual precocity variation in Brazilian herds (Nepomuceno et al., 2017; Ferraz Jr et al., 2018). The Brazilian beef cattle are raised predominantly on pasture with a wide range of environmental conditions with seasonal variations in forage production and different nutritional levels. In this scenario, genotype by environment (GxE) interactions could be an important source of phenotypic variation (Santana et al., 2015), mainly to sexual precocity traits in Nellore cattle (Ashworth et al., 2009; Chiaia et al., 2015). Thus, the genotype-environment interactions (GxE) could occur, causing changes on animal's performance under different environments, affecting the genetic variance and/or the rank of the animals (Burrow, 2012).

Differential animal responses across environmental conditions have been assessed through reaction norm models (RNM), describing the animal sensitivity to environmental changes (Kolmodin et al., 2003; Calus and Veerkamp, 2003). Genomic information has been used to assess the GxE interaction representing a promising alternative to evaluate traits strongly influenced by environmental factors (Lillehammer et al., 2008, 2009; Silva et al., 2014; Mota et al., 2016; Zhang et al., 2019). According to Hayes et al. (2016), the inclusion of genomic information in RNM has the potential to overcome the low number of phenotypic information from sires with offspring in a wide range of environments, given that the genotyped animals are well distributed across the environments.

The inclusion of genomic information has improved the estimation of GxE effects and increasing the accuracy of predictions when compared to the pedigree relationship (Silva et al., 2014). Combining the information of genotyped and non-genotyped animals in RNM may be important to aid the development of breeding programs in different environmental conditions. Thus, this study was carried out to verify the occurrence of GxE interaction in Nellore sexual precocity indicator traits by RNM using pedigree and combining pedigree and genomic relationship aiming to

increase accuracy of prediction applying different scheme of training and validation. In addition, the effect of GxE interaction on SNP markers in different environmental conditions was also evaluated.

2.2 Material and methods

2.2.1 Phenotypic and Genotypic Data

Phenotypic data of Nellore animals were provided by three breeding programs: DeltaGen, Paint and Cia de Melhoramento. Animals were born between 1984 and 2014, in 220 farms distributed in three regions of Brazil (Midwest, Southeast and Northeast). The sexual precocity indicator traits evaluated were, heifer early pregnancy (HP) and young bull's scrotal circumference (SC). The HP was defined attributing a value of 1 (success) to heifers calving until 31 months of age and a value of 0 (failure) to heifers calving after 31 months of age. The SC (cm) was measured in young bulls at yearling with 18 ± 1.5 months of age. In some herds, heifers with age of 16 to 18 months were exposed to reproduction for 60 days in an anticipated breeding season from February to April in order to identify the sexually precocious heifers. To those heifers not conceiving in this anticipate season, a second chance was given, in the normal breeding season, when they were about 24 month of age.

The contemporary groups (CG) for HP and SC were formed combining information of year and season of birth, farm (at birth, weaning and yearling) and management group (at birth, weaning and yearling). The CG for HP in which animals showed the same response category (0 or 1), i.e. without variability, were excluded from the analysis. Phenotypic information for SC outside the interval between ± 3.5 standard deviations of the CG average were excluded from the dataset. Only CG with more than five records were kept on the final dataset. The percentage of animals with unknown sire in the phenotypic file were of 30.86% for HP and 26.58% for SC. The structure of the phenotypic data is shown in Table 1.

Table 1. Structure of data set for heifer early pregnancy (HP) and scrotal circumference (SC).

Trait	Nº of observations	Nº of Sires	Nº of dams	Nº of CG	Mean (SD)
HP (%)	85,874	1,950	69,626	3,386	26.45 (0.66)*
SC (cm)	151,553	4,051	110,270	10,495	29.64 (2.34)

HP – Heifer early pregnancy; SC – scrotal circumference; CG – contemporary groups; SD - Standard deviation, * Represents percentage of success for HP.

A total of 1,900 heifers, 1,500 young bulls and 850 sires were genotyped using the Illumina BovineHD BeadChip assay (770k, Illumina Inc., San Diego, CA, USA) and 1,950 young bulls using GeneSeek® Genomic Profiler HDi 75K (GeneSeek In/c., Lincoln, NE). The young bulls genotyped with panel 75K were imputed to the HD panel using FImpute version 2.2 (Sargolzaei et al., 2014). Autosomal markers presenting minor allele frequency (MAF) lower than 0.05, p-value for Hardy–Weinberg equilibrium test less than 10^{-5}, and a call rate for markers and samples less than 0.95 were removed. After genotype quality control, 1,800 heifers, 3,343 young bulls, 826 sires, and 412,528 SNPs remained in the dataset.

2.2.2 Statistical Modeling

2.2.2.1 Animal model

The traditional linear animal model (BLUP), with pedigree information (A matrix), and the single-step GBLUP (ssGBLUP) combining pedigree and genomic information (H matrix), were used for SC. For HP, nonlinear animal models (threshold), using A or H matrix were also applied. In matrix notation, the equation model for each trait was:

$$y=X\beta+Z\alpha+e,$$

where, y is the vector of phenotypic observations (SC and HP); β is a vector of fixed effects (CG for HP and SC) and age at recording as linear co-variable for SC; α is the vector of additive genetic effects; X and Z are the incidence matrices relating y to the effects β and a, respectively, and e is a vector of random residual effects. The HP was analyzed using the threshold model, assuming an underlying distribution as follows: $f(HP|l_i) = \prod_{i=1}^{n_i} 1(l_i < t_i)1(HP = 0) + 1(l_i > t_i)1(HP = 1)$, where HP is the binary trait (0 or 1), l_i is the underlying liability of observation i, t_i is the threshold that

defines the category response for HP and n_i is the number of information (n = 85,874).

The additive genetic and residual effects were considered normally distributed: a={a_i}~$N(0, A\sigma_a^2)$ or a={a_i}~$N(0,H\sigma_a^2)$ for BLUP and ssGBLUP, respectively, and $e = \{e_{ij}\}$~$N(0,I\sigma_{eij}^2)$, where σ_a^2 and σ_{eij}^2 are the variance components for the additive genetic and residual effects, respectively; *A* represents the relationship matrix based on pedigree information, *I* is an identity matrix and *H* is a matrix combining pedigree and genomic relationship. The inverse, H^{-1} , is calculated as follows (Aguilar et al., 2010):

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

where A^{-1} is the inverse of relationship matrix based on pedigree information, A_{22}^{-1} represents the inverse of pedigree-based relationship matrix for genotyped animals and, G^{-1} is the inverse of the genomic relationship matrix obtained according to vanRaden (2008).

2.2.2.2 Reaction Norm

A reaction norm model in two-steps was considered in the present study (Calus and Veerkamp, 2003; Silva et al., 2014). In this study, the yearling body weight (YBW) was used to define the production level, given that the YBW has a great impact on sexual precocity in young bulls (Barth et al., 2008) and heifers (Nepomuceno et al., 2017).

In the first step, the environmental condition (EC) to describe the production level, were based on the CG solutions for YBW. The CG was defined by animals born in the same year and season of birth, from the same farm (from birth to yearling) and management group (from birth to yearling). These solutions were obtained with an animal model using the ssGBLUP, separately for males and females as defined above. The CG solutions were standardized to a mean of 0 and variance of 1 and with values ranging from -3 to +3 standard deviations. The phenotypic trend and number of information across EC levels are shown on Supplementary information S1 Figure.

In the second step, the single-trait reaction norm model was:

$$y_{ji} = X\beta + \varphi_f \boldsymbol{\Phi}_f(EC_i) + \alpha_j \boldsymbol{\Phi}_f(EC_i) + e_{ji},$$

where, y_{ji} is the vector for the phenotypic records (SC and HP) of animal *j* recorded in the level *i* of EG; β is the vector with the fixed effects including CG and for SC the linear effect of age at recording and X is the incidence matrix for fixed effects, φ_f are the fixed regression coefficients on Φ_f; Φ_f is the Legendre polynomial for each *EC* level *i* (EC_i); α_{jf} are the random regression coefficients of additive effects for each animal *j* on the EC_i, and e_{ji} is the random residual. The HP was analyzed assuming a threshold-reaction norm model assuming an underlying distribution as defined above.

Five classes were used to model residual variances as follows: class 1= EC level lower than -1.5; class 2 = -1.5 ≤ EC level ≤ -0.5; class 3 = -0.5 < EC level ≤ 0.5; class 4 = 0.5 < EC level ≤ 1.5 and, class 5 = EC level higher than 1.5. The additive genetic and residual effects were considered normally distributed: $\mathrm{a} = \{\mathrm{a_j}\}$~N(0, A$K_{ab}$) or $a = \{a_j\}$~$N(0, HK_{ab})$ and $\mathrm{e} = \{\mathrm{e_{ij}}\}$~N(0, I$R$), where K_{ab}= $\begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}$ is the (co)variance matrix for the reaction norm coefficients - intercept (a) and the slope (b) - considering A matrix or H matrix; R is the diagonal residual variance matrix considering 5 classes of variances defined above.

Genetic variance ($\hat{\sigma}^2_{aEC_i}$) and heritability ($\hat{h}^2_{EC_i}$) estimates for the traits in each EC level were calculated by the following equations: $\hat{\sigma}^2_{aEG_i}$=$\Phi_f K_{ab} \Phi_f'$; where Φ_f is the Legendre polynomial estimate corresponding to the EC levels. Thus, the heritability ($\hat{h}^2_{EC_i}$) for each EC level was determined as follows: $\hat{h}^2_{EC_i} = \frac{\hat{\sigma}^2_{aEC_i}}{\hat{\sigma}^2_{aEC_i} + \hat{\sigma}^2_{eEC_i}}$; $\hat{\sigma}^2_{aEG_i}$ and $\hat{\sigma}^2_{eEG_i}$ are genetic and residual variances, respectively, in environment *i*. The genetic correlation across EC levels ($r_{ECi,ECin}$) was determined as follows: $r_{ECi,ECin} = \sigma_{ECi,ECin} / \sqrt{\hat{\sigma}^2_{aEG_i} * \hat{\sigma}^2_{aEG_{in}}}$, where $\sigma_{ECi,ECin}$ is the covariance between EC level *i* and EC level *in*, estimated the same way as the genetic variance in each EC level.

The reaction norm model described above was fitted: 1) considering the relationship among individuals based on the pedigree (RNM_A) and 2) genomic information matrix combined with pedigree relationship matrix (RNM_H).

2.2.3 Model inference and comparison

Samples of the posterior distributions of the genetic parameters were obtained using a Bayesian approach and Gibbs sampler, applying GIBBS3f90 program for SC and the THRGIBBS3f90 program for HP (Misztal et al., 2015). Bayesian analysis consisted of a single chain of 1,000,000 cycles, considering a burn-in of 100,000 iterations with samples stored at every 10 cycles. Thus, 90,000 samples were used to obtain posterior parameter estimates. The convergence was evaluated through visual inspection and using the Bayesian Output Analysis (Smith, 2007) and Geweke test (Geweke, 1992).

The models were compared by the deviance information criterion (DIC) (Spiegelhalter et al., 2002):

$$DIC=\bar{D}(\theta)+p_D=2\bar{D}(\theta)-D(\bar{\theta})$$

in which $\bar{D}(\theta)=E_{\theta|y}[D(\theta)]$ is the posterior expectation of Bayesian deviance; and $D(\theta) = -2ln_p(y|\theta)$. The effective number of parameters represents a penalty from the increase in model complexity: $p_D=\bar{D}(\theta)-D(\bar{\theta})$ where θ represents the model parameter vector; and $D(\bar{\theta})$ is the Bayesian deviance obtained as the posterior mean of the parameters.

2.2.4 Breeding values and single polymorphism effects across EC levels

The estimated breeding values (EBV) and the genomic breeding value (GEBV) associated with each EC level were obtained using the following equation: $\hat{g}_{jEC_i}=\alpha_j \Phi_f'$; where α_j is the estimated additive genetic values for intercept and slope estimates of animal *j* and Φ_f' is the transpose vector of the Legendre polynomials estimates for the *EC* level *i*.

The SNP marker effects were estimated in different EC levels using linear prediction equation (Silva et al., 2014) as follows: $\hat{u}_{kEC_i}=\alpha_k \Phi_f'$; where α_k is the estimated effect for intercept and slope for each SNP marker *k*; Φ_f' is the vector of the Legendre polynomials estimates for the *EC* level *i*; which allows the estimation of a vector of marker effects across EC levels. To assess the SNP effect pattern across EC level, a total of 30 SNP markers that presented the greatest absolute effect in medium EC level (0.0), were selected. The percentage of genetic variance explained by each SNP marker (σ^2_{SNPECi}) in each EC levels was estimated as follows: $\sigma^2_{SNPECi}=$

$\frac{2p_iq_i\hat{u}^2_{kECi}}{\sigma^2_{aECi}} * 100$, where $\hat{u}^2_{kECi}$ is the square of the SNP marker effect in the EC level *i*, p_i and q_i are the allele frequencies, and $\hat{\sigma}^2_{aEG_i}$ is the additive genetic variance in the EC level *i*.

2.2.5 Predictive ability across EC levels

The predictive ability was assessed partitioning the data set in two scenarios as follows:

Scenario 1 (*younger scheme*): Training and validation data were subsets from the population based on year of birth, in such a way that the validation set was composed by animals born from 2008 to 2012 for HP and from 2011 to 2014 for SC.

Scenario 2 (*environment-specific scheme*): Training and validation were subsets from animals classified according to the environmental levels: a) animals in the lowest EC levels (-3.0 to -1.0) were used as training set and those in EC levels of -0.5 to 3.0 as validation; (b) animals in the highest EC levels, of 1.0 to 3.0, for training and those in EC levels of -3.0 to 0.5 for validation. This scenario will show how well predictions would perform across-environments.

In order to verify the distribution of the genotyped animals in the training and validation sets, a principal component analysis was used (Figure 1). The predictive ability was computed as the Pearson correlation between the breeding values obtained using A or H matrices and the phenotypes adjusted for fixed effects (y^*), as follow: $\hat{r}_i = cor(y^*, (G)EBV_{EGi})$, where y^* ($y^* = y - (X\hat{\beta})$) is the vector of the adjusted phenotypes for fixed effects ($X\beta$) and $(G)EBV_{EGi}$ is the vector of the genomic breeding values (GEBV) for genotyped animals and breeding value (EBV) in each EC level.

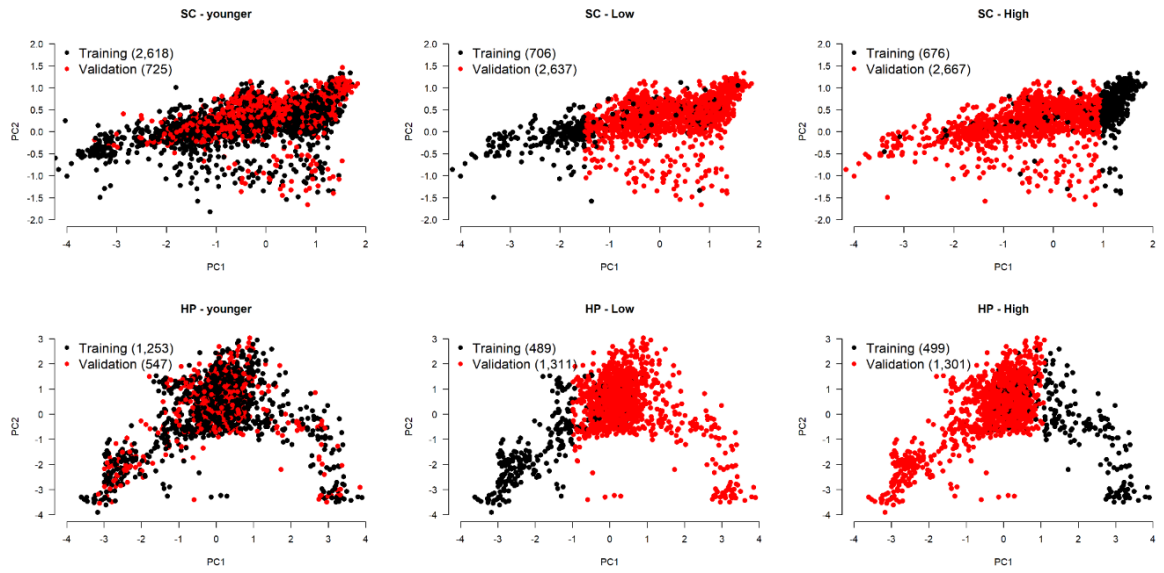


Figure 1. Distribution of animals in training (black) and validation (red) groups in each validation scenario, obtained by principal component analysis based on the genomic matrix. Younger validation - partition into training and validation sets was based on year of birth; Environment-specific: Low - dataset was partitioned based in animals classified in poor environmental condition (of -3.0 to -1.0) for training set and those environmental conditions (of -0.5 to 3.0) for validation; High - dataset was partitioned based in animals in better environmental condition (of 1.0 to 3.0) for training and those environmental condition (of -3.0 to 0.5) for validation.

2.3 Results and Discussion

2.3.1 Comparison of models

The results for DIC criterion pointed out that reaction norm models (RNM_A and RNM_H) presented a better fit than BLUP models (BLUP and ssGBLUP). Using RNM models instead of BLUP decreased DIC value in 5.71% for SC and 43.55% for HP (Table 2). Hence, models taking into account the GxE interaction showed a better adjustment of the data compared to the conventional models, as already showed by Mulder (2016).

Table 2. Deviance information (DIC) criterion for models considering pedigree information (RNM_A and BLUP) and genomic information matrix combined with pedigree (RNM_H and ssGBLUP).

Trait	Models			
	RNM_A	RNM_H	BLUP	ssGBLUP
SC	5,860	5,501	6,139	5,907
HP	2,189	1,949	3,998	3,352

SC – scrotal; HP - heifer early pregnancy; RNM_A – reaction norm model considering pedigree information; RNM_H – single-step reaction norm model; BLUP – animal model; ssGBLUP - single-step GBLUP

Including genomic information in ssGBLUP and RNM_H reduced DIC, on average, in 9.97 and 8.60%, respectively, compared to BLUP and RNM_A (Table 2). These results agree with Mota et al. (2016) that evaluated the GxE interaction for tick counts. The authors reported a reduction of 15.8% in DIC when using reaction norm with H matrix. Therefore, GxE interactions should be included in genetic evaluation models and the inclusion of genomic information in reaction norm models allows capturing genetic relationships between animals more accurately compared to evaluations considering only pedigree (Yu et al., 2017).

2.3.2 (Co) Variance Components

The inclusion of genomic information in the RNM increased the genetic variance of intercept and slope in 18.55% and 23.00%, respectively, on average (Table 3). The increase in genetic variance estimates by using genomic information has been previously described (Silva et al., 2014). In addition to resulting higher variance components estimates, the inclusion of H matrix in reaction norm models presented lower standard deviations (SD), i.e. more accurate predictions were achieved. These results are in agreement with previous studies in pigs (Silva et al., 2014) and cattle (Mota et al., 2016). According to Forni et al. (2011) and Yu et al. (2017), using the H matrix instead of A matrix for genetic evaluation will reduce the error of prediction due to a better estimation of kinship coefficient.

Table 3. Variance component estimates and standard deviation (values in parentheses) for parameters of reaction norm model using pedigree (RNM_A) and genomic information matrix combined with pedigree matrix (RNM_H), for Nellore sexual precocity indicator traits, scrotal circumference (SC) and heifer early pregnancy (HP).

Variance component	SC		HP	
	RNM_A	RNM_H	RNM_A	RNM_H
Intercept (σ^2_a)	2.56 (0.131)	2.87 (0.082)	0.12 (0.076)	0.15 (0.041)
Slope (σ^2_b)	1.42 (0.114)	1.60 (0.092)	0.09 (0.049)	0.12 (0.023)
Covariance (σ_{ab})	1.20 (0.127)	1.25 (0.052)	0.08 (0.054)	0.09 (0.038)
Slope / Intercept ratio	0.55 (0.090)	0.56 (0.072)	0.75 (0.082)	0.80 (0.066)
Correlation (r_{ab})	0.63 (0.104)	0.54 (0.086)	0.79 (0.124)	0.62 (0.075)
Residual class 1 (σ^2_{e1})	1.91 (0.110)	1.87 (0.070)	0.09 (0.027)	0.09 (0.012)
Residual class 2 (σ^2_{e2})	2.15 (0.082)	2.10 (0.051)	0.11 (0.014)	0.10 (0.013)
Residual class 3 (σ^2_{e3})	2.38 (0.058)	2.35 (0.033)	0.13 (0.043)	0.12 (0.026)
Residual class 4 (σ^2_{e4})	2.64 (0.071)	2.58 (0.059)	0.13 (0.052)	0.12 (0.031)
Residual class 5 (σ^2_{e5})	3.01 (0.095)	2.98 (0.046)	0.12 (0.020)	0.11 (0.015)

Residual class 1: EC levels lower than -1.5; Residual class 2: -1.5 ≤ EC level ≤ -0.5; Residual class 3: -0.5 ≤ EC level ≤ 0.5; Residual class 4: 0.5 < EC level ≤ 1.5; Residual class 5: EC level from > 1.5.

For both traits, the GxE interaction effects, given by the slope/intercept ratios, were higher when using RNM_H instead of RNM_A (Table 3). In this work, the genetic correlation estimates between intercept and slope were higher when using RNM_A model than RNM_H model, reinforcing the improvement in GxE estimation by using RNM_H for HP and SC (Table 3). However, the genetic correlation between intercept and slope were positive and of high magnitude, for both traits, indicating that animals with higher intercept (average breeding value) respond more when the environmental conditions become less restrictive. Some studies indicated that a high correlation between intercept and slope suggests the occurrence of animals re-ranking when evaluated in different environments (Su et al., 2006; Oliveira et al., 2016; Zhang et al., 2019). Oliveira et al. (2016) observed that this occurs by scale effect in the EBV. The correlation estimate between intercept and slope was similar to that observed by Santana Jr et al. (2018) in Nellore cattle for HP (0.65) and higher than those reported by Santana et al. (2013) and Santana et al. (2015) for SC of 0.001 and 0.14, respectively.

The residual variance estimates decreased around 1.10 to 10.50% using RNM_H instead of RNM_A (Table 3). The same trend was reported using H matrix in

RNM for body weight and visual scores in Nellore (Gordo et al., 2016; Oliveira et al., 2016) and for total number born in pigs using G matrix (Silva et al., 2014). In the present study, the addition of genomic information in RNM (RNM_H) also resulted in a decrease in the genetic parameter estimations standard deviations (SD) 35.42% for SC and 37.31% for HP, on average (Table 3). Our results agree with those from Forni et al. (2011), Gordo et al. (2016) and Mota et al. (2016) who reported reductions on variance components SD of 31.82%, 31.80% and 22.63%, respectively, when combining genomic and pedigree relationship.

2.3.3 Genetic Parameters

Heritability estimates obtained with RNM_H model were from 0.28 to 0.56 for SC and from 0.26 to 0.49 for HP (Figure 2). The heritability estimates obtained with the RNM_H were around 8.14 and 12.30% higher for SC and HP, respectively, regarding those from RNM_A. The inclusion of genomic information may improve the genetic connectedness across herds, through the better genetic relationship between animals, which will affect the heritability estimates (Bérénos et al., 2014).

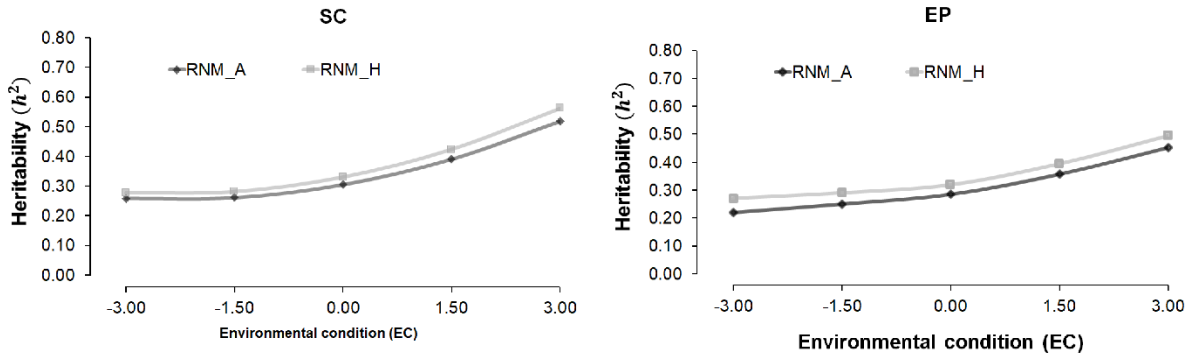


Figure 2. Heritability estimates for scrotal circumference (SC) and heifer early pregnancy (HP) in Nellore cattle across different environmental conditions by the RNM_A (reaction norm model considering pedigree information) and RNM_H (reaction norm considering the combination of pedigree and genomic information).

The heritability estimates for SC and HP traits, from both RN models showed the same trend, increasing with the environmental improvement, i.e. under less restrictive EC levels the genetic variance estimates increased. This tendency was previously described for SC (Santana et al., 2013; Chiaia et al., 2015) and HP (Santana Jr et al., 2018) in Nellore. The difference in heritability estimates in different EC levels occurs as consequence of the GxE interaction (Fordyce, 2006).

Using reaction norm models, the heritability estimates for SC and HP, in the intermediate EC level (0.0), were similar to those obtained with models disregarding GxE. These estimates using RNM_A and RNM_H were, respectively, 0.31 and 0.33 for SC, and 0.29 and 0.32 for HP, while with BLUP and ssGBLUP the values were 0.31 and 0.34 for SC, and 0.30 and 0.33 for HP (data not shown). The heritability estimates were lowest of those described in the literature for SC (0.51 to 0.67) (Santana et al., 2013; Chiaia et al., 2015) and HP (0.40 to 0.76) (Santana Jr et al., 2018) in Nellore cattle considering GxE interaction. These differences in heritability estimates are directly associated with population and mainly by the environmental conditions used to define the production level used in RNM.

Genetic correlation estimates for each trait under different EC levels ranged from -0.12 to 0.98 for SC, and from 0.30 to 0.99 for HP indicating the presence of GxE (Figure 3). For SC the genetic correlations estimated using H or A matrix were similar (Figure 3), as also found by Mota et al. (2016) and Oliveira et al. (2018) using genomic information in reaction norm for beef cattle. However, for HP, larger differences were observed in genetic correlation estimates when using the H matrix, with values ranging from 0.30 to 0.98, instead of 0.60 to 0.98 with A matrix. The effect of including genomic information seems even more important for HP trait.

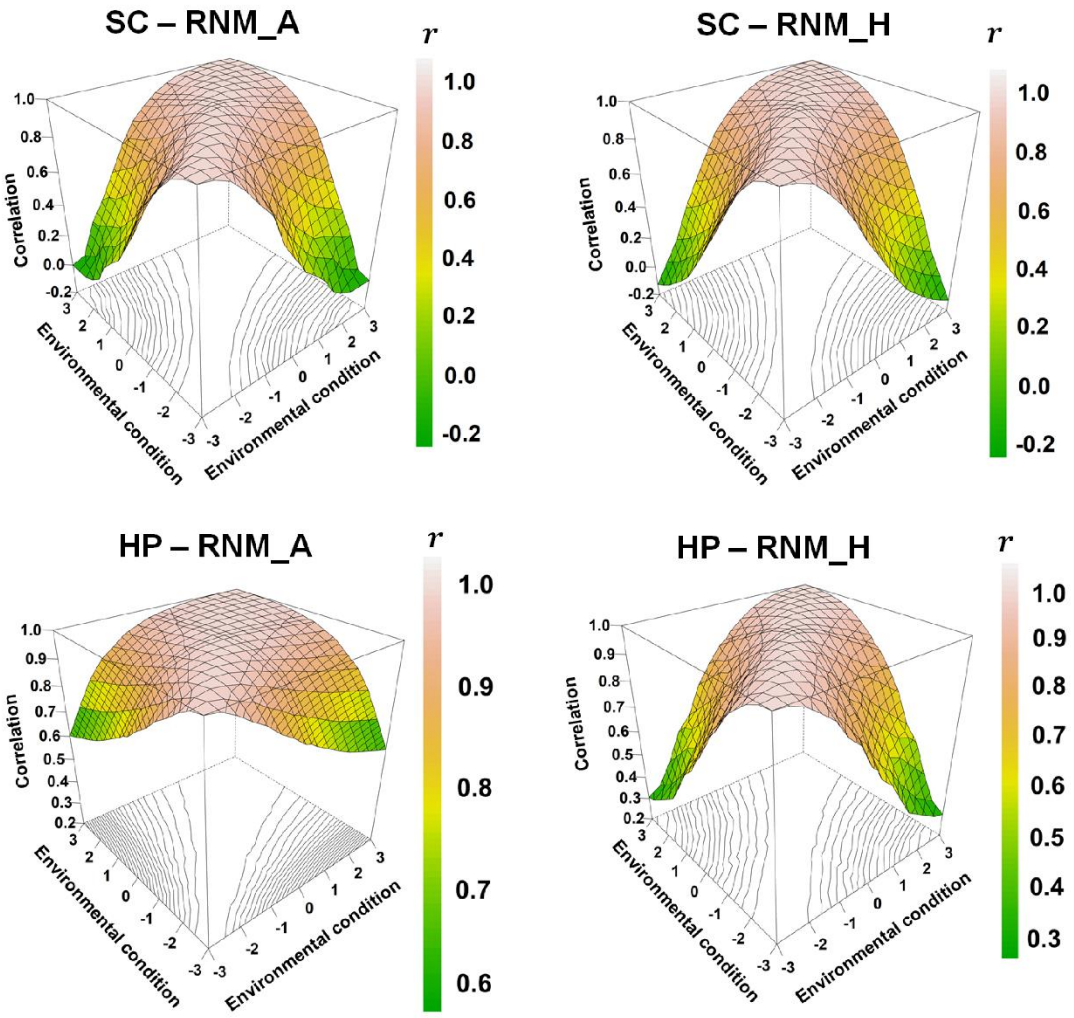


Figure 3. Genetic correlation estimates for scrotal circumference (SC) and for heifer early pregnancy (HP) across different environmental condition considering A (pedigree information) and H (genomic and pedigree information) relationship matrices.

The genetic correlation estimates between the extremes EC levels, using the H matrix, were of low magnitude and negative for SC (-0.12), and positive for HP (0.30; Figure 3). Therefore, in this case, a stronger effect of GxE interaction can be expected, leading to an important re-ranking of animals, based on EBV, across EC levels. The genetic correlations have been used to assess the degree of animal sensitivity to environmental changes (Robertson, 1959). These estimates are in agreement with other results reported by Santana et al. (2015), Chiaia et al. (2015) and Santana Jr et al. (2018), in which the magnitude of genetic correlation estimates decreased when differences between EC levels increased, indicating occurrence of GxE interaction with re-ranking of EBVs across EC levels.

2.3.4 GxE Interactions

Reaction norm for 30 animals with higher GEBV in medium EC level (0.0) showed large re-ranking of animals (Figure 4). The effect of GxE interaction on animal sensitivity to changes in EC levels was expected as indicated by the genetic correlation estimates across EC levels lower than 0.80 and moderate between intercept and slope. The positive correlation between intercept and slope (Table 3) indicates that animals with a higher intercept estimate generally have a higher slope, increasing their performance when EC levels become less restrictive (Figure 4).

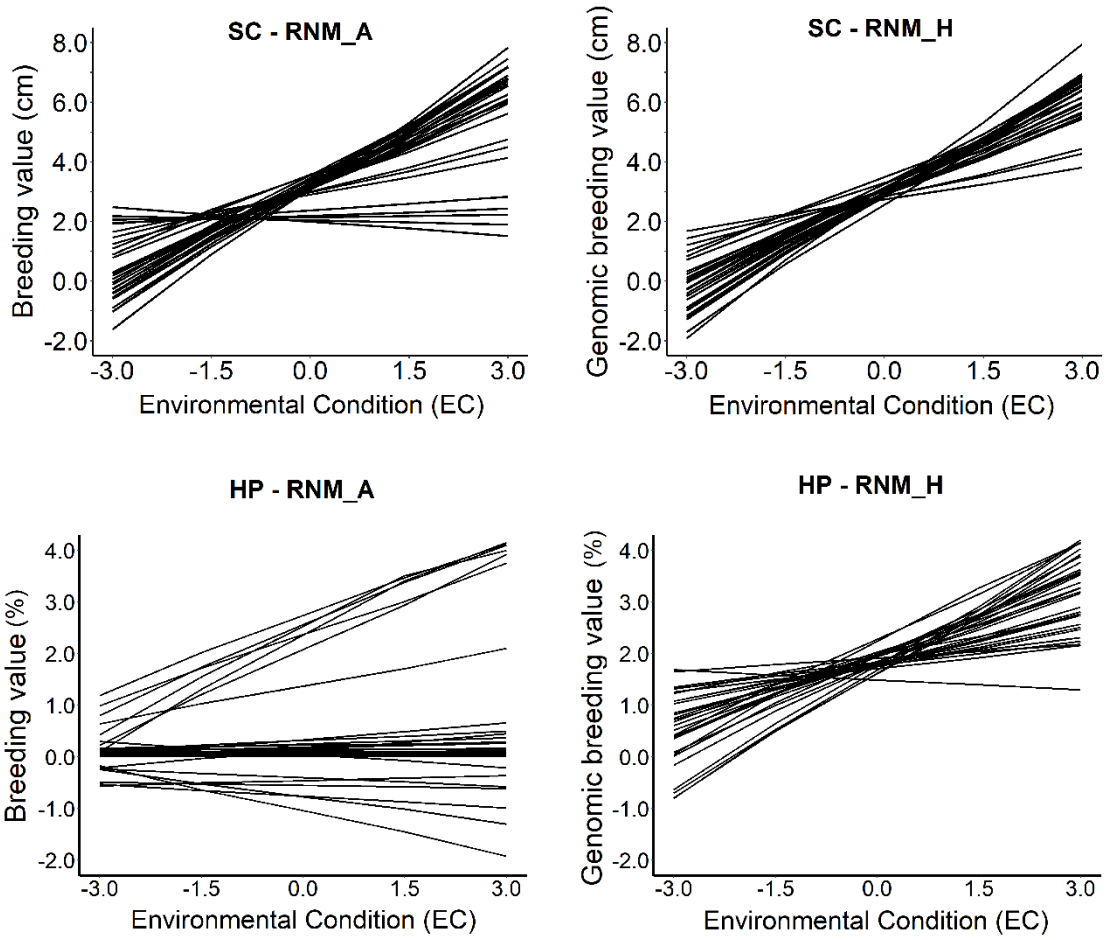


Figure 4. Reaction norms for scrotal circumference (SC) and heifer early pregnancy (HP) considering RNM_A (left – breeding value) and RNM_H (right – genomic breeding value) relationship matrices, considering the 30 animals with highest genomic breeding values in the medium environmental condition (EC = 0.0), predicted with RNM_H.

The breeding values for SC across EC levels, obtained with both models (RNM_A and RNM_H), presented similar trends (Figure 4). However, for HP, large differences between the two models were observed, as already indicated by the genetic correlation estimates (Figure 3). These differences could be due to the fact that heifers show a lower connectedness among EC levels compared to young bulls. Including genomic information in the model not only improves animal's relationship but also provides a better genetic connectedness driven by the degree of genomic links across EC levels. According to Hayes et al. (2016), with genomic information, the estimation of GxE interaction is improved since it has the potential to reduce bias when genotyped animals are well sampled across the environments.

The animal sensitivity to environmental changes affected the GEBV distribution for both traits, leading to an increase on their average and distribution when the EC levels became less restrictive (Supplementary Information: S2 Figure). Under low environmental conditions (-3.0), the GEBV average for SC and HP were, respectively, 0.97 cm and 0.25%, and, for high EC level (3.0), these values were 3.23 cm and 1.07%. According to Fordyce (2006), animal plasticity has an important effect on the mean and variance of the trait in different environmental conditions. Notably, the sexual precocity indicator traits (SC and HP) exhibited significant changes in the mean and genetic variation according to EC levels and animal selection should be carefully assessed. The findings of the present study showed similar trend to those reported in studies for reproductive traits (Santana Jr et al., 2018; Araujo Neto et al., 2018), indicating that GxE interaction affects significantly the animal re-ranking across environmental conditions.

Remarkably differences in (G)EBV and genetic correlation estimates, for Nellore sexual precocity indicator traits, were observed across EC levels (Figure 4) and indicate these traits are strongly influenced by GxE interaction. When GxE interaction occurs, the differences in (G)EBV of a trait in different environments would be determined by different genomic region. In the present study, the SNP with the greatest effect for Nellore sexual precocity indicator traits showed different contribution across EC levels, affecting the animal sensitivity (Figure 5). It was noticed that the SNP effects are environmentally dependent (Figure 5 A and B).

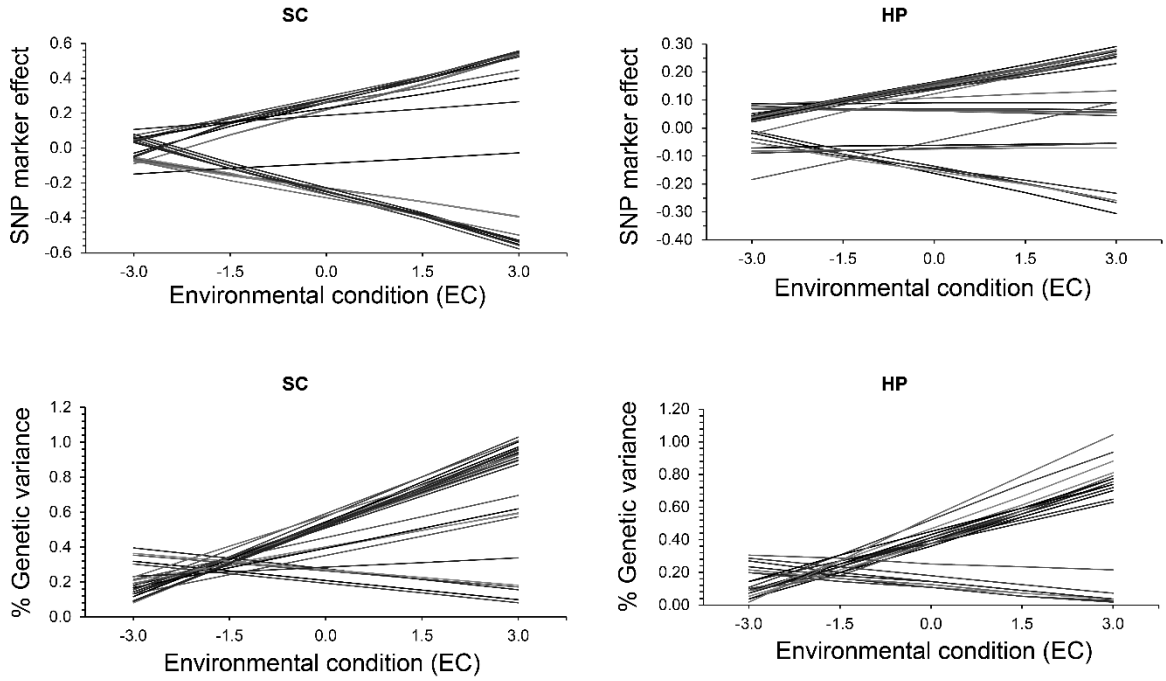


Figure 5. SNP marker effects reaction norms and percentage of the genetic variance explained by each SNP marker, for scrotal circumference (SC) and heifer early pregnancy (HP), considering the 30 SNP with the greatest absolute effects in medium environmental condition (EC = 0.0).

The environmentally dependent SNP markers showed a great contribution for genetic variance explained in some environments, although with only a small percentage of genetic variance explained in other environments (Figure 5 C and D). This strong SNPxE interaction indicates that genomic regions show a striking effect for Nellore sexual precocity indicator at a certain EC level and not in others, with changes not only in magnitude but also in direction (Figure 5). A higher dispersion of SNP marker effects was observed when EC level became less restrictive, and the highest percentage of the genetic variance explained by a SNP was 1.029% for SC and 1.04% for HP. Mota et al. (2017) observed the same trend using RNM, i.e. the range and variance of the SNP effects increased with the EC level. Lillehammer et al. (2008), using a random regression model for QTL identification, reported that the largest variation on SNP effects occurred when environmental level became less restrictive. These authors observed re-ranking of markers across different environmental conditions highlighting that genomic regions with major effect in a determinate EC level do not show the same significance in another EC level. Studies evaluating GxE interaction effects on dairy cattle (Lillehammer et al., 2009; Hayes et al., 2009), beef cattle (Mota et al., 2017) and swine (Silva et al., 2014), revealed that

genomic region effects were environment-dependent, causing an important GxE interaction effect on phenotypic expression of traits. A possible explanation for this is that animals raised under less restrictive EC levels are able to express their genetic potential for sexual precocity. van Gestel and Weissing (2016) reported that animal adaptation and plasticity are controlled by genomic regions involved on major physiological process.

2.3.5 Assessment of predictive ability across EC levels

For both RNM_A and RNM_H models in *younger scheme*, the predictive ability, assessed by Pearson correlations between phenotype corrected (y*) and (G)EBV, increased with the EC levels, being higher in the most favorable environment (EC level = 3.0) (Figure 6). The low predictive abilities of (G)EBV for SC and EP, under restrictive EC levels, i.e. EC levels from -3.0 to -0.5, are perhaps in part due to their lower heritability estimates, compared to less restrictive environmental conditions (Figure 2). Moser et al. (2010) reported that there was a strong association between the predictive ability of genomic prediction and the heritability of the trait. Thus, a trait with a low heritability requires larger number of records to achieve a predictive ability similar to that obtained for traits with high heritability. In this context, Silva et al. (2014), evaluating number of pigs born, observed that the highest predictive ability occurred at intermediate levels, which showed the lowest heritability estimate, but the highest number of individuals evaluated.

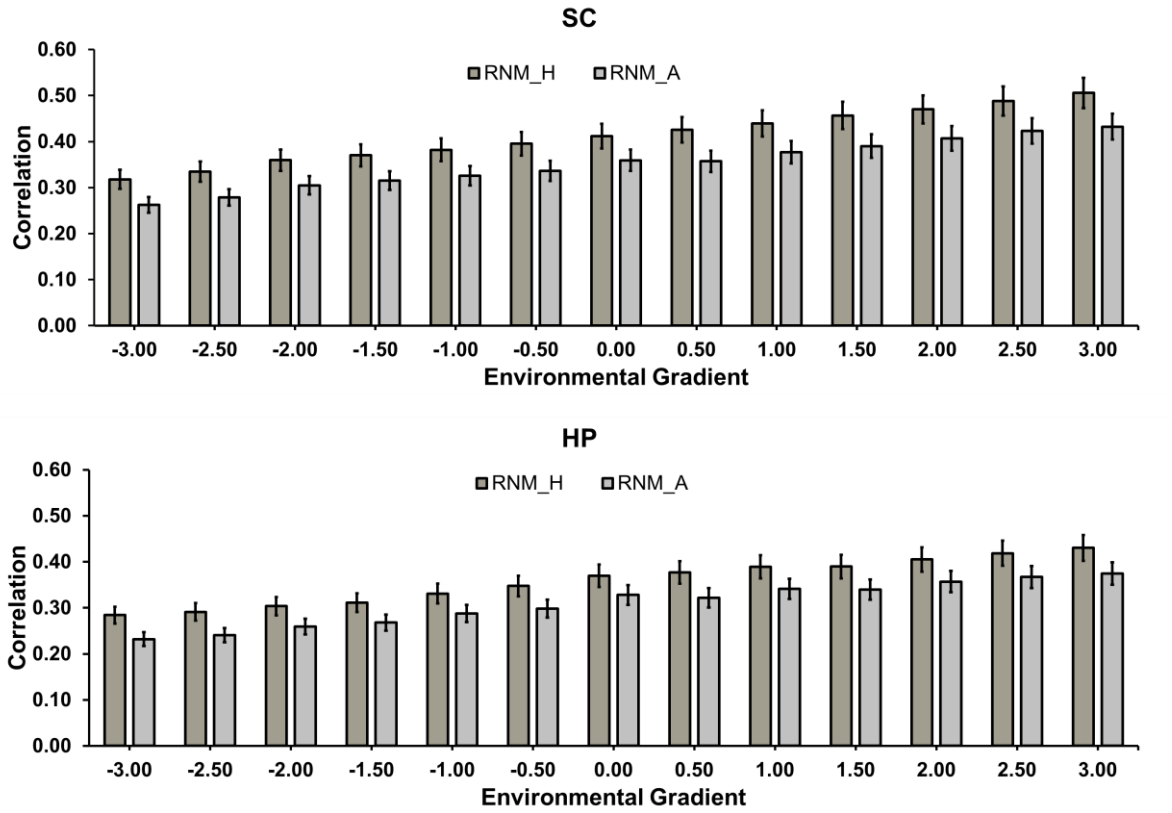


Figure 6. Predictive ability assessed in *younger scheme* by Pearson's correlations between adjusted phenotypes and genomic breeding values or breeding values for scrotal circumference (SC) and heifer early pregnancy (HP) in Nellore cattle across different environments, using pedigree-based reaction norm model (RNM_A) and single-step reaction norm model (RNM_H).

Using younger scheme and the RNM_H model, (G)EBV predictive abilities for both traits were higher than those based on the RNM_A model, increasing from 14.61 to 21.18% for SC and from 12.74 to 22.56% for HP (Figure 6). The (G)EBV predictive abilities obtained using RNM_H were moderate (from 0.32 to 0.51 for SC and 0.28 to 0.43 for HP). Our results are similar to that found by Oliveira et al. (2018) using a model including GxE interaction for Nellore body weight. The authors reported an increase in the predictive ability of 7.9% by using RNM_H instead of RNM_A. These results indicate that combining genomic and pedigree information could be considerably advantageous.

Compared to RNM_A, the RNM_H model provided the greatest predictive ability across environmental levels (Figure 6), and an explanation could be the lower standard error for slope of reaction norm (Supplementary Information: S3 Figure). The H matrix allows exploiting the genetic similarity among pedigree-based unrelated individuals but that are genetically related to each other based on genomic

information, and it should contribute to increasing the predictive ability by improving the overall genetic connectedness among EC levels.

For the *environment-specific scheme*, only RNM_H was applied. Using animals from the unfavorable environment as reference population and those from the favorable environment as validation population and vice-versa, instead of the *younger scheme*, reduced the overall predictive ability for both traits (Figure 6 and 7). These reductions were, on average, 19.30% for SC, and 12.02% for HP in Low scheme and 13.90% for SC and 6.00% for HP in High scheme. One of the major factors on decreasing predictive ability observed for *environment-specific scheme* could be, partially, due to the small number of genotyped animals in the reference population compared to validation (Figure 1). Daetwyler et al. (2010) observed that the number of animals in reference compared to validation is one of the key factors in genomic prediction accuracy.

Reduction in model predictive ability could be also due to GxE interaction effect on re-ranking of animals occurring, mainly, between those environments with genetic correlations lower than 0.80 (Figure 3). Moreover, with this *environment-specific scheme*, animals from training and validation are less genetically related than in the *younger scheme* (Supplementary Information: S4 Figure). Our results agree with those from Nirea and Meuwissen (2017). The authors observed that considering a sampling of animals from different environments as reference population resulted in higher predictive ability than using an *environment-specific scheme*.

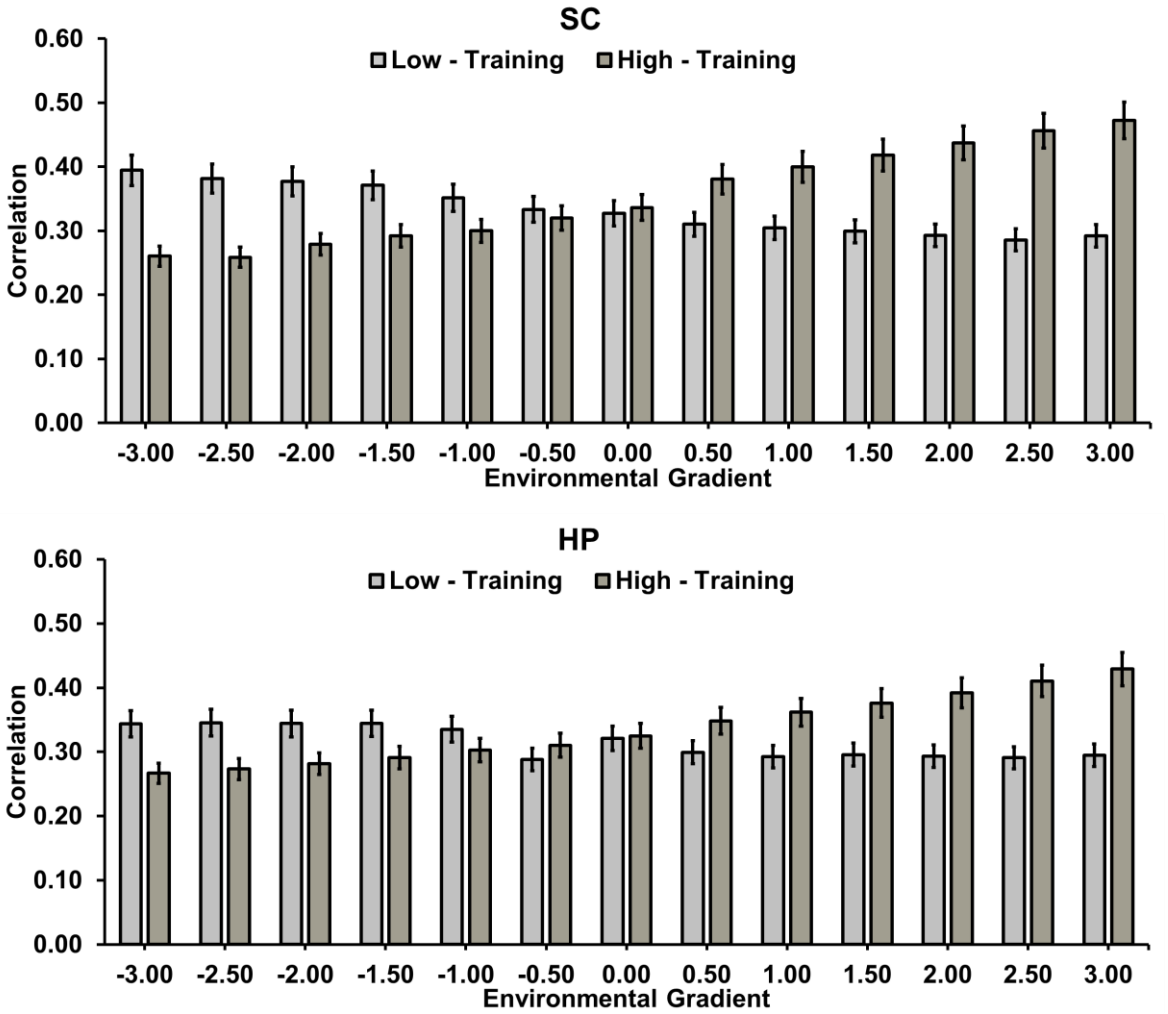


Figure 7. Predictive ability of genomic selection obtained for *environment-specific scheme*, partitioning dataset according to EC levels. Low training - animals from poor environmental condition (-3.0 to -1.0) used as training set and those environmental conditions (-0.5 to 3.0) for validation set; and High training - animals from better environmental condition (1.0 to 3.0) used as training set and those from environmental condition (-3.0 to 0.5) used as a validation set.

Higher predictive ability was achieved with the *younger scheme* because it allowed a better relationship among reference and validation population (Figure 1 and Supplementary Information: S4 Figure). Several studies have shown that the predictive ability depends on the degree of genomic relationship between individuals in training and validation subsets (Clark et al., 2012; Pryce et al., 2012; Lee et al., 2017). The predictive ability results indicated the *younger scheme* as the best scheme in this work, since it combines animals across EC levels leading to a better sampling of genotyped animals.

2.4 Conclusions

Our results indicate the occurrence of genotype-environment interaction effects on Nellore sexual precocity indicator traits (scrotal circumference and early heifer pregnancy). Using a reaction norm model, combining pedigree and genomic information to evaluate genotype-environment interaction, increases breeding values predictive abilities compared to a model considering pedigree only.

The genomic breeding values for heifer pregnancy and scrotal circumference were sensitive to changes in environmental levels demonstrating an important effect of the GxE interaction. Strong SNP marker-environment interaction was detected for Nellore sexual precocity indicator traits with changes in SNP effect and genetic variance estimates across environmental levels.

Validation schemes allowing sampling animals across the EC levels for training and validation subsets, keeping a close relationship between them, are indicated.

2.5 References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) 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.

Araujo Neto FR de, Pegolo NT, Aspilcueta-Borquis RR, Pessoa MC, Bonifácio A, Lobo RB, Oliveira HN (2018) Study of the effect of genotype–environment interaction on age at first calving and production traits in Nellore cattle using multi-trait reaction norms and Bayesian inference. **Animal Science Journal** 89:939–945.

Ashworth CJ, Toma LM, Hunter MG (2009) Nutritional effects on oocyte and embryo development in mammals: implications for reproductive efficiency and environmental sustainability. **Philosophical Transactions of the Royal Society B: Biological Sciences** 364:3351–3361.

Barth AD, Brito LFC, Kastelic JP (2008) The effect of nutrition on sexual development of bulls. **Theriogenology** 70:485–494.

Bérénos C, Ellis PA, Pilkington JG, Pemberton JM (2014) Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. **Molecular Ecology** 23:3434–3451.

Burrow HM (2012) Importance of adaptation and genotype × environment interactions in tropical beef breeding systems. **Animal** 6:729–740.

Calus MPL, Veerkamp RF (2003) Estimation of Environmental Sensitivity of Genetic Merit for Milk Production Traits Using a Random Regression Model. **Journal of Dairy Science** 86:3756–3764.

Chiaia HLJ, de Lemos MVA, Venturini GC, Aboujaoude C, Berton MP, Feitosa FB, Carvalheiro R, Albuquerque LG, de Oliveira HN, Baldi F (2015) Genotype × environment interaction for age at first calving, scrotal circumference, and yearling weight in Nellore cattle using reaction norms in multitrait random regression models. **Journal of Animal Science** 93:1503–1510.

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** 44:1–9.

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. **Genetics** 185:1021–31.

Ferraz Jr MVC, Pires A V, Santos MH, Silva RG, Oliveira GB, Polizel DM, Biehl M V, Sartori R, Nogueira GP (2018) A combination of nutrition and genetics is able to reduce age at puberty in Nelore heifers to below 18 months. **animal** 12:569–574.

Fordyce JA (2006) The evolutionary consequences of ecological interactions mediated through phenotypic plasticity. **Journal of Experimental Biology** 209:2377–2383.

Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution** 43:1–7.

Geweke J 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Bayesian Statistics 4, 169–193.

Gordo DGM, Espigolan R, Tonussi RL, Júnior GAF, Bresolin T, Braga Magalhães AF, Feitosa FL, Baldi F, Carvalheiro R, Tonhati H, de Oliveira HN, Chardulo LAL, Albuquerque LG (2016) Genetic parameter estimates for carcass traits and

visual scores including or not genomic information. **Journal of Animal Science** 94:1821–1826.

Hayes BJ, Bowman PJ, Chamberlain AJ, Savin K, van Tassell CP, Sonstegard TS, Goddard ME (2009) A Validated Genome Wide Association Study to Breed Cattle Adapted to an Environment Altered by Climate Change. **Plos one** 4:e6676.

Hayes BJ, Daetwyler HD, Goddard ME (2016) Models for Genome × Environment Interaction: Examples in Livestock. **Crop Science** 56:2251–2259.

Kolmodin R, Strandberg E, Jorjani H, Danell B (2003) Selection in the presence of a genotype by environment interaction: response in environmental sensitivity. **Animal Science** 76:375–385.

Lee SH, Clark S, van der Werf JHJ (2017) Estimation of genomic prediction accuracy from reference populations with varying degrees of relationship. **PLOS ONE** 12:1–22.

Lillehammer M, Goddard ME, Nilsen H, Sehested E, Olsen HG, Lien S, Meuwissen THE (2008) Quantitative Trait Locus-by-Environment Interaction for Milk Yield Traits on Bos taurus Autosome 6. **Genetics** 179:1539–1546.

Lillehammer M, Hayes BJ, Meuwissen THE Goddard ME (2009) Gene by environment interactions for production traits in Australian dairy cattle. **Journal of Dairy Science** 92:4008–4017.

Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z (2015) Manual for BLUPF90 family of programs. University of Georgia, Athens, USA.

Moser G, Khatkar MS, Hayes BJ, Raadsma HW (2010) Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. **Genetics Selection Evolution** 42:1–15.

Mota RR, Lopes PS, Tempelman RJ, Silva FF, Aguilar I, Gomes CCG, Cardoso FF (2016) Genome-enabled prediction for tick resistance in Hereford and Braford beef cattle via reaction norm models. **Journal of Animal Science** 94:1834–1843.

Mota RR, Silva FF, Lopes PS, Tempelman RJ, Sollero BP, Aguilar I, Cardoso FF (2017) Analyses of reaction norms reveal new chromosome regions associated with tick resistance in cattle. **Animal** 12:205–214.

Mulder HA (2016) Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. **Frontiers in Genetics** 7:1–11.

Nepomuceno DD, Pires A V., Ferraz MVC, Biehl M V., Gonçalves JRS, Moreira EM, Day ML (2017) Effect of pre-partum dam supplementation, creep-feeding and post-weaning feedlot on age at puberty in Nellore heifers. **Livestock Science** 195:58–62.

Nirea KG, Meuwissen THE (2017) Improving production efficiency in the presence of genotype by environment interactions in pig genomic selection breeding programmes. **Journal of Animal Breeding and Genetics** 134:119–128.

Oliveira DP, Lourenco DAL, Tsuruta S, Misztal I, Santos DJA, de Araújo Neto FR, Aspilcueta-Borquis RR, Baldi F, Carvalheiro R, de Camargo GMF, Albuquerque LG, Tonhati H (2018) Reaction norm for yearling weight in beef cattle using single-step genomic evaluation. **Journal of animal science** 96:27–34.

Oliveira MM, Santana ML, Cardoso FF (2016) Multiple-breed reaction norm animal model accounting for robustness and heteroskedastic in a Nelore–Angus crossed population. **animal** 10:1093–1100.

Pryce JE, Arias J, Bowman PJ, Davis SR, Macdonald KA, Waghorn GC, Wales WJ, Williams YJ, Spelman RJ, Hayes BJ (2012) Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. **Journal of Dairy Science** 95:2108–2119.

Robertson A (1959) The Sampling Variance of the Genetic Correlation Coefficient. **Biometrics** 15:469–485.

Robinson JJ, Ashworth CJ, Rooke JA, Mitchell LM, McEvoy TG (2006) Nutrition and fertility in ruminant livestock. **Animal Feed Science and Technology** 126:259–276.

Santana ML, Eler JP, Bignardi AB, Menéndez-Buxadera A, Cardoso FF, Ferraz JBS (2015) Multi-trait linear reaction norm model to describe the pattern of phenotypic expression of some economic traits in beef cattle across a range of environments. **Journal of Applied Genetics** 56:219–229.

Santana ML, Eler JP, Cardoso FF, Albuquerque LG, Ferraz JBS (2013) Phenotypic plasticity of composite beef cattle performance using reaction norms model with unknown covariate. **Animal** 7:202–10.

Santana Jr ML, Eler JP, Bignardi AB, Boligon AA, Ferraz JBS (2018) Genetic correlation between growth and reproductive performance of beef females depends on environment. **Animal Production Science** 58:1201–1209.

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. **BMC genomics** 15:1–12.

Silva FF, Mulder HA, Knol EF, Lopes MS, Guimaraes SEF, Lopes PS, Mathur PK, Viana JMS, Bastiaansen JWM (2014) Sire evaluation for total number born in pigs using a genomic reaction norms approach. **Journal of Animal Science** 92:3825–3834.

Smith BJ (2007) boa : An R Package for MCMC Output Convergence Assessment and Posterior Inference. **Journal of Statistical Software** 21:1–37.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)** 64:583–639.

Su G, Madsen P, Lund MS, Sorensen D, Korsgaard IR, Jensen J (2006) Bayesian analysis of the linear reaction norm model with unknown covariates. **Journal of animal science** 84:1651–7.

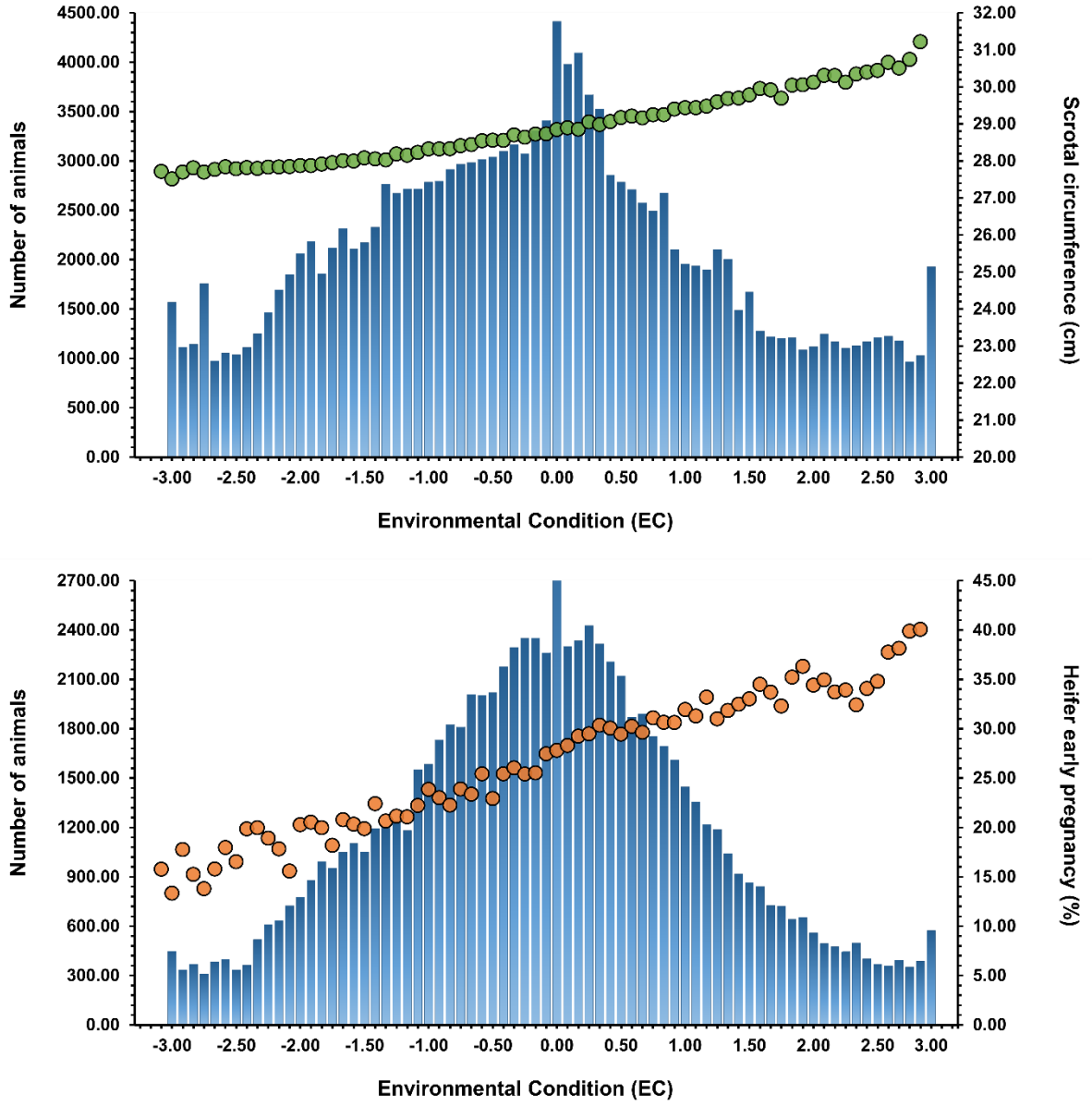
van Gestel J, Weissing FJ (2016) Regulatory mechanisms link phenotypic plasticity to evolvability. **Scientific reports** 6:24524.

VanRaden PM 2008. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science** 91:4414–4423.

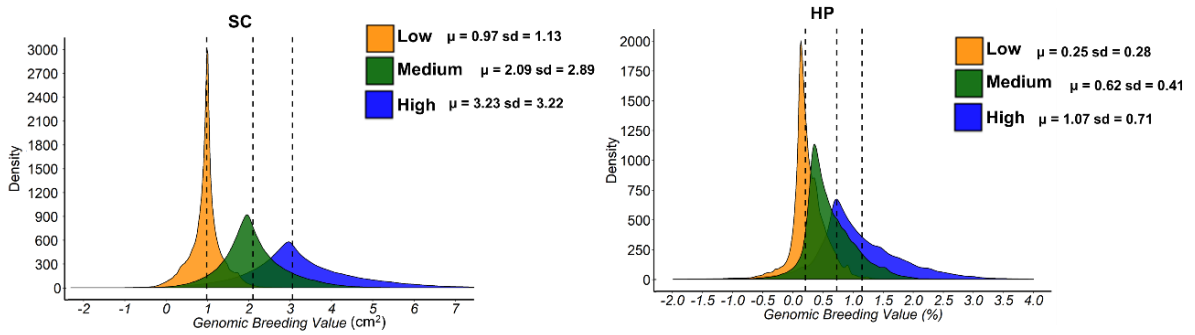
Yu H, Spangler ML, Lewis RM, Morota G (2017) Genomic Relatedness Strengthens Genetic Connectedness Across Management Units. **G3: GENES, GENOMES, GENETICS** 7:3543–3556.

Zhang Z, Kargo M, Liu A, Thomasen JR, Pan Y, Su G (2019) Genotype-by-environment interaction of fertility traits in Danish Holstein cattle using a single-step genomic reaction norm model. **Heredity** 122:1–13.

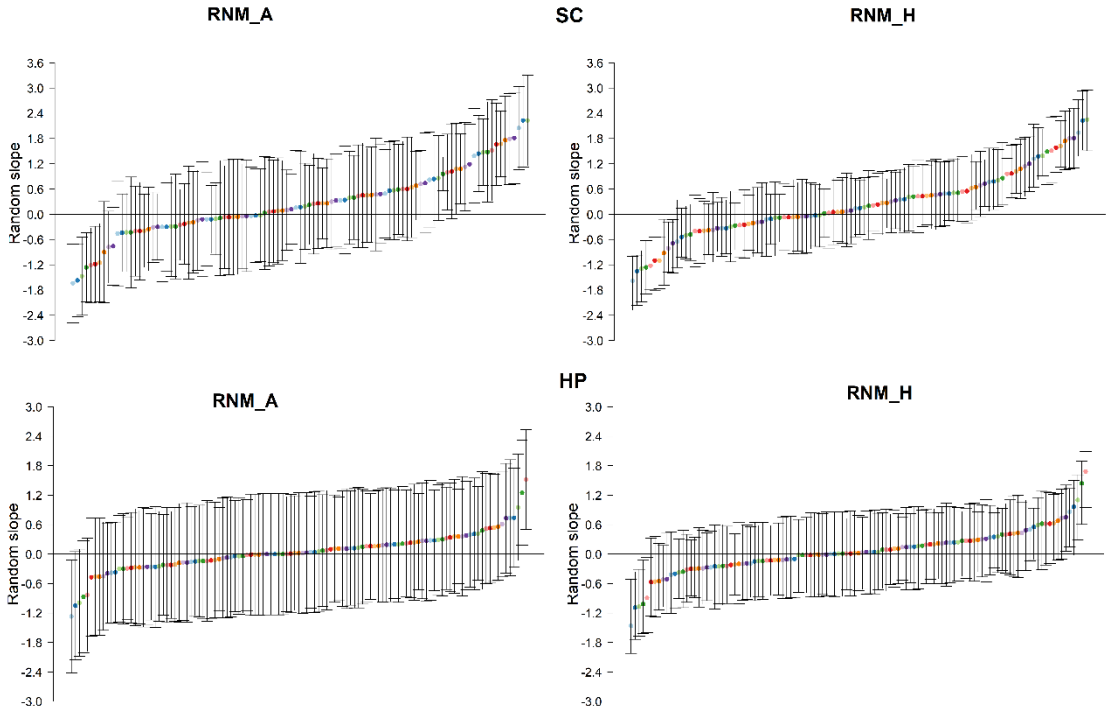
2.6 Supplementary Information



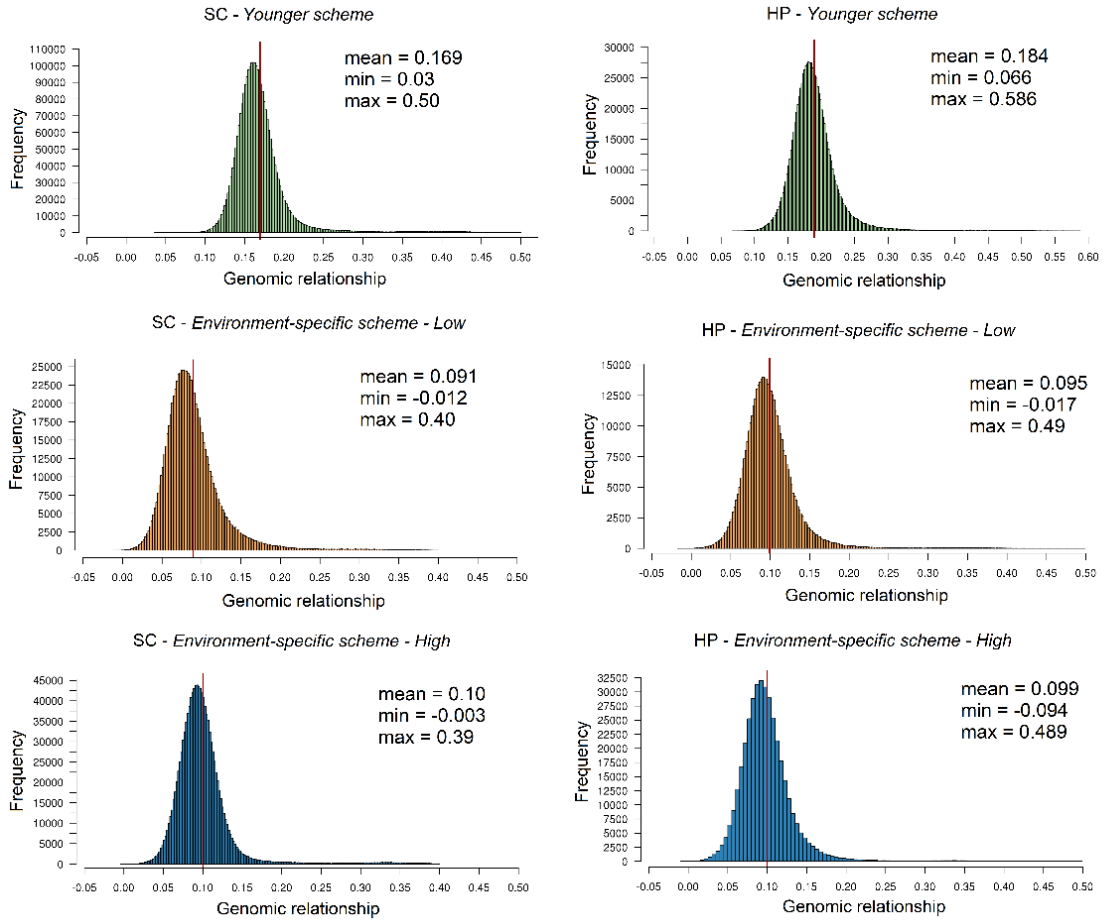
S1 Figure. Number of records across environmental conditions and phenotypic average for Nellore sexual precocity indicator traits, [A] scrotal circumference (SC) and [B] heifer early pregnancy (HP).



S2 Figure. Probability density distribution of genomic breeding values estimated using results of reaction norm model for three environmental conditions. Low (dark yellow color) environmental condition = -3.0; Medium (dark green color) environmental condition = 0.0 and High (dark blue color) environmental condition = 3.0, for Nellore scrotal circumference (SC) and heifer early pregnancy (HP).



S3 Figure. Random slope estimation and standard error (SE) for Nellore scrotal circumference (SC) and Nellore heifer early pregnancy (HP). RNM_A – reaction norm model considering pedigree information and RNM_H – single-step reaction norm model.



S4 Figure. Genomic relationships among animals in training and reference subsets for *younger scheme* and *environment-specific scheme* used to assess the predictive ability for scrotal circumference (SC) and heifer early pregnancy (HP) in Nellore cattle. The red line indicates the mean of genomic relationship among animals.

Chapter 3 – Pleiotropic regions identified as key-modulators in Nellore reproductive traits in three environmental conditions

Abstract – Genome-wide association studies (GWAS) are an effective approach to identify genetic variants associated with reproductive traits. Despite significant advances in understanding genetic determinants on sexual precocity in Nellore cattle, there is still a lack of insights about the genetic variation in response to environmental changes. Hence, multi-trait meta-analysis method for GWAS was carried out to identify candidate genomic regions affecting multiple traits associated with reproductive traits in Nellore cattle, raised under different environmental conditions (EC). Phenotypic records of age at first calving (AFC), heifer early pregnancy (HP), heifer rebreeding (HR), and scrotal circumference (SC) measured on 128,994; 85,339; 90,831 and 151,053 animals, respectively, were used. From those, 1800 heifers, 3050 young bulls, and 800 sires were genotyped with BovineHD BeadChip. A reaction norm model was applied to estimate the animals' response to environmental changes. The SNP effects for reproductive traits were estimated using a linear transformation of the genomic breeding values. The meta-analysis GWAS were performed on SNP effects for each trait under the three EC levels (Low = -3.0, Intermediate = 0, and High = 3.0). Both single-trait and the statistical combination of single-trait GWAS pointed out genomic regions on BTA 5, 14, and 29 with major pleiotropic effect for reproductive traits under EC levels. The statistical combination resulted in 56 significant SNP markers, at an empirical threshold of $p < 1 \times 10^{-6}$ (FDR = 1%). Within 0.2 Mb of the significant regions, candidate genes significantly overrepresented on biological factors affecting scrotal circumference, age at puberty, conception rate, heifer pregnancy, interval to the first estrus after calving and ovulation rate were found. From the gene network analyses, the key genes *PLAG1, IGF1, IGF2, INS, TH, LYN,* and *HRAS* were found to shelter pleiotropic variants affecting reproductive traits on different EC levels as regulators of reproductive pathways. The identification of such shared genomic regions under different EC levels and with pleiotropic effects in reproductive traits, contribute to a better understanding of the molecular mechanisms affecting reproduction in Nellore cattle raised under different environmental conditions.

Keywords: genotype-environment interaction, GWAS, genomics, meta-analysis, pleiotropic effect

3.1 Introduction

Reproductive performance plays an important role in the economic efficiency of beef cattle systems because these traits are paramount to the profitability of beef production systems. However, reproductive efficiency is mostly affected by environmental conditions defined by differences in heat stress and nutrition factors, with striking effects in beef cattle (Robinson et al., 2006; Ashworth et al., 2009). Sexual precocity phenotypic variation has often shown an association with nutritional factors, due to its effect in the energy metabolism regulating the major reproductive hormones (Roche et al., 2011; Lucy et al., 2014).

Nellore cattle in Brazil are raised on pasture in a wide range of environments, mostly in seasonally poor nutrition, high temperature, and humidity. These different environmental conditions could lead to genotype-environment (GxE) interaction (Burrow, 2012) and can be an important source to the genetic variation in reproductive traits, affecting animals ranking and population genetic gain (Rauw and Gomez-Raya, 2015; Mulder, 2016).

Reproductive traits have a complex inheritance controlled by different genomic regions (Robinson et al., 2006; Lucy, 2008; D'Occhio et al., 2019) and can be affected by environmental factors (Zhang et al., 2019). The environmental conditions may often lead to changes on metabolic or endocrine functions associated to onset of puberty (Moriel et al., 2012; Samadi et al., 2014; D'Occhio et al., 2019). Thus, when GxE interaction occurs, it is expected a changing in genomic regions with a significant effect on reproductive traits, according to the environmental conditions (Zhang et al., 2019). The understanding of this biologic process could be important to design strategies for reproductive traits selection.

Single-trait genome wide scan (GWAS) has been successfully used for identification of genetic variants associated to several reproductive traits (Cole et al., 2011; Hawken et al., 2012; Fortes et al., 2013; Irano et al., 2016; Melo et al., 2018). In this context, GWAS, through a reaction norm model, permits to identify environmental-dependent genomic regions associated to reproductive traits (Lillehammer et al., 2007; Zhang et al., 2019).

Usually, traits used to select animals for reproductive efficiency, as age at first calving, heifer early pregnancy, heifer rebreeding and scrotal circumference, are associated to each other and are affected by GxE interaction (Chiaia et al., 2015; Santana et al., 2017; Santana Jr et al., 2018). In addition, combining the results from

single-trait GWAS across environments, can improve the power for identifying shared genomic variants across EC levels and the existence of possible pleiotropic effects on different reproductive traits (Evangelou and Ioannidis, 2013; Bolormaa et al., 2014; Melo et al., 2018). Hence, the objective of this study was carried out to identifying major pleiotropic genomic regions associated to age at first calving (AFC), heifer early pregnancy (HP), heifer rebreeding (HR), and scrotal circumference (SC) traits, in Nellore cattle raised in different environmental conditions (EC).

3.2 Material and Methods

3.2.1 Ethical approval

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

3.2.2 Phenotypic information

Nellore heifers and young bulls' phenotypic information were provided by three breeding programs: DeltaGen, Paint and Cia de Melhoramento. The animals were born between 1984 and 2014 in 220 farms located in the Midwest, Southeast, and Northeast of Brazil. Reproductive traits were: age at first calving (AFC), heifer early pregnancy (HP), heifer rebreeding (HR) and scrotal circumference (SC).

The AFC was computed as the difference between the date of first calving and heifer date of birth, in days. HP was defined attributing a value of 1 (success) to heifers calving until 31 months of age and 0 (failure) otherwise. The HR was computed attributing a value of 1 (success) or 0 (failure) to females that presented or not a second calf, respectively. The SC (cm) was measured at 18 ± 1.5 months of age.

Contemporary groups (CG) were defined by animals born in the same year, raised in the same farm (at birth, weaning, and yearling), and belonging to the same management group (at birth, weaning and yearling). For HR, the systematic effect of early pregnancy and calf sex were included in the CG. In some of the herds, two breeding seasons are adopted: 1) an anticipated breeding season from February to April in which all heifers are exposed to reproduction for 60 days and 2) a breeding season starting around November and lasting for 90 days in which all cows and heifers, which did no conceive in the first breeding season, are exposed.

Phenotypic data for HP and HR, in which all animals showed the same binary response (0 or 1) within CG were excluded. For AFC and SC, phenotypic data outside of the interval of 3.5 standard deviations below and above the CG mean were excluded. For all traits, CG with less than five phenotypic records were excluded. A summary of the descriptive statistics is presented in Table 1.

Table 1 – Descriptive statistics for reproductive traits, and heritability estimates (standard errors) using an animal model and single-step GBLUP.

Trait	Number of CG	Number of information	Mean	S.D.	Heritability
AFC (days)	10495	128994	1042.00	110.91	0.12 (0.006)
SC (cm)	9154	151053	28.67	2.35	0.34 (0.008)
HP (%)*	3786	85339	26.45	0.66	0.33 (0.004)
HR (%)*	7798	90831	69.85	0.55	0.27 (0.003)

AFC – Age at first calving; SC – scrotal circumference; HP – heifer early pregnancy; HR – heifer rebreeding; S.D. – standard deviation; * Mean represents the percentage of success for HP and HR.

3.2.3 Genotypes

A total of 1,900 heifers, 1,500 young bulls and 850 sires were genotyped using the Illumina BovineHD BeadChip assay (770k, Illumina Inc., San Diego, CA, USA) and 1,750 young bulls using GeneSeek® Genomic Profiler HDi 75K (GeneSeek Inc., Lincoln, NE). The young bulls genotyped with panel 75K were imputed to the HD panel using FImpute v.2.2 (Sargolzaei et al., 2014). The imputation accuracy of animals genotyped with lower density are expected to be higher than 0.98 (Carvalheiro et al., 2014). Autosomal markers presenting minor allele frequency (MAF) less than 0.03, significant deviation from Hardy–Weinberg equilibrium ($P \leq 10^{-5}$) and call rate of markers and samples less than 0.95 were removed. After quality control, 1,800 heifers, 3,050 young bulls, 800 sires and 451,554 SNP remained for further analyses.

3.2.4 Statistical Modeling

The environmental conditions (EC), describing production levels, were based on the solutions of CG for yearling body weight (YBW). Yearling weight is an indicator of nutritional conditions across herds. Differences in EC affecting body weight could reduce reproductive performance due to changes in metabolic homeostasis (Samadi et al., 2014). In this study, two-steps were used to analyze GxE interaction effects on reproductive.

3.2.4.1 First Step

In the first step, the EC levels were determined by the CG solutions for YBW obtained using animal models and the single-step GBLUP (ssGBLUP) method, separated for males and females, as follows:

$$y = X\beta + Za + e$$

where y is the vector of phenotypic information for YBW; β is the fixed effect for CG (defined by animals born in the same year and season, in the same farm from birth to yearling and management group from birth to yearling) and age at recording as a linear co-variable; a is the vector of additive genetic effect assuming the normal distribution $N(0, H\sigma_a^2)$ in which σ_a^2 is the additive genetic variance and H is the pedigree-genomic relationship matrix; e is the residual effect following the $N(0, I\sigma_e^2)$, I is the identity matrix and σ_e^2 is the residual variance. The X and Z are the incidence matrices related to fixed and random effects.

In the ssGBLUP model the combined pedigree-genomic relationship matrix (H) was used and its inverse (H^{-1}) was calculated as follows (Aguilar et al., 2010): $H^{-1}=A^{-1}+\begin{bmatrix}0 & 0\\ 0 & G^{-1}- A_{22}^{-1}\end{bmatrix}$ where A_{22}^{-1} represents the subset of the inverse of the pedigree-based relationship matrix for the genotyped animals, and G^{-1} is the inverse of the genomic relationship matrix, according to VanRaden (2008).

The CG solutions means obtained separately for males and females to define the EC levels were standardized to a mean of 0 and variance of 1, with values ranging from -3.0 to +3.0 standard deviations. The phenotypic trend and number of phenotypic information across EC levels are shown in supplementary information S1 Figure.

3.2.4.2 Second Step

In order to obtain the random regression coefficient estimates for each animal (intercept and slope) to estimate the genomic breeding value across EC levels, to be used as pseudo-phenotypes on GWAS, the follow single-trait reaction norm model was used:

$$y_{ij} = X\beta + \varphi_f \boldsymbol{\Phi}_f(EC_j) + \alpha_{if} \boldsymbol{\Phi}_f(EC_j) + e_{ij},$$

where, y_{ij} is the vector for phenotypic record (AFC, HP, HR, and SC) of animal *i* recorded in the level *j* of EC; $X\beta$ is the fixed effects including CG and, for SC, the linear effect of age at recording, φ_f are the fixed regression coefficients on Φ_f; Φ_f is the Legendre polynomial for each *EC* level (EC_j); α_{if} are the random regression additive effect coefficients corresponding to animal *i* on EC level *j*, and e_{ij} is the random residual. Heifer early pregnancy (HP) and heifer rebreeding (HR) were analyzed assuming threshold-reaction norm models, considering an underlying distribution as follows: $f(y|l_i) = \prod_{i=1}^{n_i} 1(l_i < t_i)1(y = 0) + 1(l_i > t_i)1(y = 1)$, where y is the binary trait HP or HR, l_i is the underlying liability of binary observation *i*, t_i is the threshold that defines the category response for y scale and n_i is the number of information for HP or HR.

There were considered five classes of EC: class 1: EC level lower than -1.5; class 2: -1.5 ≤ EC level ≤ -0.5; class 3: -0.5 ≤ EC level ≤ 0.5; class 4: 0.5 < EC level ≤ 1.5; and class 5: EC level higher than 1.5. The model described above was fitted using the single-step GBLUP (ssGBLUP) approach, where the genomic information is combined with pedigree information (Aguilar et al., 2010).

The additive genetic and residual effects were considered normally distributed: $a=\{a_j\}\sim N(0,HK_{ab})$ and $e=\{e_{ij}\}\sim N(0,\mathcal{R})$, where $K_{ab}=\begin{bmatrix}\sigma_a^2 & \sigma_{ab}\\ \sigma_{ab} & \sigma_b^2\end{bmatrix}$ is the (co)variance matrix attributed to random regression coefficients of the intercept (a) and the slope (b) and $\mathcal{R}$ is the diagonal residual variance matrix considering 5 classes of heterogeneity. The *H* matrix is given by: where A corresponds to the relationship matrix based on pedigree information, $H = A + \begin{bmatrix}0 & 0\\ 0 & G - A_{22}\end{bmatrix}$, in which A_{22} represents of subset of pedigree-based relationship matrix for genotyped animals and G is the genomic relationship matrix obtained according to VanRaden (2008).

The analyses were performed using GIBBS3f90 program for AFC and SC and the THRGIBBS3f90 program for HP and HR (Misztal et al., 2015). A single chain of 1,000,000 cycles was used, considering a burn-in of 100,000 iterations and the results being saved at every 10 cycles. A total of 90,000 samples were used to obtain posterior parameter estimates and their convergence was evaluated through

visual inspection and using the Bayesian Output Analysis (Smith, 2007) and Geweke test (Geweke, 1992)

3.2.5 Estimates of SNP effects

In the GWAS analyses, animals with accuracies (based on prediction error variance [PEV] acc = $\sqrt{1\text{-PEV}/\sigma_a^2}$) lower than 0.40 for reaction norm parameters (intercept and slope) were excluded. Therefore, a total 1,778 heifers, 2,985 young bulls and 745 sires were considered for genome wide association analyses.

Animals GEBVs were estimated across EC levels using the following equation: $\hat{g}_{EC} = \alpha_i \Phi_f{}'$, where α_{if} is the vector of estimated additive genetic values for regression coefficients (intercept and slope) of animal *i* and $\Phi_f{}'$ is the transpose vector of the Legendre polynomial for the *EC* level. Animal GEBVs estimated in three EC levels -3.0 (Low), 0.0 (Medium) and 3.0 (High) were used to estimate SNP marker effects.

The SNP effects ($\hat{u}_k$) were estimated for the reproductive traits (AFC, HP, HR and SC) separately in the three EC levels: Low (AFC_L, HP_L, HR_L and SC_L) Medium (AFC_M, HP_M, HR_M and SC_M) and High (AFC_H, HP_H, HR_H and SC_H) from a linear transformation of estimated GEBV (Gualdrón Duarte et al., 2014) using: $\hat{u}_k = Z' G^{-1} \hat{g}_{EC}$. The Z (*n x m*) (where *m* and *n* are the numbers of SNPs and animals, respectively) is the SNP matrix assuming 0, 1, and 2 for genotypes AA, AB, and BB, respectively, and G^{-1} is the inverse of genomic relationship matrix, according to VanRaden (2008) and $\hat{g}_{EC}$ is a vector of the GEBV estimates in Low, Medium and High EC level.

The variance for each SNP effect ($\sigma^2_{\hat{u}_k}$) from a linear transformation of estimated GEBV was estimated according Gualdrón Duarte et al. (2014) as follows:

$$\sigma^2_{\hat{u}_k} = Z' G^{-1} Z \sigma_a^2 - Z' G^{-1} C^{aa} G^{-1} Z,$$

where σ_a^2 is the genetic variance for reproductive traits evaluated in three EC levels defined above and C^{aa} was obtained from the inverse of coefficient matrix of the mixed model equations as follows:

$$C^{aa} = \sigma_e^2 (I - X(X'X)^{-1}X' + G^{-1}\lambda)^{-1}$$

where σ_e^2 is the residual variance for each reproductive trait evaluated in three EC levels; I is the identity matrix; X is the incidence matrix related to genotypes and GEBV and λ is the residual variance and genetic variance ratio (σ_e^2 / σ_a^2) for each reproductive trait in three EC level.

3.2.6 Detection of pleiotropic candidate genomic regions

To identify the pleiotropic regions affecting reproductive traits in the three EC levels Low (AFC_L, HP_L, HR_L and SC_L), Medium (AFC_M, HP_M, HR_M and SC_M) and High (AFC_H, HP_H, HR_H and SC_H), the statistical combination method, described by Bolormaa et al. (2014), was used. Statistical tests were performed standardizing the SNP effect in each EC level as follows: $t_k = \frac{\hat{u}_k}{SE(\hat{u}_k)}$. In this way, t_k is the *t-values* for the SNP marker effect; $\hat{u}_k$ is the SNP effects for each trait in each EC level and $SE(\hat{u}_k)$ is the standard error for SNP effect ($\hat{u}_k$).

Multi-trait statistic test was used to evaluate the association and influence of SNP effects for reproductive traits across different environments and traits (Bolormaa et al., 2014). This statistic test summarizes single-marker statistics following a χ^2 distribution with *k* degrees of freedom, where *k* is the number of traits included in the multi-trait statistic test. For each SNP marker (total of 451,554 SNP markers) the statistic was: $Multi-trait\ \chi^2 = t_k' V^{-1} t_k$, where t_k is a vector 12 x 1 of the signed t-value of SNP_k for the 4 reproductive traits in the 3 environmental conditions, t_k' is a transpose of vector t_k; V^{-1} is an inverse of the *t-values* correlation matrix (12 x 12). The V^{-1} was corrected by adding the average correlation of each trait to their respective diagonal element. This correction was used because some traits show higher correlations than others and may lead to highly significant composite scores even when single-trait analyses have lower evidence of the association (Pereira et al., 2016).

The assessment of possible GWAS predictions inflation were calculated based on the inflation/deflation factor (λ): $\lambda = median\ (p-value) / 0.456$, where values of λ ranging between 1.0 and 1.1 were considered acceptable (Devlin and Roeder, 1999). The *p-value* was adjusted for multiple tests using the false discovery rate (FDR) test (Qu et al., 2010): $fdr = m * \alpha / s$, where m the total number of SNP markers in the analyze (*n* = 451,554), α is the significance threshold and s is the

number of significant markers with p-value < α. To select a value for α that results a false discovery rate lower than 1% the p-value rank position proceeding as described by Benjamini and Hochberg (1995).

3.2.7 Gene mapping of significant SNP for GWAS statistical combination

The SNP markers from GWAS statistical combination were deemed significant when $-log_{10}(p\text{-}value) > 6.0$, where the significance was based on the FDR test. Genes located within 200 kb of each SNP were identified using the Ensembl gene 93 database (Aken et al., 2016) for cattle using the UMD v3.1 assembly, as well as the BioMart R package (Durinck et al., 2009). The animal QTL database (QTLdb) for cattle (Hu et al., 2016) was used to identify if significant genomic regions were in overlapped with described quantitative trait loci (QTL). Biological mechanisms and pathways (Gene Ontology - GO) involving the candidate genes were identified using the clusterProfiler R Package (Yu et al., 2012). GO terms test was based on hypergeometric distribution using the false discovery rate (FDR) multiple testing (p-value < 0.01).

Aiming to gain additional insights regarding the genes, based on biological process, clusterProfiler R Package (Yu et al., 2012) was used to classify the GO in groups, highlighting the biological key role. Interactions between protein-coding genes were predicted using the STRING database with default settings according to Szklarczyk et al (2015). The networks were graphed using gephi 0.9.2 (available at: http://gephi.github.io/) from results of String and were used the modularity score to identify sub-network structure (Blondel et al., 2008).

3.3 Results

3.3.1 Single-trait GWAS for reproductive traits in three EC levels

With the single-trait GWAS analysis, genomic variants significantly associated (p-value < 5×10^{-6}) with reproductive traits and located on different chromosomes were identified (Additional file 2: Figure S2). The quantile-quantile (Q-Q) plot shows that the inflation-factor (λ) estimates ranged from 1.01 to 1.19 (Additional file 3: Figure S3).

Reproductive traits in High EC level showed the highest number of significant genomic regions (Figure 1). Based on the single-trait GWAS results, a total of 95 SNP for AFC_H, 96 for HP_H, 146 for HR_H and 90 for SC_H were deemed as

significant (Figure 1 A). Shared genomic regions varied according to the genetic correlation between reproductive traits in different EC level (Figure 1 B and Additional file 4: Table S1). As expected, the same trait in high and medium EC levels (highly correlated; Additional file 4: Table S1), showed a large number of shared genomic regions, a total of 49 for SC (r = 0.83), 35 for HP (r = 0.81) and 28 for HR (r = 0.75; Figure 1 B). Major significant pleiotropic variants, including peaks higher than the FDR threshold ($-log_{10}(p\text{-}value) > 6.0$), were detected on BTA 5, 7, 10, 14, 16, and 29 (Additional file 2: Figure S2).

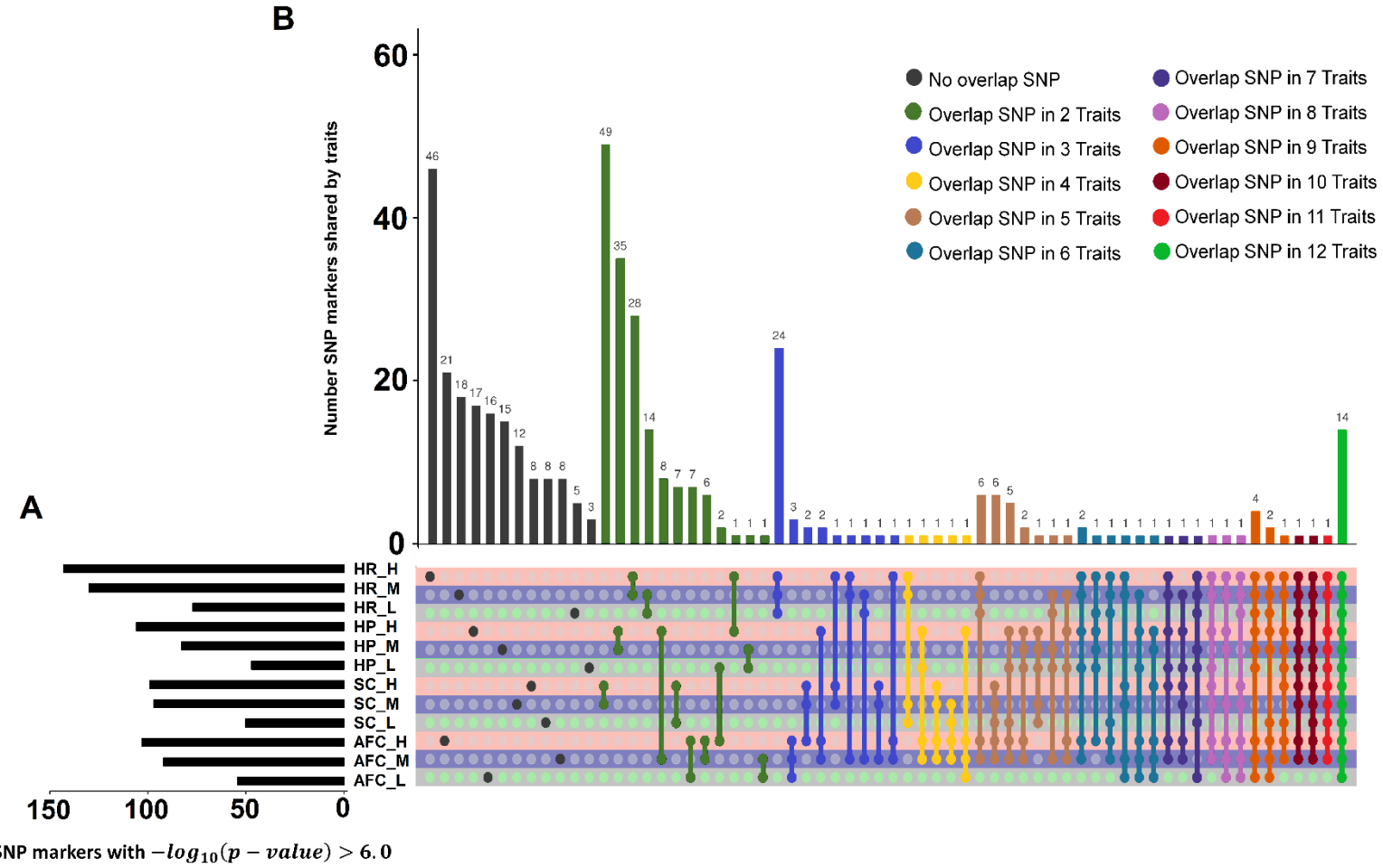


Figure 1. Genomic regions significantly ($-\log_{10}(\mathrm{p}-\mathrm{value}) > 6.0$) associated to reproductive traits - age at first calving (AFC), scrotal circumference (SC), heifer early pregnancy (HP) and heifer rebreeding (HR) - in three EC levels low (L), medium (M) and high (H). **A** – Total number of SNP markers identified and **B** - Overlap of significant SNP markers indicating the pleiotropic and shared regions. Colors represent number of 12 traits (4 traits x 3 environmental conditions) with shared regions and bars represent the number of SNP in overlap.

3.3.2 Multi-trait statistical combination to detect pleiotropic regions

Results from Multi-trait statistical combination analysis showed inflation-factors (λ) of 1.083 and a total of 56 SNPs with a significance level of $-log_{10}(p\text{-}value) > 6.0$, resulting in a $fdr = 0.01$ (Figure 2). These results highlight 13 regions, distributed in six regions chromosomes, as pleiotropic regions for reproductive traits in different EC levels (Figure 2 and Table 2). These 13 regions were also mapped in single-trait GWAS in the same positions within chromosomes 5, 7, 10, 14, 16 and 29 (Figure 3). Combining GWAS results enabled to identify pleiotropic effects for reproductive traits in EC levels, explaining their associations.

A linkage disequilibrium (LD) analysis was performed indicating a strong association (LD from 0.40 to 0.80) around the top scoring SNP markers within these regions (Figure 4). This is an indication that these chromosome segments, most likely, show similar genetic signal in the different environment levels.

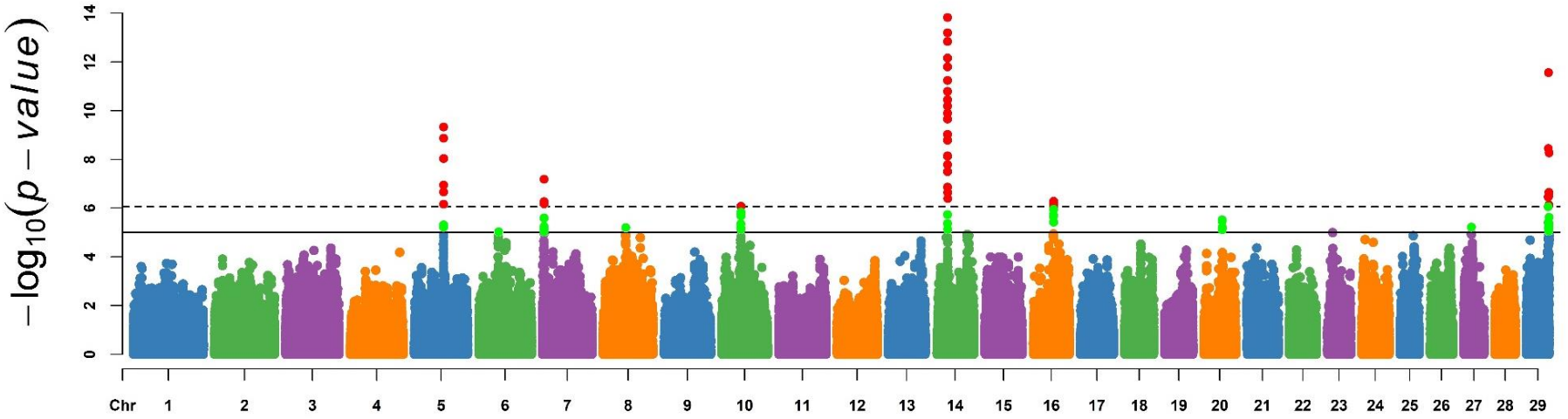


Figure 2**.** Manhattan plot for statistical combination of genome-wide results for loci affecting age at first calving (AFC), scrotal circumference (SC), heifer early pregnancy (HP), and heifer rebreeding (HR) in three environmental conditions Low (L), Medium (M) and High (H). Manhattan plot with significant ($-log_{10}(p\text{-}value) > 6.0$) associations where highlighted in red and suggestive associations ($5 < -log_{10}(p\text{-}value) < 6$) in light green.

Table 2 – Identification of genes within the 200 kb of 56 SNP markers identified in multi-trait statistical combination ($-log_{10}(p\text{-}value) > 6.0$) for reproductive traits in three environmental condition in Nellore heifers and young bulls.

Chromosome	Genomic region[1]		Number of SNP[2]	Gene symbol
	Initial position	End position		
5	66436841	66522516	8	PARPBP, PMCH, IGF1
7	3640197	3710059	5	ATP13A1, CILP2, GATAD2A, GMIP, LPAR2, NDUFA13, PBX4, TSSK6
10	43641433	43641433	1	ATL1, SAV1
14	25009960	25199512	12	LYN, RPS20, MOS, PLAG1, CHCHD7, SDR16C5, SDR16C6, PENK
14	25211447	25254540	8	SDR16C5, SDR16C6, PENK
14	25312775	25521782	6	SDR16C6, PENK, IMPAD1
16	45808109	45887895	7	RERE, SLC45A1
29	49110300	49174899	2	OSBPL5, CARS, NAP1L4
29	49588566	49588566	1	KCNQ1
29	50053108	50053108	1	ASCL2, TH, INS, IGF-2, bta-mir-48
29	50385393	50602910	1	PDDC1, POLR2L, CD151, PNPLA2, RPLP2, PIDD1, DRD4, TSPAN4
29	50702910	50717364	2	MOB2, RASSF7, TMEM80, LRRC56, HRAS, bta-mir-2410, bta-mir-210
29	51087878	51288335	2	MUC5B, MUCAC

[1] Refers to SNP markers position within 200 kb used for search genes using the Ensembl gene 93 (UMD v3.1 assembly)

[2] Number of significant SNP markers in this region.

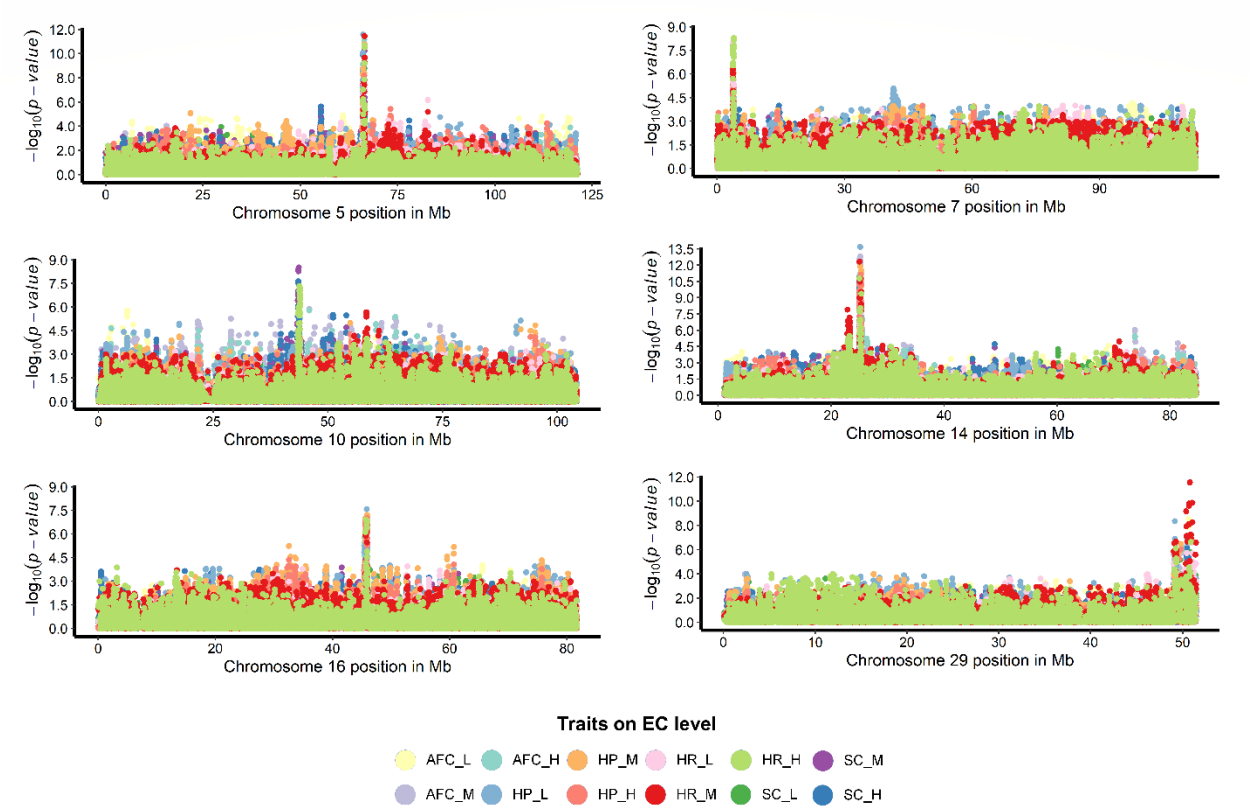


Figure 3. Single trait GWAS results Manhattan plot for age at first calving (AFC), scrotal circumference (SC), heifer early pregnancy (HP), and heifer rebreeding (HR) in three environmental conditions Low (L), Medium (M) and High (H) for the significant regions found with the Multi-trait statistical combination method.

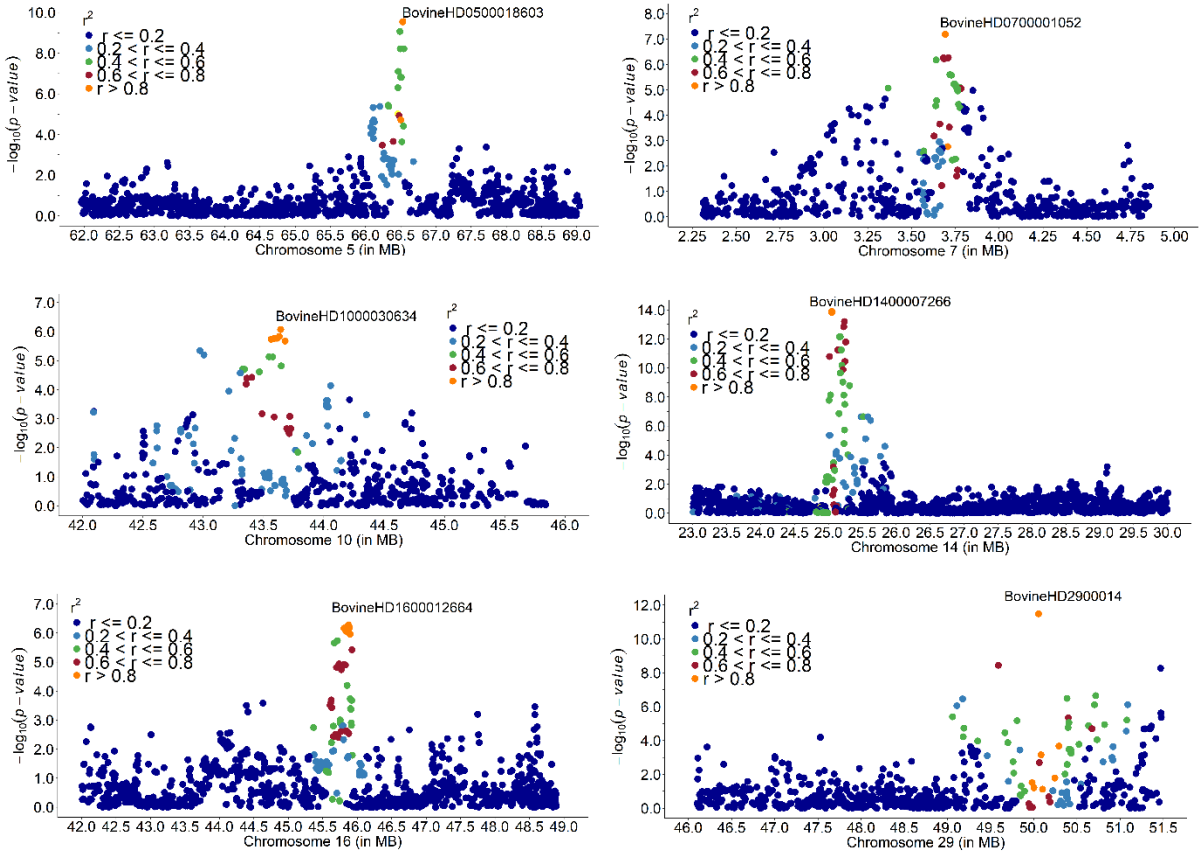


Figure 4. Association plot and linkage disequilibrium for six genomic significant regions on chromosome 5, 7, 10, 14, 16 and 29. LD (r^2) with the top scoring marker is represented according to the indicated color scale.

The effects of the 56 significant SNP markers varied according to the EC level (Figure 5), indicating a strong SNP x E interaction. Changes in the SNP effects with EC levels were both in ranking and variance. The percentages of genetic variance explained by the genomic regions were: 16.89%, 19.86% and 22.85% for AFC; 15.96%, 19.82% and 23.03% for HP; 19.04%, 22.15% and 22.51% for HR; and 15.84%, 18.50% and 21.53% for SC, respectively, in Low, Medium and High environmental conditions. Differences in SNP effects occurred as response to environmental changes, mainly for AFC, HP and SC (Figure 5). For all traits, the highest dispersion of SNP effects and the highest percentage of the genetic variance explained by SNP markers were observed in the High EC level.

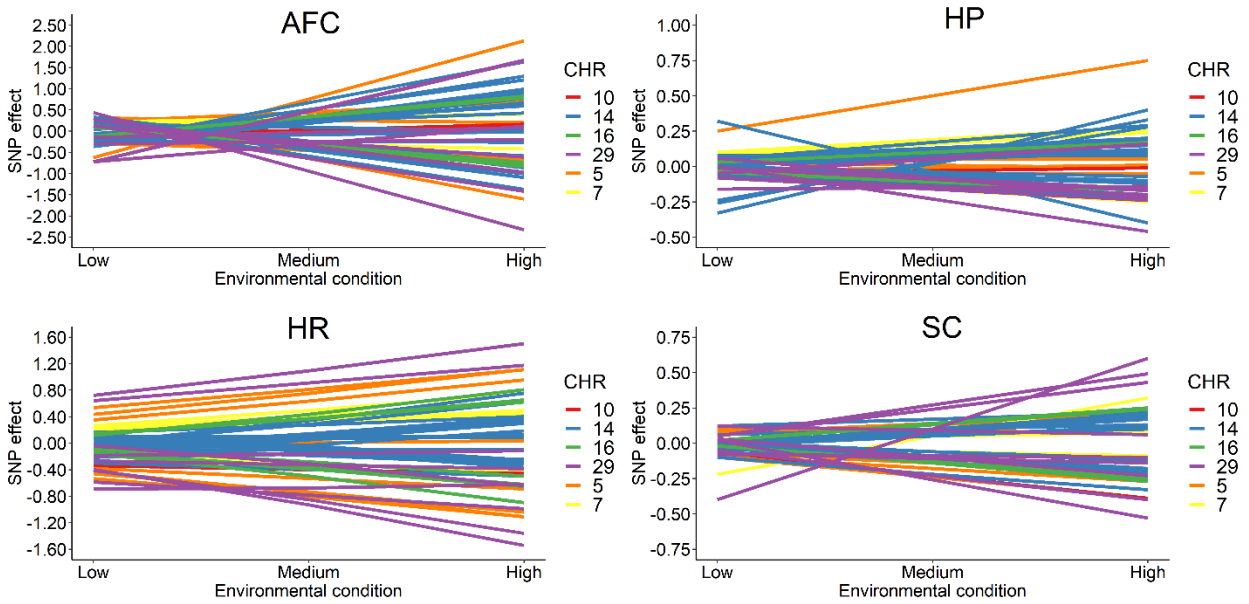


Figure 5. SNP marker effects of genomic regions significantly associated (*-log$_{10}$(p-value)>6.0*) to age at first calving (AFC), heifer early pregnancy (HP), heifer rebreeding (HR) and scrotal circumference (SC) on three EC levels: Low, Medium and High. The color represents the chromosome (CHR) and line represents the 56 SNP markers.

3.3.3 Gene-set enrichment annotation

With the meta-analysis genome scan 56 SNP markers significantly associated with male and female reproductive traits in Nellore cattle, were found. From those, a total of 45 genes were mapped within 200 kb and show a striking effect on reproduction traits (Figure 6).

Significant regions detected on BTA5 harbors three putative candidate genes for sexual precocity (*IGF1, PARPBP and PMCH*) and are associated to biological processes affecting growth and reproduction (Additional file 5: Table S2). A region on BTA7 harbors eight candidate genes (Table 2) affecting physiological processes with important effects in lipid signaling. In addition, one significant region on BTA10 harbors two genes (*ATL1* and *SAV1)* playing a key role in the immune systems, affecting cells activities and ovarian follicle growth.

The genomic region that explains the highest effect on reproductive traits was located on BTA14 (25.00 – 25.25 MB), harboring eight candidate genes (Table 2). This gene set has been associated with biological processes affecting reproductive traits,

mainly by the association with a reduction in *IGF1* levels, fat depth and beginning of puberty in females and males (Additional file 5: Table S2). The BTA16 harbors two genes (*RERE* and *SLC45A1*) playing a key role in the physiological process affecting estrogens hormone. The BTA29 harbors 23 genes (Table 2) and overall, show associations with insulin and glucose concentrations and present an important role as a metabolic signal to GnRH secretion.

The regions identified by multi-trait statistical combination analysis have been overlapped to QTLs previously reported in the literature (Figure 6 – a). Remarkably, some regions are closely associated to body weight at different ages (BTA 5, 7, 14 and 29), insulin-like growth factor 1 level (BTA 5 and 14), fat deposition (BTA 5 and 14), height (BTA 5 and 14), average daily gain (BTA 14 and 29) and body condition score (BTA 5). These regions have also been associated directly with scrotal circumference (BTA 7, 14 and 16), sperm motility (BTA 29), reproductive performance affecting age at puberty (BTA 14), conception rate (BTA 29), heifer pregnancy (BTA 7), interval to first estrus after calving (BTA 14 and 29), ovulation rate (BTA 5) and gestation length (BTA 5, 14 and 29).

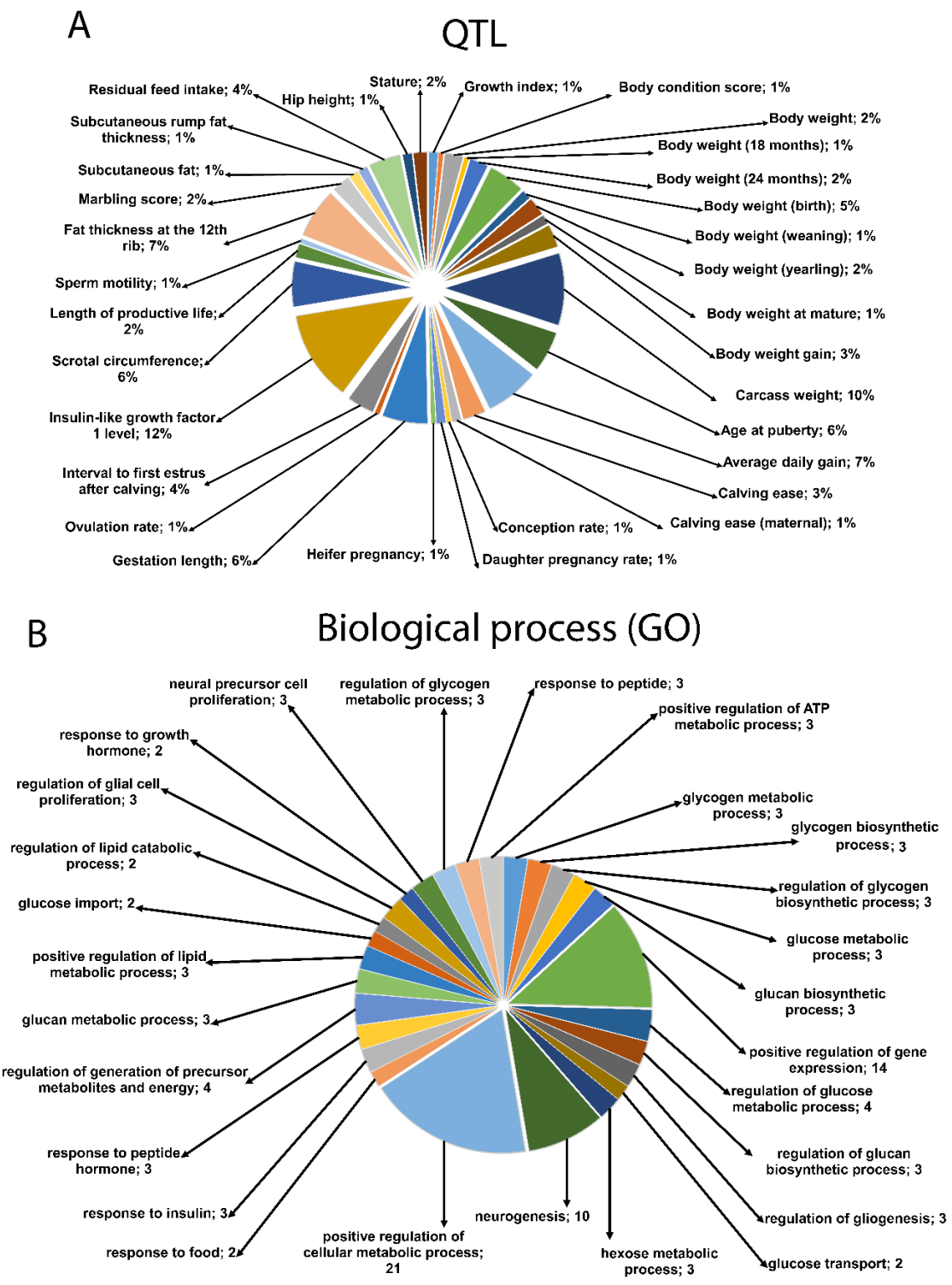


Figure 6. A) Functional annotation for quantitative trait locus (QTL). QTL percentage represents numbers of QTL overlapped with the significant regions. B) Biological process hits within the 200kb region of the SNP markers ($-log_{10}(p\text{-}value)>6.0$), the number represents the genes amount associated to biological functions.

Enrichment analyses identified biological processes and pathways that may affect reproductive traits in cattle (Figure 6 – b). A total of 18 groups of GO for biological process showed significant overrepresentation (P-value < 0.001) on reproduction and growth traits in animals (Figure 6 b). The GO terms were related with responses to growth hormone (GO:0060416), peptide hormone (GO:0043434) and insulin (GO:0032868), resulting in change state or activity of a cell, related to secretion, enzyme production, and gene expression.

With the pathway analysis was possible to identify genes involved in biological process with effect on reproductive processes, acting on ovarian follicle cell-cell adhesion (GO:0071840), negative regulation of oocyte maturation through regulation of glucose metabolism (GO:0010906) and GnRH signaling in response to peptide hormone (GO:0043434).

From the 47 candidate genes identified (Table 2), 36 have been annotated to protein-protein interaction on the STRING database (Figure 7). A total of 30 genes were involved in a single network and 3 smaller networks were formed by 2 genes (Figure 7). The key genes identified are involved in reproduction and growth such as *IGF1* (insulin-like growth factor 1), *IGF2* (insulin-like growth factor 2), *HRAS* (HRas Proto-Oncogene, GTPase), *INS* (insulin), *PLAG1* (pleomorphic adenoma gene 1), PMCH (pro-melanin-concentrating hormone), *ASCL2* (Achaete-Scute Family BHLH Transcription Factor 2) and *TSSK6* (Testis Specific Serine Kinase 6) were found to shelter pleiotropic variants in different EC levels. Using a modularity score to measure network structures it was observed that the genes *IGF2*, *IGF1, INS*, and *HRAS* have strong connection with each other, and are associated with ovarian steroidogenesis pathway (04913), at the biological process affecting response to growth hormone (GO:0060416) and response to insulin (GO:0032868; Figure 7).

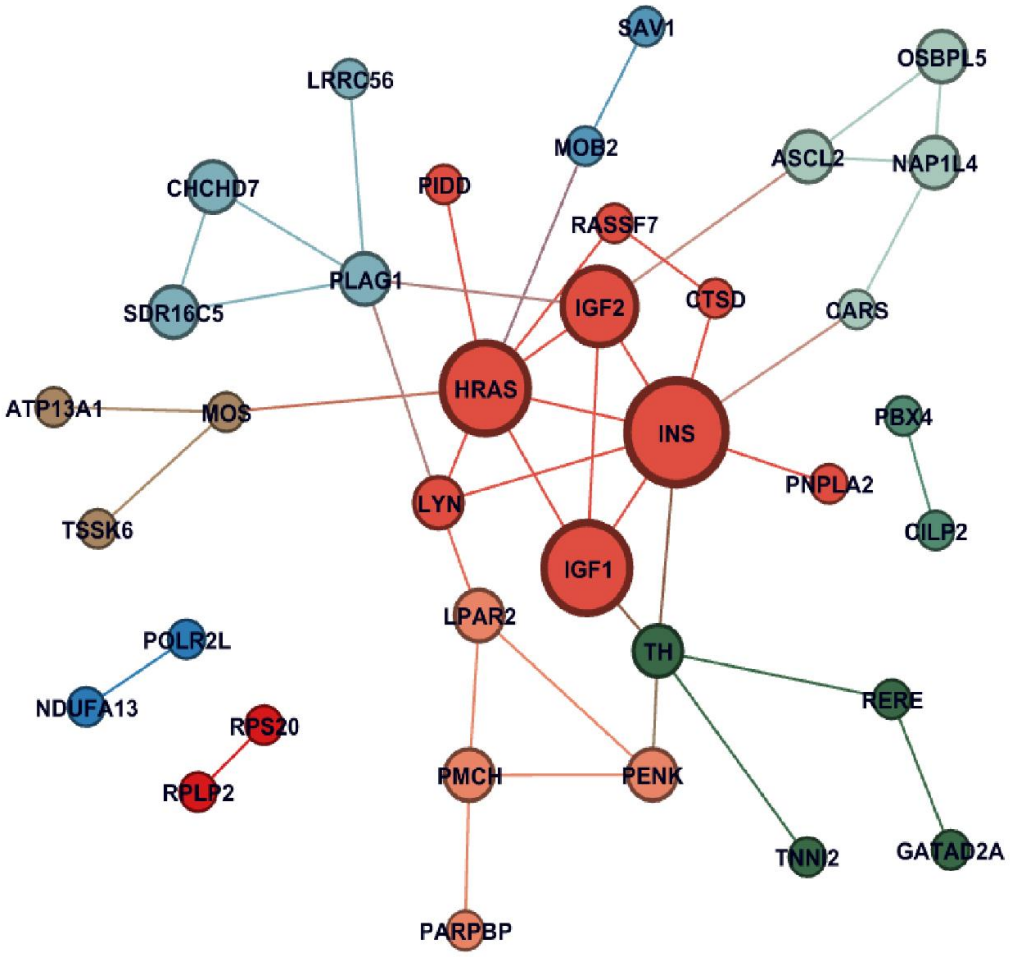


Figure 7. Network of candidate pleiotropic genes for reproductive traits on three EC levels in Bos indicus (Nellore) cattle. The network was built from known protein-protein interactions (edges) between gene products (nodes). Size of nodes represents the number of connections between genes and color represents the associations between genes in the network.

3.4 Discussion

Results from the GWAS identified 56 genetic variants associated with correlated traits, and such associations have been related to pleiotropic effects on reproductive traits. Significant genomic regions harboring variants that contribute to multiple reproductive traits (AFC, SC, HP and HR), under different EC levels, provide valuable insights into the underlying physiological mechanisms changes in response to animals' adaptation to environmental conditions.

The negative correlation between AFC in low (AFC_L) and in high (AFC_H) EC levels suggests genomics regions leading a pathway that regulates these traits in opposite direction (Additional file 4: Table S1). Therefore, changes in EC levels lead to differences in patterns of SNP effect that differentially contribute to animal fitness (Solovieff et al., 2013), and to a better understanding of the genetic association between reproductive traits on different EC levels. Therefore, different genomic variants can show specific effect in the expression of reproductive traits according to environmental condition.

The number of overlapping genomic variants associated with multiple correlated traits varied according to genetic correlation (Figure 1), indicating that the genetic cause of this correlation is mainly the pleiotropic effect existing among reproductive traits across EC levels. In this sense, a reduction of the sharing genetic variants between traits leads to a reduction in the genetic correlations (Evangelou and Ioannidis, 2013). Genetic correlations between phenotypes under different EC levels can lead to an assessment of biological pleiotropy, where the genetic variants are associated with multiple traits (Solovieff et al., 2013).

In this work 6 regions identified on BTA 5, 7, 10, 14, 16 and 29, using a statistical combination of GWAS analysis, showed a pleiotropic effect on reproductive traits in three environmental conditions (Figure 2). The SNP markers identified in the present work have been reported as associated with reproductive traits in previous GWAS studies in cattle (Velazquez et al., 2008; Hawken et al., 2012; Nicolini et al., 2013; Fortes et al., 2013a; Wolfe et al., 2014; Bolormaa et al., 2014; Smith et al., 2015). Several QTLs, genes and variants affecting growth and reproduction in cattle (Kirkpatrick et al., 2000; Snelling et al., 2010; Karim et al., 2011; Hawken et al., 2012; Fortes et al., 2013a) surrounding most of SNP markers were found, with effects on ovulation rate, age at puberty, gestation length, length of productive life and postpartum anestrous interval.

The region BTA5 (66.43 – 66.52 MB) could be explored to identify sexually precocious animals. In this region, the IGF1 is a candidate gene affecting directly QTLs associated with ovulation rate, age at puberty, gestation length, length of productive life, postpartum anestrous interval and calving ease (Kirkpatrick et al., 2000; Mullen et al.,

2011; Lirón et al., 2012; Wei et al., 2013). For example, the *PMCH* (Pro Melanin-Concentrating Hormone) gene on BTA5 (66.43 – 66.52 MB), encodes three neuropeptides (*MCH, NEI* and *NGE*) that play a role in the regulation of metabolism and feed intake in mammals (Murray et al., 2001; Mul et al., 2010; Seo et al., 2016). *PMCH* variants have been associated with scrotal circumference and fat deposition in cattle (Helgeson and Schmutz, 2008; Walter et al., 2014). Changes in energy homeostasis caused by *PMCH* and *MCH* genes have been associated with reproduction traits due to their effect on ovarian steroid feedback, with major effect on *GnRH*, *LH* and *FSH* levels (Murray et al., 2001; Pissios et al., 2006; Wu et al., 2009; Naufahu et al., 2013).

In the same region on BTA5 (66.43 – 66.52 MB), the *PARPBP* (Poly[ADP-Ribose] Polymerase 1 binding protein) positively regulates the activity of *PARP1* gene, which plays an important role in the regulation of gene expression associated to cell cycle and metabolism (Kraus, 2008; Ke et al., 2017). According to Wei et al. (2013), *PARP1* is associated with the process of oocyte nest breakdown, primordial follicle formation, and transition to primary follicles. The insulin-like growth factor 1 (*IGF1*) has been associated with sexual precocity in cattle (Fortes et al., 2013c) and reproductive performance, e.g. calving interval, post-partum resumption of ovarian cyclicity, first service conception rate (Mullen et al., 2011). Lirón et al. (2012), reported QTL in the *IGF1* gene associated with age at puberty in cattle. This gene also shows association with genetic factors contributing to an earlier puberty onset in precocious heifers and regulation of testicular growth in young bulls (Lirón et al., 2012; Johnston et al., 2014). The *IGF1* gene changes the reproductive efficiency in cattle, mainly by controlling the production of *GnRH* or the secretion of gonadotropin; as well as, it has a direct effect on the ovary susceptibility to *FSH* and LH (Velazquez et al., 2008; Akers et al., 2005).

The genes identified on BTA7 (36.40 - 37.10 MB), exhibited genes playing a role in energy metabolism regulation (GO:0031325). This region was associated with reproductive traits such as heifer early pregnancy (Irano et al., 2016) and scrotal circumference (McClure et al., 2010). The *GATA* zinc finger domain containing 2A (*GATAD2A*) gene is a member of *GATAD* family and has target action on delays puberty, acting as a neurobiological brake that controls the *GnRH* pattern pulse, necessary for proper reproduction (Leclerc et al., 2008; Lomniczi et al., 2015). The

TSSK6 (Testis Specific Serine Kinase 6) gene has been reported in Tssk6-KO mouse model to play an essential role on spermatogenesis with impact on production and sperm function, affecting the fertility due to morphologically abnormal sperm (Jha et al., 2017). The *NDUFA13* gene shows an indispensable role to assemble and function of mitochondrial complex I (Hou et al., 2017). The mitochondrial complex I represents a major target for compounds that improve insulin sensitivity through increased AMPK activity (Turner et al., 2008). Its effect on reproduction traits in males and females could be mediated by the complex I effect on AMPK that affects signal pathways controlling the interactions between energy balance and reproduction (Bertoldo et al., 2015).

Significant SNP markers on BTA10 (43.64 MB) have been associated with QTLs, (Figure 6), affecting daughter pregnancy rate and length of productive life in cattle (Cole et al., 2011). The main gene in such region was *SAV1*, which has an important function in early follicular development, granulosa cell proliferation and differentiation, and oocyte maturation during ovulation associated to Hippo signaling pathway (Lyu et al., 2016) with effect in metabolic and cellular process (GO:0008152 and GO:0009987), which play a role on ovarian follicle growth (Hsueh et al., 2015).

The major pleiotropic variant found on BTA14, targeting the traits in the three EC levels, was mapped on the position 24.5 to 25.5 Mb and harbors genes such as *RPS20, MOS, PLAG1, LYN, CHCHD7* and *SDR16C5* (*RDHE2*). Such location on BTA14 has been pointed out as an important pleiotropic region in cattle affecting traits as growth, fat deposition, age at puberty, and the negative regulation of luteinizing hormone secretion (Nishimura et al., 2012; Fortes et al., 2013a; Juma et al., 2016; Takasuga, 2016; Melo et al., 2018). These pleiotropic regions play a key role on the regulation of many genes and pathways, including *IGF2*, *IGF1R* pathways, and multiple levels of the hypothalamic–pituitary–gonadal (HPG) axis (Van Dyck et al., 2007). Changes in the effect of this region (Figure 5), lead to genetic differences in reproductive performance, affecting mostly growth pathways, due to differences in *IGF1* and *IGF2* hormonal levels (Velazquez et al., 2008; Fortes et al., 2013b). Thus, this region on BTA14 affects the endocrine hormone secretion (GO:0060986) regulating the endocrine system process involved in both secretion and response to endocrine hormones. The gene set on BTA14 (Table 2), plays an important role on regulation pathways involved in age at

puberty in males and females indicating a relevant association of growth with reproductive traits (Fortes et al., 2013b; Juma et al., 2016).

The *PENK* gene identified on BAT14 plays a biological role in response to stress (GO:0050896) and homeostasis (Karim et al., 2011). The *PENK* region has been associated with fertility traits in cattle because it presents a major role in embryonic development, estrous cycle, and early pregnancy, which match with its effect on neural tissues, regulating hormone secretion (Cánovas et al., 2014). The *MOS* gene has been associated with *Mos/MAPK* Pathway affecting female fertility acting in oocyte maturation (Brisard et al., 2014). Differences in *Mos/MAPK* signaling pathway in oocyte maturation lead to lower fertility rates in cattle (Brisard et al., 2014). Hence, factors targeting oocytes maturation and embryonic development, which might be regulated by energy metabolism-related genes, prostaglandin synthesis and lipid metabolism might reduce fertility in heifers. The *IMPAD1* gene shows SNPs associated with detection of the first corpus luteum and SC in Brahman cattle (Fortes et al., 2012).

The region on BTA16, (45.60 – 45.98 Mb), showed high association with SC and HP in different EC levels and has been, previously, associated with QTLs affecting SC (McClure et al., 2010). The *RERE* gene, mapped in this region, plays a role in embryogenesis and embryonic survival in mouse (Zoltewicz et al., 2004). Genes affecting fertilization rate and embryonic survival rate have been related to conception rate in female (Khatib et al., 2009). In this framework, changes on physiological events required for heifers' normal estrous cycle lead to differences in oocyte and spermatozoa development, ovulation, fertilization, embryo survival and early pregnancy (Robinson et al., 2006; Roche et al., 2011).

The functional candidate genes surrounding the peak on BTA29 at 49.01 – 51.18 Mb, showed a high association with reproduction traits on the three EC levels. Surround this region, there are QTLs affecting gestation length, the interval to the first estrus after calving and sperm motility. The *TH* (Tyrosine Hydroxylase) gene plays an important role in the physiology of adrenergic neuron involved in the conversion of tyrosine to dopamine and regulates physiological events affecting reproduction, via hypothalamic-pituitary-gonadal (HPG) axis (Parillo et al., 2014). Thus, the *TH* gene could be an important neurotransmitter that mediates stimulus to preovulatory *GnRH/LH* patterns

(Pau et al., 2000; Serova et al., 2008). The Insulin-like growth factor 2 (*IGF2*) and insulin (*INS*) genes located on the BTA29 are associated with regulation of glucose metabolism (GO:0010907), glycogen metabolic process (GO:0070875), regulation of lipid metabolic process (GO:0045834), response to insulin stimulus (GO:1900076). The regulation of the bioavailability of insulin-like growth factors (*IGFs)* plays a role in growth and animal development (Berkowicz et al., 2012). These genes act regulating embryo development during estrus cycle and early pregnancy (Robinson et al., 2000), affecting directly antral follicle growth, functioning as a potent stimulator for estradiol production (Meikle et al., 2004). In addition, *IGF2* and *INS* are candidate genes for mediating the effects of dietary deficiencies affecting energy balance on reproductive function (Robinson et al., 2006). The levels of insulin concentration affected by genes *IGF2* and *INS*, and it has been associated to the time of re-initiation of ovarian cyclicity after calving, acting indirectly on ovarian activity by insulin signaling (GO:0032868, response to insulin) during early embryonic development (Llewellyn et al., 2007). These biological processes are key regulators of energy homeostasis in cattle (Chagas et al., 2007; Lucy et al., 2014). The energy homeostasis, mainly the glucose and insulin metabolism, has been recognized as a major regulator of energy metabolism, although it plays a role in reproduction by their neuroendocrine function with effect on *GnRH/LH* secretion (Sliwowska et al., 2014).

Overall, candidate pleiotropic genes network (Figure 7) provides insights about reproductive pathways with physiological functions during estrous and early pregnancy in heifers and spermatogenesis in young bulls. This finding suggests that common variation in Nellore cattle reproductive traits raised under harsh conditions is essential to controlling reproductive precocity and reproductive efficiency.

The gene network (Figure 7) identified the potential genes associated with physiological mechanisms controlling reproductive traits and energy metabolism. Noticeably, the identified gene set (*IGF1, IGF2, INS, NDUFA13, PNPLA2, PMHC*, *LYN and HRAS*) has major action on feedback-regulated systems controlling reproductive and somatotropic axis. This provides a link between nutritional and metabolic aspects into the reproductive process. In this context, these genes are associated with the metabolism of energy balance affecting the onset of puberty caused by the direct effect

on oocyte maturation (Gong, 2002). Energy metabolism play a striking effect during the oocyte maturation, spermatozoa development, ovulation, fertilization, embryo survival and the establishment of pregnancy (Robinson et al., 2000; Meikle et al., 2004; Chagas et al., 2007; Lucy, 2008; Berkowicz et al., 2012; Lirón et al., 2012). Usually, these processes require a lot of energy from the different metabolic pathways, as of lipid, glucose and insulin metabolism (Gong, 2002; Ashworth et al., 2009). In dairy cattle, the negative energy balance causes impaired reproductive performance caused by the changes in levels of insulin, glucose, *IGF-1* and non-esterified fatty acids (Meikle et al., 2004; Roche et al., 2011; Lucy et al, 2014) and the gene set identified (*IGF-1, IGF-2, MOS, ATP13A1, TSSK6, PMCH, NDUFA13, PENK, PLAG1, LAPR2, SAV1* and *INS*) shows association with these physiological changes

3.5 Conclusion

Our results show shared genomic regions and metabolic pathways associated with reproductive traits in different environmental conditions. For all traits, SNP effects in these regions indicated a strong SNP x environmental condition interaction. The pleiotropic genomic regions, on BTA 5, 7, 10, 14, 16 and 29, harbor functional candidate genes affecting sexual precocity and reproductive traits in Nellore heifers and young bulls. The major pleiotropic genes identified are directly implicated in the regulation of reproductive pathways. These genes include a striking number of energetic metabolism regulators controlling reproductive traits. These results contributed to a better understanding of the physiological mechanisms involved in Nellore reproductive traits under different environmental conditions.

3.6 References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) 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.

Aken BL, Ayling S, et al. (2016) The Ensembl gene annotation system. **Database : the journal of biological databases and curation** 2016:1–19.

Akers RM, Ellis SE, Berry SD (2005) Ovarian and IGF-I axis control of mammary development in prepubertal heifers. **Domestic Animal Endocrinology** 29:259–267.

Ashworth CJ, Toma LM, Hunter MG (2009) Nutritional effects on oocyte and embryo development in mammals: implications for reproductive efficiency and environmental sustainability. **Philosophical Transactions of the Royal Society B** 364:3351–3361.

Benjamini Y and Hochberg Y 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society** 57, 289–300.

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.

Bertoldo MJ, Faure M, Dupont J, Froment P (2015) AMPK: a master energy regulator for gonadal function. **Frontiers in Neuroscience** 9:1–11.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. **Journal of Statistical Mechanics: Theory and Experiment** 10:1–12.

Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, Tier B, Savin K, Hayes BJ, Goddard ME 2014. A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. **PLOS Genetics** 10:e1004198.

Brisard D, Desmarchais A, Touzé JL, Lardic L, Freret S, Elis S, Nuttinck F, Camous S, Dupont J, Uzbekova S (2014) Alteration of energy metabolism gene expression in cumulus cells affects oocyte maturation via MOS-mitogen-activated protein kinase pathway in dairy cows with an unfavorable ‘Fertil-’ haplotype of one female fertility quantitative trait locus. **Theriogenology** 81:599–612.

Burrow HM (2012) Importance of adaptation and genotype × environment interactions in tropical beef breeding systems. **Animal** 6:729–740.

Cánovas A, Reverter A, et al. (2014) Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. **PLOS ONE** 9:e102551 1-17.

Chagas LM, Bass JJ, et al. (2007) New Perspectives on the Roles of Nutrition and Metabolic Priorities in the Subfertility of High-Producing Dairy Cows. **Journal of Dairy Science** 90:4022–4032.

Cole JB, Wiggans GR, Ma L, et al. (2011) Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. **BMC Genomics** 12:1–17.

Devlin B, Roeder K (1999) Genomic Control for Association Studies. **Biometrics** 55:997–1004.

Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. **Nature Protocols** 4:1184–1191.

Evangelou E, Ioannidis JPA (2013) Meta-analysis methods for genome-wide association studies and beyond. **Nature Reviews Genetics** 14:379–389.

Fortes MRS, Lehnert SA, Bolormaa S, Reich C, Fordyce G, Corbet NJ, Whan V, Hawken RJ, Reverter A (2012) Finding genes for economically important traits: Brahman cattle puberty. **Animal Production Science** 52:143–150.

Fortes MRS, Kemper K, et al. (2013a) Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. **Animal Genetics** 44:636–647.

Fortes MRS, Reverter A, Kelly M, McCulloch R, Lehnert SA (2013b) Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. **Andrology** 1:644–650.

Fortes MRS, Li Y, Collis E, Zhang Y, Hawken RJ (2013c) The IGF1 pathway genes and their association with age of puberty in cattle. **Animal Genetics** 44:91–95.

Gualdrón Duarte JL, Cantet RJ, Bates RO, Ernst CW, Raney NE, Steibel JP (2014) Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. **BMC Bioinformatics** 15:1–11.

Gong JG (2002) Influence of metabolic hormones and nutrition on ovarian follicle development in cattle: Practical implications. **Domestic Animal Endocrinology** 23:229–241.

Hawken RJ, Zhang YD, et al. (2012) Genome-wide association studies of female reproduction in tropically adapted beef cattle. **Journal of Animal Science** 90:1398–1410.

Helgeson SC, Schmutz SM (2008) Genetic variation in the pro-melanin-concentrating hormone gene affects carcass traits in Bos taurus cattle. **Animal Genetics** 39:310–315.

Hou W-L, Yin J, Alimujiang M, Yu X-Y, Ai L-G, Bao Y-Q, Liu F, Jia W-P 2017. Inhibition of mitochondrial complex I improves glucose metabolism independently of AMPK activation. **Journal of Cellular and Molecular Medicine** 22:1316–1328.

Hu Z-L, Park CA, Reecy JM (2016) Developmental progress and current status of the Animal QTLdb. **Nucleic Acids Research** 44:D827–D833.

Hsueh AJW, Kawamura K, Cheng Y, Fauser BCJM (2015) Intraovarian Control of Early Folliculogenesis. **Endocrine Reviews** 36:1–24.

Irano N, de Camargo GMF, Costa RB, Terakado APN, Magalhães AFB, Silva RMDO, Dias MM, Bignardi AB, Baldi F, Carvalheiro R, Oliveira HN, Albuquerque LG (2016) Genome-Wide Association Study for Indicator Traits of Sexual Precocity in Nellore Cattle. **PLOS ONE** 11:1–14.

Jha KN, Tripurani SK, Johnson GR (2017) TSSK6 is required for γH2AX formation and the histone-to-protamine transition during spermiogenesis. **Journal of cell science** 130:1835–1844.

Johnston DJ, Corbet NJ, Barwick SA, Wolcott ML, Holroyd RG (2014) Genetic correlations of young bull reproductive traits and heifer puberty traits with female reproductive performance in two tropical beef genotypes in northern Australia. **Animal Production Science** 54:74–84.

Juma AR, Damdimopoulou PE, Grommen SVH, Van de Ven WJM, De Groef B (2016) Emerging role of PLAG1 as a regulator of growth and reproduction. **Journal of Endocrinology** 228:R45–R56.

Karim L, Takeda H, et al. (2011) Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. **Nature Genetics** 43:405–413.

Ke Y, Han Y, Guo X, Wen J, Wang K, Jiang X, Tian X, Ba X, Boldogh I, Zeng X (2017) PARP1 promotes gene expression at the post-transcriptiona level by modulating the RNA-binding protein HuR. **Nature Communications** 8:1–15.

Khatib H, Huang W, Wang X, Tran AH, Bindrim AB, Schutzkus V, Monson RL, Yandell BS (2009) Single gene and gene interaction effects on fertilization and embryonic survival rates in cattle. **Journal of Dairy Science** 92:2238–2247.

Kirkpatrick BW, Byla BM, Gregory KE (2000) Mapping quantitative trait loci for bovine ovulation rate. **Mammalian Genome** 11:136–139.

Kraus WL (2008) Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation. **Current Opinion in Cell Biology** 20:294–302.

Leclerc GM, Bose SK, Boockfor FR (2008) Specific GATA-binding elements in the GnRH promoter are required for gene expression pulse activity: role of GATA-4 and GATA-5 in this intermittent process. **Neuroendocrinology** 88:1–16.

Lillehammer M, Hayes BJ, Meuwissen THE, Goddard ME (2009) Gene by environment interactions for production traits in Australian dairy cattle. **Journal of Dairy Science** 92:4008–4017.

Lirón JP, Prando AJ, Fernandez ME, Ripoli M V, Rogberg-Muñoz A, Goszczynski DE, Posik DM, Peral-García P, Baldo A, Giovambattista G (2012) Association between GNRHR, LHR and IGF1 polymorphisms and timing of puberty in male Angus cattle. **BMC Genetics** 13:1–6.

Llewellyn S, Fitzpatrick R, Kenny DA, Murphy JJ, Scaramuzzi RJ, Wathes DC 2007. Effect of negative energy balance on the insulin-like growth factor system in pre-recruitment ovarian follicles of post partum dairy cows. **Reproduction** 133:627–39.

Lomniczi A, Wright H, Castellano JM, Matagne V, Toro CA, Ramaswamy S, Plant TM, Ojeda SR (2015) Epigenetic regulation of puberty via Zinc finger protein-mediated transcriptional repression. **Nature Communications** 6:1–16.

Lucy M (2008) Functional Differences in the Growth Hormone and Insulin-like Growth Factor Axis in Cattle and Pigs: Implications for Post-partum Nutrition and Reproduction. **Reproduction in Domestic Animals** 43:31–39.

Lucy MC, Butler ST, Garverick HA (2014) Endocrine and metabolic mechanisms linking postpartum glucose with early embryonic and foetal development in dairy cows. **Animal** 8:82–90.

Lyu Z, Qin N, Tyasi TL, Zhu H, Liu D, Yuan S, Xu R (2016) The Hippo/MST Pathway Member SAV1 Plays a Suppressive Role in Development of the Prehierarchical Follicles in Hen Ovary. **PLOS ONE** 11:e0160896.

McClure MC, Morsci NS, Schnabel RD, Kim JW, Yao P, Rolf MM, McKay SD, Gregg SJ, Chapple RH, Northcutt SL,Taylor JF (2010) A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. **Animal Genetics** 41:597–607.

Meikle A, Kulcsar M, Chilliard Y, Febel H, Delavaud C, Cavestany D, Chilibroste P (2004) Effects of parity and body condition at parturition on endocrine and reproductive parameters of the cow. **Reproduction** 127:727–737.

Melo TP, Fortes MRS, Bresolin T, Mota LFM, Albuquerque LG, Carvalheiro R (2018) Multitrait meta-analysis identified genomic regions associated with sexual precocity in tropical beef cattle. **Journal of Animal Science** 96:4087–4099.

Mul JD, Yi C-X, et al. (2010) Pmch expression during early development is critical for normal energy homeostasis. AJP: **Endocrinology and Metabolism** 298:E477–E488.

Mulder HA (2016) Genomic Selection Improves Response to Selection in Resilience by Exploiting Genotype by Environment Interactions. **Frontiers in Genetics** 7:1–11.

Mullen MP, Berry DP, Howard DJ, Diskin MG, Lynch CO, Giblin L, Kenny DA, Magee DA, Meade KG, Waters SM (2011) Single Nucleotide Polymorphisms in the Insulin-Like Growth Factor 1 (IGF-1) Gene are Associated with Performance in Holstein-Friesian Dairy Cattle. **Frontiers in genetics** 2:1–9.

Murray JF, Adan RA, Walker R, Baker BI, Thody AJ, Nijenhuis WA, Yukitake J, Wilson C A (2001) Melanin-Concentrating Hormone, Melanocortin Receptors and Regulation of Luteinizing Hormone Release. **Journal of Neuroendocrinology** 12:217–223.

Naufahu J, Cunliffe AD, Murray JF (2013) The roles of melanin-concentrating hormone in energy balance and reproductive function: Are they connected? **Reproduction** 146:R141-50.

Nicolini P, Carriquiry M, Meikle A (2013) A polymorphism in the insulin-like growth factor 1 gene is associated with postpartum resumption of ovarian cyclicity in Holstein-Friesian cows under grazing conditions. **Acta Veterinaria Scandinavica** 55:1–8.

Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, Sugimoto Y, Takasuga A (2012) Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. **BMC Genetics** 13:1–11.

Parillo F, Maranesi M, Mignini F, Marinelli L, Di Stefano A, Boiti C, Zerani M (2014) Evidence for a dopamine intrinsic direct role in the regulation of the ovary reproductive function: In Vitro study on rabbit corpora lutea. **PLoS ONE** 9:e104797.

Pau KYF, Hess DL, Kohama S, Bao J, Pau CY, Spies HG (2000) Oestrogen upregulates noradrenaline release in the mediobasal hypothalamus and tyrosine hydroxylase gene expression in the brainstem of ovariectomized rhesus macaques. **Journal of Neuroendocrinology** 12:899–909.

Pereira AGT, Utsunomiya YT, Milanesi M, Torrecilha RBP, Carmo AS, Neves HHR, Carvalheiro R, Ajmone-Marsan P, Sonstegard TS, Sölkner J, Contreras-Castillo CJ, Garcia JF (2016) Pleiotropic Genes Affecting Carcass Traits in Bos indicus (Nellore) Cattle Are Modulators of Growth. **PLOS ONE** 11:1–13.

Pissios P, Bradley RL, Maratos-Flier E (2006) Expanding the Scales: The Multiple Roles of MCH in Regulating Energy Balance and Other Biological Functions. **Endocrine Reviews** 27, 606–620.

Qu H-Q, Tien M, Polychronakos C (2010) Statistical significance in genetic association studies**. Clinical and investigative medicine** 33:E266-70.

Rauw WM, Gomez-Raya L (2015) Genotype by environment interaction and breeding for robustness in livestock. **Frontiers in Genetics** 6:1–15.

Robinson RS, Mann GE, Gadd TS, Lamming GE, Wathes DC (2000). The expression of the IGF system in the bovine uterus throughout the oestrous cycle and early pregnancy. **The Journal of endocrinology** 165:231–43.

Robinson JJ, Ashworth CJ, Rooke JA, Mitchell LM, McEvoy TG (2006) Nutrition and fertility in ruminant livestock. **Animal Feed Science and Technology** 126:259–276.

Roche JR, Burke CR, Meier S, Walker CG (2011) Nutrition x reproduction interaction in pasture-based systems: is nutrition a factor in reproductive failure? **Animal Production Science** 51:1045–1066.

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. **BMC genomics** 15:1–12.

Seo M, Kim K, Yoon J, Jeong JY, Lee HJ, Cho S, Kim H (2016) RNA-seq analysis for detecting quantitative trait-associated genes. **Scientific Reports** 6:24375.

Serova LI, Gueorguiev V, Cheng S-Y, Sabban EL 2008. Adrenocorticotropic hormone elevates gene expression for catecholamine biosynthesis in rat superior cervical ganglia and locus coeruleus by an adrenal independent mechanism. **Neuroscience** 153:1380–9.

Smith BJ (2007) boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. **Journal of Statistical Software** 21:1–37.

Smith LC, Therrien J, Filion F, Bressan F, Meirelles F V (2015) Epigenetic consequences of artificial reproductive technologies to the bovine imprinted genes SNRPN, H19/IGF2, and IGF2R. **Frontiers in genetics** 6:1–6.

Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TPL, Sonstegard TS, Thallman RM, Bennett GL (2010) Genome-wide association study of growth in crossbred beef cattle. **Journal of Animal Science** 88:837–848.

Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex traits: challenges and strategies. **Nature Reviews Genetics** 14:483–495.

Sliwowska JH, Fergani C, Gawałek M, Skowronska B, Fichna P, Lehman MN (2014) Insulin: Its role in the central control of reproduction. **Physiology & Behavior** 133:197–206.

Szklarczyk D, Franceschini A, et al. (2015) STRING v10: Protein-protein interaction networks, integrated over the tree of life. **Nucleic Acids Research** 43:D447–D452.

Takasuga A (2016) PLAG1 and NCAPG-LCORL in livestock. **Animal Science Journal** 87:59–167.

Turner N, Li J-Y, et al. (2008) Berberine and Its More Biologically Available Derivative, Dihydroberberine, Inhibit Mitochondrial Respiratory Complex I: A Mechanism for the Action of Berberine to Activate AMP-Activated Protein Kinase and Improve Insulin Action. **Diabetes** 57:1414–1418.

Van Dyck F, Declercq J, Braem C V, Van De Ven WJM (2007) PLAG1, the prototype of the PLAG gene family: Versatility in tumour development. **International Journal of Oncology** 30:765–774.

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science** 91:4414–4423.

Velazquez MA, Spicer LJ, Wathes DC (2008) The role of endocrine insulin-like growth factor-I (IGF-I) in female bovine reproduction. **Domestic Animal Endocrinology** 35, 325–342.

Walter LJ, Gasch CA, Mcevers TJ, Hutcheson JP, Defoor P, Marquess FLS, Lawrence TE (2014) Association of pro-melanin concentrating hormone genotype with beef carcass quality and yield. **Journal of Animal Science** 92:325–331.

Wei Q, Ding W, Shi F (2013) Roles of poly (ADP-ribose) polymerase (PARP1) cleavage in the ovaries of fetal, neonatal, and adult pigs. **Reproduction** 146:593–602.

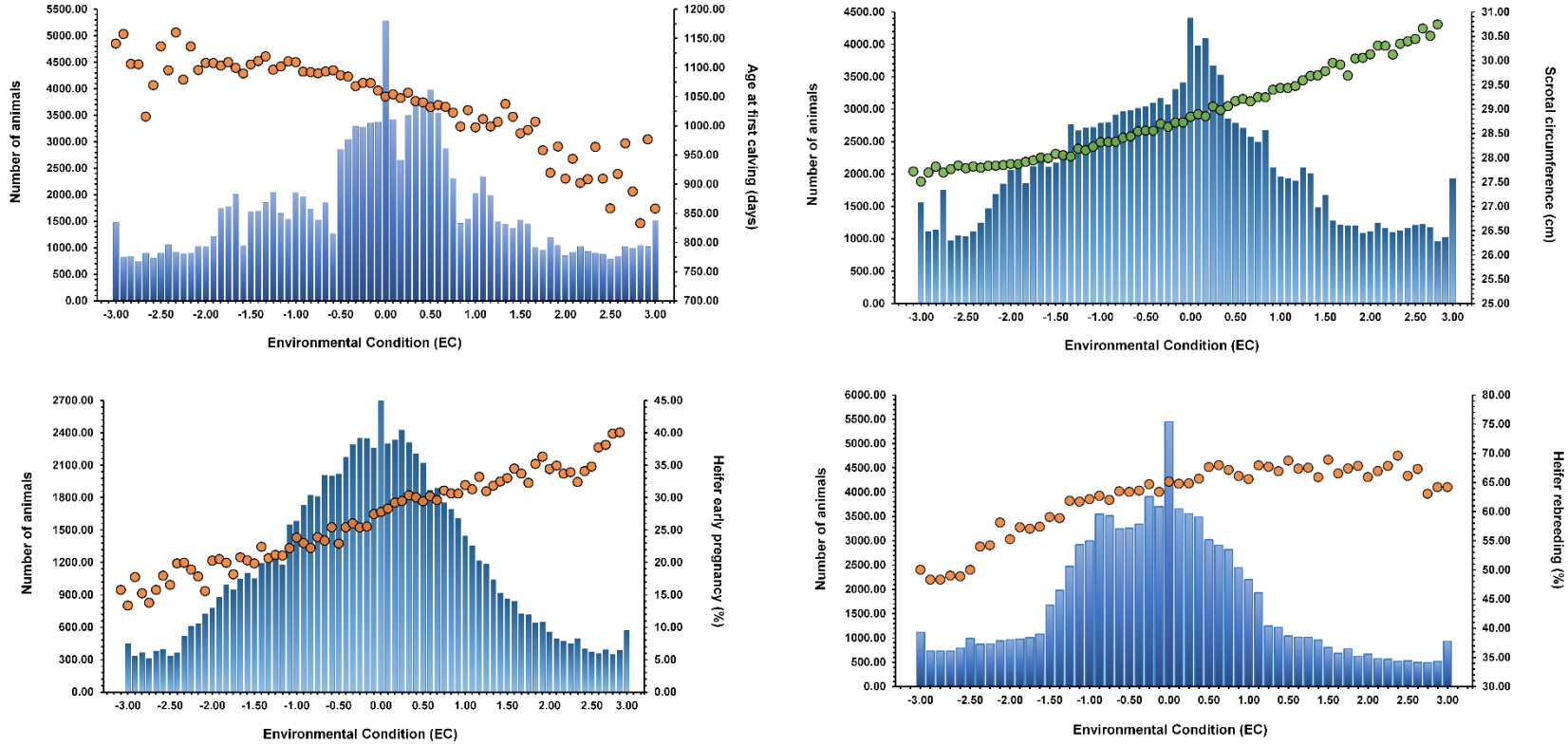
Wolfe A, Divall S, Wu S (2014) The regulation of reproductive neuroendocrine function by insulin and insulin-like growth factor-1 (IGF-1). **Frontiers in Neuroendocrinology** 35:558–572.

Wu M, Dumalska I, Morozova E, van den Pol A and Alreja M 2009. Melanin-concentrating hormone directly inhibits GnRH neurons and blocks kisspeptin activation, linking energy balance to reproduction. **Proceedings of the National Academy of Sciences** 106:17217–17222.

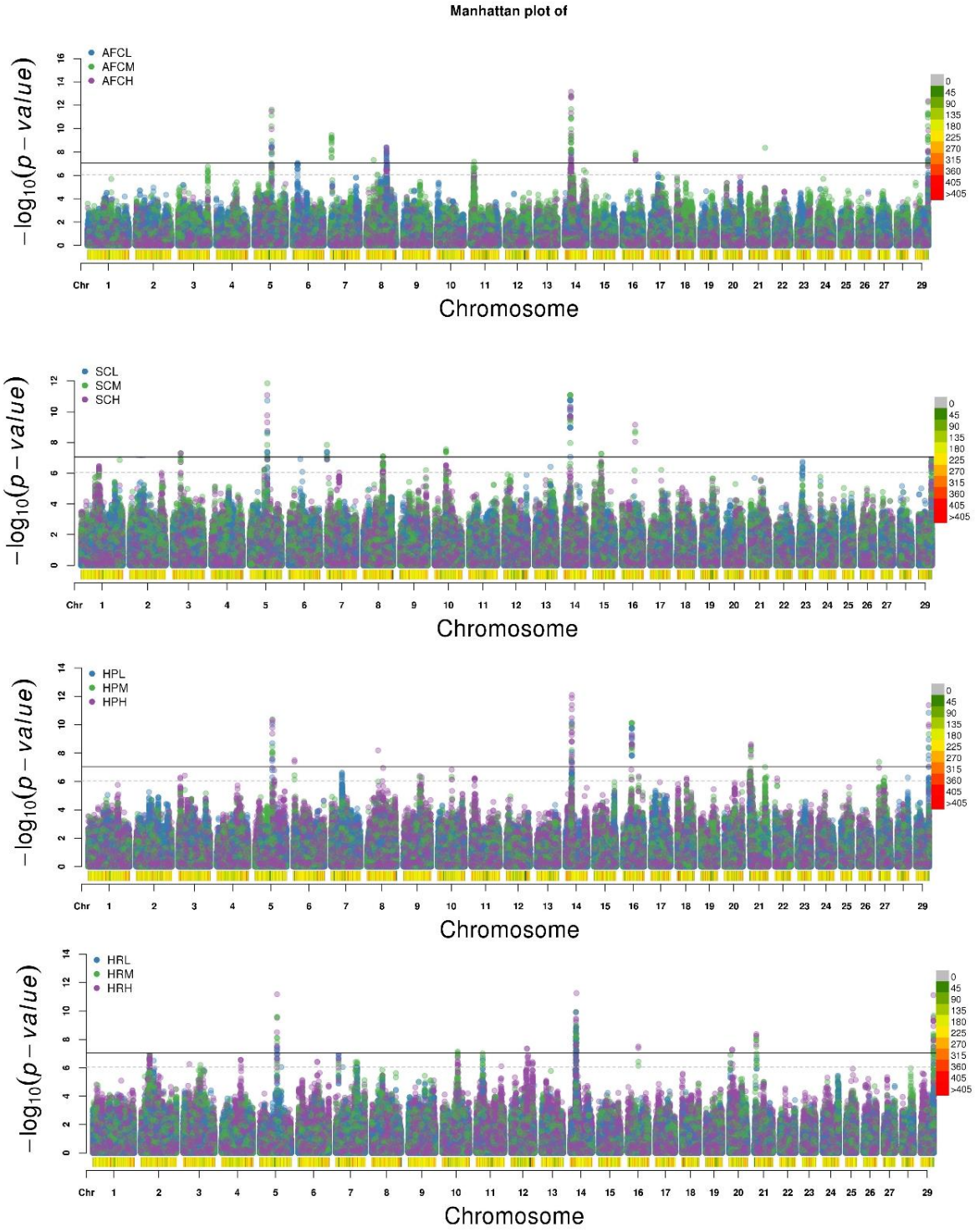
Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. **OMICS** 16:284–287.

Zoltewicz JS, Stewart NJ, Leung R, Peterson AS (2004) Atrophin 2 recruits histone deacetylase and is required for the function of multiple signaling centers during mouse embryogenesis. **Development** 131:3–14.

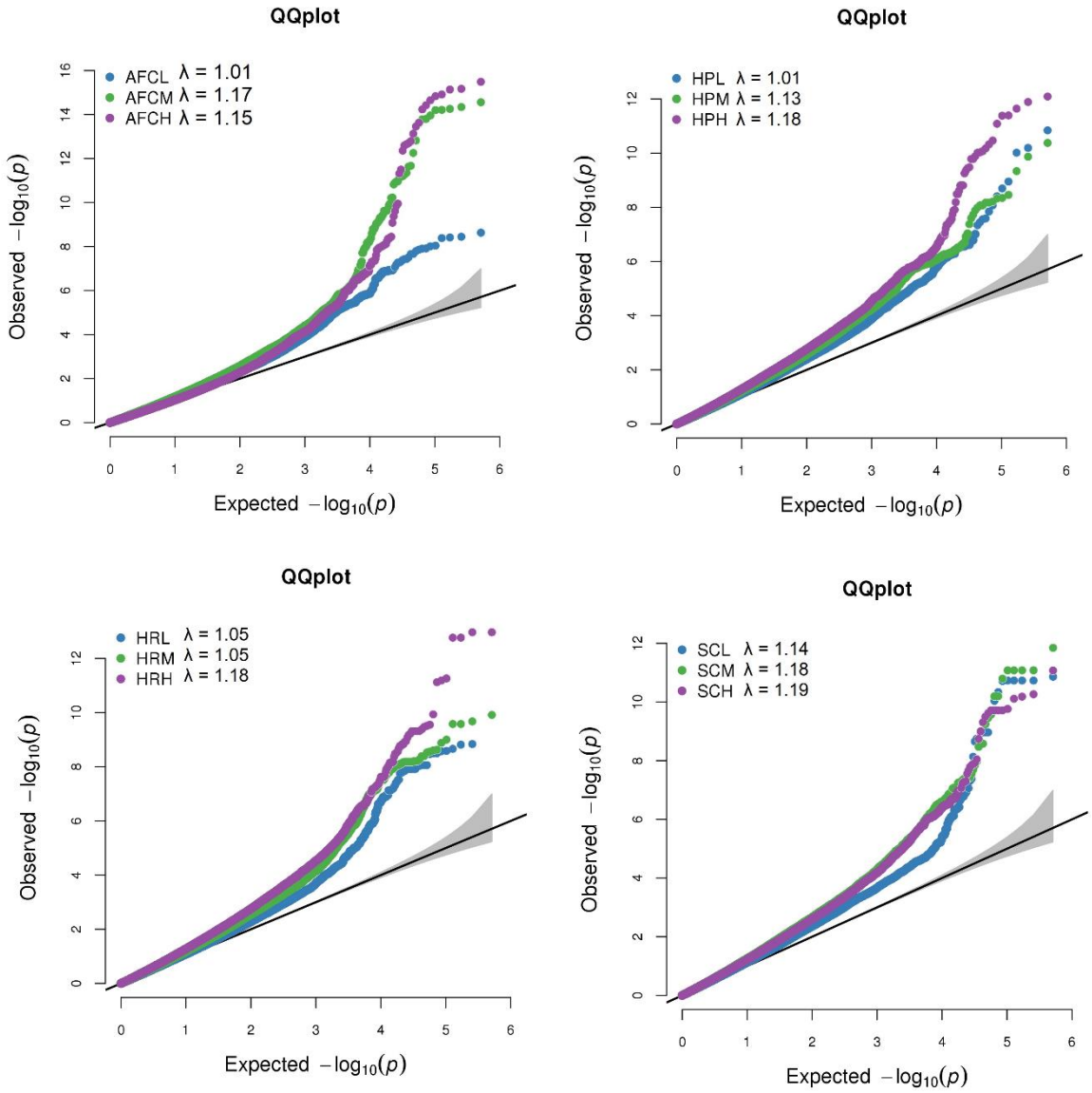
3.7 Supplementary Information



Additional file 1: Figure S1. Number of records across environmental conditions and phenotypic average for Nellore reproductive traits, age at first calving (AFC), scrotal circumference (SC), heifer early pregnancy (HP) and heifer rebreeding (HR).



Additional file 2: Figure S2. Manhattan plots for reproductive traits in Low (AFCL, SCL, HPL and HRL), Medium (AFCM, SCM, HPM and HRM) and High (AFCH, SCH, HPH and HRH) EC levels. The horizontal line indicates the genome-wide significance level of *-log$_{10}$ (p-value) > 7*.



Additional file 3: Figure S3. Quantile-quantile plot for SNP-based p-value by the chi-square test statistics (χ^2) used in the association analysis for reproductive traits in Low (AFCL, SCL, HPL and HRL), Medium (AFCM, SCM, HPM and HRM) and High (AFCH, SCH, HPH and HRH) EC levels.

Additional file 4: Table S1 – Estimates of Pearson correlations between estimated genomic breeding values (below diagonal) and heritability (on diagonal) in Nellore reproductive traits in three EC levels.

Trait	AFC_L	AFC_M	AFC_H	HP_L	HP_M	HP_H	HR_L	HR_M	HR_H	SC_L	SC_M	SC_H
AFC_L	0.23											
AFC_M	0.22	0.29										
AFC_H	-0.10	0.79	0.37									
HP_L	-0.71	-0.73	-0.45	0.26								
HP_M	-0.54	-0.80	-0.67	0.61	0.36							
HP_H	-0.37	-0.78	-0.76	0.31	0.81	0.41						
HR_L	-0.68	-0.30	-0.27	0.63	0.42	0.38	0.20					
HR_M	-0.33	-0.40	-0.32	0.52	0.51	0.46	0.59	0.25				
HR_H	-0.25	-0.37	-0.35	0.44	0.44	0.50	0.32	0.79	0.39			
SC_L	-0.37	-0.51	-0.27	0.54	0.45	0.40	0.19	0.27	0.24	0.30		
SC_M	-0.35	-0.49	-0.32	0.52	0.52	0.48	0.19	0.27	0.26	0.47	0.37	
SC_H	-0.28	-0.40	-0.33	0.42	0.43	0.49	0.16	0.24	0.28	0.20	0.83	0.38

AFC_L – age at first calving on low EC level; AFC_M – age at first calving on medium EC level; AFC_H – age at first calving on high EC level; HP_L – heifer early pregnancy on low EC level; HP_M – heifer early pregnancy on medium EC level; HP_H – heifer early pregnancy on high EC level; HR_L – heifer rebreeding on low EC level; HR_M – heifer rebreeding on medium EC level; HR_H – heifer rebreeding on high EC level; SC_L – scrotal circumference on low EC level; SC_M – scrotal circumference on medium EC level; SC_H – scrotal circumference on high EC level.

Additional file 5: Table S2 - Gene Ontology for biological process related with gene set identified in Multiple-trait meta-analysis.

ID	Description	count	p-value	q-value	Gene name
GO:0015758	glucose transport	2	0.0032123	0.0064245	INS; IGF1
GO:0034637	cellular carbohydrate biosynthetic process	2	0.0015887	0.0031775	IGF1; IMPAD1
GO:0032094	response to food	2	0.0015887	0.0031775	INS; IGF1
GO:0046323	glucose import	2	0.0012683	0.0025367	INS; IGF1
GO:0006006	glucose metabolic process	3	0.0004299	0.0008597	INS; IGF1; IGF2
GO:0019318	hexose metabolic process	3	0.0007904	0.0015809	INS; IGF1; IGF2
GO:0050994	regulation of lipid catabolic process	2	0.0012683	0.0025367	INS; PNPLA2
GO:0060416	response to growth hormone	2	0.0004253	0.0008505	INS; IGF1
GO:0032868	response to insulin	3	0.0002888	0.0005775	INS; IGF1; IGF2
GO:1901652	response to peptide	3	0.001003	0.002006	INS; IGF1; IGF2
GO:0043434	response to peptide hormone	3	0.0008304	0.0016608	INS; IGF1; IGF2
GO:0061351	neural precursor cell proliferation	3	0.0025806	0.0051612	IGF1; LYN; RERE
GO:0022008	neurogenesis	10	0.0017911	0.0035822	AP2A2; HRAS; IGF1; LYN; ASCL2; RERE; ATL1; PLAG1; TH; MOB2
GO:0045834	positive regulation of lipid metabolic process	3	0.0017194	0.0034388	INS; LYN; PNPLA2
GO:0010906	regulation of glucose metabolic process	3	0.001311	0.0026221	IGF1; IGF2; INS
GO:1903580	positive regulation of ATP metabolic process	2	0.0009151	0.0018302	IGF1; INS
GO:0043467	regulation of generation of precursor metabolites and energy	3	0.0008232	0.0016463	IGF1; IGF2; INS
GO:0044042	glucan metabolic process	3	0.0006427	0.0012854	IGF1; IGF2; INS
GO:0005977	glycogen metabolic process	3	0.0006192	0.0012384	IGF1; IGF2; INS
GO:0010628	positive regulation of gene expression	11	0.0006086	0.0012172	HRAS; IGF1; IGF2; INS; RERE; PLAG1; POLR2L; PIDD1; SAV1; TNNI2; PBX4
GO:0014013	regulation of gliogenesis	3	0.0005963	0.0011925	LYN; ASCL2; PLAG1
GO:0070873	regulation of glycogen metabolic process	3	0.000052	0.000104	IGF1; IGF2; INS
GO:0031325	positive regulation of cellular metabolic process	17	0.0000313	0.0000626	HRAS; IGF1; IGF2; INS; LYN; MOS; RERE; NDUFA13; PLAG1; POLR2L; PIDD1; PNPLA2; SAV1; TNNI2; PBX4; MOB2; LPAR2

GO:0010962	regulation of glucan biosynthetic process	3	0.000032	0.000064	IGF1; IGF2; INS
GO:0005979	regulation of glycogen biosynthetic process	3	0.000032	0.000064	IGF1; IGF2; INS
GO:0045725	positive regulation of glycogen biosynthetic process	3	0.0000041	0.0000081	IGF1; IGF2; INS
GO:0060251	regulation of glial cell proliferation	3	0.0000061	0.0000121	LYN; ASCL2; PLAG1
GO:0032868	response to insulin	3	0.0000008	0.0000016	LYN; HRAS; INS; IGF2