

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

**AVALIAÇÃO GENÔMICA EM BOVINOS DA RAÇA GIR DE BRASIL E  
COLÔMBIA**

**Alejandra Maria Toro Ospina**

Zootecnista

**2020**

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

**AVALIAÇÃO GENÔMICA EM BOVINOS DA RAÇA GIR DE BRASIL E  
COLÔMBIA**

**Alejandra Maria Toro Ospina**

Orientador: **Prof. Dr. Josineudson Augusto II de Vasconcelos Silva**

Coorientador: **Prof. Dr. Luis Gabriel Gonzales Herrera**

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

**2020**

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| O83a | <p>Ospina, Alejandra Maria Toro</p> <p>Avaliação genômica em bovinos da raça Gir de Brasil e Colômbia / Alejandra Maria Toro Ospina. -- Jaboticabal, 2020</p> <p>74 p. : il., tabs.</p> <p>Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal</p> <p>Orientador: Josineudson Augusto de Vasconcelos Silva</p> <p>Coorientador: Luis Gabriel Gonzales Herrera</p> <p>1. Runs of Homozigosity (ROH). 2. Accuracy imputation. 3. Interação genotipo ambiente. 4. Predição genômica. I. Título.</p> |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

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: AVALIAÇÃO GENÔMICA EM BOVINOS DA RAÇA GIR DE BRASIL E COLÔMBIA

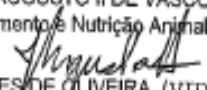
**AUTORA: ALEJANDRA MARIA TORO OSPINA**

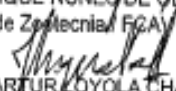
**ORIENTADOR: JOSINEUDSON AUGUSTO II DE VASCONCELLOS SILVA**

**COORIENTADOR: LUIS GABRIEL GONZALEZ HERRERA**

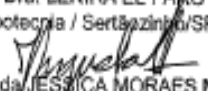
Aprovada como parte das exigências para obtenção do Título de Doutora em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:

  
Prof. Dr. JOSINEUDSON AUGUSTO II DE VASCONCELLOS SILVA  
Departamento de Melhoramento e Nutrição Animal / FMVZ/UNESP - Botucatu/SP

  
Prof. Dr. HENRIQUE NUNES DE OLIVEIRA (VIDEOCONFERÊNCIA)  
Departamento de Zootecnia / FCAV / Unesp - Jaboticabal

  
Prof. Dr. LUIS ARTUR LOYOLA CHARDULO (VIDEOCONFERÊNCIA)  
Departamento de Melhoramento e Nutrição Animal / FMVZ/UNESP - Botucatu

  
Pesquisadora Dra. LENIRA EL FARO ZADOR (VIDEOCONFERÊNCIA)  
Instituto de Zootecnia / Sertãozinho/SP

  
Pós-Doutoranda JÉSSICA MORAES MALHEIROS (VIDEOCONFERÊNCIA)  
EMBRAPA / São Carlos/SP

Jaboticabal, 06 de maio de 2020

## **DADOS CURRICULARES DO AUTOR**

**Alejandra Maria Toro Ospina** – Filha de Whiston Líder Toro Garcia e Maria del Carmen Ospina Sanchez, nascida no dia 11 de janeiro de 1992, no município de Trujillo Valle del Cauca, Colômbia. Zootecnista, graduada pela universidad de Santa Rosa de Cabal UNISARC, Colômbia, em dezembro de 2014, com trabalho de grado titulado “ efecto del reemplazo del 50% de plasma seminal em toros Gyr durante la congelacion sobre el acrosoma y la endosmosis del espermatozoide” sob a orientação do Dr. Prof. German Dario Saavedra. Em julho de 2014 realizou um estagio na área de genética, na Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista (UNESP) sob orientação do Dr. Prof. Humberto Tonhati. Em fevereiro de 2017, concluiu o Curso de Mestrado do Programa de Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias – UNESP - Câmpus de Jaboticabal, sob orientação do Dr. Prof. Josineudson Augusto II de Vasconcelos Silva. Durante o curso recebeu apoio financeiro da CAPES. Em março de 2017, Ingressou no Curso de Doutorado do mesmo programa e orientação do Prof. Dr. Josineudson Augusto II de Vasconcelos Silva e co-orientação Prof. Dr. Luis Gabriel Gonzales Herrera, da Universidad Nacional de Colômbia. Em agosto de 2018, realizou estágio no Instituto Nacional de Investigacion Agropecuaria INIA, Uruguai, durante período de quatro meses sendo supervisionado pelo Pesquisador Dr. Ignacio Aguilar.

Dedico esta tese a meus Pais, irmãos e sobrinhos pelo amor e apoio incondicional, sempre acreditando em mim e me incentivando a ir mais longe para realizar meus sonhos.

Vocês são minha inspiração para ser melhor cada dia, eu não poderia ter feito isso sem seu apoio, amo muito vocês.

**Dedicatória**

## **Agradecimentos**

Em primeiro lugar, gostaria de agradecer a Deus pela vida e força para realizar meu sonho e a minha família pelo amor incondicional.

Sou grata a todas as pessoas que ajudaram nessa fase tão importante de minha vida Profissional, principalmente agradeço a meu orientador, Prof. Augusto por acreditar em mim, pela paciência, confiança e amizade nesses anos, sempre preocupado por ajudar e aconselhar Profissional e pessoalmente. Muito obrigada por tudo.

A meus amigos e companheiros de pós-graduação, Luiz Eduardo, Ricardo, Matheus, Eduardo, Pablito, Ana beatriz obrigada pela amizade, apoio e muita paciência.

A minhas amigas Valentina, Aline, Ana, Gleize pelo apoio sempre junto ajudando em momentos difíceis e suportando-me kkk amo vocês. A minhas amigas da republica #familia Aconcheco “Fran, Mari, Aline, Gleize, Raquel” nunca vou esquecer os momentos juntas, obrigada pela amizade.

Agradeço a UNESP-FCAV pela oportunidade realizar meu doutorado e ao CAPES pela concepção da Bolsa de estudos.

Aos professores Arthur Chardulo, Rogerio Curi e Guilherme Pereira do departamento de Melhoramento e Nutrição da FMVZ, pelo apoio acadêmico, a Dra Jessica Malheiros pelo apoio e amizade na fase de doutorado. Aos professores participaram no projeto de pesquisa internacional PRISA Prof. Luis Gabriel Gonzáles (Colômbia) e Profa Ana Carolina Espansadin (Uruguai).

Esta pesquisa teve o apoio financeiro da FAPESP # 2015/06686-7, Agence Universitaire de la Francophonie (PRISA) # 1140/2018 e CNPq # 304047/2017-1.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## Sumário

|                                                                                                                                          |           |
|------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Resumo.....                                                                                                                              | 11        |
| Abstract .....                                                                                                                           | 13        |
| <b>Capítulo 1 - Considerações gerais.....</b>                                                                                            | <b>15</b> |
| 1. Introdução.....                                                                                                                       | 15        |
| 2. Objetivos.....                                                                                                                        | 16        |
| 2.1. Objetivo geral.....                                                                                                                 | 16        |
| 2.2. Objetivo específico .....                                                                                                           | 16        |
| 3. Revisão de literatura.....                                                                                                            | 17        |
| 3.1. Raça Gir .....                                                                                                                      | 17        |
| 3.2. Ilhas do ROH (Runs of Homozygosity) e coeficiente de endogamia pelo FROH .....                                                      | 18        |
| 3.3. Interação genótipo ambiente (IGA) .....                                                                                             | 19        |
| 4. Referencias .....                                                                                                                     | 21        |
| <b>Chapter 2. Genome-wide identification of runs of homozygosity islands in the Gyr breed (<i>Bos indicus</i>). .....</b>                | <b>24</b> |
| 1. Introduction .....                                                                                                                    | 24        |
| 2. Material and methods .....                                                                                                            | 26        |
| 2.1. Animals and phenotype.....                                                                                                          | 26        |
| 2.2. Imputation accuracy .....                                                                                                           | 29        |
| 2.3. Runs of homozygosity .....                                                                                                          | 29        |
| 2.4. ROH islands .....                                                                                                                   | 30        |
| 3. Results .....                                                                                                                         | 30        |
| 3.1. Runs of homozygosity (ROH) and genomic inbreeding ( $F_{ROH}$ ) .....                                                               | 30        |
| 3.2. ROH islands .....                                                                                                                   | 32        |
| 4. Discussion.....                                                                                                                       | 34        |
| 4.1. Identification of candidate genes within ROH islands.....                                                                           | 36        |
| 5. Conclusion.....                                                                                                                       | 38        |
| 6. Acknowledgments.....                                                                                                                  | 39        |
| 7. References .....                                                                                                                      | 39        |
| <b>Chapter 3: Assessing accuracy of imputation in Gyr breed using different SNP panels <i>Bos taurus</i> and <i>Bos indicus</i>.....</b> | <b>50</b> |
| Abstract .....                                                                                                                           | 50        |
| 1. Introduction .....                                                                                                                    | 50        |
| 2. Material e methods.....                                                                                                               | 51        |
| 3. Results and discussion.....                                                                                                           | 53        |



|                                                                                                                        |           |
|------------------------------------------------------------------------------------------------------------------------|-----------|
| 4. Acknowledgments.....                                                                                                | 56        |
| 5. References .....                                                                                                    | 56        |
| <b>Capítulo 4: Interação genótipo ambiente para produção de leite em bovinos da raça Gir do Brasil e Colômbia.....</b> | <b>59</b> |
| Resumo.....                                                                                                            | 59        |
| 1. Introdução.....                                                                                                     | 60        |
| 2. Material e Métodos .....                                                                                            | 61        |
| 2.1 Dados fenotípicos.....                                                                                             | 61        |
| 2.2 Genotipagem, imputação e controle de qualidade.....                                                                | 62        |
| 2.3 Análise genético-quantitativa.....                                                                                 | 62        |
| 2.4 Análises de Associação Genômica (GWAS).....                                                                        | 65        |
| 3. Resultados.....                                                                                                     | 65        |
| 3.1 Componentes de Variância .....                                                                                     | 66        |
| 3.2 Precisão dos valores de genéticos .....                                                                            | 67        |
| 3.3 ssGWAS para PL305BR e PL305COL .....                                                                               | 68        |
| 4. Discussão.....                                                                                                      | 71        |
| 4.1 Componentes de variância estimados.....                                                                            | 71        |
| 4.2 Precisão dos valores genéticos preditos.....                                                                       | 71        |
| 4.3 Mapeamento de regiões genômicas associadas à intercepto e inclinação da norma de reação .....                      | 72        |
| 5. Conclusão .....                                                                                                     | 74        |
| 6. Referências .....                                                                                                   | 74        |

# AVALIAÇÃO GENÔMICA EM BOVINOS DA RAÇA GIR DE BRASIL E COLÔMBIA

## Resumo

A raça Gir (*Bos indicus*) é importante recurso genético para produção de carne e leite no Brasil e em países tropicais. Estudos genômicos entre populações são de interesse para identificar regiões genômicas importantes e aplicar na seleção de características de produção. As corridas de homozigose (ROH) são regiões homozigotas contíguas do genoma, utilizadas na identificação de genes associados a características de interesse econômico, bem como na obtenção dos coeficientes de endogamia. A interação genótipo ambiente (IGA), também representa papel importante no estudo de populações de bovinos e na seleção dos melhores reprodutores para os diferentes ambientes. Com isto, os objetivos do presente estudo foram analisar o comprimento e a distribuição das ilhas ROH do genoma, com identificação dos genes presentes, além de avaliar a acurácia da imputação utilizando diferentes painéis comerciais e a interação genótipo ambiente da produção de leite em bovinos da raça Gir do Brasil e Colômbia. Na avaliação das ilhas ROH foram utilizados dados genotípicos de 173 animais selecionados para produção de carne e 291 animais selecionados para produção leiteira e obtidos os resultados via programa Plink. Análise da acurácia utilizando diferentes painéis de SNPs (GGP Bovine 30K, GGP indicus 35K e HD 777K) de 464 animais do Brasil e da Colômbia foram avaliados e os resultados comparados via correlação simples (CS) e taxa de concordância (CR). Na análise da IGA foi utilizado modelo de norma de reação bi-característica para produção de leite até 305 dias no Brasil (PL305BR) e Colômbia (PL305COL) usando os programas GIBBS3F90 e POSTGSf90, e incluía dados genotípicos (27.505 SNPs) de 464 animais. A análise ROH apresentou diferenças no número de ROH curto (ROH1-2 Mb) entre as duas populações, enquanto o número de ROH longo (ROH > 16 Mb) foi semelhante. As ilhas ROH com frequência maior que 0,5 sobrepueram vários segmentos do genoma das duas populações. Os genes identificados foram associados à produção de leite, crescimento, reprodução, resposta imune e características de resistência. A maior porcentagem de acurácia dos quatro cenários estudados, demonstrou o painel de 30K com maiores valores de CS (94%) e CR (96%), portanto, a densidade de 30K foi indicada para ser utilizada na imputação dos dados das análises posteriores que incluía dados do Brasil e Colômbia. No modelo

de norma de reação, os resultados das características PL305BR e PL305COL evidenciaram IGA pela reclassificação dos touros entre os dois países, demonstrando diferenças ambientais e a importância da aplicação desta metodologia na avaliação e seleção dos touros. Além disto, foram identificadas regiões genômicas afetando o intercepto e inclinação da norma de reação e influenciando a expressão de genes relacionados à produção de leite, imunidade e fertilidade. Os resultados contribuem no entendimento da seleção para diferentes fins (leite e carne) moldando segmentos do genoma e aumentando a frequência de regiões em homozigose. Quanto a imputação, se verificou a importância da escolha do painel para posterior imputação dos dados genotípicos de animais da raça Gir. O uso de metodologia adequada considerando a IGA se faz necessário em análise incluindo dados de PL305 do Brasil e Colômbia, e evita reclassificação dos touros nos diferentes ambientes. Conclui-se, aplicação de programa de melhoramento em animais da raça Gir com informações genômicas permite controlar os coeficientes de endogamia e identificar alterações de frequência e tamanho de regiões relacionadas as características de interesse econômico, devido ao direcionamento da seleção, além da importância da escolha de painéis de forma a obter a maior acurácia da imputação e da aplicação de modelos considerando a IGA, possibilitando maior precisão na escolha dos reprodutores.

**Palavras-chave:** *Bos taurus indicus*, bovinos de corte, duplo proposito, genes, GWAS, endogamia, leite, norma de reação, predição genômica.

## GENOMIC EVALUATION IN GIR BREED CATTLE FROM BRAZIL AND COLOMBIA

### Abstract

The Gir breed (*Bos indicus*) is an important genetic resource for meat and milk production in Brazil and in tropical countries. Genomic studies between populations are of interest to identify important genomic regions and apply in the selection of production traits. Runs of Homozygosity (ROH) are contiguous homozygous regions of the genome, used to identify genes associated with characteristics of economic interest, as well as to obtain inbreeding coefficients. The environment genotype (IGA) interaction also plays an important role in the study of bovine populations and in the selection of the best breeders for different environments. Thus, the objectives of the present study were to analyze the length and distribution of the ROH islands of the genome, with identification of the present genes, in addition to assessing the accuracy of imputation using different commercial panels and the environment genotype interaction of milk production in cattle of the Gir breed from Brazil and Colombia. In the evaluation of the ROH islands, genotypic data from 173 animals selected for beef production and 291 animals selected for milk production were used and the results were obtained through the Plink program. Accuracy analysis using different SNP panels (GGP Bovine 30K, GGP indicus 35K and HD 777K) from 464 animals from Brazil and Colombia were evaluated and the results were compared using simple correlation (CS) and concordance rate (CR). The IGA analysis used a bi-traits reaction standard model for milk production up to 305 days in Brazil (PL305BR) and Colombia (PL305COL) using the GIBBS3F90 and POSTGSf90 programs, and included genotypic data (27,505 SNPs) from 464 animals. The ROH analysis showed differences in the number of short ROH (ROH1-2 Mb) between the two populations, while the number of long ROH (ROH> 16 Mb) was similar. ROH islands with a frequency greater than 0.5 overlapped several segments of the genome of the two populations. The identified genes were associated with milk production, growth, reproduction, immune response and resistance characteristics. The highest percentage of accuracy of the four scenarios studied, demonstrated the panel of 30K with higher values of CS (94%) and CR (96%), therefore, the density of 30K was indicated to be used in the imputation of the data of the later analyzes that included data from Brazil and Colombia. In the reaction norm model, the results of the traits PL305BR and PL305COL showed IGA by the

reclassification of sires between the two countries, demonstrating environmental differences and the importance of applying this methodology in the evaluation and selection of sires. In addition, genomic regions were identified, affecting the intercept and slope of the reaction norm and influencing the expression of genes related to milk production, immunity and fertility. The results contribute to the understanding of the selection for different purposes (milk and meat) shaping segments of the genome and increasing the frequency of homozygous regions. As for the imputation, it was verified the importance of choosing the panel for later imputation of the genotypic data of Gir animals. The use of appropriate methodology considering the IGA is necessary in analysis including data from PL305 from Brazil and Colombia, and avoids reclassification of bulls in different environments. In conclusion, application of a breeding program in Gir animals with genomic information allows controlling the inbreeding coefficients and identifying changes in frequency and size of regions related to the traits of economic interest, due to the direction of selection, in addition to the importance of choice of panels in order to obtain the highest accuracy of imputation and application of models considering the IGA, enabling greater precision in the choice of sires.

**Keywords:** *Bos taurus indicus*, beef cattle, dual purpose, genes, GWAS, inbreeding, milk, reaction norm, genomic prediction.

## Capítulo 1 - Considerações gerais

### 1. Introdução

A raça Gir (*Bos indicus*) e seus cruzamentos são a base, junto com as outras raças zebuínas, para a sustentação da pecuária (leite e corte) de países tropicais e subtropicais de centro e sul America. De acordo com Verneque et al. (2000) e Ledic (2009) o Gir Leiteiro é a primeira raça leiteira do Brasil e a primeira raça zebuína do mundo a participar de programas delineados de melhoramento genético, a ter touros provados para leite pelo teste de progênie, além de proceder a avaliação genética de vacas quanto à capacidade de produzir leite. Além disso, os touros são anualmente avaliados para características de conformação e de manejo, baseando-se em medidas realizadas nas suas progênies.

Segundo Guimarães e Ledic. (2010), a raça Gir exerce destacado papel no contexto da comercialização de sêmen por ser uma raça que incorpora rusticidade, produtividade e longevidade. Animais da raça Gir com aptidão leiteira são eficientes na produção de leite a baixo custo, sendo grandes contribuintes destas características para as vacas mestiças. Apesar de representar uma parcela tão pequena da população, impressionante é o fato de o Gir Leiteiro ter vendido 508.978 doses de sêmen ate o ano 2010 (ASBIA, 2010), que correspondendo a 49,04% de todas as doses das raças leiteiras nacionais comercializadas.

No foco de incrementar e melhor a raça Gir e seus cruzamentos com os avanços científicos, foram realizados o sequenciamento do genoma da raça Gir Leiteiro no ano 2016. O feito tem importância histórica, já que é o primeiro sequenciamento do genoma de um mamífero feito por equipe 100% brasileira. O avanço científico também traz perspectivas muito otimistas para o setor produtivo, pois completa a outra metade do quebra-cabeça que forma a genética do Girolando. A cadeia leiteira detém o maior faturamento do agronegócio do Brasil e é a que mais gera emprego (Embrapa 2017). As diferentes pesquisas feitas na raça Gir são de importância para o progresso da mesma, e também para auxiliar nas pesquisas do híbrido Girolando que altamente usado comercialmente.

Apesar da crescente expansão da raça nos países tropicais, ainda é uma população pequena e precisa de monitoramento constante para reduzir a taxa de endogamia (Ledic,

2009). Os acasalamentos entre os animais importados para a criação das primeiras linhagens de Gir Leiteiro conduziu a um aumento da endogamia na população (REIS FILHO, 2006). Santana et al. (2016) estudou a estrutura genética da população de animais desta raça selecionada para leite e avaliou o efeito da endogamia sobre as características produtivas da mesma, e, constatou um aumento significativo da taxa de endogamia por geração.

Com a globalização e comercialização de sêmen e embriões da raça Gir entre países, levou a maior número de descendentes de touros provados do Brasil em diferentes países tropicais. Assim, os diferentes ambientes podem determinar a extensão em que o fenótipo é expresso. Portanto, é possível que o melhor genótipo em determinado ambiente possa não ser em outros (Warwick e Legates, 1980). Este efeito é chamado de interação genótipo ambiente, a expressão do genótipo em ambientes diferentes. Dada a importância da raça no Brasil e Colômbia há necessidade de estudar o genoma do Gir para expandir o conhecimento genético e aplicação na melhoria da raça.

## **2. Objetivos**

### **2.1. Objetivo geral**

Estudos genômicos aplicados para a raça Gir em diferentes populações

### **2.2. Objetivo específico**

- (1) Analisar o comprimento e a distribuição das ilhas ROH (Runs of Homozygosity) e identificar genes nessas regiões em bovinos da raça Gir selecionados para corte e leite.
- (2) Avaliar a precisão da imputação da população da raça Gir obtida com painéis de diferentes densidades, incluindo a raça Nelore na população de referência para realizar as imputações.

- (3) Estudar a interação genótipo ambiente para produção de leite com modelos de norma de reação com informação genômica na raça Gir de Brasil e Colômbia e identificar regiões genômicas que estão associadas a produção de leite.

### **3. Revisão de literatura**

#### **3.1. Raça Gir**

O gado Gir (*Bos indicus*) é uma raça Zebu de dupla aptidão (leite e carne) que é usada principalmente para a produção de laticínios em climas tropicais (Queiroz e Lôbo, 1993). Juntamente com os cruzamentos Holstein x Zebu, a raça Gir representa mais de 80% do leite produzido no Brasil (Embrapa, 2015). O Gir chegou ao Brasil em 1911, e tornou-se uma das raças mais comumente utilizadas na produção de leite no país (ABCGIL, 2015). O gado Gir Leiteiro sob controle oficial apresenta produção média de 3.233 kg, o que está 290% acima da média nacional. Tal resultado indica a potencialidade do Gir Leiteiro para as condições tropicais, sendo alternativa para incrementar a pecuária leiteira, seja como raça pura ou utilizada como base em diversos sistemas de cruzamento (Verneque et al., 2000).

Na Colômbia, as primeiras importações do Gir vieram do Brasil e tem sido realizadas a mais de 30 anos por meio das associações dos criadores. Nas importações recentes, a escolha foi por touros provados e matrizes com mais de 4.500 litros/lactação, com a finalidade de melhorar a produção das fazendas na Colômbia (ASOCEBU, 2015).

O Gir está presente em quase todo o Brasil, respondendo por mais de 80% de rebanhos leiteiros como raças puras ou como cruzamentos com animais da raça Holandesa, esta difusão deve-se principalmente ao fato de que esses animais podem ser criados em pastagem, possuem resistência aos endo- e ectoparasitas e à alta temperatura, com produção máxima de leite de 14 kg de leite por dia (Pereira et al., 2012). Na Colômbia, a raça Gir se encontra em climas temperados e terras altas, mas deslocou-se para áreas abaixo de 2.000 metros acima do nível do mar, onde foi descoberto que pode produzir até 20 litros de leite por dia com 4,0% de gordura e 3,5% de proteína, em alguns ambientes com melhor adaptação (Ganaderia sostenible Colômbia, 2015).



O Gir Leiteiro é a primeira raça leiteira do Brasil a participar de programas de melhoramento genético, a ter touros provados para leite pelo teste de progênie (Verneque et al. 2000). Com o intuito de avançar foi implantado o Programa Nacional de Melhoramento do Gir Leiteiro (PNMGL) da ABCGIL/EMBRAPA no ano 1985, com o objetivo de promover o melhoramento genético da raça Gir por meio da identificação e seleção de touros geneticamente superiores para as características de produção (leite, gordura, proteína, lactose e sólidos totais), conformação e manejo (Martinez et al., 2005).

### **3.2. Ilhas do ROH (Runs of Homozygosity) e coeficiente de endogamia pelo FROH**

O ROH são regiões contínuas do genoma em homozigoses, causada pelos haplotipos idênticos por descendência (IBD) derivados de um ancestral (Lencz et al., 2007). O coeficiente de endogamia é tradicionalmente estimado principalmente a partir do informações derivadas de dados de pedigree (Wright, 1922). O recente desenvolvimento de painéis de SNP de alta densidade levou a um interesse crescente no cálculo de coeficientes endogamia com informação molecular em animais (Purfield et al., 2012; Marras et al., 2014; Zavarez et al., 2015); dados moleculares são mais eficazes para estimar a auto-zigosidade e detectar os efeitos de consanguinidade que dados de pedigree (Keller et al., 2011). Com a informação molecular os coeficientes de consanguinidade, podem ser estimadas a partir genótipos homozigotos ROH (FROH). A homozigose observada pode ser definida como a proporção de locos homozigotos no indivíduo ou na população. As estimativas da FROH podem identificar historia da consanguinidade com base no comprimento da ROH, que é uma ferramenta importante genética populacional (Peripolli et al., 2016).

A seleção artificial tem sido amplamente aplicada a espécies de animais para obter fenótipos desejáveis e lucrativos. O processo de seleção resulta em raças especializadas para produção de leite ou carne bovina ou raças de dupla finalidade em diferentes regiões do mundo (Ledic, 1996). A busca por animais superiores via seleção reduziu a diversidade de haplótipos e aumentou a homozigose, gerando alta frequência de ROH em regiões com pressão de seleção (Zhang et al., 2015). As regiões genômicas que são marcas de seleção tendem a gerar Ilhas ROH, definidas como regiões genômicas com

diversidade genética reduzida e, conseqüentemente, alta homozigose ao redor do locus selecionado em comparação com o restante o genoma (Pemberton et al., 2012). É provável que estratégias de seleção artificial pressionem regiões genômicas específicas que controlam características importantes, como adaptação, reprodução e resistência a doenças.

### **3.3. Interação genótipo ambiente (IGA)**

Os potenciais produtivos e reprodutivos dos animais são expressos apenas na medida em que as suas condições ambientais o permitam. O ambiente não altera a constituição genética do indivíduo, mais os fatores ambientais determinam a extensão em que o fenótipo é expresso; Portanto, é possível que o melhor genótipo em determinado ambiente possa não ser em outros (Warwick e Legates, 1980). Assim, as interações podem envolver mudanças na classificação do genótipo ou mudanças na magnitude das variâncias genéticas, ambientais e fenotípicas em diferentes ambientes (Cerón-Muñoz et al., 2004). O problema se reflete na seleção dos animais, que não considera a influência do meio ambiente como um dos principais fatores que pode afetar a produção dos seus descendentes. Este efeito é chamado de interação genótipo ambiente (IGA), a expressão do genótipo em ambientes diferentes. Esta variável causa efeito sobre o fenótipo, causando variações nas características produtivas e reprodutivas que serão transmitidas aos seus descendentes.

A IGA é comumente analisada como uma característica que é medida nos diferentes ambientes e são tratadas como características diferentes, mas correlacionadas (Hayes et al., 2009). Entre as análises para identificar uma IGA estão às correlações genéticas onde valores  $< 0.8$  mostram que os ambientes são muito diferentes (Paterson, 1987), por exemplo, tropical vs. temperado. Outra análise utilizado é as normas de reação (NR) que descrevem o fenótipo de um indivíduo como uma função do ambiente. Um modelo de norma de reação é útil para descrever características que mudam gradual e continuamente ao longo de um gradiente ambiental, por exemplo, temperatura ou qualidade da alimentação. A inclinação de uma norma de reação linear mede a sensibilidade de um indivíduo em direção a uma mudança no ambiente (VanRaden, 2008; Yang et al., 2010). Lillehammer et al. (2009) descrevem que este efeito pode ser

quantificado utilizando dados genômicos e serem incluídos em programas de seleção de fêmeas e reprodutores que formarão a base do rebanho.

Os efeitos de genes dependentes do ambiente podem ser detectados por meio da inclusão do efeito da interação QTL e ambiente (Jarquín et al., 2014). Com o desenvolvimento da genotipagem e tecnologias de sequenciamento, em conjunto com as informações de registros disponíveis se realiza a modelagem. Como o rebanho colombiano de Gir é de origem brasileira, é importante utilizar esta informação de forma integrada no estudo genômico da raça, considerando a identificação dos efeitos da interação genótipo ambiente na produção de leite. Alguns estudos têm relatado interações genótipo ambiente para a produção de leite em gado leiteiro (Calus et al., 2002; Kolmodin et al., 2002). Os genes são afetados por mudanças ambientais (alimentação, manejo, sistema produtivo) isto implica que alguns genes têm efeitos diferentes de acordo com o ambiente e pode ser de interesse a detecção destes genes para aumentar a produtividade dos animais (Lillehammer et al., 2008).

A utilização da informação genotípica no melhoramento genético animal tornou-se área de pesquisa abrangente e altamente prioritária nos últimos anos. O principal foco tem sido o desenvolvimento de métodos para estimar os valores genéticos com base na informação genotípica, fenotípica e de pedigree (Meuwissen et al., 2001). O valor genético calculado a partir dos efeitos estimados dos marcadores é muitas vezes chamado de GEBV e o valor combinado entre DGEVBV e o valor genético tradicional (EBV), frequentemente chamado de valor genético genômico (GEBV). A acurácia da DGEVBV e GEBV depende de: 1) o nível do DL entre os marcadores e os QTL (loci para característica quantitativa); 2) o número de animais na população de treinamento; 3) a estimativa de herdabilidade da característica e 4) a distribuição dos efeitos dos QTL (Hayes et al., 2009).

Aguilar et al. (2010) implementou e comparou três métodos para obtenção dos valores genéticos utilizando a informação genômica. Após a verificação da acurácia dos métodos foi proposto que a avaliação genética nacional de único-passo com a matriz de relacionamento pelo pedigree incluindo a informação genômica fornece predições mais acuradas e viés comparado ao procedimento de múltiplos passos e ainda permite levar em consideração a estrutura de pedigree da população.

#### 4. Referencias

Aguilar I.; Misztal I.; Johnson D.L.; Legarra A.; Tsuruta S.; Lawlor T.J. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **J. Dairy Sci.**, v.93, p.743–752. 2010.

Associação brasileira dos criadores de Gir leiteiro. ABCGIL 2016  
<http://girleiteiro.org.br/>

Asociacion colombiana de criadores de ganado cebu ASOCEBU. 2015.  
<http://www.asocebu.com/index.php/el-cebu/razas/gyr>

ASBIA. Associação Brasileira de Inseminação Artificial. A ASBIA apresenta o relatório de comercialização de sêmen do ano 2009. Disponível em: <<http://www.asbia.org.br/novo/relatorios/>>. Acesso em: set.2017.

Calus M.P.; Groen A.F.; Jong, G. Genotype x environment interaction for protein yield in Dutch dairy cattle as quantified by different models. **J Dairy Sci.** v.11, p.3115-3123, 2002.

Cerón-Muñoz, M.F.; Tonhati, H.; Costa, C.N.; Maldonado-Estrada, J.; Rojas-Sarmiento, D. Genotype - Environment Interaction for Age at First Calving in Brazilian and Colombian Holstein. **J Dairy Sci.** v.87, p.2455–2458, 2004.

Empresa Brasileira De Pesquisa Agropecuária (EMBRAPA). Pesquisadores brasileiros sequenciam o genoma do Gir Leiteiro. **Produção animal**, 2015.

Falconer, D.S. Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. **Genetical Research**, v.56, p.57-70, 1990.

Falconer, D.S.; Mackay, T.F.C. Introduction to quantitative genetics. 4th ed. Essex: Longman: Pearson, 1996. 464p

Ferreira, M. B. D. Programa núcleo MOET da EPAMIG, Disponível em: <<http://www.portaldogir.com/site/artigos.php?tla=2&cod=1881&pag=0>>. Acesso em: out. 2017

Goddard, M.E.; Hayes, B.J. Genome-Wide Association Studies and Linkage Disequilibrium in Cattle, in *Bovine Genomics* (ed J.E. Wornack), Wiley-Blackwell, Oxford, UK. 2012.

Guimarães, L. S.; Ledic, I. L. Programas nacionais de melhoramento do Gir leiteiro: uma abordagem crítica. FAZU em Revista, **Uberaba**, n.7, p. 191 - 198, 2010

Jarquín, D.; Crossa, J.; Lacaze, X.; Cheyron, P.D.; Daucourt, J.; Lorgeou, J.; Piraux, F.; Guerreiro, L.; Pérez, P.; Calus, M.; Burgueño, J.; Campos, G. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. **Theor Appl Genet.** v.127, p.595–607, 2014.

Kolmodin R.E.; Strandberg, E.; Madsen, Per.; Jensen, Just.; Jorjani, H. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. *Acta Agric. Scand.* **A Anim. Sci.** v.52, p.11-24, 2002.

Ledic, I. L.; Reis Filho, J. C.; Vercesi Filho, A. E.; Verneque, R.S. Avaliações Genéticas. Diferenças nas PTAs entre diferentes estimativas. **Revista Gir Leiteiro**, Uberaba, v. 9, n. 9, pag.110-116, out. 2009.

Legarra, A.; Aguilar, I.; Misztal, I. A relationship matrix including full pedigree and genomic information. **J Dairy Sci.** v.92, p.4656-4663, 2009.

Lillehammer, M.; Goddard, M. E.; Nilsen, H.; Sehested, E.; Olsen, H.G.; Lien, S.; Meuwissen, T. Quantitative Trait Locus-by-Environment Interaction for Milk Yield Traits on Bos taurus Autosome 6. **Genetics.** v.179, p.1539–1546, 2008.

Lillehammer, M.; Odegård, J.; Meuwissen T. Reducing the bias of estimates of genotype by environment interactions in random regression sire models. **Genet Sel Evol.** v.41, p.30-36, 2009.

Meuwissen T.H.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics.** v.157, p.1819-1829, 2001.

Misztal, I. FAQ for genomic selection. *Journal of Animal Breeding and Genetics*, v.128, p.245–246, 2011.

Pereira, R.J.; Verneque, R.S.; Lopes P.S.; Santana, M.L.; Lagrotta J.R.M.R.; Torres, R.A.; Vercesi Filho, A.E.; Machado, M.A. Milk yield persistency in Brazilian Gyr cattle based on a random regression model. **Genet. Mol. Res.** v.11, p.1599-1609, 2012.

Regitano, L.C.A.; Veneroni G.B. Marcadores moleculares e suas aplicações no melhoramento animal. In: Simpósio de Biologia Molecular Aplicada à Produção Animal. São Carlos - SP. 2009.

Santana, M.L, Pereira, R.J, Bignardi , A.B, Ayres, D.R, Menezes, G.R.O, Silva, L.O.C, Leroy ,G, CHC Machado, LA Josahkian, Albuquerque, L.G. Structure and genetic diversity of Brazilian Zebu cattle breeds assessed by pedigree analysis.2016. Livestock Science.

Vanraden, P.M. Efficient methods to compute genomic predictions. **J Dairy Sci.** v.91, p.4414-4423, 2008.

Verneque, R.S.; Ledic, I. L.; Martinez, M. L.; Teodoro, R. L.; Paula, L. R. De O.; Cruz, M.; Campos, J. De P. Programa de Melhoramento do Gir Leiteiro, In: SIMPÓSIO NACIONAL DE MELHORAMENTO ANIMAL, 3, 2000. Anais... Belo Horizonte: SBMA, 2000, p. 212-218.

Warwick, E.J.; Legates, J.E. Cria y mejora del ganado. McGraw-Hill, Mexico 1980. 145p.

## **Chapter 2. Genome-wide identification of runs of homozygosity islands in the Gyr breed (*Bos indicus*).**

### **Abstract**

Runs of homozygosity (ROH) are contiguous homozygous regions of the genome. These regions can be used to identify genes associated with traits of economic interest, as well as inbreeding levels. The aim of the present study was to analyze the length and distribution of ROH islands in Gyr cattle and to identify genes within these regions. A population of 173 animals selected for beef production and a population of 291 animals selected for dairy production were used. Differences in the number of short ROH ( $ROH_{1-2 \text{ Mb}}$ ) were observed between the two populations, while the number of long ROH ( $ROH_{>16 \text{ Mb}}$ ) was similar. ROH islands with the highest incidences ( $>0.50$ ) overlapped in several segments of the genome in the two populations. The genes identified were associated with milk production, growth, reproduction, immune response, and resistance traits. Our results contribute to the understanding of how selection can shape the distribution of ROH and ROH islands within the same breed when animals are selected for different purposes such as dairy or beef production.

**Keywords:** beef, dairy, dual-purpose, genes, inbreeding

### **1. Introduction**

The Zebu cattle breed (*Bos taurus indicus*) is an important genetic resource for meat and milk production in Brazil and in tropical countries. Gyr cattle are a dual-purpose Zebu breed (milk and beef) that is mainly used for dairy production in tropical climates (Queiroz and Lôbo, 1993). Together with the Holstein x Zebu crossbreed, the Gyr breed accounts for more than 80% of the milk produced in Brazil (Embrapa, 2015). It is therefore essential to apply available genomic tools to identify regions of genes that affect traits of economic interest for milk and meat production in the Gyr genome.

Artificial selection has been widely used in livestock species in order to achieve more desirable and profitable phenotypes. The selection process resulted in specialized breeds for milk or beef production or dual-purpose breeds in different regions of the world (Ledic, 1996). It is likely that artificial selection strategies will put pressure on specific genomic regions that control important traits such as adaptation, reproduction, and disease resistance.

The unique genetic patterns that persist in the genome of individuals under natural and/or artificial selection are defined as selection signatures. These signatures are generally regions of the genome that harbor functionally important sequence variants (Qanbari et al., 2014). The detection of selection signatures is a relevant issue since it has the potential to elucidate the identities of genes and mutations associated with phenotypic traits. Furthermore, this knowledge is important to better understand the process of evolution and the characteristics that have been exposed to intensive artificial selection (Brito et al., 2017). This information can thus be used as selection criteria in animal breeding programs, especially of dual-purpose breeds so that the breeder can define whether to select for milk or beef.

Runs of homozygosity (ROH) are contiguous homozygous regions of the genome, which are caused by the transmission of haplotypes that are identical by descent (IBD) (Lencz et al., 2007). The information obtained from SNP microarrays can be used to accurately determine whether a particular SNP belongs to a ROH (M. Ferenčaković et al., 2013a). Pemberton et al. (2012) found that the patterns of local levels of autozygosity are not uniform and called regions with an extremely high frequency of ROH ( $>0.50$ ) islands. The ROH islands can be used to identify selection signatures in populations under intensive selection for traits of economic interest (Kim et al., 2013). ROH have been used to identify regions of genes in dairy and beef populations (M. Ferenčaković et al., 2013b).

ROH studies have been performed on *Bos taurus* breeds selected for milk and beef production (Purfield et al., 2012). According to the authors, ROH quantify characteristics of the genomic variation that can be used to obtain information on the history of a population and for association with important production traits using ROH islands as selection signatures. Peripolli et al. (2018) identified segments of ROH islands that suggest strong selection for milk production and environmental adaptation traits in the Gyr genome (Matukumalli et al., 2009). However, the use of genomic tools such as ROH permits to identify shared segments that are more frequent in the population, establishing similarities and differences that the genome presents in the same breed selected for different purposes.

Few studies involving dual-purpose breeds have used ROH for the identification of genomic regions applicable as selection criteria. Maiorano et al. (2018) studied the genetic architecture and selection signatures of dual-purpose Gyr cattle using genomic information, but it would be important to investigate dual-purpose breeds using the new



ROH approach. Therefore, the aim of the present study was to analyze the length and distribution of ROH islands in Gyr cattle and to identify genes within these regions.

## 2. Material and methods

The experiment was conducted in accordance with animal welfare guidelines according to EU Directive 2010/63/EU for animal experiments and State Law No. 11.977 of Sao Paulo state, Brazil. All animal procedures were approved by the Ethics and Animal Handling Committee of the Instituto de Zootecnia, Nova Odessa, SP, Brazil.

### 2.1. Animals and phenotype

The genotyped animals are from two population samples selected for different traits, growth and milk traits (Table 1). The Beef production (Bp) population sample consisted of 173 animals (52 sires) from the breeding program of Instituto de Zootecnia (IZ), Sertãozinho, SP, Brazil. This program is considered unique in the Gyr breed and is aimed at improving beef production traits in a closed herd scheme. The IZ started with the introduction of founder sires in 1976 to increase genetic variability. The choice of founders was based on the body weight at 550 days of age (Razook et al., 1998). Next, the breeding program established the selection of young sires and females using the body weight at 378 days and at 550 days, respectively, as selection criteria. For this herd, the generation interval was 5.65 years, and genetic gains were 2.88 kg/year for the weight of sires after the weight gain test and 2.80 kg/year for females considering the adjusted weight at 550 days.

**Table 1.** Distribution of the number of animals (number of complete equivalent generations) according to the different characteristics of production and genotyping panel used in the Gyr breed

|        | Gyr population           |                       |                          |                       |
|--------|--------------------------|-----------------------|--------------------------|-----------------------|
|        | Beef production          |                       | Dairy production         |                       |
|        | BovineSNP30 <sup>a</sup> | BovineHD <sup>b</sup> | BovineSNP30 <sup>a</sup> | BovineHD <sup>b</sup> |
| Male   | 47 (8.6)                 | 5 (9.7)               | 15 (1,7)                 | 18 (1,7)              |
| Female | 108 (9.4)                | 13 (10.1)             | 273 (1,4)                | 0                     |

|           |           |          |           |          |
|-----------|-----------|----------|-----------|----------|
| Sub-total | 155 (9.0) | 18 (9.9) | 289 (1,6) | 18 (1,7) |
| Total     | 173 (9.5) |          | 306 (1.7) |          |

a=BeadChip (Illumina) that contains 30,609 SNPs.

b=BeadChip (Illumina) that contains 777,962 SNPs

The Dairy production (Dp) population sample consisted of 306 animals (33 sires) selected for milk traits that belonged to five farms located in the states of Minas Gerais and Sao Paulo in Brazil and are part of the population evaluated by the Brazilian Dairy Gyr Breeding Program (PMNGL). The PMNGL was established in 1985 and is run by the Associação Brasileira de Criadores de Gir Leiteiro (ABCGIL) and Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). The dairy animals were selected to identify genetically superior bulls based on progeny tests for milk production, milk composition (fat and protein), conformation, and management traits. The mean 305-day milk yield of the dairy Gyr population is  $3,000 \pm 1,500$  kg (Queiroz and Lôbo, 1993).

Genomic DNA was extracted from blood cells using the Illustra Blood GenomicPrep Mini Spin kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK). The DNA concentration in the samples was adjusted to 40 to 60 ng/μl. Genotyping was performed with the HD and LD panels (Illumina, Inc., San Diego, CA, USA) using the HiScan system (Illumina, Inc.). The animals sampled from each population (Table 1) were selected as follows: in the Bp population, 18 individuals (5 sires and 13 dams) were selected from the 173 animals evaluated based on the largest number of progenies and a greater kin relationship with the population. These animals were genotyped with the HD panel (Illumina, Inc.) that contains 777,962 SNPs and were used as reference for imputation. The remaining individuals, 155 descendants (47 males and 108 females) of these 18 parents/animals, were genotyped with the LD panel (Illumina, Inc.) that contains 30,609 SNPs. The quality of the pedigree data of Bp was evaluated based on the number of equivalent complete generations (EqG) as indicated by Maignel et al. (1996) using the Endog 4.8v software (Gutierrez and Goyache, 2005). The mean EqG was 8.6 and 10.1 for animals genotyped with the LD and eHD panels, respectively. The kinship coefficient (Goyache et al., 2003; Gutiérrez et al., 2003) between Bp animals ranged from 19.1% (between females genotyped with LD and HD) to 39.3% (between males genotyped with HD panel).

Eighteen animals (all sires) were selected from the Dp population based on the largest number of progenies and low additive relationship between them. These animals

were genotyped with the BovineHD BeadChip (Illumina, Inc.) that contains 777,962 SNPs and were used as reference for imputation. The remaining animals of this population, 289 animals (15 sires and 273 dams), were genotyped with the BovineSNP30 BeadChip (Illumina, Inc.) that contains 30,609 SNPs. The 273 dairy cows genotyped using the LD panel were chosen because they were daughters or granddaughters of the 18 reference sires, with a similar distribution of daughters per sire, and were considered a representative sample of the female population that formed the PMNGL dataset. The 15 sires were genotyped to contribute to the accuracy of imputation since they were fathers of the cows and were related to the total population. The mean EqG was 1.4 and 1.7 for animals genotyped with the BovineSNP30 and BovineHD panels, respectively. The kinship coefficient between Dp animals ranged from 1.8% (between males and females genotyped with BovineSNP30 and BovineHD, respectively) to 8.2% (between males genotyped with BovineSNP30). The kinship of Dp was lower because animals from five farms were evaluated, while animals of Bp originated from the same closed herd, increasing the kinship between animals (Table 1).

According to principal component analysis of the Gyr breed (beef and dairy) depicted in Figure 1 of the study of Maiorano et al. (2018), PC2 shows a more compact cluster in the Bp population, possibly because the population belongs to a closed herd in which few sires are used for mating, increasing the level of kinship and permitting accurate imputation of the genotypes of this population. The Dp population exhibited a more disperse cluster based on PC2, indicating a greater genetic distance and lower kinship between individuals of the population compared to Bp. Principal component analysis probably showed a more disperse pattern as the animals belonged to different commercial herds. Thus, 18 sires representative of each population were chosen and genotyped with the HD panel, data not included in the cited study (Maiorano et al., 2018). The aim of including the HD genotypes was to minimize differentiation in the populations studied and to perform genotype imputation in order to ensure greater accuracy.

Genotype imputation was performed with the FImpute program (Sargolzaei et al., 2014). This program considers a family-based approach that uses the information of short and long haplotypes, linkage disequilibrium, and relationships among individuals (Sargolzaei et al., 2014). The gains in imputation accuracy depend on the structure of the reference population and the level of relationship of the imputed animals with the

reference animals Ventura et al. (2016). Before imputation, SNPs of sex chromosomes were excluded so that only autosomes were considered in the genotype imputation process. SNPs of the LD panel that were not present in the HD panel were also excluded. Quality control of the genotypes was performed after imputation adopting a  $MAF < 0.01$  (Hong et al., 2015; Wang et al., 2016; Vergara et al., 2018) and a call rate  $< 0.90$  for SNPs. After genotype imputation and quality control, 418,086 SNPs were obtained for analysis.

## **2.2. Imputation accuracy**

Statistical analysis to obtain the performance metrics of imputation was performed using each of the 18 animals genotyped with the HD panel, removing genotypes of the HD panel that were not present in the LD panel and imputing the genotype of this animal based on the other 17 animals. This analysis was repeated 18 times for each animal genotyped with the HD panel, calculating the accuracy of the imputed SNPs with the true SNPs of the HD panel.

Two performance metrics were calculated to capture different characteristics of imputation accuracy and quality. The concordance rate, a measure of imputation accuracy (HUANG et al., 2009; CHARLES et al., 2011), was calculated as the percentage of imputed genotypes that correspond to the true genotype. In addition, simple correlations between each imputed genotype and its underlying true genotype were calculated and the mean of all correlations of imputed SNPs was obtained. As measures of accuracy were 96% and 97% (bp populations) and 95% 97% (dp population), for concordance rate and simple correlation, respectively (Toro Ospina et al., 2019).

## **2.3. Runs of homozygosity**

ROH were calculated with the PLINK 1.9 program (Purcell et al., 2007) using the following parameters: (i) 30 or more consecutive homozygous SNPs, (ii) 1 heterozygous per window; (iii) 0 missing calls in homozygous windows (iv) minimum proportion of homozygous windows in which each SNP occurs of 0.05 (Curik et al., 2014; Zavarez et al., 2015). (v) the minimum length of a ROH was set to be 1,000 kb; (vi) the required minimum SNP density considered was 1 SNP per 50 kb; (vii) the maximum gap between two consecutive SNPs was set to be 500 kb (Karimi, 2013). The

ROH were analyzed in different length classes (ROH1-2Mb, ROH2-4Mb, ROH4-8Mb, ROH8-16Mb and ROH>16Mb).

The FROH of each animal was calculated as (KELLER et al., 2011; BJELLAND et al. 2013):

$$FROHi = \frac{\sum t \text{ length}(ROHt)}{\text{lengthtotal}}$$

where,  $ROHt$  is the  $t^{\text{th}}$  ROH in genome of individual  $i$  and length total is 2,612,820 kb, which represents the total length of the bovine genome (ZIMIN et al., 2009). The ROH segments were detected using the PLINK 1.07 software (PURCELL et al., 2007). For each animal FROH ( $F_{ROH1-2Mb}$ ,  $F_{ROH2-4Mb}$ ,  $F_{ROH4-8Mb}$ ,  $F_{ROH8-16Mb}$  and  $F_{ROH>16Mb}$ ).

## 2.4. ROH islands

ROH islands were calculated with the PLINK 1.9 program (Purcell et al., 2007) using the `--homozyg-group--` command. Segments with a frequency higher than 50% were classified as the main regions indication of possible ROH islands throughout the genome. Average linkage disequilibrium was estimated with the PLINK 1.9 program using  $r^2$ , which was the most accurate measure (Espigolan et al., 2013; Hayes et al., 2004) for all SNP pair combinations within each ROH island.

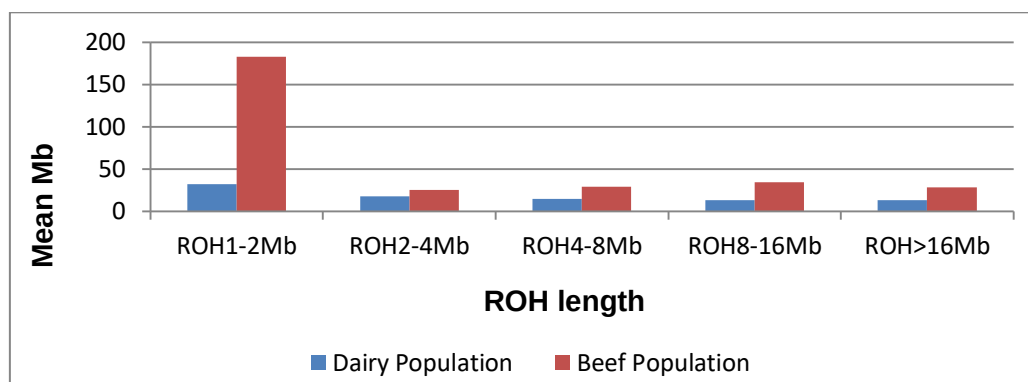
The genes located within the standard limits and the orthologs in humans were identified using the Biostatistical RStudio program (RStudio Team, 2015), using the Bioconductor BiomaRt (Ensemble Genome) R package (Durinck S et al., 2009). The genes were also analyzed with the Panther Classification System (<http://www.pantherdb.org>) to identify significant ( $P<0.05$ ) Gene Ontology (GO) terms.

## 3. Results

### 3.1. Runs of homozygosity (ROH) and genomic inbreeding ( $F_{ROH}$ )

The frequency of the different ROH length categories differed between populations (Fig. 1). The Bp population had a larger mean portion of their genome (183.6 Mb) covered by shorter length ROH ( $ROH_{1-2Mb}$ ) and the mean portion of the  $ROH_{8-16Mb}$  category was 34 Mb, indicating recent inbreeding. In the Dp population, the smallest category also covered a larger mean portion of the genome (32.04 Mb), but the value

was lower than in the Bp population (Fig. 1). For the other categories, the mean portion in the Dp population ranged from 13 to 18 Mb.



**Figure 1.** Average sum of runs of homozygosity (ROH) per animal in the Dairy and Beef populations. The sum of ROH was measured in megabase (Mb) within each length category of ROH in the two populations.

**Table 2.** Number of genotyped animals (n) and genomic inbreeding coefficients based on runs of homozygosity (FROH) for different lengths (FROH 1-2 Mb, FROH 2-4 Mb, FROH 4-8 Mb, FROH 8-16 Mb, and FROH >16 Mb).

| Inbreeding Coefficient  | Beef population and Dairy population |       |       |        |       |         |       |         |       |         |
|-------------------------|--------------------------------------|-------|-------|--------|-------|---------|-------|---------|-------|---------|
|                         | Classes                              | Mean  |       | Median |       | Minimum |       | Maximum |       | N°      |
| $F_{ROH1-2\text{ Mb}}$  |                                      | 0,178 | 0,016 | 0,169  | 0,016 | 0,152   | 0,013 | 0,172   | 0,017 | 167 285 |
| $F_{ROH2-4\text{ Mb}}$  |                                      | 0,018 | 0,006 | 0,012  | 0,006 | 0,008   | 0,005 | 0,014   | 0,007 | 167 285 |
| $F_{ROH4-8\text{ Mb}}$  |                                      | 0,013 | 0,002 | 0,010  | 0,002 | 0       | 0     | 0,013   | 0,002 | 167 281 |
| $F_{ROH8-16\text{ Mb}}$ |                                      | 0,010 | 0,003 | 0,008  | 0,003 | 0       | 0     | 0,015   | 0,003 | 167 242 |
| $F_{ROH>16\text{ Mb}}$  |                                      | 0,013 | 0,008 | 0,012  | 0,007 | 0       | 0     | 0,013   | 0,007 | 156 150 |

In the Bp population, greater average inbreeding (18%) was observed for the  $F_{ROH1-2\text{ Mb}}$  class (Table 2) in 167 animals, indicating ancient inbreeding. This percentage was probably influenced by the founders, in which some animals of the recent population still share ROH segments from the ancestors. A decrease in the percentage of inbreeding was observed for the  $F_{ROH4-8\text{ Mb}}$ ,  $F_{ROH8-16\text{ Mb}}$  and  $F_{ROH>16\text{ Mb}}$  classes, indicating the control of inbreeding in recent matings. In the Dp population, lower average inbreeding (1.6%) was observed for the  $F_{ROH1-2\text{ Mb}}$  class in 285 animals when compared to Bp. The mean

percentages in the other  $F_{ROH}$  classes were also reduced, probably because of the presence of many founders in the population studied, indicating low inbreeding and greater variability in the population.

### 3.2. ROH islands

The descriptive linkage disequilibrium ( $r^2$ ) was estimated for all SNP pair combinations within each ROH island (Supplemental Table S1). The Dp population showed intermediate levels of linkage disequilibrium in the ROH islands ranging from 0.080 to 0.290, with the highest value being observed for BTA16 (0.145). A high level of linkage disequilibrium in the ROH islands ranging from 0.135 to 0.203 was found in the Bp population, with the highest value being observed for BTA10 (0.135).

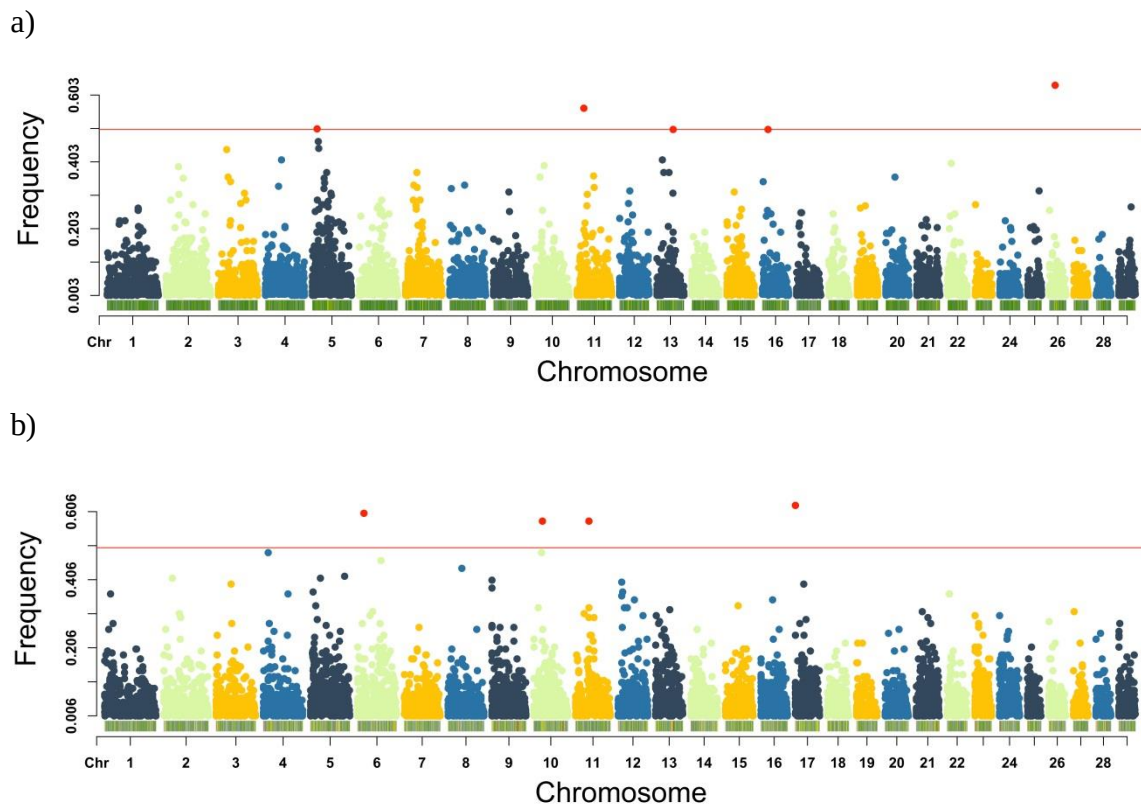


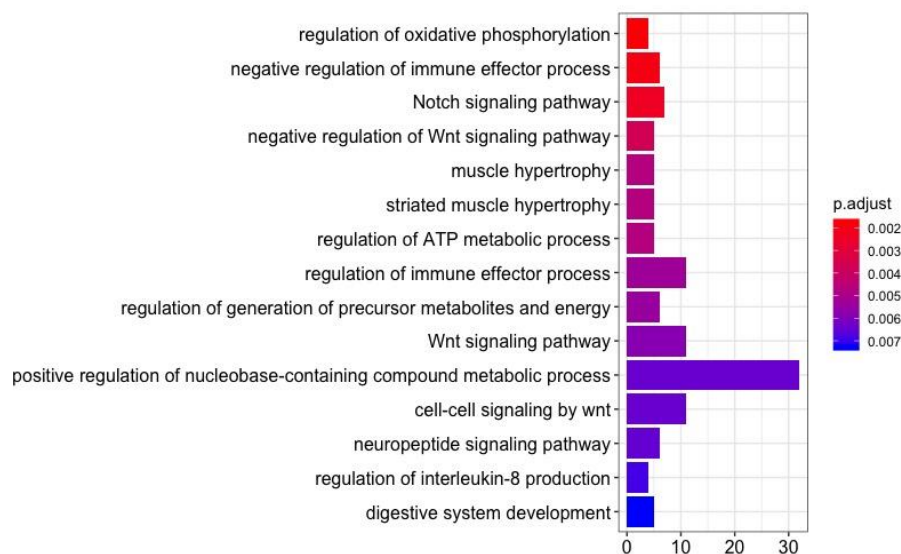
Figure 2. Manhattan plot of the distribution of ROH islands in both populations Dairy (a) and Beef (b) the Gyr breed.

ROH island segments were more frequently classified in the top regions  $> 50\%$ . The genomic distribution of ROH segments was non-uniform across chromosomes, regardless of the population. Fig. 2 We identified 4 ROH islands on chromosomes BTA

6, 10, 11 and 17 in the Bp population and 5 islands on BTA 5, 11, 13, 16 and 26 in the Dp population. The highest frequency of ROH islands in the Bp and Dp populations was found on chromosomes 17 and 26, with 65% and 61% of individuals sharing SNP positions in ROH. The second strongest pattern was observed on chromosome 6 in the Bp population, in which 60% of individuals were involved. The most common pattern was located on chromosome 11, exceeding 35% of individuals of both populations. Differences in the segments of ROH islands on the chromosomes were identified between the two populations. According to the results of principal component analysis and discriminant analysis of principal components shown in of the study of Maiorano et al. (2018), there is a genetic difference between the Bp and Dp populations and this difference reflects the variation that exists between the populations.

In the Dp and Bp populations, GO enrichment analysis was performed and genes with statistically significant functions ( $P > 0.05$ ) were identified in the categories molecular functions, cellular components (Supplemental Fig S1), and biological processes. Figure 3 shows the 15 main biological processes, including the regulation of cold-induced thermogenesis (GO:0120161), cytokine secretion (GO:0050663), fatty acid oxidation (GO:0019395), regulation of calcium-mediated signaling (GO:0050848), regulation of generation of precursor metabolites and energy (GO:0043467), muscle hypertrophy (GO:0014896), regulation of ATP metabolic process (GO:1903578), digestive system development (GO:0055123), and regulation of immune effector process (002697).

a)



b)



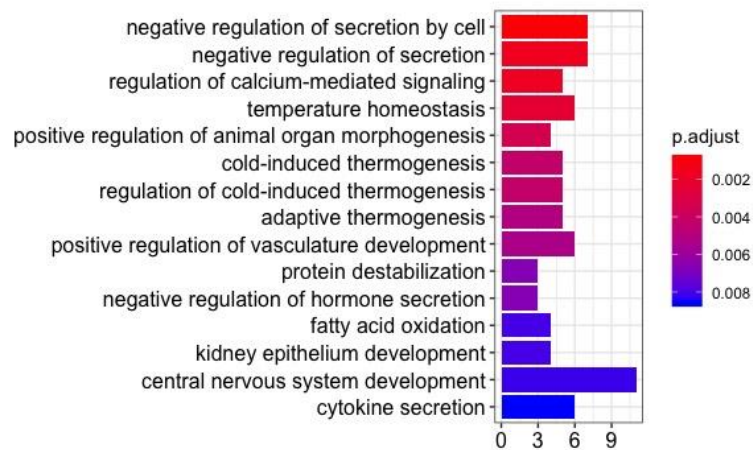


Figure 3. The 15 main biological processes (GO) terms, based on runs of homozygosity islands in populations of Dairy (a) and Beef (b) in Gyr breed.

#### 4. Discussion

Analyzing the Bp population (Fig. 1), a high frequency of  $ROH_{1-2\text{ Mb}}$  was observed compared to the other lengths studied. This finding may be due to a strong ancient relationship between animals that can be attributed to the use of few sires during the mating season, with a consequent loss of genetic variability. Similar results have been reported for native breeds from Austria and Norway (Hamzić, 2011; Purfield et al., 2012). According to Zhang et al. (2015), a high frequency of short ROH is possibly derived from ancient haplotypes that became fixed in the population. Similar inbreeding rates for the  $F_{ROH1-2Mb}$  class in the Bp population have been reported for taurine breeds (Ferencaković et al., 2012; Marras et al., 2014; Gurgul et al., 2016). Evaluation of long ROH ( $>16\text{ Mb}$ ) showed a low frequency, but a higher frequency was observed in some animals. This presence of long homozygous segments indicates recent inbreeding, mainly due to the intensive and recent selection probably resulting from the use of reproductive biotechnology (M. Ferencaković et al., 2013; Kirin et al., 2010). The selection process can cause high levels of homozygosity (Ku et al., 2011) and decrease genetic diversity, which is essential as a gene pool for evolution and reproduction.

Analysis of  $ROH_{1-2Mb}$  in the Dp population showed a low frequency of short ROH, indicating that this lineage has different populations in the country, with a greater variability of ancestors and different selection pressures. Peripolli et al. (2018) studied dairy Gyr animals and obtained an  $F_{ROH1-2Mb}$  of 1.7%, a value similar to that observed in the present study, indicating low inbreeding in the Gyr breed. The genotypes identified

in this study have not been reported in any other study, but most of the parents (sires) are the same as those of the study of Peripolli et al. (2018), which provided a similar ancient inbreeding value and shows that the two Gyr populations are derived from a small group of founders.

Evaluation of long ROH (>16 Mb) showed a low frequency. The inbreeding coefficient for the  $F_{\text{ROH}8-16\text{MB}}$  and  $F_{\text{ROH} >16\text{MB}}$  classes was considerably reduced, probably as a result of the implementation of the Dairy Gyr Breeding Program (PNMGL) in 1985, which led to the emergence of different proven sire lines that were widely used for mating, increasing genetic variability and reducing average inbreeding (Santana Junior et al., 2015). The  $F_{\text{ROH}}$  levels provided a good indication of the individual inbreeding coefficient and information about inbreeding from remote ancestors, permitting the estimation of inbreeding when pedigree data are limited or missing. This fact makes it important to study the relative levels of genetic diversity within and between populations using metrics such as inbreeding, heterozygosity (Purfield et al., 2012), and mean frequency of minor alleles to identify differences among populations of the same breed.

The ROH islands were distributed across chromosomes in the two populations (Fig. 2). Only one genomic region located on chromosome BTA11 was overlapped on the ROH islands and shared by the two populations, probably because of the recent divergence between the two lineages (dairy and beef), with ancestral segments of the breed remaining on chromosome 11. This may indicate a strong selection for different traits such as beef and milk production (Brito et al., 2017). Maiorano et al. (2018) studied the same populations (Bp and Dp) and demonstrated proximity between these populations indicated by the low fixation index ( $F_{\text{st}} = 0.003$ ), as they belong to the same breed and originated from the same ancestral population. Thus, many alleles are shared and are fixed in the two populations (Qanbari et al., 2011).

The most frequent ROH islands in the Bp population were located on BTA6 and BTA17. These regions have also been described by Szmatoła et al. (2016) in the Simmental breed, by Sölkner et al. (2014) in the Angus breed, and by Mastrangelo et al. (2018) in the Italian Reggiana breed. In the Dp population, ROH islands were most frequently identified on BTA11 and BTA26. These islands were also reported in Simmental and Limousine breeds (Szmatoła et al., 2016) and in the Italian Holstein breed (Gaspa et al., 2014; Mastrangelo et al., 2018).

The high degree of linkage disequilibrium (0.28) found in the ROH islands of the Bp population can be attributed to intensive selection that preserves ROH regions originating in the IBD segments because it is a small population in which recombination in SNPs is possibly low. Peripolli et al. (2018) found high degrees of linkage disequilibrium (0.29) across ROH islands in the Gyr breed selected for milk production.

The use of two populations of the same breed selected for different purposes, as reflected in the ROH islands, may indicate a specific selection history for each breeding objective rather than race-specific selection histories. These scenarios could facilitate the interpretation of the selection signatures. However, the inconsistencies in the criteria used to define the ROH islands impair the comparison between studies because of the lack of consensus regarding the best parameter and different signals are therefore obtained (Peripolli et al., 2017). The results obtained for the ROH,  $F_{ROH}$  and ROH islands are consistent with the history of the two populations, which underwent different breeding programs that focused on specific phenotypes through artificial selection, altering allele frequencies and producing ROH islands or selection signatures in the populations.

#### **4.1. Identification of candidate genes within ROH islands**

An enrichment of genes involved in several GO terms was performed for both Gyr populations and we focused on selected genes in highly enriched GO terms that, on the basis of the literature, showed associations with several livestock-specific traits. Therefore, the functions of candidate genes within ROH islands play an important role in cattle. Analyzing the Bp population, ROH islands were identified on BTA6, 10, 11 and 17. Identified in the biological process function on BTA17, the *UCP1* and *SIRT4* genes are related to the absorption of fatty acids, which are important for carcass quality in cattle (Jiang et al., 2009; Gui l et al., 2019). In addition, *PRKAB1* encodes the AMPK-beta1 regulatory subunit that plays a key role in glycogen metabolism, a function of enzymes involved in meat tenderness. Genes related to fertility such as *PPP3CA* and *ERP29*, which are important for sexual precocity and sperm maturation in cattle (Dias et al., 2015), and *GABRA4*, which is important for ovarian follicle development in cattle (Zielak-Steciwko et al., 2014), were also identified. On BTA6, the *TLR6-TLR10* genes were identified, which are located within a QTL region related to milk production traits and immune system process during lactation (Russell et al., 2012). Another important gene for immunity is *IL1A*, an interleukin that mediates

inflammatory processes in the first post-partum week. The proinflammatory cytokine IL1A is mainly located in the epithelium of the endometrium (Herath et al., 2009). These genes demonstrate the selection of the Bp population for growth traits over a period of more than 30 years and the correlated response in other traits. The candidate genes found in the ROH islands indicate strong selection for traits of economic importance such as growth, meat quality, immunity, fertility, and reproductive traits in beef cattle.

ROH islands in the Dp population were identified on BTA5, 11, 13, 16, and 26. On BTA16, *TGFB2* is a candidate gene that affects fertility and is involved in follicular development since it interacts with the LH and FSH reproductive hormones (Minozzi et al., 2013). In addition, the *NPPA* gene related to pregnancy (Iso-Touru et al., 2016) and the *PPM1B* gene, which is associated with the non-return rates of cows and heifers and calving to first service interval (Höglund et al., 2015), were identified. These genes are related to reproduction and fertility, traits of interest for milk production in Gyr cattle. Some growth genes were also identified such as *GLRX3*, which has been associated with feeding behavior and growth in a study on beef cattle (Liu et al., 2013), and *DDIT3* that functions as an anti-adipogenic factor, increasing intramuscular fat in cattle (Lee et al., 2013).

We identified important candidate genes for milk production and composition. The *P2RX4* and *PKD2* genes are strong candidates in which differences in the expression of the genes can affect milk water content and indirectly reduce the percentages of fat and protein in milk (Lopdell et al., 2017). *VAMP8* is one of the most important genes expressed in secretory tissues such as the mammary gland. Its expression increases the production of milk by secretory cells of the mammary gland at the beginning of lactation (Farhadian et al., 2018). In addition, *LZTS2* was found to be involved in milk fat composition in a study on Italian Holstein cows (V.Palombo et al., 2018). The *SLC25A33* gene was identified on BTA16. This gene plays an important role in immunity related to tropical adaptation (Chan et al., 2010). The genes identified highlight the good adaptation and rusticity of the Gyr breed for milk production in tropical climates, as well as its potential for the production and quality of milk solids, a fact rendering this breed an option to increase the milk demand in tropical countries.

Considering both Gyr populations, ROH islands were identified on chromosome 11. This chromosome probably preserves segments in high homozygosis characteristic of the ancient selection of the breed. On BTA11, the *ANXA4* and *POMC* genes were

identified, which are important for muscle development and carcass quality (Chaze et al., 2008; Deobald and Buchanan, 2011; Liu et al., 2013). Another gene identified, *HADHB*, mainly expresses oxidative fibers and plays a role in beef tenderness (ref). *SDC1*, involved in growth (length and hip height), is associated with morphometric traits in cattle (Liu et al., 2010). Another gene identified, *COX7A2L*, is involved in sperm motility and fertility. Several enriched GO terms were genes related to reproduction, growth traits, immune response, and resistance to diseases in the two populations studied. However, they differed in GO terms related to milk production and udder development for the Dp population and to growth and body development for the Bp population. Kim et al. (2013) found that several genomic regions within ROH were associated with economically important traits, such as milk, fat and protein production traits in the population selected for milk, and growth, fertility and adaptability traits in the population selected for beef. Therefore, the annotated genes that were assigned to these ROH islands were perceived as being exposed to intensive selection.

This indicated that the individuals analyzed may have experienced selective pressure on their genomes for these specific traits. Some genomic regions may be fixed in individuals within a population as a result of artificial or natural selection for reasons such as adaptability or productivity. The Gyr breed is well adapted to hostile environments and exhibits high levels of resistance to infections and diseases, good maternal skills, and high-quality milk production traits. The genes that are involved in these traits were detected in our study using the ROH approach and were consistent with the phenotypic characteristics of both populations.

## **5. Conclusion**

In this study, we examined the distribution of ROH and ROH islands across the genomes of two Gyr populations selected for different purposes. We confirmed that ROH levels varied in the two populations (Bp and Dp) of the same breed and that they experienced historical and recent events of inbreeding. The ROH islands on the chromosomes were not uniform in either population. Most of the genes associated with ROH islands are potentially involved in economically important traits undergoing selection and are consistent with the phenotypic traits of each population (Bp and Dp) and breed. The results suggest the use of ROH islands as possible selection signatures in breeding programs for each population (Bp and Dp) of the Gyr breed.

## 6. Acknowledgments

The authors wish to thank Instituto de Zootecnia, SP, Brazil, and Associação Brasileira de Criadores de Gir Leiteiro for providing the data used in this study. We also acknowledge the financial support of Fundação de Apoio a Pesquisa do Estado de São Paulo (FAPESP, 2015/06686-7), Agence Universitaire de la Francophonie (PRISA, 1140/2018), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 304047/2017-1), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

## Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## 7. References

- Bjelland, D.W., Weigel, K.A., Vukasinovic, N., Nkrumah, J.D., 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* <https://doi.org/10.3168/jds.2012-6435>
- Brito, L.F., Kijas, J.W., Ventura, R. V., Sargolzaei, M., Porto-Neto, L.R., Cánovas, A., Feng, Z., Jafarikia, M., Schenkel, F.S., 2017. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics*. <https://doi.org/10.1186/s12864-017-3610-0>
- Chan, E.K.F., Nagaraj, S.H., Reverter, A., 2010. The evolution of tropical adaptation: Comparing taurine and zebu cattle. *Anim. Genet.* <https://doi.org/10.1111/j.1365-2052.2010.02053.x>
- Charles, B.A., Shriner, D., Doumatey, A., Chen, G., Zhou, J., Huang, H., Herbert, A., Gerry, N.P., Christman, M.F., Adeyemo, A., Rotimi, C.N., 2011. A genome-wide association study of serum uric acid in African Americans.

BMC Med. Genomics. <https://doi.org/10.1186/1755-8794-4-17>

Chaze, T., Meunier, B., Chambon, C., Jurie, C., Picard, B., 2008. In vivo proteome dynamics during early bovine myogenesis. *Proteomics*. <https://doi.org/10.1002/pmic.200701101>

Curik, I., Ferenčaković, M., Sölkner, J., 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livest. Sci.* <https://doi.org/10.1016/j.livsci.2014.05.034>

Deobald, H.M., Buchanan, F.C., 2011. Characterization of two Pro-opiomelanocortin gene variants and their effects on carcass traits in beef cattle. *BMC Genet.* <https://doi.org/10.1186/1471-2156-12-2>

Dias, M.M., Souza, F.R.P., Takada, L., Feitosa, F.L.B., Costa, R.B., Diaz, I.D.P.S., Cardoso, D.F., Tonussi, R.L., Baldi, F., Albuquerque, L.G., Oliveira, H.N., 2015. Study of lipid metabolism-related genes as candidate genes of sexual precocity in nellore cattle. *Genet. Mol. Res.* <https://doi.org/10.4238/2015>.

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4, 1184–1191. <https://doi.org/10.1038/nprot.2009.97>

Embrapa, 2015. Pesquisadores brasileiros sequenciam o genoma do Gir Leiteiro [WWW Document]. URL <https://www.embrapa.br/busca-de-noticias/-/noticia/2634883/pesquisadores-brasileiros-sequenciam-o-genoma-do-gir-leiteiro>

Espigolan, R., Baldi, F., Boligon, A.A., Souza, F.R.P., Gordo, D.G.M., Tonussi, R.L., Cardoso, D.F., Oliveira, H.N., Tonhati, H., Sargolzaei, M., Schenkel, F.S., Carneiro, R., Ferro, J.A., Albuquerque, L.G., 2013. Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics*. <https://doi.org/10.1186/1471-2164-14-305>

Farhadian, M., Rafat, S.A., Hasanpur, K., Ebrahimi, M., Ebrahimie, E., 2018. Cross-species meta-analysis of transcriptomic data in combination with supervised machine learning models identifies the common gene signature of lactation process. *Front. Genet.* <https://doi.org/10.3389/fgene.2018.00235>

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T.R., Klemetsdal, G., Curik, I., Sölkner, J., 2013. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. *J. Anim. Breed. Genet.* <https://doi.org/10.1111/jbg.12012>

Ferenčaković, Maja, Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genet. Sel. Evol.* <https://doi.org/10.1186/1297-9686-45-42>

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C., Jensen, L.J., 2013. STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gks1094>  
Gaspa, G., Marras, G., Sorbolini, S., Ajmone-Marsan, P., Williams, J.L., Valentini, A., Dimauro, C., Macciotta, N.N.P., 2014. Genome-wide homozygosity in Italian Holstein cattle using HD panel, in: 10th World Congress of Genetics Applied to Livestock Production.

Goyache, F., Gutiérrez, J.P., Fernández, I., Gomez, E., Alvarez, I., Díez, J., Royo, L.J., 2003. Using pedigree information to monitor genetic variability of endangered populations: The Xalda sheep breed of Asturias as an example. *J. Anim. Breed. Genet.* 120, 95–105. <https://doi.org/10.1046/j.1439-0388.2003.00378.x>

Gui, L., Wu, H., Raza, S.H.A., Schreurs, N.M., Shah, M.A., 2019. The effect of haplotypes in the promoter region of SIRT4 gene on the ultrasound traits in Qinchuan cattle. *Trop. Anim. Health Prod.* <https://doi.org/10.1007/s11250-019-01881-7>



Gurgul, A., Szmatoła, T., Topolski, P., Jasielczuk, I., Żukowski, K., Bugno-Poniewierska, M., 2016. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. *J. Appl. Genet.*  
<https://doi.org/10.1007/s13353-016-0337-6>

Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J., Piedrafita, J., 2003. Pedigree analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.* 35, 43–63. <https://doi.org/10.1051/gse>

Gutiérrez, J.P., Goyache, F., 2005. A note on ENDOG: A computer program for analysing pedigree information. *J. Anim. Breed. Genet.* 122, 172–176.  
<https://doi.org/10.1111/j.1439-0388.2005.00512.x>

Hamzić, E., 2011. Levels of Inbreeding Derived from Runs of Homozygosity: A Comparison of Austrian and Norwegian Cattle Breeds. University of Natural Resources and Life Sciences Vienna.

Hayes, B.J., Visscher, P.M., McPartlan, H.C., Goddard, M.E., 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* <https://doi.org/10.1101/gr.387103>

Herath, S., Lilly, S.T., Santos, N.R., Gilbert, R.O., Goetze, L., Bryant, C.E., White, J.O., Cronin, J., Sheldon, I.M., 2009. Expression of genes associated with immunity in the endometrium of cattle with disparate postpartum uterine disease and fertility. *Reprod. Biol. Endocrinol.*  
<https://doi.org/10.1186/1477-7827-7-55>

Höglund, J.K., Buitenhuis, B., Guldbbrandtsen, B., Lund, M.S., Sahana, G., 2015. Genome-wide association study for female fertility in Nordic Red cattle. *BMC Genet.* <https://doi.org/10.1186/s12863-015-0269-x>

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* <https://doi.org/10.1038/nprot.2008.211>

Iso-Touru, T., Tapio, M., Vilkki, J., Kiseleva, T., Ammosov, I., Ivanova, Z., Popov, R., Ozerov, M., Kantanen, J., 2016. Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Anim. Genet.* <https://doi.org/10.1111/age.12473>

Jiang, Z., Michal, J.J., Chen, J., Daniels, T.F., Kunej, T., Garcia, M.D., Gaskins, C.T., Busboom, J.R., Alexander, L.J., Wright, R.W., MacNeil, M.D., 2009. Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. *Int. J. Biol. Sci.* <https://doi.org/10.7150/ijbs.5.528>

Karimi, Z., 2013. Runs of Homozygosity Patterns in Taurine and Indicine Cattle Breeds. BOKU - University of Natural Resources and Life Sciences.

Keller, M.C., Visscher, P.M., Goddard, M.E., 2011. Quantification of Inbreeding Due to Distant Ancestors and its Detection Using Dense SNP Data. *Genetics.* <https://doi.org/10.1534/genetics.111.130922>.

Kim, E.S., Cole, J.B., Huson, H., Wiggans, G.R., Van Tassel, C.P., Crooker, B.A., Liu, G., Da, Y., Sonstegard, T.S., 2013. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. *PLoS One.* <https://doi.org/10.1371/journal.pone.0080813>

Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., Mckeigue, P.M., Wilson, J.F., 2010. Genomic runs of homozygosity record population history and consanguinity. *PLoS One.* <https://doi.org/10.1371/journal.pone.0013996>

Ku, C.S., Naidoo, N., Teo, S.M., Pawitan, Y., 2011. Regions of homozygosity and their impact on complex diseases and traits. *Hum. Genet.* <https://doi.org/10.1007/s00439-010-0920-6>

Zavarez L.B., Utsunomiya Y.T., Carmo A.S., Neves H.H., Carvalheiro R., Ferencaković M., Pérez O&#39;Brien A.M., Curik I, Cole J.B., Van Tassell C.P., da Silva M.V., Sonstegard T.S., Sölkner J., Garcia J.F., 2015. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP

genotypes. *Front. Genet.* <https://doi.org/http://dx.doi.org/10.3389/fgene.2015.00005>

Ledic, I., 1996. Prova de ganho de peso na raça Gir leiteiro visando obter animal duplo aprovado, in: SMBA.

<http://sbmaonline.org.br/anais/i/trabalhos/pdfs/it24.pdf>

Lee, H.J., Jang, M., Kim, Hyeongmin, Kwak, W., Park, W.C., Hwang, J.Y., Lee, C.K., Jang, G.W., Park, M.N., Kim, H.C., Jeong, J.Y., Seo, K.S., Kim, Heebal, Cho, S., Lee, B.Y., 2013. Comparative Transcriptome Analysis of Adipose Tissues Reveals that ECM-Receptor Interaction Is Involved in the Depot-Specific Adipogenesis in Cattle. *PLoS One*.

<https://doi.org/10.1371/journal.pone.0066267>

Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Morgan, T. V., Kane, J.M., Kucherlapati, R., Malhotra, A.K., 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc. Natl. Acad. Sci.*

<https://doi.org/10.1073/pnas.0710021104>

Liu, M., Zhang, Y., Yang, M., Tian, Z., Ren, L., Zhang, S., 2012. Abundance and distribution of tetracycline resistance genes and mobile elements in an oxytetracycline production wastewater treatment system. *Environ. Sci. Technol.*

<https://doi.org/10.1021/es301145m>

Liu, Y.F., Zan, L. Sen, Li, K., Zhao, S.P., Xin, Y.P., Lin, Q., Tian, W.Q., Wang, Z.W., 2010. A novel polymorphism of GDF5 gene and its association with body measurement traits in *Bos taurus* and *Bos indicus* breeds. *Mol. Biol. Rep.*

<https://doi.org/10.1007/s11033-009-9604-5>

Lopdell, T.J., Tiplady, K., Struchalin, M., Johnson, T.J.J., Keehan, M., Sherlock, R., Couldrey, C., Davis, S.R., Snell, R.G., Spelman, R.J., Littlejohn, M.D., 2017. DNA and RNA-sequence based GWAS highlights membrane-transport genes as key modulators of milk lactose content. *BMC Genomics*.

<https://doi.org/10.1186/s12864-017-4320-3>

Maignel, L., Boichard, D., Verrier, E., 1996. Genetic Variability of French Dairy

Breeds Estimated From Pedigree Information, in: *Interbull Bull.* pp. 48–56.

Maiorano, A.M., Lourenco, D.L., Tsuruta, S., Ospina, A.M.T., Stafuzza, N.B., Masuda, Y., Filho, A.E.V., Cyrillo, J.N. dos S.G., Curi, R.A., Silva, J.A.I. de V., 2018. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. *PLoS One* 13, e0200694. <https://doi.org/10.1371/journal.pone.0200694>

Mastrangelo, S., Sardina, M.T., Tolone, M., Di Gerlando, R., Suter, A.M., Fontanesi, L., Portolano, B., 2018. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. *Animal*. <https://doi.org/10.1017/S1751731118000629>

Mastrangelo, S., Tolone, M., Sardina, M.T., Sottile, G., Suter, A.M., Di Gerlando, R., Portolano, B., 2017. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genet. Sel. Evol.* <https://doi.org/10.1186/s12711-017-0360-z>

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S., Van Tassell, C.P., 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS One*. <https://doi.org/10.1371/journal.pone.0005350>

Minozzi, G., Nicolazzi, E.L., Stella, A., Biffani, S., Negrini, R., Lazzari, B., Ajmone-Marsan, P., Williams, J.L., 2013. Genome wide analysis of fertility and production traits in Italian Holstein cattle. *PLoS One*. <https://doi.org/10.1371/journal.pone.0080219>

Palombo, V., Milanesi, M., Sgorlon, S., Capomaccio, S., Mele, M., Nicolazzi, E., Ajmone-Marsan, P., Pilla, F., Stefanon, B., D'Andrea, M., 2018. Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. *J. Dairy Sci.* <https://doi.org/10.3168/jds.2018-14413>

Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A., Li, J.Z., 2012. Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* <https://doi.org/10.1016/j.ajhg.2012.06.014>

Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R., Baldi, F., 2017. Runs of homozygosity: current knowledge and applications in livestock. *Anim. Genet.* <https://doi.org/10.1111/age.12526>

Peripolli, E., Stafuzza, N.B., Munari, D.P., Lima, A.L.F., Irgang, R., Machado, M.A., Panetto, J.C. do C., Ventura, R.V., Baldi, F., da Silva, M.V.G.B., 2018. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC Genomics.* <https://doi.org/10.1186/s12864-017-4365-3>

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* <https://doi.org/10.1086/519795>

Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G., 2012. Runs of homozygosity and population history in cattle. *BMC Genet.* <https://doi.org/10.1186/1471-2156-13-70>

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G., Simianer, H., 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics.* <https://doi.org/10.1186/1471-2164-12-318>

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T.M., Fries, R., Nielsen, R., Simianer, H., 2014. Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLoS Genet.* <https://doi.org/10.1371/journal.pgen.1004148>

Queiroz, S.A., Lôbo, R.B., 1993. Genetic relationship, inbreeding and

generation interval in registered Gir cattle in Brazil. *J. Anim. Breed. Genet.*  
<https://doi.org/10.1111/j.1439-0388.1993.tb00734.x>

Razook, A.G., Figueiredo, L.A., Bonilha Neto, L.M., 1998. Selection for yearling weight in Nelore and Guzará zebu breeds: selection applied and response in 15 years of progeny, in: 6th World Congress on Genetics Applied to Livestock Production. pp. 133–136.

RStudio Team, 2015. RStudio: Integrated Development for R. RStudio.

Russell, C.D., Widdison, S., Leigh, J.A., Coffey, T.J., 2012. Identification of single nucleotide polymorphisms in the bovine Toll-like receptor 1 gene and association with health traits in cattle. *Vet. Res.*  
<https://doi.org/10.1186/1297-9716-43-17>

Santana, M.L., Pereira, R.J., Bignardi, A.B., El Faro, L., Tonhati, H., Albuquerque, L.G., 2014. History, structure, and genetic diversity of Brazilian Gir cattle. *Livest. Sci.* <https://doi.org/10.1016/j.livsci.2014.02.007>  
Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics.* <https://doi.org/10.1186/1471-2164-15-478>

Seabury, C.M., Oldeschulte, D.L., Saatchi, M., Beever, J.E., Decker, J.E., Halley, Y.A., Bhattarai, E.K., Molaei, M., Freetly, H.C., Hansen, S.L., Yampara-Iquise, H., Johnson, K.A., Kerley, M.S., Kim, J.W., Loy, D.D., Marques, E., Neibergs, H.L., Schnabel, R.D., Shike, D.W., Spangler, M.L.,

Weaber, R.L., Garrick, D.J., Taylor, J.F., 2017. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics.* <https://doi.org/10.1186/s12864-017-3754-y>

Sölkner, J., Ferenc`aković, M., Karimi, Z., Perez O'Brien, A.M., Mészáros, G., Eaglen, S., Boison, S.A., Curik, I., 2014. Extremely non-uniform: patterns of runs of homozygosity in bovine Populations, in: 10th World Congress of

Genetics Applied to Livestock Production. Vancouver.

<https://doi.org/10.13140/2.1.1652.8960>

Szmatoła, T., Gurgul, A., Ropka-Molik, K., Jasielczuk, I., Zabek, T., Bugno-Poniewierska, M., 2016. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. *Livest. Sci.*

<https://doi.org/10.1016/j.livsci.2016.04.006>.

Toro-Ospina, Alejandra M., Maiorano Amanda Marchi., Curi, Rogério A., Pereira, Guilherme L., Zerlotti-Mercadante, Maria Eugenia., Cyrillo, Joslaine Noely dos S, Gonçalves., Aspilcueta-Borquis, Rusbel R., Silva, Josineudson A. II de V. Linkage disequilibrium and effective population size in Gir cattle selected for yearling weight. *Reproduction in domestic animals*.

<https://doi.org/10.1111/rda.13559>.

Ventura, R. V., Miller, S.P., Dodds, K.G., Auvray, B., Lee, M., Bixley, M., Clarke, S.M., McEwan, J.C., 2016. Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. *Genet. Sel. Evol.*

<https://doi.org/10.1186/s12711-016-0244-7>

Wang, X., Ma, P., Liu, J., Zhang, Q., Zhang, Yuan, Ding, X., Jiang, L., Wang, Y., Zhang, Yi, Sun, D., Zhang, S., Su, G., Yu, Y., 2015. Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. *BMC Genet.*

<https://doi.org/10.1186/s12863-015-0263-3>

Wise, A.L., Gyi, L., Manolio, T.A., 2013. eXclusion: toward integrating the X chromosome in genome-wide association analyses. *Am. J. Hum. Genet.*

<https://doi.org/10.1016/j.ajhg.2013.03.017>

Zhang, Q., Chen, H., Zhao, S., Zhang, Li, Zhang, Liangzhi, Wang, X., 2010. Polymorphisms in the promoter region of bovine PRKAB1 gene. *Mol. Biol. Rep.*

<https://doi.org/10.1007/s11033-009-9612-5>

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M.S., Sahana, G., 2015. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genomics. <https://doi.org/10.1186/s12864-015-1715-x>

Zielak-Steciwko, A.E., Browne, J.A., McGettigan, P.A., Gajewska, M., Dzięcioł, M., Szulc, T., Evans, A.C.O., 2014. Expression of microRNAs and their target genes and pathways associated with ovarian follicle development in cattle. Physiol. Genomics. <https://doi.org/10.1152/physiolgenomics.00036.2014>

Zimin, A. V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J.A., Salzberg, S.L., 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. <https://doi.org/10.1186/gb-2009-10-4-r42>.



### **Chapter 3: Assessing accuracy of imputation in Gyr breed using different SNP panels *Bos taurus* and *Bos indicus***

#### **Abstract**

The objective of this study was to evaluate the imputation accuracy achieved with SNP chips of different densities in Gyr populations, including the Nelore breed in the reference population to carry out the imputations. The database consisted of the genotypes of 18 Gyr animals and 604 Nelore animals obtained with the Illumina BovineHD BeadChip containing 777,962 SNPs. A total of 289 Gyr females from Brazil were genotyped with the GGP Bovine LDv4 chip containing 30,000 SNPs and 158 Gyr females from Colombia with the GGP indicus chip containing 35,000 SNPs. Principal component analysis (PCA) was computed for the genomic matrix and sample-specific imputation accuracies were calculated using Pearson's correlation (CS) and the concordance rate (CR) for imputed genotypes. PCA of the Nelore, Colombian Gyr and Brazilian Gyr populations demonstrated differences in the genomic relationship between the two breeds. The CS and CR ranged from 0.88 to 0.94 and from 0.93 to 0.96, respectively, for imputation of the four scenarios studied. The present results highlight the importance of the choice of chip for imputation in Gyr breed. However, the variation in SNPs may reduce the imputation accuracy even when the chip of the *Bos indicus* subspecies is used.

**Keywords:** Accuracy Imputation, genomic analyzes, tropical breeds.

#### **1. Introduction**

Zebu breeds (*Bos indicus*) are widely used in tropical and subtropical countries such as Brazil and Colombia for meat and milk production. Gyr cattle are a Zebu breed used for dairy production in tropical climates (Toro Ospina et al., 2020). The implementation of genomic selection after 2000 has permitted the development of different single nucleotide polymorphism (SNP) panels arrays for genotyping and cost reduction. Most commercial bovine chips were developed based on *Bos taurus* data, a

fact that resulted in a considerable percentage of monomorphic SNPs in *Bos indicus* populations. Although Zebu cattle play an important role in beef and dairy production in Latin America, their success depends on the availability of adequately designed SNP arrays for *Bos indicus* animals (Montaldo et al., 2012).

Research in the area of genomics requires the improvement of strategies for the use of data in genetic breeding programs. One such tool is the imputation of genotypes, which is performed based on the linkage disequilibrium of a population and on family relationships (Sargolzaei et al., 2014; Boison et al., 2015). The reliability of imputation for use of the genotypes in subsequent analyses is very important. For this purpose, accuracy tests are performed using the correlation between true and imputed SNPs, concordance rate (CR), and allelic imputation error. Studies investigating imputation strategies using different taurine breeds (Binsbergen et al., 2014; Oliveira Júnior et al., 2017) and Zebu breeds such as Nelore (Chud et al., 2015) are available in the literature.

Carvalho et al. (2014) highlighted the importance of evaluating the accuracy of imputation in order to define the best method of imputation and the most appropriate panel for genomic applications to Nelore breed. However, the remaining Zebu breeds (Gyr, Brahman and Guzará) are also of economic importance for tropical and subtropical countries and genomic studies on imputation strategies in these Zebu breeds are necessary to improve the genetic estimates of breeds for which less information is available.

In February 2017, only 4.38 million SNPs out of a total of 99.71 million dbSNP database was deposited for Indian breeds, that is, more than 96% of published SNPs are of Taurine breeds (Iqbal et al., 2019). Genotyping companies recently introduced low- and medium-density markers for Indica breeds, but still suffer from verification bias (Boison et al., 2015). Therefore, the objective of this study was to evaluate the imputation accuracy of Gyr breed population achieved with chips of different densities, including the Nelore breed in the reference population to carry out the imputations.

## **2. Material e methods**

The Brazilian dairy Gyr population used in this study belongs to five farms in the states of Minas Gerais and São Paulo and is part of the population evaluated by the Program of the Brazilian Association of Dairy Gyr Breeders (ABCGIL in the Portuguese acronym). The data of the Colombian dairy Gyr population were provided by the Colombian Association of Zebu Cattle Breeders (ASOCEBU in the Spanish acronym). The data of the Nelore population were provided by the Nelore Qualitas breeding program, Brazil.

The database consisted of the genotypes of 18 Gyr animals and 604 Nelore animals obtained with the Illumina BovineHD BeadChip (HD) (Illumina, San Diego, CA) that contains 777,962 SNPs distributed across the genome. In another group, 289 Brazilian Gyr females were genotyped with the GGP Bovine LDv4 chip (Illumina, San Diego, CA) (LD 30K) that contains 30,000 SNPs and 158 Colombian Gyr females were genotyped with the GGP indicus chip (LD 35IND) (Illumina, San Diego, CA) that contains 35,000 SNPs. The latter chip mainly comprises information from Nelore animals (69% of the data), Brahman, Guzará, Gyr and from crossbreeds such as Santa Gertrudis, Droughtmaster and tropical crosses.

Quality control of the genotypes was performed for autosomal SNPs using the preGSf90 program (Aguilar et al., 2010). The filters applied to each scenario included the elimination of SNPs with a call rate  $< 0.90$ , a minor allele frequency (MAF)  $< 0.02$ , and a maximum difference between observed and expected frequency (Hard-Weinberg equilibrium)  $> 0.15\%$ . Principal component analysis (PCA) was performed to account for stratification of the Colombian and Brazilian Gyr and Nelore populations. PCA was computed for the genomic matrix using the plotpca function of the preGSf90 program (Aguilar et al., 2010).

The population genotyped with the HD chip had SNPs removed in order to present the format of the LD 30K and LD 35IND chips. For the study, a lower density chip was created that only contained SNPs common to the LD 30K and LD 35IND chips, comprising 9,109 SNPs (LD 9K). Four imputation scenarios were investigated using different reference populations (RPop): 1) LD9K\_30KGIR with RPop containing 306 Gyr animals; 2) LD9K\_35INDGIR with RPop containing 176 Gyr animals; 3) LD9K\_30KGIR\_NEL with RPop containing 909 Gyr and Nelore animals; 4) LD9K\_35INDGIR\_NEL with RPop containing 780 Gyr and Nelore animals. Genotype

imputation was performed with the FImpute program (Sargolzaei et al., 2014). Four imputations were performed per scenario considering only  $\frac{1}{4}$  of the RPop escolhidos aleatoriamente do total da RPop. Each run was repeated 30 times to ensure that all animals were included in the repetitions. The imputation accuracy was obtained using custom programs in Linux and Python. The figures were constructed with the Biostatistical RStudio package (<http://www.R-project.org>).

The imputation accuracies were evaluated based on Pearson's correlation (CS) between true and imputed SNPs and on CR represented by the proportion of correctly imputed genotypes. The CS was calculated as values of observed and imputed genotypes, respectively. The CR with total number of markers of observed (true) genotype (0, 1, 2) for SNP of individual and the imputed genotype (0, 1, 2) for SNP of individual (Boison et al., 2015).

### **3. Results and discussion**

Principal component analysis of the Nelore, Brazilian Gyr and Colombian Gyr populations revealed genomic differences (Figure 1), demonstrating that the two Gyr populations are related through the use of semen from Brazilian sires in Columbian dams (Asocebu, 2018). This analysis also showed a difference between the Nelore and Gyr breeds, although both belong to the subspecies *Bos indicus*. The result confirms the findings of Villalobos-Cortés et al. (2015) who studied the genetic structure and diversity of Zebu breeds and described independence between the Gyr, Nelore and Guzerá breeds, supporting the diverse origin and different models of introgression in South America.

The CS between the four scenarios ranged from 0.88 to 0.94 and the CR from 0.93 to 0.96 (Table 1). The highest CS (0.94) was observed for the LD9K\_30KGIR scenario, which included the Brazilian Gyr population and RPop of 306 animals. According to Bolormaa et al. (2015), the accuracy of imputation depends on the genetic distance between animals of the reference population and of the imputed population, and allowing high accuracy in imputation. However, in the LD9K\_30KGIR\_NEL scenario with RPop of 909 animals that included the Nelore population, the CS (0.93) and CR

(0.93) were lower, although this scenario comprised the largest number of reference animals.

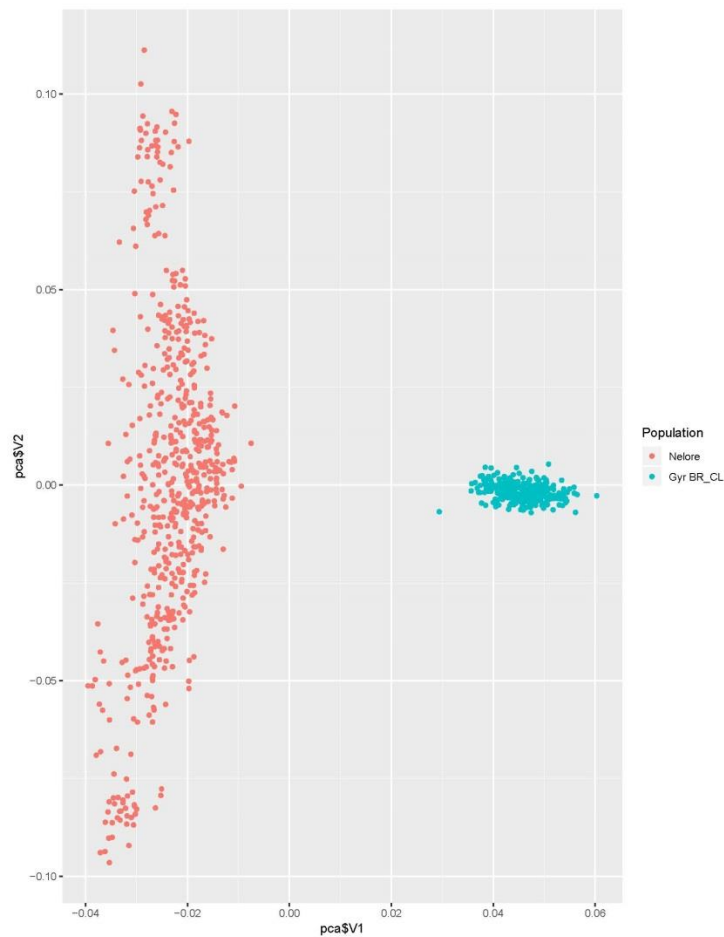


Fig 1. Principal component analysis of the genomic relationship between the Nelore and Gyr breeds (Colombia and Brazil).

Table 1. Information of the genotyping chips and imputation accuracy (SD). CS (correlation ), CR (concordance rate)

| Scenarios      | ref/imput | Imputation accuracy |                 |
|----------------|-----------|---------------------|-----------------|
|                |           | CS                  | CR              |
| LD9K_30KGIR    | 306/272   | 0.9488 (0.0498)     | 0.9646 (0.0337) |
| LD9K_35IND_GIR | 176/158   | 0.8771 (0.1356)     | 0.9315 (0.0768) |

|                    |         |                 |                 |
|--------------------|---------|-----------------|-----------------|
| LD9K_30KGIR_NEL    | 909/272 | 0.9341 (0.0754) | 0.9548 (0.0497) |
| LD9K_35IND_GIR_NEL | 780/158 | 0.8917 (0.1184) | 0.9316 (0.0669) |

---

Some studies found no significant differences in imputation accuracy when crossbred animals were used and there was no gain in accuracy when different breeds were included in the reference population (Piccoli et al., 2014). Berry et al. (2014) reported that the accuracy of imputation was reduced when different breeds are used in the reference population because of the low linkage disequilibrium between SNPs and short haplotypes of different breeds. These findings may explain the reduction in CS when the Nelore population was included as RPop in Gyr imputation.

The LD9K\_35IND\_GIR scenario exhibited the lowest CS (0.87) and CR (0.93) of the four scenarios evaluated (Table 1), probably because of the smaller RPop (176 animals) and the use of the indicus chip. However, Boison et al. (2015), performing imputation in different scenarios, obtained higher CS (0.91) and CR (0.92) values when genotypes were imputed from the 3K to the Illumina 50K chip using 171 reference animals of the Gyr breed. A possible explanation for the low imputation accuracy in the LD9K\_35IND\_GIR scenario might be the use of the indicus chip, which was customized based on the analysis of genotyped Nelore animals (69.32%), with a small percentage (7.0%) of SNPs from Gyr animals (Ferraz et al., 2018), i.e., the SNPs customized based on Nelore animals may be monomorphic for the Gyr breed. LD9K\_35IND\_GIR\_NEL scenario included in Nelore population reference observe that did not improve the accuracy of imputation. i.e., the inclusion of other animal studies with Zebu populations Gyr not helped imputation accuracy, even when tested commercial chip indicus.

Different factors influence the accuracy of imputation, such as the presence of monomorphic SNPs, low linkage disequilibrium, and a small reference population, which reduce the accuracy of imputations (ventura et al., 2016). Furthermore, the chips for Bos indicus use as a reference genotyped animals of the Nelore breed because this breed is more representative in Brazil and has a higher percentage of SNPs related to the selection objective (meat production – beef cattle). This fact results in low accuracy

when breeds with different selection objectives, such as dairy and dual-purpose breeds, are used for imputation.

In addition, the performance of different marker density panels (such as LD or HD) and different genotyping platforms need to be investigated as they present differences where some have a high determination bias in relation to the taurine breeds, showing differences in the indicus breeds (Utsunomiya et al., 2014). The present results highlight the importance of the choice of chip for imputation in Zebu breeds. However, the variation in SNPs may reduce accuracy even when the chip of the *Bos indicus* subspecies is used considering the poor representation of the other Zebu breeds in the assembly of the chips, suggesting the need for a greater representation of SNPs containing genotype data from Zebu animals other beyond the Nelore breed in the design of new chips.

#### **4. Acknowledgments**

The authors wish to thank Associação Brasileira de Criadores de Gir Leiteiro (ABCGIL), Asociación Colombiana de Criadores de Ganado Cebu (ASOCEBU), and the Nelore Qualitas breeding program, Brazil, for providing the data used in this study. We also acknowledge the financial support of Fundação de Apoio a Pesquisa do Estado de São Paulo (FAPESP, 2015/06686-7) e (FAPESP, 17/50024-4) Agence Universitaire de la Francophonie (PRISA, 1140/2018), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 304047/2017-1), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

#### **5. References**

Asociacion Colombiana de Criadores de Ganado Cebu ASOCEBU. 2018. <http://www.asocebu.com/index.php/el-cebu/razas/gyr>.

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor T. J. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, v.93, p.743:752. 2010.

Boison, S. A., D. J. A. Santos, A. H. T. Utsunomiya, R. Carneiro, H. H. R. Neves, A. M. P. O'Brien, J. F. Garcia, J. Sölkner, and M. V. G. B. da Silva. 2015. Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (*Bos indicus*) dairy cattle: Comparison of commercially available SNP chips. *J. Dairy Sci.* 98:4969–4989. [https:// doi .org/ 10 .3168/ jds .2014 -9213](https://doi.org/10.3168/jds.2014-9213).

Bolormaa, S., K. Gore, J. H. J. Van Der Werf, B. J. Hayes, and H. D. Daetwyler. 2015. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Anim. Genet.* 46:544–556. [https:// doi .org/ 10.1111/ age .12340](https://doi.org/10.1111/age.12340).

Carvalho, R., S. A. Boison, H. H. R. Neves, M. Sargolzaei, F. S. Schenkel, Y. T. Utsunomiya, A. M. P. O'Brien, J. Sölkner, J. C. McEwan, C. P. Van Tassell, T. S. Sonstegard, and J. F. Garcia. 2014. Accuracy of genotype imputation in Nelore cattle. *Genet. Sel. Evol.* 46:69. [https:// doi .org/ 10 .1186/ s12711 -014 -0069 -1](https://doi.org/10.1186/s12711-014-0069-1).

Piccoli, M. L., J. Braccini, F. F. Cardoso, M. Sargolzaei, S. G. Larmer, and F. S. Schenkel. 2014. Accuracy of genome-wide imputation in Braford and Hereford beef cattle. *BMC Genet.* 15:157 [https:// doi.org/ 10 .1186/ s12863 -014 -0157 -9](https://doi.org/10.1186/s12863-014-0157-9).

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478. [https:// doi .org/ 10 .1186/ 1471-2164 -15 -478](https://doi.org/10.1186/1471-2164-15-478).

Chud, T. C. S., R. V. Ventura, F. S. Schenkel, R. Carvalho, M. E. Buzanskas, J. O. Rosa, M. de A. Mudadu, M. V. G. B. da Silva, F. Mokry, C. R. Marcondes, L. C. A. Regitano, and D. P. Munari. 2015. Strategies for genotype imputation in composite beef cattle. *BMC Genet.* 16:99. [https:// doi .org/ 10 .1186/ s12863 -015 -0251 -7](https://doi.org/10.1186/s12863-015-0251-7).

Ventura, R. V., S. P. Miller, K. G. Dodds, B. Auvray, M. Lee, M. Bixley, S. M. Clarke, and J. C. McEwan. 2016. Assessing accuracy of imputation using different SNP panel



densities in a multibreed sheep population. *Genet. Sel. Evol.* 48:71. [https:// doi .org/10 .1186/ s12711 -016 -0244 -7](https://doi.org/10.1186/s12711-016-0244-7).

Berry, D. P., M. C. McClure, and M. P. Mullen. 2014. Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes. *J. Anim. Breed. Genet.* 131:165–172. [https:// doi .org/ 10 .1111/ jbg .12067](https://doi.org/10.1111/jbg.12067).

Oliveira Júnior, G. A., Chud, T.C. S., Ventura, R.V., Garrick , D. J., Cole, J.B., Munari D.P., Ferraz, J.B.S., Mullart, E., DeNise, Sue., Smith, S. and Silva da, M.V. G. B. 2017. Genotype imputation in a tropical crossbred dairy cattle population. *J. Dairy Sci.* 100:1–12. <https://doi.org/10.3168/jds.2017-12732>.

Ferraz, J.B.S., Wu, X., Li, H., Xu, J., Ferretti, R., Simpson, B., Walker, J., Silva, L.R., Garcia, J.F., Tait Jr, R.G., Bauck. S. 2018. Design of a low-density SNP chip for *Bos indicus*: GGP *indicus* technical characterization and imputation accuracy to higher density SNP genotypes. Header: Proceedings of the World Congress on Genetics Applied to Livestock Production 2018.

Villalobos-Cortés, Axel., Martínez, A., Vega-Pla, J.L., Landi, V., Quiroz, J, Marques, J.R., Delgado, J.V. 2015 Genetic Relationships Among Five Zebu Breeds Naturalized in America Accessed with Molecular Markers, *Italian Journal of Animal Science*, 14:2, 3280, DOI: 10.4081/ijas.2015.3280

Montaldo HH, Casas E, Ferraz JBS, Vega-Murillo VE, and Román-Ponce SI. 2012.

Opportunities and challenges from the use of genomic selection for beef cattle breeding in Latin America. *Anim Front.* 2: 23-29.

Binsbergen, R., Bink, M.C, Calus, M.P.L., Fred A van Eeuwijk, Hayes, B.J., Hulsege I and Veerkamp, R.F. 2014. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. *Genetics Selection Evolution* 201446:41 <https://doi.org/10.1186/1297-9686-46-41>

Toro Ospina AM, da Silva Faria RA, Vercesi Filho AE, et al. Genome-wide identification of runs of homozygosity islands in the Gyr breed (*Bos indicus*). *Reprod Dom Anim.* 2020;00:1–10. [https ://doi.org/10.1111/rda.13639](https://doi.org/10.1111/rda.13639)

## **Capítulo 4: Interação genótipo ambiente para produção de leite em bovinos da raça Gir do Brasil e Colômbia**

### **Resumo**

A interação genótipo ambiente (IGA) podem desempenhar um papel importante nas populações de bovinos e devem ser consideradas nos programas de melhoramento para selecionar os melhores reprodutores para diferentes ambientes. O objetivo do presente estudo foi investigar a interação genótipo ambiente para produção de leite em bovinos da raça Gir do Brasil e Colômbia com aplicação de modelo de norma de reação genômica, bem identificar regiões genômicas associadas a produção de leite nos dois países. O banco de dados de Brasil e Colômbia incluiu 464 animais (273 fêmeas, 33 reprodutores do Brasil e 158 fêmeas de Colômbia) com 27.505 SNPs. Foi utilizado o modelo animal de bi-característica para produção de leite até 305 dias no Brasil (PL305BR) e Colômbia (PL305COL), em função do país de origem dos animais e incluindo a informação genômica via modelo de norma de reação genômica em passo único (single-step). Foi usando os programas AIREMF90, GIBBS3F90 e POSTGSf90. Os resultados das características PL305BR e PL305COL evidenciam IGA pela reclassificação dos touros entre os dois países, demonstrando que Brasil e Colômbia apresentam diferenças ambientais e se torna importante considerar ambiente na escolha do reprodutor, para que sua progênie possa desempenhar adequadamente. Com isto, a reclassificação dos touros e os diferentes SNPs associados à produção de leite nos dois países, sugere que a interação genótipo ambiente é efeito importante a ser considerado na avaliação genética do Gir leiteiro do Brasil e Colômbia. Portanto, programa de melhoramento do Gir pode ser otimizado ao escolher o ambiente de seleção que possibilitará o máximo progresso genético para a produção de leite nos diferentes ambientes dentro e entre países.

**Palavras chaves:** Bos taurus indicus, GWAS, Norma de reação, predição genômica.

## 1. Introdução

Os zebuínos são a base da pecuária de países tropicais e subtropicais como Brasil, Colômbia, México, Peru e Venezuela (ASOCEBU 2015; ABCGIL 2017), apesar de bem adaptados sofrem influências das condições ambientais na expressão do potencial produtivo e reprodutivo. A interação genótipo ambiente (IGA) caracteriza-se por respostas diferentes dos genótipos dos animais às variações ambientais, ocasionando alterações nos fenótipos, e nas magnitudes das variações genéticas e ambientais (Falconer e Mackay, 1996).

A IGA em bovinos tem sido amplamente investigada ao comparar genótipos entre subespécies (*Bos taurus* vs. *Bos indicus*), raças e indivíduos da mesma raça (Oliveira et al., 2018; Hayes et al., 2016). A IGA pode ser avaliada considerando o ambiente discreto (por exemplo, tropical vs. temperado) ou contínuo (por exemplo, temperatura e umidade) (Santana et al., 2013; Streit et al., 2013; Hayes et al., 2016). Neste sentido, a correlação genética fornece um meio de quantificar a interação com o objetivo de prever respostas à seleção, uma baixa correlação genética significa que o genótipo reage de maneira diferente e tem linhas de regressão com diferentes inclinações, ou seja, os indivíduos têm diferentes sensibilidades ambientais (Falconer e Mackay, 1996). O modelo de norma de reação modela a trajetória do desempenho do animal em função do gradiente ambiental (GA) e, portanto, o valor genético de um animal é particionado em uma parte independente do ambiente (intercepto) e uma parte dependente do ambiente (inclinação) (Falconer e Mackay, 1996).

A inclusão de informações genotípicas no modelo de norma de reação podem fornecer estimativas e previsões mais acuradas dos valores genéticos, por integrar informações de animais genotipados e não genotipados por meio do uso da matriz genômica e realização da análise em passo único, via ssGBLUP (Misztal et al. 2009) que permite também obter estimativas mais acuradas dos efeitos dos SNPs para determinada característica na análise de GWAS .

Assim, com a globalização da indústria leiteira (Hammami et al., 2009), uma preocupação do criador é a escolha do reprodutor adequado ao seu sistema produtivo, considerando que a IGA ocasiona reclassificação dos reprodutores em distintos sistemas

de produção (Fikse et al., 2003; Hammami et al., 2009). Análise considerando a importância e as implicações da IGA na produção de bovinos leiteiros no mundo foi constituída pela Interbull, entidade que realiza avaliação genética internacional de touros, que uniu informações das filhas de touros de diferentes países participantes e disponibiliza o valor genético do touro para cada um dos países (Hayes et al., 2016).

Na comercialização de sêmen das raças zebuínas (*Bos indicus*), o Brasil é o principal provedor de genética avaliada de touros da raça Gir e Girolando para países tropicais da América Latina. A compreensão da importância da IGA em bovinos leiteiros da raça Gir por parte dos criadores que realizam intercâmbio genético entre Brasil e Colômbia é de suma importância para os sistemas de produção. Os touros Gir Leiteiro Brasileiros podem não produzir os mesmos resultados em termos de produção de leite quando utilizados nos ambientes de produção Colombianos. Com isso, o objetivo do presente estudo foi investigar a interação genótipo ambiente para produção de leite em bovinos da raça Gir do Brasil e Colômbia com aplicação de modelo de norma de reação genômica, bem identificar regiões genômicas associadas a produção de leite nos dois países.

## **2. Material e Métodos**

### **2.1 Dados fenotípicos**

O banco de dados utilizado no presente estudo foi processado no Departamento de Melhoramento e Nutrição Animal da Universidade Estadual Paulista “Júlio de Mesquita Filho” – Faculdade de Medicina Veterinária e Zootecnia, em Botucatu, São Paulo, Brasil. Foram utilizadas informações fenotípicas da produção de leite acumulada até 305 dias (PL305), considerando as três primeiras lactações de 1.022 animais da raça Gir provenientes de fazendas localizadas nos estados de São Paulo e Minas Gerais, controladas pela Associação Brasileira de Criadores de Gir Leiteiro (ABCGIL). Os dados da Colômbia compreenderam informação de PL305 de 586 vacas provenientes de fazendas das regiões do Caribe, Andina e Amazônia controladas pela Associação Colombiana de Criadores de Gado Zebu (ASOCEBU).

## **2.2 Genotipagem, imputação e controle de qualidade**

A genotipagem foi realizada com três diferentes painéis da Illumina, sendo 18 touros do Brasil com o painel SNP de alta densidade (BovineHD BeadChip assay 777K, Illumina, San Diego, CA), 273 vacas e 15 touros do Brasil com BovineSNP30 BeadChip (30K) e 158 vacas da Colômbia com GGP indicus chip (35K).

Devido as diferentes densidades dos painéis, foi realizada a imputação dos genótipos, sendo considerado os SNPs em comum (9.109K) dos painéis de baixa densidade e implementado utilizando o programa FImpute v2.2 (Sargolaiez et al., 2014). Foram realizadas estratégias de imputação para cada cenário: 1) 9.109 SNPs do painel 30K com 273 vacas e 15 touros do Brasil; 2) 9.109 SNPs do painel de 35K com 158 vacas da Colômbia e 18 touros do Brasil. A precisão da imputação dos genótipos das populações foi medida pela taxa de concordância e pela correlação entre os SNPs. O cenário de maior precisão de imputação consistiu no painel de 30K, obtendo correlação de 94% e taxa de concordância de 96%. Deste modo, a imputação de todos os animais do presente estudo foi realizada para o painel de 30K.

Após a imputação dos genótipos, foram excluídos os SNPs with a call rate < 0.90, a minor allele frequency (MAF) < 0.02, and a maximum difference between observed and expected frequency (Hard-Weinberg equilibrium) > 0.15%. O arquivo final incluiu 464 animais genotipados e 27.505 SNPs.

A análise de componentes principais (PCA) foi realizada para explicar a estratificação das populações de Gir de Colombia e Brasil. O PCA foi calculado para a matriz genômica usando a função plotpca do programa preGSf90 (Aguilar et al., 2010).

## **2.3 Análise genético-quantitativa**

A interação genótipo ambiente (IGA) foi abordada por meio da utilização de modelo de norma de reação genômica em duas etapas. A primeira etapa consistiu em análises por meio do modelo animal padrão desconsiderando a existência da IGA para estimar as soluções dos grupos de contemporâneos (GC). Os GC para produção de leite acumulada até 305 dias foram definidos pela combinação das informações de país, fazenda e ano de parto. Além do GC, a época de parição (meses: 1: jan, fev, mar; 2: abri, maio, jun; 3: jul, ago, set; 4: out, nov, dez) e a lactação (1, 2 e 3) foram

considerados como efeitos fixos classificatórios, a idade da vaca ao parto como covariável linear e os efeitos aleatórios de animal e resíduo.

A segunda etapa consistiu na análise dos GC como gradiente ambiental (covariável ambiental) com uso de modelo animal bi-característica para produção de leite até 305 dias no Brasil (PL305BR) e Colômbia (PL305COL), em função do país de origem dos animais e incluindo a informação genômica via modelo de norma de reação genômica em passo único (single-step).

No modelo de norma de reação, os valores genéticos aditivos dos animais foram regredidos sobre o gradiente ambiental. Desta forma, o efeito genético aditivo foi composto pelo intercepto (nível geral de produção sobre todos os ambientes), inclinação (resposta ambiental específica aos ambientes) e a covariância entre esses termos. O modelo em termos matriciais foi:

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

em que:  $y$  é o vetor de observações para produção de leite das lactações;  $\beta$ , vetor dos efeitos sistemáticos;  $a$ , representa o vetor dos efeitos genéticos aditivos para os coeficientes da norma de reação (intercepto e inclinação);  $\varepsilon$ , vetor dos efeitos residuais;  $X$  e  $Z$ , matrizes de incidência que relaciona  $\beta$  e  $a$  ao vetor de observações.

Assumir-se-á que:

$$E \begin{bmatrix} y \\ a \\ e \end{bmatrix} = \begin{bmatrix} x'_i \beta \\ 0 \\ 0 \end{bmatrix}; V \begin{bmatrix} a \\ e \end{bmatrix} = \begin{bmatrix} K_a \otimes A & 0 \\ 0 & R \end{bmatrix}$$

em que:  $K_a$  é a matriz de (co)variância entre os coeficientes de regressão aleatórios para os efeitos genético aditivo direto,  $A$  é a matriz de parentesco entre os indivíduos,  $\otimes$  é o produto de Kronecker entre matrizes, e  $R$  representa a matriz bloco diagonal, contendo as variâncias residuais.

No método single-step a inversa da matriz de parentesco aditivo ( $A$ ) foi substituída pela matriz  $H$  (Aguilar et al., 2010):

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

em que; H é a matriz de parentesco com base tanto na informação genômica quanto na informação de pedigree; A é a matriz dos coeficientes de parentesco com base no pedigree de todos os animais; A<sub>22</sub> é a matriz dos coeficientes de parentesco com base no pedigree dos animais genotipados, e "G" é a matriz de parentesco genômica.

A matriz G foi determinada segundo VanRaden (2008):

$$G = \frac{ZZ'}{2 \sum_{j=1}^J p_j(1 - p_j)}$$

sendo: Z obtida da subtração (M – P); M corresponde à matriz de genótipos, com colunas representando os marcadores e as linhas os indivíduos. Cada elemento em M<sub>ij</sub> foi codificado como 0, 1 ou 2, caso o genótipo do indivíduo i para o SNP<sub>j</sub> seja homozigoto para o primeiro alelo, heterozigoto, ou homozigoto para o segundo alelo, respectivamente e P corresponde à matriz contendo a frequência do segundo alelo p<sub>j</sub>, expresso como 2p<sub>j</sub> (VanRaden, 2008).

Os componentes de variância foram obtidos por máxima verossimilhança restrita nos modelos animal tradicional e single-step, usando o software AIREMLF90 (Misztal et al., 2016).

Tabela 1. Descrição da base de dados fenotípicos dos animais da raça Gir do Brasil e Colômbia.

|                                                                     | <b>Brasil</b>                 | <b>Colômbia</b> |
|---------------------------------------------------------------------|-------------------------------|-----------------|
| Nº de vacas                                                         | 1.022                         | 586             |
| Nº de lactações (1 <sup>a</sup> ; 2 <sup>a</sup> e 3 <sup>a</sup> ) | 1.922                         | 1.081           |
| Nº grupo contemporâneo                                              | 81                            | 36              |
| Nº vacas com genótipo                                               | 273                           | 158             |
| Pedigree                                                            | 2.889 (565 pais e 1.959 mães) |                 |
| Nº touros com genótipos                                             | 33                            |                 |

Os dados foram analisados usando inferência bayesiana por meio dos programas GIBBS3F90 e POSTGIBBSF90 (Misztal et al., 2002), em cadeia amostral de 1.000.000 ciclos, com período de *burn-in* de 20.000 ciclos e intervalo amostral a cada 20 ciclos.

O valor genético genômico foi obtido com a seguinte expressão:

$$VG_i = intercept_i + (inclinação_i * gradiente ambiental_{117})$$

Em que:  $VG_i$  é o valor genético genômico de cada animal, o intercepto é o valor genético aditivo do animal  $i$ , inclinação é resposta ambiental do animal  $i$ , o gradiente ambiental é 117 níveis no gradiente ambiental composto por soluções de GC para PL305.

A acurácia dos valores genéticos foi obtida pela expressão:

$$acurácia\ real = \sqrt{1 - \frac{SE_i^2}{(1 + F_i) + \sigma_a^2}}$$

onde:  $SE$  é o erro padrão da estimativa do valor genético do animal  $i$ ;  $F$  é a endogamia do animal  $i$ ,  $\sigma_a^2$  é a variância genética aditiva da característica.

## 2.4 Análises de Associação Genômica (GWAS)

As análises de GWAS foram conduzidas para estimar os efeitos dos SNP, bem como janelas adjacentes de 10 SNP por meio do software POSTGSf90 (version 1.38) (Misztal et al., 2016; Wang et al., 2012). Os resultados das análises de GWAS foram apresentados com base na proporção da variância explicada para cada janela de SNP. As janelas que capturaram a maior proporção da variância explicada pelos SNPs associados aos genes candidatos obtidos por meio do intercepto ou inclinação foram identificados utilizando o pacote BioMart (Durinck et al. 2005), por meio do mapa de referência UMD 3.1.

## 3. Resultados

Os componentes principais foram derivados da matriz de parentesco genômicas entre as populações da raça Gir do Brasil e Colômbia (Figure 1), que descreve o grau de estratificação das populações. O parentesco entre as populações Gir dos dois países se deve, principalmente a aquisição por parte dos criadores da Colômbia de material genético (animais, sêmen e embriões) proveniente do Brasil. Desta forma, permite a integração e análise conjunta das informações dos países.



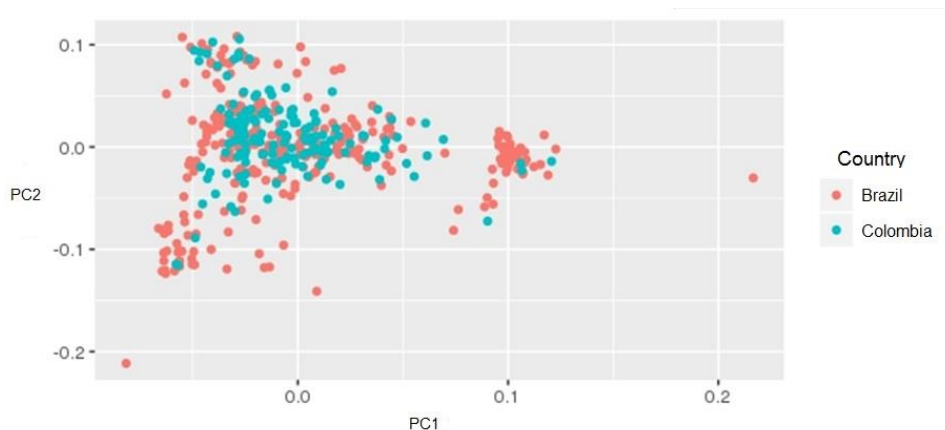


Figura1. Análises de componentes principais com informação da matriz de parentesco genômica nas populações da raça Gir do Brasil e Colômbia.

### 3.1 Componentes de Variância

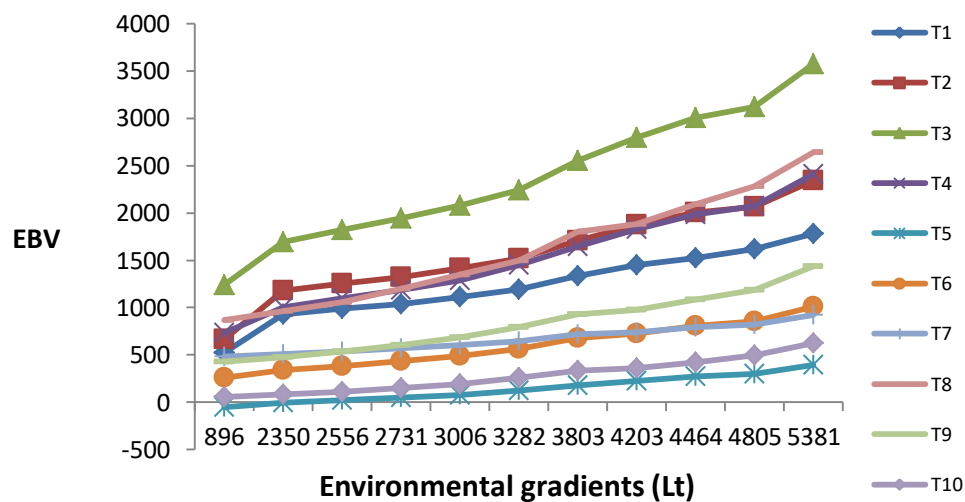
As estimativas de herdabilidade das características e PL305COL com informação somente do pedigree foram  $0,45 \pm 0,019$  e  $0,44 \pm 0,026$  e com uso da matriz genômica foram  $0,49 \pm 0,015$  e  $0,46 \pm 0,023$ , respectivamente. As estimativas de herdabilidade apresentaram alta a moderada magnitude, e demonstra que a variância genética aditiva é responsável por considerável parcela da variância fenotípica, a qual indica a possibilidade de obter ganhos genéticos significativos para a característica PL\_305 por meio da seleção.

A correlação genética das características PL305COL e PL305BR com a matriz de pedigree (A) e informação genômica (H) foram  $0,66 \pm 0,041$  e  $0,71 \pm 0,040$ , respectivamente. Além disto, os resultados das correlações genéticas do presente estudo apresentaram-se abaixo de 0,80, indicando segundo Robertson (1959) a presença de interação genótipo ambiente. Esses resultados demonstram a variação ambiental entre Colômbia localizada em clima tropical e tropical com altitude, e o Brasil localizado em clima tropical e subtropical, que proporcionaram ambientes diferentes para a produção animal.

### 3.2 Precisão dos valores de genéticos

Os resultados das correlações genéticas entre as características PL305BR e PL305COL, usando o modelo animal tradicional matriz A e modelo animal com matriz H, identificaram maior correlação usando a matriz H, sendo escolhido este modelo para realizar os análises e estimar os valores genéticos dos touros. Os valores genéticos (EBVs) dos 10 (T1 –T10) principais touros reprodutores utilizados nos dois países (Brasil e Colômbia) foram calculados ao longo do GA (Figura 2). Na característica PL305BR (Figura 2a), o touro T3 foi superior na classificação nas gradientes ambientais. Os touros T5 e T10 apresentaram menores valores genéticos nas GA, os touros foram sensíveis ao meio ambiente, porque seus valores genéticos variavam do ambiente menos ideal ao ambiente de alta qualidade. Isso indica que variações no GA influenciaram a expressão fenotípica de PL305.

a)



b)

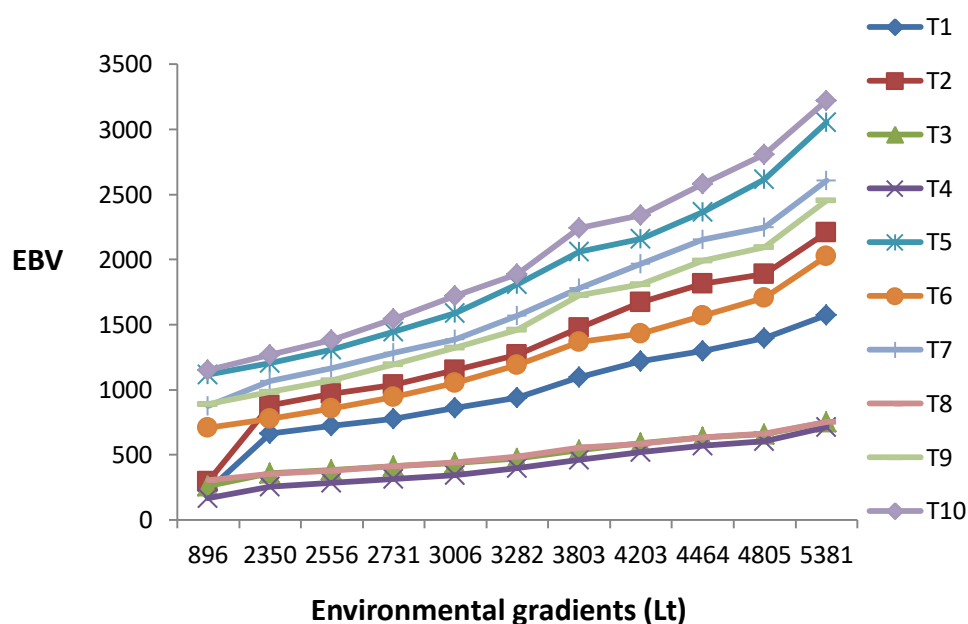


Figura 2. Os dez principais touros (T1 a T10) reprodutores usados nos dos países com base nos valores genéticos genômicos (EBVs) para produção de leite acumulada até 305 dias nos gradientes ambientais: a) PL305BR e b) PL305COL.

Na caraterística PL305COL (Figura 2b), o touro T10 apresentou valor superior de EBVs ao longo dos gradientes ambientais, em comparação com os demais touros. Os touros T3 e T4 apresentaram os menores EBVs nas GA, o que demonstra um aumento constante no desempenho fenotípico conforme melhora o ambiente, porém esse fato ocorre em um ritmo mais lento e apresentam maior plasticidade em comparação aos demais animais. A acurácia média dos valores genéticos (EBVs) foi 0,60 (de 0,56 a 0,69) para as duas características.

### 3.3 ssGWAS para PL305BR e PL305COL

Os resultados do ssGWAS demonstram uma sobreposição entre as 10 principais janelas dos parâmetros da norma de reação para as características PL305BR e PL305COL das populações de animais da raça Gir (Tabela 2). Na caraterística PL305BR, 9 das 10 principais janelas foram observadas no mesmo cromossomo, e na característica PL305COL, 7 das 10 janelas superiores do intercepto e inclinação também

se apresentaram no mesmo cromossomo, no entanto, explicam diferentes proporções de variação.

Na característica PL305BR, os cromossomos 5 e 21 foram observados na primeira e segunda região e explicam a variação do intercepto, enquanto que os mesmos cromossomos foram apresentados na segunda e sexta região, os quais explicam a variação de inclinação. Na característica PL305COL, o cromossomo 21 foi apresentado na primeira e a terceira região para explicar as variações de intercepto e inclinação, respectivamente. Já o cromossomo 20 foi classificado em sexta e primeira região para explicar as variações de intercepto e inclinação, respectivamente. Deste modo, apesar dos cromossomos 5, 20 e 21 se apresentarem nas janelas superiores do intercepto e inclinação, os mesmos explicaram diferentes proporções da variação genética, onde possivelmente distintos genes podem estar associados a produção de leite dos bovinos Gir no Brasil e Colômbia.

Tabela 2. Cromossomos (Chr), posições iniciais, posições finais e variância genética (%) explicada por marcadores moleculares das normas de reação para PL305BR (a) e PL305COL (b) em bovinos leiteiros da raça Gir.

a)

| Intercepto |          |          |         | Inclinação |          |           |         |
|------------|----------|----------|---------|------------|----------|-----------|---------|
| Chr        | inicio   | final    | %       | Chr        | inicio   | final     | %       |
| 5          | 34446901 | 35385581 | 0.65715 | 13         | 313308   | 898514    | 0.45054 |
| 21         | 58743987 | 59526690 | 0.54659 | 5          | 53863852 | 54711045  | 0.40156 |
| 20         | 38880048 | 39727100 | 0.51236 | 6          | 14814828 | 15978059  | 0.38768 |
| 21         | 69740387 | 70155265 | 0.49376 | 3          | 32988739 | 33008170  | 0.31552 |
| 5          | 31104780 | 31964893 | 0.48365 | 6          | 1,09E+08 | 109719477 | 0.31534 |
| 14         | 1651311  | 1967325  | 0.45387 | 21         | 22655425 | 23755490  | 0.30044 |
| 24         | 61995052 | 62238372 | 0.43496 | 2          | 1,35E+08 | 135179024 | 0.29403 |
| 17         | 70460874 | 70935589 | 0.42321 | 14         | 2596928  | 2826632   | 0.28770 |
| 19         | 35618689 | 36433251 | 0.39528 | 5          | 34446901 | 35385581  | 0.27711 |
| 23         | 24372201 | 25074632 | 0.38981 | 12         | 86414606 | 86989516  | 0.27015 |

b)

| Intercepto |          |          |         | Inclinação |          |          |         |
|------------|----------|----------|---------|------------|----------|----------|---------|
| Chr        | inicio   | final    | %       | Chr        | inicio   | final    | %       |
| 21         | 8859292  | 9494265  | 0.59349 | 20         | 38880048 | 39727100 | 0.44058 |
| 4          | 77074551 | 77958922 | 0.49170 | 14         | 1651311  | 1967325  | 0.41276 |

|    |          |          |         |    |          |          |         |
|----|----------|----------|---------|----|----------|----------|---------|
| 8  | 4358764  | 4830615  | 0.48318 | 21 | 69740387 | 70155265 | 0.39486 |
| 14 | 1651311  | 1967325  | 0.46062 | 17 | 70460874 | 70935589 | 0.39380 |
| 24 | 62004962 | 62250785 | 0.45495 | 21 | 58743987 | 59526690 | 0.39329 |
| 20 | 955427   | 1716974  | 0.39985 | 23 | 24372201 | 25074632 | 0.36935 |
| 9  | 10705314 | 10748367 | 0.39562 | 19 | 35618689 | 36433251 | 0.36578 |
| 24 | 60845041 | 61205792 | 0.39474 | 5  | 31104780 | 31964893 | 0.33580 |
| 16 | 41794521 | 42954110 | 0.39443 | 5  | 57030766 | 58140204 | 0.31648 |
| 16 | 25917821 | 25940729 | 0.34969 | 3  | 32988739 | 33008170 | 0.31552 |

Na inclinação da característica PL305BR os cromossomos 13 e 5 explicaram a maior variância e para PL305COL os cromossomos 20 e 14 foram evidenciados, assim os resultados demonstraram que o ambiente influencia diretamente na expressão de genes em diferentes cromossomos para a produção de leite de bovinos da raça Gir. Além disso, também foi possível observar que os cromossomos 3, 5 e 21 apresentaram-se comumente entre as duas características, no entanto, explicam diferentes proporções da variância genética devido a influência do ambiente, a qual pode ocasionar diferentes expressões do genótipo. No entanto, apesar de identificarmos algumas diferenças de variância explicadas para a intercepto inclinação em diferentes regiões do genoma nas duas características, a variância explicada por essas regiões foi pequena, sendo responsável por 0,65 e 0,45% da PL305BR e 0,59 e 0,44% PL305COL, da variância genética total.

A análise de ssGWAS permitiu identificar SNPs significativos para PL305BR nos cromossomos 1, 3, 5, 19 e 21 para o intercepto. Os SNPs significativos, BovineHD0500009996, BovineHD1900010366 e BovineHD2100016917, foram associados aos genes *MGST1* no BTA5, *CCDC57* no BTA19 e *TGLA122* no BTA21, respectivamente. Os SNPs significativos BovineHD1500008597 no cromossomo 15 para inclinação no presente estudo foi associado ao gene *SPS115*. Na característica PL\_305COL no intercepto os SNPs significativos ARS-BFGL-NGS-97163, BovineHD070001873 e BovineHD1200005209 foram associados ao gene *SOX14* no BTA1, e o SNP BovineHD0500025415 foi associado ao gene *SLC36A2* no BTA7 e ao gene *IRS2* no BTA12. Na inclinação o gene *SLO1C1* no BTA5 foi observado no presente estudo.

#### **4. Discussão**

Os resultados mostraram que as herdabilidades e correlações genéticas entre as características PL305BR e PL305COL apresentaram diferenças com inclusão de informações genômicas aumentando a precisão dos componentes de variância. Além disso, alguns genes candidatos que influenciam o intercepto e inclinação da norma de reação para PL305BR e PL305COL também foram identificados.

##### **4.1 Componentes de variância estimados**

As estimativas das herdabilidades observadas no presente estudo foram superiores aos valores relatados por Prata et al. (2015) para produção de leite aos 305 dias na raça Gir (0,33) e por Hammami et al. (2009) na raça Holandesa (0,40). As diferentes estimativas de herdabilidade obtidas para PL305 entre Brasil e Colômbia evidenciam distintas respostas a seleção dependendo do ambiente de criação. As correlações genéticas estimadas pela matriz A foram ligeiramente inferiores às estimadas por matriz H para PL305BR e PL305COL. Em recentes estudos Oliveira et al. (2018) e Zhang et al. (2019) observaram maior acurácia no uso da matriz H em comparação com a matriz A nas raças Nelore e Holandesa, respectivamente.

Além do ambiente em si, as práticas de seleção no Brasil e Colômbia podem ter contribuído para as diferenças dos parâmetros genéticos das características nas populações. O principal critério de seleção do Brasil é com base no teste de progênie, enquanto os criadores da Colômbia escolhem os animais com base em um índice que inclui longevidade, tipo leiteiro (conformação morfológica) e produção de leite. O Brasil possui programa de melhoramento para a raça Gir desde 1985 (Santana et al., 2014). No entanto, a Colômbia ainda não apresenta programa de melhoramento genético estruturado.

##### **4.2 Precisão dos valores genéticos preditos**

Nos resultados dos valores genéticos das características PL305BR e PL305COL (figura 2), podemos observar que tem re-ranking dos touros entre os dois países, ou seja, o melhor touro no Brasil é o pior na Colômbia, e vice-versa. Os melhores e os piores animais eram sensíveis ao meio ambiente, porque seus valores genéticos variavam do

ambiente menos ideal ao ambiente de alta qualidade, evidenciando a presença da IGA nas populações de Gir.

Segundo Panetto et al. (2018), a raça Gir leiteiro do Brasil apresenta IGA para a característica produção de leite até 305 dias, os autores observaram rearranjos na classificação dos touros em relação ao ambiente do Brasil. No entanto, este é o primeiro estudo que relata a avaliação da IGA na população colombiana de bovinos da raça Gir, e nossos resultados demonstram diferenças entre os touros de acordo com os gradientes ambientais, e ressalta a importância de incluir o efeito ambiental nas avaliações genéticas dentro de cada país.

Os resultados das características PL305BR e PL305COL evidenciam IGA pela reclassificação dos touros entre os dois países, demonstrando que Brasil e Colômbia apresentam diferenças ambientais e se torna importante considerar ambiente na escolha do reprodutor, para que sua progênie possa desempenhar adequadamente. As associações da raça Gir podem enfrentar o problema de otimizar o programa de melhoramento para responder a múltiplos requisitos ambientais dentro e entre os países. Embora as populações do Brasil e Colômbia apresentem distintos sistemas de produção, ambas mantêm laços genéticos importantes. Assim, a consideração da IGA pode ajudar na concepção de programas de melhoramento mais efetivos. Contudo, países importadores como Colômbia, os quais não apresentam programas de melhoramento próprio, necessitam de conhecimento sobre a sensibilidade ambiental para proceder escolha apropriada de material genético e obter progresso satisfatório.

#### **4.3 Mapeamento de regiões genômicas associadas à intercepto e inclinação da norma de reação**

Este estudo realizou o ssGWAS para localizar regiões particulares ao longo do genoma associadas a produção de leite até 305 dias. Algumas janelas genômicas identificadas eram comuns tanto para o intercepto quanto para a inclinação das duas características (Tabela 2). Os genes associados à característica PL305BR para o intercepto, o gene *MGST1* foi associado à produção e percentual de gordura do leite e a características da composição do leite de vacas leiteiras (Littlejohn et al., 2016; Jiang et al., 2019). A expressão do gene *CCDC57* foi observada no desenvolvimento da glândula mamária de vacas (Medrano et al., 2010), e relacionado ao rendimento e

porcentagem de gordura aos 305 dias de produção em três raças francesas de gado leiteiro (Marete et al., 2018). O gene *TGLA122* foi associado à adaptação ambiental de bovinos da raça Gir (Curi et al., 2002). No entanto, apesar de sua importância para o presente estudo, o gene *TGLA122* não foi relatado em estudos que investigaram diretamente a produção de leite em bovinos. Os SNPs significativos BovineHD1500008597 no cromossomo 15 para inclinação no presente estudo foi associado ao gene *SPS115*, identificado relacionado à adaptação de zebuínos na região tropical (Curi et al., 2002; Montoya et al., 2010). Portanto, os genes observados no intercepto do presente estudo são consistentes com a característica de produção de leite, pois são independentes ao ambiente. Além disso, os resultados demonstram que os SNPs se apresentam distintos na expressão da genética do intercepto e inclinação, identificando genes associados com adaptação e resistência aos trópicos.

Os genes associados à característica de PL305COL, o gene *SOX14* pertencente a Sox Family of transcription factors possui evidências de estar envolvido na regulação do desenvolvimento embrionário (Stanisavljevic et al., 2017), porém, ainda não foi relatada associação com produção de leite em bovinos. O gene *SLC36A2* foi correlacionado a composição da proteína do leite em bovinos Chinese Holstein (Zhou et al., 2019), e o gene *IRS2* codifica o substrato do receptor de insulina 2, e a infusão de insulina aumenta a produção de leite e proteínas, além de reduzir o teor de gordura em cabras (Peripolli et al 2018). Na inclinação o gene *SLO1C1* no BTA5 foi observado no presente estudo. Além disso, esse gene foi associado com produção de gordura no leite em bovinos Holandeses (Nayeri et al., 2016). A este respeito, os resultados do presente estudo sugerem que os genes observados para característica no intercepto PL305COL e inclinação estão relacionados a produção de leite, proteína e gordura.

As principais regiões genômicas que afetaram a inclinação da norma de reação incluem genes observados nas populações Gir do Brasil e Colômbia relacionados à imunidade (*TGLA122*, *SPS115*) e sólidos totais (*SLO1C1*, *SLC36A2*), enquanto as principais regiões que influenciam a intercepto da norma de reação incluíram genes relacionados produção de leite (*MGST1*, *CCDC57*, *IRS2*) e fertilidade (*SOX14*). Diferentes genes desempenham seus papéis biológicos com importância distinta na expressão do genótipo para o intercepto e inclinação para as duas características PL305. Com isto, a reclassificação dos touros e os diferentes SNPs associados à produção de leite nos dois países, sugere que a interação genótipo ambiente é efeito importante a ser



considerado na avaliação genética do Gir leiteiro do Brasil e Colômbia. Portanto, programa de melhoramento do Gir pode ser otimizado ao escolher o ambiente de seleção que possibilitará o máximo progresso genético para a produção de leite nos diferentes ambientes dentro e entre países.

## 5. Conclusão

A IGA é um fator importante a ser considerado na estimativa de parâmetros genéticos e avaliação genética entre Brasil e Colômbia da característica de produção de leite da raça Gir.

Além disso, a produção de leite de bovinos Gir do Brasil e Colômbia depende, pelo menos em parte, da expressão de genes relacionados à produção de leite, imunidade e fertilidade, os quais foram identificados nas regiões genômicas que afetam a intercepto e inclinação da norma de reação.

## 6. Referências

Aguilar I.; Misztal I.; Johnson D.L.; Legarra A.; Tsuruta S.; Lawlor T.J. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.*, v.93, p.743–752.

Aken, B.L.; Ayling, S.; Barrell, D.; et al. (2016) The Ensembl gene annotation system. Database (Oxford). 2016;2016:baw093. doi:10.1093/database/baw093

Associação Brasileira Dos Criadores De Gir Leiteiro. ABCGIL 2016 <http://girleiteiro.org.br/>

Asociacion Colombiana De Criadores De Ganado Cebu. ASOCEBU. 2017. <http://www.asocebu.com/index.php/el-cebu/razas/gyr>

Curi, Rogério Abdallah and Lopes, Catalina Romero. 2002. Evaluation of nine microsatellite loci and misidentification paternity frequency in a population of Gyr breed bovines. *Braz. J. Vet. Res. Anim. Sci.* vol.39.

Falconer & Mackay (1996). Introduction to quantitative genetics. Longman group ltd 1996. Fourth edition.

Fikse, W.F.; Rekaya, R.; Weigel, K.A. (2003). Genotype environment interaction for milk production in Guernsey cattle. *J. Dairy Sci.* v.86, p.1821–1827.

Hammami, H., B. Rekik, H. Soyeurt, C. Bastin, E. Bay, J. Stoll, and N. Gengler. 2009. Accessing genotype by environment interaction using within- and across-country test-day random regression sire models. *J. Anim. Breed. Genet.* 126:366–377.

Hayes, B. J.; Daetwyler, H.D. and Goddard, M. E. Models for Genome Environment Interaction: Examples in Livestock. **crop science**, vol. 56, 2016.

Littlejohn M. D., Tiplady K., Fink T. A., Lehnert K., Lopdell T., Johnson T., et al. (2016). Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. *Sci. Rep.* 6:25376. 10.1038/srep25376.

Marete, A; Lund, MS; Boichard, D; Ramayo-Caldas Y (2018). A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. **PLoS ONE** 13(7):e0199931.

Medrano JF, Rincon G, Islas-Trejo A. 2010. Comparative analysis of bovine milk and mammary gland transcriptome using RNA-Seq. 9th World Congress on Genetics applied to Livestock Production; Leipzig, Germany.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D.H. 2002. BLUPF90 and related programs (BGF90). In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August, 2002. Session 28. Institut National de la Recherche Agronomique (INRA). pp. 1–2.

Misztal, I., Legarra A., Aguilar, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.*, 92 (2009), pp. 4648-4655

Misztal, Ignacy. 2016. Expensive Computation of the Inverse of the Genomic Relationship Matrix in Populations with Small Effective Population Size. **Genetics** vol. 202. 401-409.

Montoya, Alba E; Cerón-Muñoz, Mario F; Manuel A Moreno; Edwin Martínez; Juan D Corrales; Juan F Tirado; Samir J Calvo. 2010. Genetic characterization of the Hartón del Valle, Angus, Brangus, Holstein, and Senepol cattle breeds in Colombia, using ten microsatellite markers. **Rev Colomb Cienc Pecu** 2010; 23:283-291.

Nayeri, S., Sargolzaei, M., Abo-Ismael, M.K. et al. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. **BMC Genet.** 17, 75 (2016).

Jiang, J., Ma, L., Prakapenka, D., VanRaden, P. M., Cole, J. B., & Da, Y. (2019). A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. *Frontiers in genetics*, 10, 412. <https://doi.org/10.3389/fgene.2019.00412>

Oliveira, D.P., Lourenco, D.A.L., Tsuruta, S., Misztal, I., Santos, D.J.A., de Araújo Neto, F.R., Aspilcueta-Borquis, R.R., Baldi, F., Carneiro, R., de Camargo, G.M.F., Albuquerque, L.G., Tonhati, H. 2018. Reaction norm for yearling weight in beef cattle using single-step genomic evaluation. *J Anim Sci.* 2018 Feb 15;96(1):27-34.

Panetto, J.C.C., Verardo, L.L., Machado, M.A., Martins, M.F., Faza, D.R.L.R. & Silva, M.V.G.B. 2018. Genotype by environment interaction in Brazilian dairy Gir cattle. *Proceedings of the World Congress on Genetics Applied to Livestock Production*, 11.

Peripolli, E., Stafuzza, N.B., Munari, D.P., Lima, A.L.F., Irgang, R., Machado, M.A., Panetto, J.C. do C., Ventura, R.V., Baldi, F., da Silva, M.V.G.B., 2018. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. *BMC Genomics*. <https://doi.org/10.1186/s12864-017-4365-3>

Prata, M.A., Faro, L.E., Moreira, H.L., Verneque, R.S., Vercesi Filho, A.E., Peixoto, M.G., Cardoso, V.L. 2015. Genetic parameters for milk production traits and breeding goals for Gir dairy cattle in Brazil. *Genet Mol Res.* 19;14(4):12585-94.

Rauw, W.M.; Gomez-Raya, L. (2015). Genotype by environment interaction and breeding for robustness in livestock. **Frontiers in genetics**, v.6, p.310.

Santana, M.L. Bignardi, A.B. Eler, J.P. Cardoso, F.F. and Ferraz, J.B.S. 2013. Genotype by environment interaction and model comparison for growth traits of Santa Ines sheep. **Journal of Animal Breeding and Genetics.**

Santana Junior ML, Pereira RJ, Bignardi AB, El Faro L, Tonhati H, Albuquerque LG. History, structure, and genetic diversity of Brazilian Gir cattle. **Livest Sci.** 2014;163:26–33.

Stanisavljevic, D., Petrovic, I., Vukovic, V., Schwirtlich, M., Gredic, M., Stevanovic, M., & Popovic, J. (2017). SOX14 activates the p53 signaling pathway and induces apoptosis in a cervical carcinoma cell line. **PloS one**, 12(9), e0184686.

Streit, M., F. Reinhardt, G. Thaller, and J. Bennewitz. 2013. Genome-wide association analysis to identify genotype × environment interaction for milk protein yield and level of somatic cell score as environmental descriptors in German Holsteins. **J. Dairy Sci.** 96:7318–7324.

Vanraden, P.M. Efficient methods to compute genomic predictions. **J Dairy Sci.** v.91, p.4414-4423, 2008.

Warwick, E.J.; Legates, J.E. Cria y mejora del ganado. McGraw-Hill, Mexico 1980. 145p.

Zhang, H., Yin, L., Wang, M., Yuan, X., Liu, X. (2019). Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. **Front. Genet.** 10, 189.

Zhou, C., Li, C., Cai, W., Liu, S., Yin, H., Shi, S., Zhang, Q., & Zhang, S. (2019). Genome-Wide Association Study for Milk Protein Composition Traits in a Chinese Holstein Population Using a Single-Step Approach. **Frontiers in genetics**, 10, 72.