

RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta tese será disponibilizado somente a partir de 08/05/2019.

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

**ESTRUTURA GENÔMICA POPULACIONAL DE BOVINOS
LEITEIROS GIR**

Tássia Souza Bertipaglia

Zootecnista

2017

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

**ESTRUTURA GENÔMICA POPULACIONAL DE BOVINOS
LEITEIROS GIR**

Tássia Souza Bertipaglia

Orientadora: Profa. Dra. Vera Fernanda Martins Hossepian de Lima

Coorientadores: Prof. Dr. Danísio Prado Munari

Dr. Rodrigo Pelicioni Savegnago

Dr. Marcos Vinícius Gualberto Barbosa da Silva

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

2017

Bertipaglia, Tássia Souza
B544e Estrutura genômica populacional de bovinos leiteiros Gir / Tássia Souza Bertipaglia. -- Jaboticabal, 2017
v, 69 p. : il. ; 29 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2017

Orientadora: Vera Fernanda Martins Hossepian de Lima

Coorientadores: Danísio Prado Munari, Rodrigo Pelicioni Savegnago, Marcos Vinícius Gualberto Barbosa da Silva

Banca examinadora: Lenira El Faro, Joslaine Noely dos Santos Gonçalves Cyrillo, Ricardo da Fonseca, João Ademir de Oliveira
Bibliografia

1. Análise multivariada. 2. Bos taurus indicus. 3. Bovino de leite. 4. Linhagens. 5. Painéis de SNP. 6. Variabilidade genética I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.2

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: ESTRUTURA GENÔMICA POPULACIONAL DE BOVINOS LEITEIROS GIR

AUTORA: TÁSSIA SOUZA BERTIPAGLIA

ORIENTADORA: VERA FERNANDA MARTINS HOSSEPIAN DE LIMA

COORIENTADOR: MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA

COORIENTADOR: DANISIO PRADO MUNARI

COORIENTADOR: RODRIGO PELICIONI SAVEGNAGO

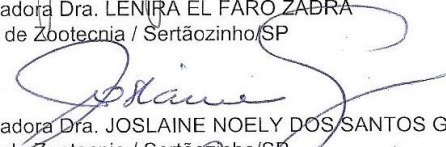
Aprovada como parte das exigências para obtenção do Título de Doutora em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:



Pós-doutorando RODRIGO PELICIONI SAVEGNAGO
Departamento de Ciências Exatas / FCAV / UNESP - Jaboticabal




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



Pesquisadora Dra. JOSLAINE NOELY DOS SANTOS GONÇALVES CYRILLO
Instituto de Zootecnia / Sertãozinho/SP



Prof. Dr. RICARDO DA FONSECA
Departamento de Zootecnia / FCAT / UNESP - Dracena



Prof. Dr. JOÃO ADEMIR DE OLIVEIRA
Departamento de Ciências Exatas / FCAV / UNESP - Jaboticabal

Jaboticabal, 08 de maio de 2017

DADOS CURRICULARES DO AUTOR

TÁSSIA SOUZA BERTIPAGLIA, nascida em Tupi Paulista – SP em 11 de agosto de 1986, filha de José Alipio Bertipaglia e Rosani Maria de Souza Bertipaglia. Iniciou o curso de Zootecnia em agosto de 2006 na Universidade Estadual Paulista “Júlio de Mesquita Filho” (FCAT/Unesp), Dracena – SP, e obteve o título de bacharel em Zootecnia em julho de 2011. Em agosto de 2011, ingressou no curso de mestrado do programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias (FCAV/UNESP), Jaboticabal - SP, sob orientação do professor Dr. Ricardo da Fonseca, tornando-se mestre em julho de 2013. Em agosto de 2013, ingressou no curso de doutorado do programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias (FCAV/ UNESP), Jaboticabal - SP, sob orientação do professora Dra. Vera Fernanda Martins Hossepian de Lima e coorientação do Prof. Dr. Danísio Prado Munari, Dr. Rodrigo Pelicioni Savegnago e Dr. Marcos Vinícius Gualberto Barbosa da Silva.

“Não é sobre ter todas as pessoas do mundo para si.
É sobre saber que em algum lugar alguém zela por ti.
É sobre cantar e poder escutar mais do que a própria voz.
É sobre dançar na chuva de vida que cai sobre nós.
É saber se sentir infinito num universo tão vasto e bonito, é saber sonhar.
Então fazer valer a pena cada verso daquele poema sobre acreditar.
Não é sobre chegar no topo do mundo e saber que venceu.
É sobre escalar e sentir que o caminho te fortaleceu.
É sobre ser abrigo e também ter morada em outros corações.
E assim ter amigos contigo em todas as situações.
A gente não pode ter tudo, qual seria a graça do mundo se fosse assim?
Por isso eu prefiro sorrisos e os presentes que a vida trouxe para perto de mim.
Não é sobre tudo que o seu dinheiro é capaz de comprar.
E sim sobre cada momento, sorriso a se compartilhar.
Também não é sobre correr contra o tempo para ter sempre mais.
Porque quando menos se espera, a vida já ficou para trás!
Segura teu filho no colo, sorria e abraça os teus pais enquanto estão aqui.
Que a vida é trem bala, parceiro, e a gente é só passageiro prestes a partir...”

Ana Vilela

Aos meus pais José Alipio e Rosani,
A minha irmã Camila, cunhado Gustavo e sobrinhos Cecília e Antonio,
Ao meu amor Fabio,
Pelo apoio, força, amor e por me guiarem nos momentos de dificuldade.

Dedico.

AGRADECIMENTOS

À Deus, pela vida, por me proteger e me guiar em toda a caminhada.

Ao Fabio, meu amor, marido, amigo, que foi essencial para este trabalho, pela ajuda, ensinamentos, paciência, companheirismo e por colaborar sempre que precisei. Muito obrigada!

Aos meus familiares, avós, tios, primos, cunhadas, sogros, e amigos, da graduação, da pós-graduação e demais amigos, que sempre torceram por mim.

Ao professor Dr. Danísio Prado Munari por me receber no seu grupo, pela orientação, ensinamentos e confiança no meu trabalho. À professora Dra. Vera Fernanda Martins Hossepian de Lima pela orientação e confiança. Aos coorientadores Dr. Rodrigo Penilioni Savegnago e Dr. Marcos Vinícius Gualberto Barbosa da Silva, pela coorientação e colaboração com o trabalho. Muito obrigada.

Aos professores da pós-graduação Dr. Adhemar Sanches, Dr. Anibal Eugênio Vercesi Filho, Dr. Danísio Prado Munari, Dr. Euclides Malheiros, Dr. Fernando Sebastian Baldi Rey, Dr. Guilherme Rosa, Dr. Henrique Nunes de Oliveira, Dr. João Ademir de Oliveira, Dra. Lúcia Galvão de Albuquerque, Dr. Ricardo da Fonseca, Dr. Roberto Carvalheiro e Dra. Sandra Aidar de Queiroz. Obrigada pelos ensinamentos.

Aos membros da banca de qualificação e defesa, Dra. Lenira El Faro, Dra. Joslaine Noely dos Santos Gonçalves Cyrillo, Dr. João Ademir de Oliveira, Dr. Ricardo da Fonseca, Dr. Rodrigo Pelicioni Savegnago, Dra. Nedenia Stafuzza. Obrigada por colaborarem com o meu trabalho. Em especial ao Dr. João Ademir, que colaborou diretamente com este trabalho. Obrigada pela disponibilidade.

Ao pessoal do grupo EAGMA, Alejandro, Ana Paula, Guilherme Bio, Jaque, Jorge, Letícia, Luara, Marcos, Natalia, Nedenia, Priscila, Rafael, Rebeka, Rodrigo, Salvador e Tati. Obrigada pela ajuda e pela companhia nos momentos do café.

À UNESP, ao Programa de Pós-graduação em Genética e Melhoramento Animal, aos funcionários da Unesp, aos queridos funcionários do Departamento de Ciências Exatas, à CAPES pela bolsa de estudos concedida.

À Embrapa CNPGL, principalmente ao Dr. Marcos Vinícius e Dr. João Claudio Panetto, e ABCGIL, por cederem os dados para as minhas análises e pela disponibilidade em colaborar com o trabalho.

A todos vocês minha gratidão!

SUMÁRIO

	Página
LISTA DE TABELAS.....	ii
LISTA DE FIGURAS.....	iii
RESUMO.....	iv
ABSTRACT.....	v
1 INTRODUÇÃO.....	1
2 REVISÃO DE LITERATURA.....	2
2.1 História da raça Gir no Brasil.....	2
2.2 Variabilidade genética.....	4
2.3 Controle de qualidade dos genótipos.....	5
2.4 Estrutura da população.....	11
2.5 Técnicas para análise de agrupamento.....	13
2.6 Análises de componentes principais (PCA).....	18
2.7 Parâmetros populacionais baseados em informações genômicas.....	21
3 MATERIAL E MÉTODOS.....	28
3.1 Dados genômicos.....	28
3.2 Controle de qualidade.....	29
3.3 Análise dos dados genômicos.....	30
4 RESULTADOS E DISCUSSÃO.....	36
5 CONCLUSÕES.....	57
6 REFERÊNCIAS.....	58
7 APÊNDICES.....	66

LISTA DE TABELAS

	Página
Tabela 1. Análise descritiva dos marcadores com base nos cromossomos	36
Tabela 2. Variâncias das distâncias genéticas dos indivíduos dos cluster 1, 2, 3 e 4 nos três primeiros componentes principais (PC1, PC2 e PC3)	39
Tabela 3. Média do desvio de parentesco genômico dentro (diagonal) e entre cluster (acima da diagonal) e desvio-padrão (entre parênteses e abaixo da diagonal)	51
Tabela 4. Média, mediana, desvio-padrão (DP), coeficiente de variação (CV), valores mínimo e máximo de heterozigosidade esperada (He), heterozigosidade observada (Ho), coeficiente de endogamia e desequilíbrio de ligação (LD) para a população estudada	52
Tabela 5. Probabilidade de indivíduos endogâmicos (x) e coeficiente de endogamia (e) de acordo com a Distribuição de Qui-Quadrado (Tabela IV).....	53
Tabela 6. Desequilíbrio de ligação (r^2) entre pares de SNP (N) localizados em diferentes distâncias (em Kb)	55

LISTA DE FIGURAS

	Página
Figura 1. Eixos do componente principal (CP1) e secundário (CP ₂), perpendicular, sendo a elipse a densidade.	20
Figura 2. Etapas utilizadas para a análise de componentes principais.....	31
Figura 3. Determinação do número ótimo de cluster obtidos pelos Métodos de Ward (a), Elbow (b), pela minimização da soma de quadrados, e Silhouette (c), pelo maior grau de ajuste.	38
Figura 4. Dendograma da matriz de distância genética de alelos compartilhados colorido para os diferentes cluster.	40
Figura 5. Projeção dos indivíduos em 3D, considerando os três primeiros componentes principais (PC1, PC2, PC3) em duas faces, anterior (a) e esquerda (b), coloridas para os diferentes cluster, e os touros mais influentes na variância dos componentes principais, enumerados.....	41
Figura 6. Proporção das variâncias dos dez primeiros componentes principais (barras) e proporção da variância acumulada de todos os componentes principais (linha contínua).	43
Figura 7. Projeção das distâncias genéticas dos indivíduos nos dois primeiros componentes principais.	43
Figura 8. Projeção das variâncias das distâncias genéticas dos indivíduos nos dois primeiros componentes principais coloridos considerando a contribuição individual.....	44
Figura 9. Projeção dos indivíduos nos três primeiros componentes principais em 3D considerando as diferentes fazendas, identificadas por diferentes colorações.	45
Figura 10. Projeção dos vinte touros mais (a) e menos (b) influentes na variância dos dois primeiros componentes principais.	46
Figura 11. Matriz de parentesco genômico entre os animais distribuídos em seus respectivos cluster.	49
Figura 12. Distribuição das estimativas de parentesco genômico dos animais nos diferentes cluster.....	50
Figura 13. Desequilíbrio de ligação (LD) ao longo das distâncias entre SNP em Megabase (Mb).....	56

ESTRUTURA GENÔMICA POPULACIONAL DE BOVINOS LEITEIROS GIR

RESUMO – Considerando painéis de SNP (do inglês *Single Nucleotide Polymorphism*, polimorfismo de nucleotídeo único) objetivou-se descrever o perfil genômico populacional de bovinos leiteiros da raça Gir por meio da caracterização de linhagens desta população e inferir sobre a variabilidade genética utilizando informações de animais provenientes de diversas fazendas do Brasil. Após o controle de qualidade de genótipos e amostras, foram obtidos 1987 SNP bialélicos em cromossomos autossomos presentes nos painéis HD e 50K SNP, em um painel resultante de SNP em comum dos dois painéis, 21K SNP. Distâncias genéticas entre os marcadores foram obtidas e, por meio da decomposição espectral, foram gerados os componentes principais. A análise multivariada de componentes principais, em um contexto de estrutura populacional, permite inferir sobre a variabilidade genética utilizando informações de parentesco entre os animais. A determinação do número de agrupamentos foi realizada com o procedimento *k-means*, utilizando a distância genética entre os animais, que permite identificar subgrupos genéticos na população, com moderada-alta qualidade de agrupamento, com coeficiente de correlação cofenética equivalente a 0,66. O parentesco, heterozigosidade esperada (H_e) e observada (H_o), coeficientes de endogamia, desequilíbrio de ligação (LD) e tamanho efetivo populacional, baseados exclusivamente em informações genômicas também foram obtidos. Os resultados evidenciaram quatro diferentes agrupamentos genéticos, definidos como linhagens da raça Gir leiteira do Brasil, indicando que o uso de poucos reprodutores ou de seus descendentes com maior intensidade e também devido a diferentes origens dos ancestrais, formou os diferentes agrupamentos genéticos, os quais são compostos por animais de várias fazendas. Há pouco parentesco entre os animais e isso se refletiu na média do coeficiente de endogamia, sendo que a média da endogamia obtida para todos os indivíduos foi de 0,017, evidenciando o fluxo gênico nas diferentes fazendas, o que evitou o acasalamento entre animais aparentados, ou ainda devido à escolha dos animais para genotipagem. O valor médio de H_e foi 0,25, próximo a média de H_o , revelando que não houve perda e nem fixação de alelos com efeitos indesejáveis. O LD foi obtido para todos os pares de SNP adjacentes com janelas de 50 Mb, revelando que o LD é maior em menores distâncias entre os marcadores, com valor médio de LD de $0,17 \pm 0,26$ para até 200 Kb, sendo assim, o LD estimado para espaçamento até 40 Kb seria suficiente para se obter associação com *Quantitative Trait Loci*. O tamanho efetivo populacional (N_e) foi 92,84 animais, sendo assim, cerca de 93 indivíduos nesta população contribuirão em termos de variância genética para a próxima geração desta população. Portanto, concluiu-se que o fluxo gênico nas propriedades mantém a variabilidade genética dos rebanhos. Esses parâmetros populacionais poderão auxiliar nas decisões de acasalamentos dos programas de melhoramento genético da raça a fim de aumentar ou manter a variabilidade genética dos rebanhos.

Palavras-chave: análise multivariada, *Bos taurus indicus*, bovino de leite, linhagens, painéis de SNP, variabilidade genética

GENOMIC STRUCTURE OF DAIRY CATTLE GIR POPULATION

ABSTRACT – The objective was to describe the genomic structure of Gir dairy cattle population through the characterization of lineages of this population and infer about the genetic variability considering information from animals from several farms in Brazil using panels of SNP (Single Nucleotide Polymorphism). After quality control of genotypes and samples, we obtained 1987 biallelic SNPs on autosomal chromosomes present in the HD panels and 50K SNP, in a resultant panel of SNPs in common from two panels, 21K SNP. Genetic distances between the markers were obtained and, through spectral decomposition, the principal components were generated. Multivariate analysis of principal components, in a context of population structure, allows inferring about genetic variability using relatedness information between animals. The determination of the number of clusters was performed using the k-means procedure, using the genetic distance between the animals, which allows the identification of genetic subgroups in the population, with moderate-high quality of grouping, with a coefficient correlation coefficient equivalent to 0.66. Expected heterozygosity (H_e) and observed (H_o), coefficients of inbreeding, linkage disequilibrium (LD) and effective population size (N_e) based exclusively on genomic information were also obtained. The results evidenced four different genetic groups, defined as breeding lineages of the Gir dairy breed of Brazil, indicating that the use of a few breeders or their descendants with greater intensity and also due to the different origins of the ancestors, formed the different genetic groups, which are composed of animals from various farms. There was little relationship between the animals and this was reflected in the average inbreeding coefficient, and the average inbreeding obtained for all the individuals was 0.017, evidencing the gene flow in the different farms, which avoided the mating between related animals, or due to the choice of animals for genotyping. The mean value of H_e was 0.25, close to the mean of H_o , revealing that there is no loss or fixation of alleles with undesirable effects. The LD was obtained for all adjacent SNP pairs with 50 Mb windows, revealing that the LD is larger at smaller distances between the markers, with a mean LD value of 0.17 ± 0.26 for up to 200 Kb, therefore, the estimated LD for spacing up to 40 Kb would be sufficient to obtain association with quantitative trait loci (QTL). The effective population size was 92.84 animals, thus, about 93 individuals in this population will contribute in terms of genetic variance for the next generation of this population. Therefore, it was concluded that the gene flow in the properties maintains the genetic variability of the herds. These population parameters may aid in the breeding decisions of breed breeding programs in order to increase or maintain the genetic variability of the herds.

Keywords: *Bos taurus indicus*, dairy cattle, genetic variability, lineages, multivariate analysis, SNP panels

1 INTRODUÇÃO

O agronegócio do leite e seus derivados ocupam uma posição de destaque na economia nacional. Além do leite ser um dos produtos mais importantes da pecuária brasileira, é uma fonte de alimento de alto valor nutritivo e representa importante papel social na geração de empregos no país. De acordo com a Organização das Nações Unidas para Alimentação e Agricultura, o Brasil é um fundamental agroexportador, principalmente em relação aos produtos do setor pecuário, sendo responsável por 40% da produção de leite da América Latina. A produção de leite no Brasil dobrou entre os anos 2000 e 2015, o que colocou o Brasil entre os cinco maiores produtores mundiais de leite de vaca no *ranking* mundial, superado pela Índia, Estados Unidos, China e Paquistão (FAO, 2015).

A raça Gir é a principal zebuína leiteira e também a mais utilizada em cruzamentos com a raça Holandesa no país (SILVA; MACHADO, 2011). Apesar da pequena população da raça quando comparada às demais implementadas no país, a contribuição genética do Gir leiteiro para os rebanhos comerciais tem aumentado. O aumento no uso de animais da raça Gir foi atribuído à implementação do Programa Nacional de Melhoramento de Gir Leiteiro (PNMGL) em 1985, com rápido crescimento desde aquele ano.

Para o desenvolvimento dos programas de melhoramento genético da raça, um fator importante a ser considerado é o conhecimento da estrutura genômica populacional. No contexto de estrutura populacional é possível inferir sobre a variabilidade genética da população, fator fundamental, pois além de fatores evolutivos, está aliada à expressão do potencial genético da população e à facilidade de seleção de animais geneticamente superiores. Além disso, o estudo da estrutura genômica populacional é uma importante ferramenta por permitir maior compreensão sobre o histórico da população como endogamia, animais que contribuem na variabilidade genética da população, introdução de novo material genético, identificar subgrupos genéticos, agrupá-los por semelhança genética, detectar desequilíbrio de ligação entre marcadores e QTL e conhecer as relações de parentesco.

O estudo da estratificação populacional para identificar possíveis linhagens, que são subestruturas genéticas, permite considerar esses efeitos para se obter maior acurácia nas análises utilizando informações genômicas, por alocar os indivíduos de diferentes grupos genéticos nos grupos de treinamento e validação nas análises genômicas. Ao considerar os diferentes subgrupos genéticos há a correção para a estratificação (PRICE et al., 2006) por ajustar o efeito dos agrupamentos e, com isso, elimina a ocorrência de falsos SNP significativos para as características de interesse econômico de forma mais eficiente, por abranger todos os grupos genéticos.

A compreensão da estrutura genômica populacional é importante para os estudos de predição genômica, seleção e associação, e poderá auxiliar nas decisões dos acasalamentos dentro dos programas de melhoramento genético da raça Gir. Portanto, o objetivo deste estudo foi descrever o perfil genômico populacional por meio da caracterização de linhagens desta população de bovinos leiteiros da raça Gir e inferir sobre a variabilidade genética, por meio das análises de componentes principais, análise de agrupamento, endogamia, heterozigosidade, desequilíbrio de ligação e tamanho efetivo populacional, baseados exclusivamente em informações genômicas contidas em painéis de SNP.

5 CONCLUSÕES

A população está estruturada pela presença de quatro agrupamentos genéticos, podendo ser definidos como linhagens desta população, as quais devem ser consideradas nas análises de GWAS e seleção genômica. Portanto, além do desequilíbrio de ligação, as análises genômicas para esta população devem levar em conta os efeitos de subgrupos genéticos.

Com base no parentesco genômico, heterozigosidade e coeficiente de endogamia, concluiu-se que há fluxo gênico dos indivíduos desta população nas

fazendas, proporcionado pelos acasalamentos entre animais pouco aparentados, mantendo a variabilidade genética nesta população, apesar do tamanho efetivo populacional pequeno. Esses resultados devem ser considerados nas decisões de escolha dos indivíduos para acasalamento nos programas de melhoramento genético da raça a fim de aumentar ou manter a variabilidade genética dos rebanhos.

6 REFERÊNCIAS

ABCGIL. Associação Brasileira de Criadores de Gir Leiteiro. 2017. Acesso em: 15 mar. 2017. Disponível em: <<http://girleiteiro.org.br/>>.

ACGZ. **Associação dos Criadores de Gaúchos de Zebu**. 2016. Disponível em: <http://www.acgz.com.br/secao_racas.php?pagina=5>. Acesso em: 3 nov. 2016.

ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, K. Fast model-based estimation of ancestry in unrelated individuals. **Genome Research, Woodbury**, v. 19, n. 9, p. 1655-1664, 2009.

AMARAL, R.; VILLARES, J. B.; FARIA, R. S.; GESTAL, R. L. Melhoramento tecnológico: estímulo alimentar, no período de lactação de Gir leiteiro. In: CONGRESSO BRASILEIRO DE PESQUISA DE ZEBU, 1, 1988, Uberaba. **Anais...** Uberaba: Empresa de Pesquisa Agropecuária de Minas Gerais, 1988.

ANNEY, R. J. L.; KENNY, E.; O'DUSHLAINE, C. T.; LASKY-SU, J.; FRANKE, B.; MORRIS, D. W.; NEALE, B. M.; ASHERSON, P.; FARAONE, S. V.; GILL, M. Non-Random Error in Genotype Calling Procedures: implications for family-based and case-control genome-wide association studies. **American Journal of Medical Genetics B: Neuropsychiatric Genetics**, Hoboken, v. 147, n. 8, p. 1379–1386, 2008.

BIEGELMEYER, P.; GULIAS-GOMES, C. C.; CAETANO, A. R.; STEIBEL, J. P.; CARDOSO, F. F. Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle. **BMC Genetics**, London, v. 17, n. 1, p. 32, 2016.

BULMER, M. G. The effect of selection on genetic variability. **The American Naturalist**, v. 105, n. 943, p. 201-211, 1971.

BUZANSKAS, M. E.; VENTURA, R. V.; CHUD, T. C. S.; BERNARDES, P. A.; SANTOS, D. J. A.; REGITANO, L. C. A.; ALENCAR, M. M.; MUDADU, M. A.; ZANELLA, R.; SILVA, M. V. B. G.; LI, C.; SCHENKEL, F. S.; MUNARI, D. P. Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. **PloS One**, San Francisco, v. 12, n. 2, p. e0171660, 2017.

CARDON, L. R.; PALMER, L. J. Population stratification and spurious allelic association. **Lancet**, London, v. 361, n. 9357, p. 598-604, 2003.

CAVALLI-SFORZA, L. L.; FELDMAN, M. W. The application of molecular genetic approaches to the study of human evolution. **Nature Genetics**, v. 33, p. 266–275, 2003.

CIRULLI, E. T.; GOLDSTEIN, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. **Nature Reviews Genetics**, London, v. 11, n. 6, p. 415-425, 2010.

COLEMAN, J. R.; EUESDEN, J.; PATEL, H.; FOLARIN, A. A.; NEWHOUSE, S.; BREEN, G. Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray. **Briefings in Functional Genomics**, Oxford, v. 15, n. 4, p. 298-304, 2016.

COX, D. G.; KRAFT, P. Quantification of the power of Hardy-Weinberg equilibrium testing to detect genotyping error. **Human Heredity**, New York, v. 61, n. 1, p. 10-14, 2006.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: Editora UFV, 2014.

ELSIK, C. G.; TELLAM, R. L.; WORLEY, K. C. The genome sequence of taurine cattle: a window to ruminant biology and evolution. **Science**, Washington, v. 324, n. 5926, p. 522-528, 2009.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to Quantitative Genetics**. London: Pearson United Kingdom, 1996. p.1-22, 57-72, 247-262.

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

FAO. Food and Agriculture Organization of the United Nations. **Statistical Pocketbook. World food and agriculture**. 2015. Disponível em: <<http://www.fao.org/3/a-i4691e.pdf>>. Acesso em: 15 mar. 2017.

FORNI, S.; AGUILAR, I.; MISZTAL, I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. **Genetics Selection Evolution**, Les Ulis, v. 43, n. 1, p. 1, 2011.

GILES, R.E.; BLANC, H.; CANN, H. M.; WALLACE, D. C. Maternal inheritance of human mitochondrial DNA. **Proceeding of the National Academy of Sciences of the United States of America**, v. 77, n. 11, p. 6715-6719, 1980.

GODDARD, M. Mapping genes for quantitative traits using linkage disequilibrium. **Genetics Selection Evolution**, Les Ulis, v. 23, n. Suppl 1, p. 131s-134s, 1991.

GONDRO, C. **Primer to analysis of genomic data using R**. Cham: Springer. 2015. cap. 3, p. 91.

GONDRO, C.; VAN DER WERF, J.; HAYES, B. **Genome-Wide Association Studies and Genomic Prediction**. Armidale: Humana Press, 2013. cap. 9, p. 218, cap. 19, p. 430-434.

GRIFFITHS, F; WESSLER, S. R.; CARROLL, S. B.; DOEBLEY, J. **Introdução em Genética**. Rio de Janeiro: Guanabara koogan, 2015. p. 170.

GUTIÉRREZ, J. P.; ALTARRIBA, J.; DÍAZ, C.; QUINTANILLA, R.; CAÑÓN, J.; PIEDRAFITA, J. Pedigree analysis of eight Spanish beef cattle breeds. **Genetics Selection Evolution**, Les Ulis, v. 35, n. 1, p. 43-64, 2003.

HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Análise multivariada de dados**. Porto Alegre: Bookman, 2005.

HARTL, D. L.; CLARK, A. G. **Princípios de Genética de Populações**. 4. ed. Porto Alegre: Artmed, 2010. cap. 9, p. 491-500.

HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Invited review: Genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, Champaign, v. 92, n. 2, p. 433-443, 2009.

HAYES, B. J.; VISSCHER, P. M.; MCPARTLAN, H. C.; GODDARD, M. E. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. **Genome Research**, Woodbury, v. 13, n. 4, p. 635-643, 2003.

HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, Heidelberg, v. 38, n. 6, p. 226-231, 1968.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, Washington, v. 24, n. 6, p. 498-520, 1933.

ILLUMINA. **GenomeStudio: An integrated platform for data visualization and analysis**. 2009. Disponível em: <<http://support.illumina.com/>>. Acesso em: 13 set. 2016.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRAN, R. **An Introduction to Statistical Learning with Applications in R**. 2. ed. New York: Springer-Verlag New York, 2013. cap. 6, p. 230-237.

JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC Genetics**, London, v. 11, n. 1, p. 94, 2010.

KOLDE, R. **Package "pheatmap"**. R-project. Disponível em: <<http://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf>>. Acesso em: 7 nov. 2016.

KRISTENSEN, T. N.; SORENSEN, A. C. Inbreeding – lessons from animal breeding, evolutionary biology and conservation genetics. **Animal Science**, Cambridge, v. 80, n. 2, p. 121-133, 2005.

LAM, A. C.; SCHOUTEN, M.; AULCHENKO, Y. S.; HALEY, C. S.; KONING, D. J. Rapid and robust association mapping of expression quantitative trait loci. **BMC Proceedings**, London, v. 1, n. 1, p. S144, 2007.

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

LEÃO, G. F. M.; PIVATO, D. R. D.; CARNIEL, H.; RODRIGUES, M. G. K.; BRAGA, R. A.; SILVA, M. R. H.; TEIXEIRA, P. P. M. Melhoramento genético em zebuínos leiteiros – uma revisão. **Agropecuária Científica no Semiárido**, Campina Grande, v. 9, n. 4, p. 9-14, 2013.

LEE, S. H.; CHO, Y. M.; LIM, D.; KIM, H. C.; CHOI, B. H.; PARK, H. S.; KIM, O. H.; KIM, S.; KIM, T. H.; YOON, D.; HONG, S. K. Linkage disequilibrium and effective population size in Hanwoo Korean Cattle. **Asian-Australasian Journal of Animal Sciences**, Seoul, v. 24, n. 12, p. 1660-1665, 2011.

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

LIN, D. Y.; HU, Y.; HUANG, B. E. Simple and efficient analysis of disease association with missing genotype data. **The American Journal of Human Genetics**, Houston, v. 82, n. 2, p. 444-452, 2008.

MACKAY, I.; POWEL, W. Methods for linkage disequilibrium mapping in crops. **Trends in Plant Science**, v. 12, n. 2, p. 57-63, 2007.

MADALENA, F. E.; TEODORO, R. L.; LEMOS, A. M.; MONTEIRO, J. B. N.; BARBOSA, R. T. Evaluation of Strategies for Crossbreeding of Dairy Cattle in Brazil. **Journal of Dairy Science**, Champaign, v. 73, n. 7, p. 1887-1901, 1990.

MAKINA, S. O.; MUCHADEYI, F. C.; VAN MARLE-KÖSTER, E.; MACNEIL, M. D.; MAIWASHE, A. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. In: **Advances in Farm Animal Genomic Resources**, Lausanne:Frontiers in Genetics, 2014. p. 92-98.

MARCHINI, J.; CARDON, L. R.; PHILLIPS, M. S.; DONNELLY, P. The effects of human population structure on large genetic association studies. **Nature Genetics**, New York, v. 36, n. 5, p. 512-517, 2004.

MASON, I. L. **Mason's World Dictionary of Livestock Breeds, Types and Varieties**. Wallingford: CAB Internacional, 2002. p. 33, 43.

MCKAY, S D; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. K.; AERTS, J.; COPPIETERS, W.; CREWS, D.; DIAS NETO, E.; GILL, C. A.; GAO, C.; MANNEN, H.; STOTHARD, P.; WANG, Z.; VAN TASSELL, C. P.; WILLIAMS, J. L.; TAYLOR, J.F.; MOORE, S. S. Whole genome linkage disequilibrium maps in cattle. **BMC Genetics**, London, v. 8, n. 1, 2007.

NCBI. National Center for Biotechnology Information Gene. **Bos indicus (zebu cattle)**. Disponível em: <<https://www.ncbi.nlm.nih.gov/genome/?term=bos+indicus%5Borgn%5D>>. Acesso em: 13 mar. 2017.

NEIVA, R. Genômica promove seleção mais veloz. **Revista XXI Ciência para a vida**, 2017, p. 21.

NEVES, H. H. R.; DESIDÉRIO, J. A.; PIMENTEL, E. C. G.; SCALEZ, D. C. B.; QUEIROZ, S. A. Preliminary study to determine extent of linkage disequilibrium and estimates of autozygosity in Brazilian Gyr dairy cattle. **Archivos de Zootecnia**, Cordoba, v. 64, n. 246, p. 99-107, 2015.

NOTTER, D. R. The importance of genetic diversity in livestock populations of the future. **Journal of Animal Science**, Champaign, v. 77, n. 1, p. 61-69, 1999.

NOVEMBRE, J.; STEPHENS, M. Interpreting principal component analyses of spatial population genetic variation. **Nature Genetics**, New York, v. 40, p. 646-649, 2008.

O'BRIEN, A. M. P.; MÉSZÁROS, G.; UTSUNOMIYA, Y. T.; SONSTEGARD, T. S.; GARCIA, F. J.; TASSELL, C. P. V.; CAVALHEIRO, R.; SILVA, M. V. B.; SOLKNER, J. Linkage disequilibrium levels in *Bos indicus* and *Bos Taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. **Livestock Science**, Amsterdam, v. 166, n. 5, p. 646-649, 2014.

PANETTO, J. C. C.; VERNEQUE, R. S.; SILVA, M. V. G. B.; MACHADO, M. A.; MARTINS, M. F.; BRUNELI, F. A. T.; PEIXOTO, M. G. C. D.; SANTOS, G. G.; ARBEX, W. A.; REIS, D. R. L.; GERALDO, C. C.; MACHADO, C. H. C.; VENTURA, H. T.; PEREIRA, M. A.; HORTOLANI, B.; VERCESI FILHO, A. E.; MACIEL, R. S.; FERNANDES, A. R. **Programa Nacional de Melhoramento do Gir Leiteiro - Sumário brasileiro de touros - Resultado do teste de progênie 7ª prova de pré-seleção de touros**. Juiz de Fora: Embrapa: CNPGL, 2016. 86 p. (Embrapa-CNPGL. Documentos, 187).

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

PEARSON, K. On Lines and Planes of Closest Fit to Systems of Points in Space, **Philosophical Magazin**, London, v. 2, p. 559-572, 1901.

PEREIRA, R. J.; SANTANA JÚNIOR, M. L.; AYRES, D. R.; BIGNARDI, A. B.; VERCESI FILHO, A. E. Depressão endogâmica na produção de leite no dia do controle de bovinos Gir leiteiro. **Pesquisa Agropecuária Brasileira**, Brasília, v. 51, n. 6, p. 751-758, 2016.

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

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

PRITCHARD, J. K.; PRZEWORSKI, M. Linkage disequilibrium in humans: models and data. **The American Journal of Human Genetics**, Houston, v. 69, n. 1, p. 1-14, 2001.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A.; BENDER, D.; MALLER, J.; SKLAR, P.; BAKKER, P. I.; DALY, M. J.; SHAM, P. C. PLINK: a tool set for whole-genome association and population-based linkage analyses. **The American Journal of Human Genetics**, Houston, v. 81, n. 3, p. 559-575, 2007.

QUEIROZ, S. A.; ALBUQUERQUE, L. G.; LANZONI, N. A. Efeito da endogamia sobre características de crescimento de bovinos da raça Gir no Brasil. **Revista Brasileira de Zootecnia**, Viçosa, v. 29, n. 4, p. 1014-1019, 2000.

R CORE TEAM. **A language and environment for statistical computing**, R Foundation for Statistical Computing. Vienna, Austria. 2016.

REIS FILHO, J. C.; LOPES, P. S.; VERNEQUE, R. S.; TORRES, R. A.; TEODORO, R. L.; CARNEIRO, P. L. S. Population structure of Brazilian Gyr dairy cattle. **Revista Brasileira de Zootecnia**, Viçosa, v. 39, n. 12, p. 2640-2645, 2010.

RELETHFORD, J. H. **Human population genetics**. Hoboken: John Wiley & Sons, 2012. p. 112-119.

ROUSSEEUW, P. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, n. 1, 1987.

SANTANA JUNIOR, M. L.; PEREIRA, R. J.; BIGNARDI, A. B.; EL FARO, L.; TONHATI, H.; ALBUQUERQUE, L. G. History, structure, and genetic diversity of Brazilian Gir cattle. **Livestock Science**, Amsterdam, v. 163, p. 26-33, 2014.

SANTOS, R.; CORRÊA, A. B. Como são feitos os testes de progênie (provas de touros). In: SIMPÓSIO NACIONAL DE MELHORAMENTO ANIMAL, 3, 2000, Belo Horizonte. **Anais eletrônicos...** Belo Horizonte: Sociedade Brasileira de Melhoramento Animal, 2000. Disponível em: <<http://sbmaonline.org.br/anais/iii/palestras/pdfs/iip26.pdf>>. Acesso em: 9 jan. 2017.

SARGOLZAEI, M.; SCHENKEL, F. S.; JANSEN, G. B.; SCHAEFFER, L. R. Extent of linkage disequilibrium in Holstein cattle in North America. **Journal of Dairy Science**, Champaign, v. 91, n. 5, p. 2106-2117, 2008.

SEIDEL, E. J.; MOREIRA JÚNIOR, F. J.; ANSUJ, A. P.; NOAL, M. R. C. Comparação entre o método Ward e o método K-médias no agrupamento de produtores de leite. *Ciência e Natura*, v. 30, n. 1, p. 7-15, 2008.

SILVA, P. B. R.; MACHADO, C. H. C. Atribuições da subdivisão do julgamento da raça Gir em: Gir leiteiro e Gir dupla-aptidão. **Cadernos de Pós-Graduação da FAZU**, 2011.

SLATKIN, M. Linkage disequilibrium-understanding the evolutionary past and mapping the medical future. **Nature Reviews Genetics**, London, v. 9, n. 6, p. 477-485, 2008.

SNEATH, P. H. A.; SOKAL, R. R. **Numerical taxonomy**. San Francisco: W. H. Freeman & Co, 1973.

SOKAL, R. R.; ROHLF, F. J. The Comparison of Dendrograms by Objective Methods. **Taxon**, v. 11, n. 2, p. 33-40, 1962.

TEO, Y. Y. Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. **Current Opinion in Lipidology**, v. 19, n. 2, p. 133-143, 2008.

TEO, Y. Y.; FRY, A. E.; CLARK, T. G.; TAI, E. S.; SEIELSTAD, M. On the usage of HWE for identifying genotyping errors. **Annals of Human Genetics**, v. 71, n. 5, p. 701-703, 2007.

TORO, M. A.; VILLANUEVA, B.; FERNÁNDEZ, J. Genomics applied to management strategies in conservation programmes. **Livestock Science**, Amsterdam, v. 166, 48-53, 2014.

TURNER, S.; ARMSTRONG, L. L.; BRADFORD, Y.; CARLSON, C. S.; CRAWFORD, D. C.; CRENSHAW, A. T.; ANDRADE, M.; DOHENY, K. F.; HAINES, J. L.; HAYES, G.; JARVIK, G.; JIANG, L.; KULLO, I. J.; LI, R.; LING, H.; MANOLIO, T. A.; MATSUMOTO, M.; MCCARTY, C. A.; MCDAVID, A. N.; MIREL, D. B.; PASCHALL J. E., PUGH, E. W.; RASMUSSEN, L. V.; WILKE, R. A.; ZUVICH, R. L.; RITCHIE M. D. Quality control procedures for genome wide association studies. **Current Protocols in Human Genetics**, Hoboken, p. 1-24, 2011.

VENTURA, R.; LARMER, S.; SCHENKEL, F. S.; MILLER, S. P; SULLIVAN, P. Genomic clustering helps to improve prediction in a multibreed population. **Journal of Animal Science**, Champaign, v. 94, n. 5, p. 1844-1856, 2016.

VERCESI FILHO, A. E.; DIAS, A. L.; CARDOSO, V. L.; EL FARO, L.; MERINGUE, G. K. F.; MEIRELLES, F. V. Caracterização de um rebanho Gir Leiteiro quanto à origem do DNA mitocondrial (mtDNA). **Boletim de Indústria Animal**, v. 67, n. 1, p. 91-95, 2010.

VILAÇA, L. F.; DINIZ, W. J. S.; MELO, T. F.; OLIVEIRA, J. C. V.; GUIDO, S. I.; BRITO, C. E. V. L.; COSTA, N. A.; SANTORO, K. R. Polimorfismos do gene BoLA-DRB3 em rebanhos bovinos leiteiros 5/8 Girolando e Holandês no estado de Pernambuco. **Archivos de Zootecnia**, Cordoba, v. 65, n. 249, p. 7-11, 2016.

WANG, D.; SUN, Y.; STANG, P.; BERLIN, J. A.; WILCOX, M. A.; QINGQIN, L. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. **BMC Proceedings**, London, v. 3, p. S109, 2009.

WRIGHT, S. Evolution in Mendelian Populations, **Genetics**, Bethesda, v. 16, n. 2, p. 97–159, 1931.

YAMAZAKI, T. The effects of overdominance on linkage in a multilocus system. **Genetics**, v. 86, n. 1, p. 227-236, 1977.

ZHANG, Q; CALLUS, M. P. L.; GULDBRANDTSEN, B.; LUND, M. S.; SAHANA, G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. **BMC Genetics**, London, v. 16, p. 1-11, 2015.

ZHAO, H.; NETTLETON, D.; DEKKERS, J. C. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. **Genetical Research**, Cambridge, v. 89, n. 1, p. 1-6, 2007.

ZIEGLER, A.; KONIG, I. R.; THOMPSON, J. R. Biostatistical aspects of genome-wide association studies. **Biometrical Journal**, Malden, v. 50, n. 1, p. 8-28, 2008.

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. A whole-genome assembly of the domestic cow, *Bos Taurus*. **Genome Biology**, London, v. 10, p. R42, 2009.