UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

CARACTERIZAÇÃO DAS CORRIDAS DE HOMOZIGOSE NO GENOMA DE BOVINOS DA RAÇA GIR

Elisa Peripolli
Zootecnista

2016

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

CARACTERIZAÇÃO DAS CORRIDAS DE HOMOZIGOSE NO GENOMA DE BOVINOS DA RAÇA GIR

Elisa Peripolli
Orientador: Prof. Dr. Fernando Sebástian Baldi Rey
Coorientadores: Prof. Dr. André Luís Ferreira Lima
Dr. Marcos Vinícius Gualberto Barbosa da Silva
Prof. Dr. Renato Irgang

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

2016

Peripolli, Elisa
P445c Caracterização das corridas de homozigose no genoma de bovinos da raça Gir / Elisa Peripolli. – – Jaboticabal, 2017
vi, 81 p. : il. ; 29 cm

Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2017
Orientador: Fernando Sebástian Baldi Rey
Banca examinadora: Ignácio Aguilar; Roberto Carvalheiro
Bibliografia

1. Autozigosidade 2. *Bos indicus* 3. Diversidade genética 4. Endogamia. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.2:636.082

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação – Serviço Técnico de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: CARACTERIZAÇÃO DAS CORRIDAS DE HOMOZIGOSE NO GENOMA DE BOVINOS DA RAÇA GIR

AUTORA: ELISA PERIPOLLI
ORIENTADOR: FERNANDO SEBASTIAN BALDI REY
COORIENTADOR: ANDRÉ LUÍS FERREIRA LIMA
COORIENTADOR: RENATO IRGANG
COORIENTADOR: MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA

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

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

Prof. Dr. IGNACIO AGUILAR (Participação por Videoconferência)
Instituto Nacional de Investigación Agropecuaria / Canelones - Uruguay

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

Jaboticabal, 17 de fevereiro de 2017

DADOS CURRICULARES DO AUTOR

Elisa Peripolli, nascida em 10 de abril de 1991 na cidade de Joinville – Santa Catarina, filha de Odilo João Peripolli e Ingrid Zimmermann Peripolli. Iniciou em março de 2009 o curso de graduação em Zootecnia na Universidade Federal de Santa Catarina, obtendo o título de Zootecnista em fevereiro de 2015. Durante a graduação foi bolsista de mobilidade acadêmica na University of Delaware – EUA pelo programa Ciências sem Fronteiras. Durante o período de mobilidade acadêmica, além de cursar as disciplinas de zootecnia, fez estágio no Laboratório de Genética da mesma instituição de fomento, sob a orientação do Prof. Dr. Behnam Abasht. Em março de 2015, ingressou no Programa de Pós-graduação em Genética e Melhoramento Animal da Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista “Júlio de Mesquita Filho”, campus de Jaboticabal, sob a orientação do Prof. Dr. Fernando Sebastián Baldi Rey, como bolsista CAPES-EMBRAPA.

“*Nunca, jamais desanimeis, embora venham ventos contrários*”

Santa Madre Paulina

Dedico

Àqueles que não medem esforços para realizar meus sonhos, que sempre estarão ao meu lado e com suas mãos sempre estendidas, meus pais Odilo João Peripolli e Ingrid Zimmermann Peripolli e aos meus irmãos Jorge e André.

AGRADECIMENTOS

À minha família que sempre apoiou minhas escolhas e compreendeu minha ausência. Aos meus pais Odilo e Ingrid pelos sábios ensinamentos e conselhos, por serem minha fortaleza e meu porto seguro e por sempre me encorajarem a seguir em frente e a nunca desistir dos meus sonhos. Aos meus irmãos Jorge e André, apesar dos “arranca rabos” na infância, eu não escolheria irmãos melhores para ter.

Ao meu orientador Prof. Dr. Fernando Sebastián Baldi Rey pela disponibilidade em me orientar e por me receber tão bem no seu grupo de pesquisa. Pela confiança em mim depositada na execução desse e de outros trabalhos, pelos seus ensinamentos que muito contribuíram para minha formação e por ser um grande amigo além de um ótimo orientador.

Aos meus coorientadores, Marcos Vinícius Gualberto Barbosa da Silva, André Luís Ferreira Lima e Renato Irgang, pelo auxílio e prontidão no decorrer do desenvolvimento deste trabalho. À Embrapa Gado de leite (Juiz de Fora - MG) pelo fornecimento dos dados.

Ao meu namorado, Hugo Borges de Quadros, por compreender minha ausência, meus momentos de estresse e por sempre estar ao meu lado me incentivando e me motivando a seguir em frente. Você foi fundamental para que eu conseguisse chegar até aqui, obrigada pelo companheirismo de uma vida inteira!

Às meninas do apartamento trinta e três, Malise e Ana Paula, por me receberem tão bem e por se tornarem minhas irmãs, pelas pipocas, filmes, risadas e gordices. À Malise por dividir sua família comigo, em especial à Margarete e ao Paulinho.

Aos meus “hermanitos”, Bianca, Fabi, Tonussi, Mari, Marcos, Hermenegildo e Medeiros, por tornarem a salinha divertida e as conversas produtivas quando não estávamos produtivos. A todos os amigos do café e aos demais do programa de Genética e Melhoramento Animal – UNESP/FCAV.

Aos amigos do coração, Max, Gabi, Lê, Elaine, Géssica e Jacke, pela amizade verdadeira e por todos os momentos que compartilhamos.

Aos membros da banca do Exame Geral de Qualificação, Dr. Roberto Carvalheiro e Dr. Nedenia Bonvino Stafuzza, pelas sugestões que muito contribuíram e acrescentaram a esse trabalho.

Aos membros da banca de defesa da dissertação, Dr. Ignacio Aguilar e Dr. Roberto Carvalheiro, pelas contribuições e disponibilidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos concedida no curso de mestrado.

SUMÁRIO

CARACTERIZAÇÃO DAS CORRIDAS DE HOMOZIGOSE NO GENOMA DE BOVINOS DA RAÇA GIR

RESUMO – As corridas de homozigose, do inglês "*Runs of homozygosity*" (ROH), são segmentos homozigóticos contínuos que estão presentes em indivíduos e populações. A habilidade desses segmentos em elucidar sobre eventos genéticos populacionais torna-os uma ferramenta capaz de prover informações valiosas a respeito da evolução demográfica de uma população ao longo do tempo. Além disso, informações amplas do genoma fornecem subsídios relevantes para compreender a constituição genética de um animal por meio da caracterização dos ROH, constituindo uma metodologia acurada para manter a diversidade genética em diversas populações animais. O objetivo deste estudo foi (i) acessar a autozigosidade do genoma de bovinos da raça Gir Leiteiro a fim de caracterizar os padrões de ROH; (ii) prospectar genes em ROH compartilhados por mais de 50% da população, e por fim (iii) comparar as estimativas de endogamia calculadas a partir da proporção genômica em homozigose (F_{ROH}), da matriz genômica de parentesco G (F_{GRM}) e do pedigree tradicional (F_{PED}). Animais da raça Gir Leiteiro foram genotipados com o BovineHD BeadChip (Illumina Inc., San Diego, CA, USA) que contêm 777.962 SNPs (n=582), BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) contendo 54.609 SNPs (n=1664) e GGP-LD *Indicus* (Geneseek® Genomic Profiler *Indicus* 30K) que contêm 27.533 SNPs (n=662). Todos os genótipos foram imputados para o painel BovineHD BeadChip (Illumina Inc., San Diego, CA, USA). SNPs sem posição definida ou mapeados nos cromossomos sexuais foram removidos do conjunto de dados. Após a edição, 2908 animais e 735,236 SNPs foram mantidos para as análises. Os ROH foram identificados por meio do software PLINK v1.07 considerando os seguintes parâmetros: uma janela deslizante de 50 SNPs, o número mínimo de SNPs consecutivos incluídos em um ROH foi 100, o comprimento mínimo de um ROH foi ajustado para 1 Mb, o intervalo máximo entre SNPs homozigóticos consecutivos foi de 500 kb, uma densidade de 1 SNP por 50 kb, cinco SNPs com genótipos faltantes e um genótipo heterozigoto. O F_{PED} foi estimado por meio do software INBUPGF90. Para cada animal foi calculado um F_{ROH} ($F_{ROH1\text{-}2\ Mb}$, $F_{ROH2\text{-}4\ Mb}$, $F_{ROH4\text{-}8\ Mb}$, $F_{ROH8\text{-}16\ Mb}$ e $F_{ROH>16\ Mb}$) com base na distribuição de ROH a partir de cinco comprimentos (ROH_j): 1-2, 2-4, 4-8, 8-16 e > 16 Mb, respectivamente. O F_{GRM} foi calculado a partir da diagonal da matriz de relação genômica (G). Os ROH foram identificados em todos os animais, apresentando um número médio de 55,12±10,37 segmentos por animal e um comprimento médio de 3,17 Mb. Segmentos curtos ($ROH_{1\text{-}2\ Mb}$) foram abundantes nos genomas, representando cerca de 60% de todos os segmentos identificados, no entanto, cobriram apenas uma pequena proporção do genoma. Nossos resultados demonstraram que em

média 7,01% (175,28 Mb) do genoma dessa população é autozigótico. As estimativas de F_{PED} variaram de 0,000 a 0,327 e as de F_{ROH} de 0,001 a 0,201. Correlações baixas a moderadas foram observadas entre as estimativas de F_{PED}-F_{ROH} e F_{GRM}-F_{ROH}, com valores entre -0,16 e 0,59. As correlações entre F_{ROH} de diferentes comprimentos e F_{PED} aumentaram com comprimento dos ROH. ROH compartilhados por mais de 50% das amostras foram classificados como ilhas de ROH. Quatorze ilhas de ROH foram identificadas e diversos genes contidos nessas ilhas foram associados com o teor de gordura do leite (*DGAT1*, *CYP11B*, *EEF1D*, *INSIG2* e *STAT1*), involução da glândula mamária (*IGFBP7*), lactação (*CHR e TRAPPC9*) e adaptação ao calor (*HSF1*). Nossos resultados sugerem que (i) as baixas correlações ($r<0.44$) entre F_{PED}-F_{ROH} para segmentos pequenos indicam que o F_{PED} não é adequado para capturar eventos remotos de endogamia. As correlações moderadas ($r>0.44$) entre segmentos grandes indica que os níveis de autozigosidade derivados dos ROH podem ser utilizados como uma estimativa acurada dos níveis de endogamia; e (ii) ROH podem ser utilizados para identificar regiões genômicas sob seleção, uma vez que várias regiões compartilhadas estavam associadas com características de leite e adaptação.

Palavras-chaves: Autozigosidade, *Bos indicus*, Diversidade genética, Endogamia

CHARACTERIZATION OF RUNS OF HOMOZYGOSITY IN GYR CATTLE GENOME

ABSTRACT – Runs of homozygosity (ROH) are continuous homozygous segments that are common in individuals and populations. The ability of these segments to give insight into a population's genetic events makes them a useful tool to provide information about the demographic evolution of a population over time. Additionally, genome-wide information provides valuable information to comprehend the animal's genome based on ROH, being and accurate tool to maintain diversity and fitness in livestock populations. The aim of this study was (i) to access genome-wide autozygosity to identify and characterize ROH patterns in Gyr dairy cattle genome; (ii) identify ROH islands for gene content and enrichment in segments shared by more than 50% of the samples, and (iii) compare estimates of molecular inbreeding calculated from ROH (F_{ROH}), GRM approach (F_{GRM}), and from pedigree-based coefficient (F_{PED}). Gyr dairy animals were genotyped with the BovineHD BeadChip (Illumina Inc., San Diego, CA, USA), that contains 777,962 bialleleic SNPs markers (n=582); the BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), containing 54,609 SNPs (n=1664); and with the GGP-LD *Indicus* (Geneseek® Genomic Profiler *Indicus* 30K), that contains 27,533 bialleleic SNPs markers (n=662). All genotypes were imputed to the BovineHD BeadChip (Illumina Inc., San Diego, CA, USA). SNPs unsigned to any chromosome and mapped to sexual chromosomes were removed from the dataset. After editing, 2908 animals and 735,236 SNPs were retained for the analyses. ROH were identified using PLINK v1.07 considering the following parameters: a sliding window of 50 SNPs, the minimum number of consecutive SNPs included in a ROH was 100, the minimum length of a ROH was set to 1 Mb, the maximum gap between consecutive homozygous SNPs was 500 kb, a density of 1 SNP per 50 kb, and a maximum of five SNPs with missing genotypes and up to one heterozygous genotype were allowed in a ROH. F_{PED} was estimated through the software INBUPGF90. For each animal F_{ROH} ($F_{ROH1\text{-}2\ Mb}$, $F_{ROH2\text{-}4\ Mb}$, $F_{ROH4\text{-}8\ Mb}$, $F_{ROH8\text{-}16\ Mb}$, and $F_{ROH>16\ Mb}$) was calculated based on ROH distribution of five minimum different lengths (ROH_j): 1-2, 2-4, 4-8, 8-16, and >16 Mb, respectively. F_{GRM} was calculated from the diagonal of the genomic relationship matrix (G). ROH were identified in all animals, with an average number of 55.12±10.37 segments and a mean length of 3.17 Mb. Short segments ($ROH_{1\text{-}2\ Mb}$) were abundant through the genomes, accounting for 60% of all segments identified, however, they just covered a small proportion of the genome. Our results suggest that on average 7.01% (175.28Mb) of the genome of this population is autozygous. F_{PED} estimates ranged from 0.000 to 0.327 and F_{ROH} from 0.001 to 0.201. Low to moderate correlations were observed between F_{PED}-F_{ROH} and F_{GRM}-F_{ROH}, with values ranging from -0.16 to 0.59. Correlations between F_{ROH} from different lengths and F_{PED} increased with ROH length. ROH shared by more than 50% of the samples were chosen as an indication of a possible ROH islands throughout the genome, and 14 regions were identified. Several genes inside those ROH islands were associated with milk fat content (*DGAT1*, *CYP11B*, *EEF1D*, *INSIG2*, and *STAT1*), mammary gland involution (IGFBP7), lactation (*CHR* and *TRAPPC9*), and heat adaptation (*HSF1*). Our

results suggest that (i) low correlations (r<0.44) between F_{PED}-F_{ROH} for small segments indicate that F_{PED} estimates are not the most suitable method to capture ancient inbreeding. Moderate correlations (r>0.44) between larger ROH indicate that the levels of autozygosity derived from ROH can be used as an accurate estimator of individual inbreeding levels, and (ii) ROH might be used to identify genomic regions under selection as several overlapping regions were associated with dairy and adaptive traits.

Keywords: Autozygosity, *Bos indicus*, Genetic diversity, Inbreeding

INTRODUÇÃO E JUSTIFICATIVA

É sabido que as condições ambientais nos trópicos geram grandes prejuízos econômicos na produção animal, uma vez que temperaturas elevadas, altas taxas de radiação solar e a elevada umidade relativa presente nessas regiões tendem a provocar mudanças fisiológicas nos animais, reduzindo a taxa de crescimento, a produção leiteira e o desempenho reprodutivo (AZEVÊDO; ALVES, 2009). Com isso, a pecuária leiteira nos países tropicais necessita de opções que permitam uma exploração mais eficiente dentro de suas realidades ambientais. Uma das estratégias adotadas para minimizar os efeitos das condições ambientais nos rebanhos localizados nessas regiões tem sido a introdução de animais zebuínos, importando animais mais adaptados às condições ambientais brasileiras para serem utilizados como puros ou em cruzamentos (MADALENA et al., 1990; MCGLOTHEN et al., 1995). Devido à sua melhor adaptação ao clima tropical (HANSEN, 2004), as raças zebuínas se difundiram amplamente no Brasil e nesse sentido, a raça Gir apresentou-se como uma raça com grande potencial para a exploração leiteira nos países tropicais.

Os primeiros animais da raça Gir foram importados para o Brasil em 1912 e a maioria dos touros entre 1914 e 1921, e então, empregados em cruzamentos (SANTIAGO, 1986). Em um primeiro momento esses animais foram utilizados para a produção de carne, porém, alguns criadores observaram que determinados exemplares se destacavam por sua capacidade leiteira, direcionando assim sua criação para a produção de leite. Os animais Gir também foram e têm sido intensamente utilizados em regiões tropicais e subtropicais como base para cruzamentos com raças leiteiras europeias, a fim de conferir à progênie cruzada uma maior rusticidade e tolerância aos ambientes adversos (QUEIROZ; LÔBO, 1993). Os acasalamentos entre os animais importados para a criação das primeiras linhagens de Gir Leiteiro conduziu a um aumento da endogamia na população, porém, neste primeiro momento, este aumento foi aliado a ganhos genéticos e fixação de alelos favoráveis na raça (REIS FILHO, 2006). Por conseguinte, o rápido crescimento e disseminação da

raça, aliada a uma base genética reduzida devido ao número limitado de animais importados da Índia, além do intenso uso de um pequeno número de touros provados com um alto valor genético (WANG, 2015) e a inexistência de um programa de melhoramento genético animal na época, invariavelmente levaram a um aumento gradual na taxa de endogamia na raça (REIS FILHO, 2006).

O interesse por animais da raça Gir está em crescente expansão, não só no Brasil, mas em vários países de clima tropical importadores da genética Gir brasileira, como a Colômbia e a Venezuela. Apesar da crescente expansão da raça, Reis Filho (2006) caracterizou 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. O autor enfatiza que, como consequência deste aumento significativo, há uma evidente perda da variabilidade genética, afetando significativamente o ganho genético e a expressão de características de interesse econômico relacionadas à produção e viabilidade, além de características produtivas, como a produção de leite e crescimento. O aumento da taxa de endogamia em populações de animais domésticos é um assunto importante e de muita preocupação por parte de técnicos e produtores. A endogamia tem como consequência direta a redução da heterozigosidade e, consequentemente, um aumento na homozigosidade e na frequência de genes recessivos deletérios que podem reduzir o desempenho fenotípico e a viabilidade dos indivíduos (AGERHOLM et al., 2001; ZHANG et al., 2015), resultando na herança de haplótipos IBD (“*Identical By Decedent*”), originando longos segmentos de corridas de homozigosidade (ROH) no genoma dos seus sucessores (KU et al., 2011).

As recentes e intensas práticas de seleção das populações animais têm alertado a comunidade científica para o estudo de estratégias para a conservação e manutenção das populações (HERRERO-MEDRANO et al., 2013), aumentando o interesse por parte dos pesquisadores em caracterizar e monitorar a autozigosidade, com o objetivo de preservar a diversidade genética em longo prazo dos programas de melhoramento genético animal (ZAVAREZ et al., 2015). Neste sentido, o estudo dos segmentos de ROH tornou-se uma

ferramenta aplicável para estudar a estrutura das populações, entender suas interações e para estimar parâmetros populacionais relevantes. A identificação e caracterização dos ROH podem informar sobre o histórico de uma população, fornecendo informações a respeito da ocorrência de gargalos populacionais e do seu histórico demográfico (BOSSE et al., 2012; PURFIELD et al., 2012; HERRERO-MEDRANO et al., 2013), além de prover informações a respeito da pressão de seleção exercida e do manejo reprodutivo adotado (BOSSE et al., 2012; PURFIELD et al., 2012; KARIMI, 2013; KIM et al., 2013; ZHANG et al., 2015).

O objetivo deste estudo é identificar e caracterizar segmentos homozigóticos presentes no genoma de bovinos da raça Gir leiteiro, bem como prospectar genes de importância econômica e produtiva em segmentos compartilhados por mais da metade dos indivíduos, além associá-los com níveis de endogamia obtidos a partir do pedigree e de dados genômicos oriundos da matriz genômica de parentesco. Deste modo, o estudo da distribuição dos alelos idênticos por descendência no genoma de bovinos da raça Gir constitui-se uma ferramenta de suma importância visando à estimação das relações de parentesco entre animais e os valores genéticos para as características produtivas, bem como na manutenção da variabilidade genética, auxiliando no delineamento e direcionamento dos sistemas de acasalamentos, evitando o aumento da endogamia nos rebanhos.

Referências

AGERHOLM J.S.; BENDIXEN C.; ANDERSEN O.; ARNBJERG J. Complex vertebral malformation in Holstein calves. **Journal of Veterinary Diagnostic Investigation**, v. 13, p. 283–289, 2001.

AZEVÊDO D.M.M.R.; ALVES A.A. **Bioclimatologia Aplicada à Produção de Bovinos Leiteiros nos Trópicos.** Teresina: EMBRAPA Meio-Norte, 2009, 83 p. (Documentos / Embrapa Meio-Norte, ISSN 0104-866X; 188).

BOSSE M.; MEGENS H-J.; MADSEN O.; PAUDEL Y.; FRANTZ L.A.F.; SCHOOK L.B; CROOIJMANS R.P.M.A.; GROENEN M.A.M. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. **PLoS Genetic**, v. 8, e1003100, 2012.

HANSEN, P.J. Physiological and cellular adaptations of zebu cattle to thermal stress. **Animal Reproduction Science**, p. 349-360, 2004.

HERRERO-MEDRANO J.; MEGENS H-J.; GROENEN M. et al. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. **BMC Genetics**, v. 14, 2013.

KARIMI, Z. **Runs of Homozygosity patterns in Taurine and Indicine cattle breeds**. 2013. 53 f. Dissertação (Major thesis animal breeding and genetics) - University of Natural Resources and Life Sciences, Vienna, 2013.

KIM E.-S.; COLE J.B.; HUSON H.; WIGGANS G.R.; VAN TASSELL C.P.; CROOKER B.A.; LIU G.; DA Y.; SONSTEGARD T.S. Effect of artificial selection on runs of homozygosity in U.S.Holstein cattle. **PLoS One**, v. 8, e80813, 2013.

KU C.S.; NAIDOO N.; TEO S.M.; PAWITAN Y. Regions of homozygosity and their impact on complex diseases and traits. **Human Genetics**, vol. 129, 2011.

MADALENA F.E.; LEMOS A.M.; TEODORO R.L.; BARBOSA R.T.; MONTEIRO J.B.N. Dairy production and reproduction in Holstein-Friesian and Guzera crosses. **Journal of Dairy Science**, v. 73, p. 1872–1886, 1990.

MCGLOTHEN M.E.; EL AMIN F.; WILCOX C.J.; DAVIS R.H. Effects on milk yield of crossbreeding zebu and European breeds in Sudan**. Brazilian Journal of Genetics**, v. 18, p. 221–228, 1995.

PURFIELD D.C.; BERRY D.; MCPARLAND S.; BRADLEY D.G. Runs of homozygosity and population history in cattle. **BMC Genetics**, v. 13, 2012.

QUEIROZ S.A.; LÔBO R.B. Genetic relationship, inbreeding and generation interval in registered Gir cattle in Brazil. **Journal of Animal Breeding and Genetics**, p. 228-233, 1993.

REIS FILHO J.C.; VERNEQUE R.S.; TORRES R.A; LOPES P.S.; RAIDAN F.S.S; TORAL F.L.B. Inbreeding on productive and reproductive traits of dairy Gyr cattle. **Revista Brasileira de Zootecnia**, v. 44, supl. 5, p. 174-179, 2015.

SANTIAGO A.A.: **O Zebu na Índia, no Brasil e no mundo**. Instituto campineiro de ensino agrícola: Campinas, 1986.

WANG Y. **Genetic and Geographic Diversity of Gyr (Bos Indicus) Cattle in Brazil**. 2015. 27 f. Dissertação (Major thesis animal breeding and genetics) - University of Natural Resources and Life Sciences, Vienna, 2015.

ZHANG Q.; GULDBRANDTSEN B.; BOSSE M.; LUND M.S.; SAHANA G. Runs of homozygosity and distribution of functional variants in the cattle genome. **BMC Genomics**, v.16, 2015.

CAPÍTULO 1 – Runs of homozygosity: current knowledge and applications in livestock [a]

Abstract

This review presents a broader approach to the implementation and study of runs of homozygosity (ROH) in animal populations; focusing on identifying and characterizing ROH and their practical implications. ROH are continuous homozygous segments that are common in individuals and populations. The ability of these homozygous segments to give insight into a population's genetic events makes them a useful tool that can provide information about the demographic evolution of a population over time. Furthermore, ROH provide useful information about the genetic relatedness among individuals, helping to minimize the inbreeding rate and also helping to expose deleterious variants in the genome. The frequency, size, and distribution of ROH in the genome are influenced by factors such as natural and artificial selection, recombination, linkage disequilibrium, population structure, mutation rate, and inbreeding level. Calculating the inbreeding coefficient from molecular information from ROH (F_{ROH}) is more accurate for estimating autozygosity and for detecting both past and more recent inbreeding effects than estimates from pedigree data (F_{PED}). The greater results of F_{ROH} suggest that F_{ROH} can be used to infer information about the history and inbreeding levels of a population in the absence of genealogical information. The selection of superior animals has produced large phenotypic changes and reshaped the ROH patterns in various regions of the genome. Additionally, selection increases homozygosity around the target locus, and deleterious variants are seen to occur more frequently in ROH regions. Studies involving ROH are increasingly common and provide valuable information about how the genome’s architecture can disclose a population's genetic background. By revealing the molecular changes in populations over time, genome-wide information is crucial to understanding antecedent genome architecture and, therefore, to maintaining diversity and fitness in endangered livestock breeds.

Keywords: Autozygosity, Genetic diversity, Homozygosity, Inbreeding, Livestock

[a] Article published in the “Animal Genetics” Journal: doi: 10.1111/age.12526

1.1 Introduction

In diploid genomes, runs of homozygosity (ROH) are continuous homozygous segments of the DNA sequence (Gibson et al., 2006). ROH have been applied to quantifying individual autozygosity (McQuillan et al., 2008; Ferenčaković et al., 2011; Keller et al., 2011; Ferenčaković et al., 2013a; Silió et al., 2013; Marras et al., 2014; Zavarez et al., 2015), given their high correlation with autozygosity ($r \approx 0.7$) (McQuillan et al., 2008), and consequently, the high accuracy with which autozygosity is detected (Keller et al., 2011). Autozygosity occurs when parents have a common ancestor and pass shared chromosomal segments on to their progeny, i.e., an individual inherits chromosomal fragments that are identical by descent (IBD) from both parents (Wright, 1922), resulting in homozygous segments in the offspring's genome that give rise to ROH (Broman and Webber et al., 1999). This results from population phenomena such as genetic drift, population bottleneck, inbreeding, and intensive artificial selection (Falconer & Mackay, 1996).

The identification and characterization of ROH can provide insight into how population history, structure, and demography have evolved over time. Such population phenomena can impact homozygosity patterns in the genome and these events can be revealed by ROH (Bosse et al., 2012; Purfield et al., 2012; Herrero-Medrano et al., 2013). ROH can identify inbreeding levels and the genetic relationships between individuals, providing support in estimating the true level of autozygosity at the individual and population levels (Ferenčaković et al., 2011; Ferenčaković et al., 2013a; Kim et al., 2015a; Zavarez et al., 2015). Selection pressure and mating scheme can also be revealed by ROH (Bosse et al., 2012; Purfield et al., 2012; Karimi, 2013; Kim et al., 2013; Zhang et al., 2015).

The use of high-density SNP arrays in scanning the genome for ROH has been proposed as an effective method for identifying IBD haplotypes (Gibson et al., 2006; Lencz et al., 2007). In this regard, SNP arrays can provide information about both past and more recent demographic variations of a population, i.e., population size reflecting founder effects and bottlenecks (Megens et al., 2009; Bosse et al., 2012), allowing a comparison of the degree of homozygosity among populations with varying degrees of isolation and inbreeding (Kirin et al., 2010). Curik et al. (2014)

presented a review of different approaches to estimate inbreeding levels using genomic information, highlighting the importance of ROH in quantifying and understanding inbreeding in livestock, humans, and plants.

The intense selection in livestock has alerted the scientific community to the necessity of strategies to preserve populations (Herrero-Medrano et al., 2013), characterize and monitor autozygosity, and maintain genetic diversity in long-term animal breeding programs (de Cara et al., 2013; Bosse et al., 2015). Studies have also shown a relationship between ROH in the genome and the occurrence of recessive disorders, mainly in humans (Lencz et al., 2007; Nalls et al., 2009; Vine et al., 2009; Szpiech et al., 2013) and more recently in livestock (Biscarini et al., 2014a; Kim et al., 2015a; Meszaros et al., 2015; Muchadeyi et al., 2015; Zhang et al., 2015). The analysis of ROH has therefore proved to be important in the design of mating systems to minimize the inbreeding rate (Toro & Varona, 2010; Biscarini et al., 2015a). Besides, it's important in mapping recessive alleles related to the occurrence of diseases, since ROH have an increased risk of carrying IBD deleterious recessive alleles, thereby reducing the viability of the organism (de Cara et al., 2013; Bosse et al., 2015).

Our review presents a broad overview of the study and application of ROH in animal populations and future prospective research areas. The addressed aspects are related to: i) a brief history of research on ROH, including the first studies on ROH in different livestock species; ii) practical implications of ROH identification, in which we discuss the software used to identify ROH and the consequences of some of the selected parameters when detecting them; iii) identification and characterization of ROH in livestock populations, wherein ROH patterns and their relationship with demographic history and inbreeding are identified; iv) the impact of molecular information on the measure of inbreeding coefficients; v) the effect on phenotype and disease risk, exploring autozygosity throughout the genome to determine whether ROH are correlated to deleterious variants; vi) the effects of selection on ROH; vii) genetic diversity and ROH, exploring genomic tools to minimize the loss of genetic diversity in livestock populations; and viii) a final discussion to recognize trends in livestock ROH studies and gaps for new research areas.

1.2. A brief history of the surge of ROH research

The first human study which recognized long homozygous chromosomal segments was performed by Broman and Weber (1999). They inferred that long chromosomal segments likely represent autozygosity, and might have implications for human health. The first human study using high-density SNP arrays was conducted by Gibson et al. (2006), whose results described different ROH lengths, frequency, and distribution across the genome. This work spurred a host of other research on ROH analysis in human population genetics (McQuillan et al., 2008; Kirin et al., 2010; Nothnagel et al., 2010). In livestock, the first studies on ROH were performed on cattle by Sölkner et al. (2010) and Ferenčaković et al. (2011). Several studies on ROH in cattle followed (Purfield et al., 2012; Bjelland et al., 2013; Ferenčaković et al., 2013a; Kim et al., 2013). In swine, the first studies on ROH were performed to highlight the influence of population relationships, demographic history and the effects of inbreeding on homozygosity (Bosse et al., 2012; Herrero-Medrano et al., 2013). Silió et al. (2013) and Saura et al. (2015) measured inbreeding and inbreeding depression in pigs from pedigree and genome-wide data. Khanshour (2013a) and Metzger et al. (2015) performed ROH analysis to reveal signatures of positive selection in horses. In sheep, research on population history and structure, and homozygosity using ROH was presented by Beynon et al. (2015) and Muchadeyi et al. (2015). Guangul (2014) characterized ROH patterns and genomic inbreeding coefficients in goats.

1.3 Criteria and software for detecting ROH

To date, few studies have examined and compared the performance of different software to identify ROH and which set of parameters within a given software is optimal for detecting them (Howrigan et al., 2011). Different studies have used different methodologies for predicting ROH (Zhang et al., 2013). A recurrent limitation in studies involving ROH is the lack of consensus in establishing the criteria to define ROH (Ku et al., 2011). The main objective in establishing criteria for defining the patterns of ROH lies in the fact that they are used to identify the

autozygosity, differentiating non-autozygotic segments that are identical by state (IBS) from autozygotic and IBD segments.

Howrigan et al. (2011) evaluated the performance of three major software programs, PLINK (Purcell et al., 2007), GERMLINE (Gusev et al., 2009) and BEAGLE (Browning & Browning, 2010). The authors observed that PLINK generated the highest proportion of significant results to detect autozygosity from distant common ancestors, outperforming the other two programs. GERMLINE, compared to PLINK, underperformed regarding the proportion of significant results to detect autozygosity due to the low resolution at the start and endpoints of ROH. The authors had hypothesized that BEAGLE would yield the best accuracy in detecting autozygosity because it incorporated linkage disequilibrium (LD), but it was not observed.

Karimi (2013) analyzed PLINK (Purcell et al., 2007), SVS (Golden Helix SNP & Variation Suite v.7.6.8) and cgaTOH (Zhang et al., 2013) to determine whether frequencies of detected ROH were significantly different. PLINK and SVS are common tools used to identify ROH patterns while cgaTOH is a new algorithm proposed by Zhang et al. (2013). The results showed that the ROH islands were located in similar regions by the three programs with only small differences in the frequency of ROH. However, cgaTOH captured a higher number of individuals that had ROH compared to the other two. The SVS and PLINK results overlap to a larger extent when compared to cgaTOH. The frequency of ROH were significantly different ($P<0.001$), but all three software identified ROH islands in similar regions. The development of accurate tools to assess genome-wide autozygosity is a prerequisite for successful research on ROH (Ku et al., 2011). In this sense, new algorithms and studies have also been developed to assess and compare individual genome-wide homozygosity (Seelow et al., 2009; Browning & Browning, 2010; Polašek et al., 2010; Marras et al., 2016).

The inconsistency among the criteria for defining ROH make it difficult to compare studies since the lack of consensus allows different thresholds across studies (Howrigan et al., 2011; Ku et al., 2011). This limitation may be responsible for bias in ROH-based estimates of autozygosity (Ferenčaković et al., 2013b). Table 1 presents some of the major works on ROH for various livestock species and the

respective parameters and thresholds used to identify ROH. It can be seen that the criteria for identification and characterization of ROH differ among and within species.

Howrigan et al. (2011) reported that some parameters and thresholds imposed during sequence analysis can impact the number and length of ROH. Results from Mastrangelo et al. (2016) showed different inbreeding coefficient estimates from ROH (F_{ROH}) depending on whether one, two or three heterozygous genotypes were allowed. Ferenčaković et al. (2013b) pointed out that the density of the SNP chip used to generate the data for ROH analysis, and also the frequency of SNP genotyping errors, can influence ROH identification in cattle. The 50k panel overestimated the number of small segments (1 to 4 Mb long). With the HD panel, the number of segments longer than 8 Mb was underestimated when limiting the number of heterozygous SNP genotypes within a single ROH. The same was observed by Purfield et al. (2012), whose findings pointed out that the 50k panel was appropriate for identifying ROH longer than 5 Mb. The minimum ROH length that can be detected depends therefore on the density of the SNP chip. In addition, allowing a number of genotyping errors in long ROH may minimize the underestimation of these segments (Ferenčaković et al., 2013b).

Given that the definition of ROH is still not unambiguous and that some studies employ more strict criteria compared to others (Ku et al., 2011), we believe that the lack of standards for ROH increases the likelihood of biased and false positive results, since one set of parameters can be chosen among many. Therefore, to overcome this situation, we suggest a cautious and critical interpretation when comparing studies, analyzing the density of the SNP chip used, the minimum length of ROH, the number of genotyping errors allowed, and the minimum number of SNP allowed in a single ROH, since they are likely to greatly affect ROH-based estimates of autozygosity. Furthermore, additional studies identifying which set of parameters is optimal for detecting ROH is needed in a wider range of livestock species, since it has not been systematically analyzed and the main published works are in cattle.

Additionally, several factors can influence the frequency, size and location of ROH in individuals. These factors may be related to recombination, LD, and mutation rate (Gibson et al., 2006). Other factors may be derived from chromosomal aberrations such as the occurrence of uniparental disomy (inheritance of a pair of

homologous chromosomes from only one parent), hemizygous deletion (one chromosome is deleted and is not repaired, resulting in the loss of half genome for a locus) or loss of heterozygosity in the genome (the inactivation of a functional allele at a heterozygous locus) (Engel, 1980; Koufos et al., 1985; Yokota et al., 1987; Huie et al., 2002; Dong, 2001; Marguerite et al., 2008).

Purfield et al. (2012) observed in cattle that 87.2% of the animals in the dataset had a ROH located on chromosome 14 (centered on the 25Mb position). This ROH was 127.3 Kb long and consisted of 28 SNPs. When analyzed using Haploview (Barrett et al., 2005), they observed that most of the SNPs in this region were in high LD with each other. Bosse et al. (2012) studied pigs and observed that longer ROH were located in regions of low recombination in the central part of the chromosome, and that smaller ROH had a relatively higher distribution toward telomeric regions. Herrero-Medrano et al. (2013) also reported a correlation among LD, ROH size and recombination rate in pigs. The authors found a positive correlation between average values of LD and number of ROH per chromosome ($\rho=0.70$, $p<0.002$), and a negative correlation between ROH size and rate of recombination ($\rho=-0.67$, $p<0.003$). Ai et al. (2013) compared Chinese and Western pig populations and observed that the Chinese population of Jinhua pigs had the highest fraction of long ROH and hence a high LD rate. As expected by the authors, Chinese pig populations that had lower LD exhibited fewer ROH.

1.4 ROH in livestock populations

After the first pioneering work by Sölkner et al. (2010), scientists studied ROH more deeply in various livestock species, focusing on their identification and characterization and on the relationship with demographic history and inbreeding.

Purfield et al. (2012) analyzed the patterns of ROH in *Bos taurus* cattle and observed that they differed markedly among breeds. For almost all breeds, most ROH were short (1-5 Mb). Holstein, Holstein-Friesian, and Friesian breeds showed the greatest coverage in the longer ROH. The three most homozygous animals had on average of 700.3 Mb of their genome as ROH. Kim et al. (2013) studied Holstein cattle with different selection intensities and observed that the mean ROH length per

animal was ≈ 6 Mb (Table 2). ROH shorter than 5 Mb long accounted for 53% of all segments identified and contributed less than 30% of the total cumulative ROH length. The ROH distribution pattern observed in Holstein cows can result from population bottlenecks during the breed formation together with a constant directional selection, resulting in high inbreeding. In this same study, non-selected animals showed a significantly lower average size of ROH.

Ferenčaković et al. (2013a) studied the autozygosity in Brown Swiss, Fleckvieh, Norwegian Red and Tyrol Grey cattle breeds, and found that distribution and frequencies of ROH varies among breeds. Brown Swiss animals had the highest average number of ROH (98.9) and the highest genome coverage with ROH (Table 2). The shortest average ROH length was found in Fleckvieh animals, whose genome was composed mostly of many short ROH. Brown Swiss animals had mostly few large ROH. Norwegian Red animals showed a similar pattern as Fleckvieh, and some Tyrol Grey animals had few ROH that, however, covered more than 630 Mb of the genome. According to the authors, the high number of long ROH observed in Brown Swiss animals is related to the importation of semen from a small number of bulls (Yoder & Lush, 1937). Fleckvieh animals showed a small autozygous proportion of the genome, consistent with a larger effective population size. The ROH patterns found in Norwegian Red animals is attributed to a high heterogeneity that has resulted from the historic admixture in the breed (Ferenčaković et al., 2013a).

An example of an extremely homozygous population is Chillingham cattle. The Chillingham breed has not been subjected to selection and for the last 350 years has not experienced migration events and has remained closed (Hall & Hall 1988; Hall et al., 2005). Williams et al. (2015) observed that 90.9% of the SNP were monomorphic, and the ROH genome coverage was 95%, indicating a reduced genetic variation and extreme homozygosity in this population.

Bosse et al. (2012) examined different European and Asian pig populations and found that ROH size and abundance in the genome varied considerably among individuals from different populations and subpopulations. Moreover, animals of the same population showed similar ROH patterns in their genomes. On average, 23% of the genome was considered to be in a region of homozygosity, and the most autozygous individual was a Japanese wild boar (78% of its genome). Overall, small

ROH were abundant throughout the genome, and even though large ROH were at most one tenth as abundant as small ROH, they still covered a higher proportion of the genome. In domesticated Asian pigs, the cumulative ROH length was dominated by large ROH, which may be indicative of a recently reduced population that originated from a larger population. The genome of the European wild pig had the greatest number of ROH and the highest genome autozygotic proportion. These results are consistent with the evidence of a greater intensity of population bottlenecks due to glaciation in Europe as compared to Asia (Groenen et al., 2012). Therefore, further degradation of genetic diversity is expected in European populations as compared to Asian populations.

Herrero-Medrano et al. (2013) studied populations of wild and domesticated pigs of the Iberian Peninsula and observed ROH in all animals analyzed. These authors reported differences between wild and domesticated populations in terms of quantity and variation of ROH length. Domesticated pigs had both the highest and lowest average proportion of the genome covered by ROH: 29% in the Chato Murciano breed and 10% in the Bisaro breed. Additionally, Chato Murciano pigs had the highest number of long ROH, emphasizing the recent inbreeding and low genetic diversity in the genome of this breed. The analyses also showed that wild pigs have a very large number of short ROH and no long ROH, which can be correlated with a reduced population size in the past and little inbreeding in recent times (Kirin et al., 2010). This pattern of ROH could be from population bottlenecks in the past century in Europe (Apollonio et al., 1988), resulting in a drastic reduction of the effective population size. Other possibilities are the formation of sub-populations and the migration of animals, random crossing with domesticated pigs, and the absence of geographical barriers in the Iberian Peninsula, which may have prevented high inbreeding in wild pigs (Ferreira et al., 2009).

Metzger et al. (2015) identified ROH in 10 horses of six different populations to estimate the genetic diversity and detect signatures of selection. Small ROH (40-49 Kb) were found in abundance and equally distributed in all animals, while ROH longer than 59 Kb showed a distinct distribution among different populations. Longer ROH (>400 Kb) and high inbreeding coefficients in Sorraia and Thoroughbred horses were attributed to a closed population, and in Arabian horse, it was attributed to a

limited genetic base (Aberle et al., 2004; Khanshour et al., 2013b). The low genetic diversity and high inbreeding in Thoroughbred horses may be a result of selective pressure on racing performance traits (Gu et al., 2009).

Beynon et al. (2015) inferred population history and structure of Welsh sheep breeds using the haplotype homozygosity (HHn) method. This method relies on the genome-wide distribution of ROH. The inference using HHn reflected a very large ancestral population size which has been sharply reduced since then. This steep reduction could reflect the Last Glacial maximum which occurred 20,000 to 30,000 years ago (Clark et al., 2009), and the domestication event that occurred 12,000 years ago (MacLeod et al., 2013).

An overview of the mean number of ROH per animal and the mean genome coverage in different livestock species is given in table 2. Horses showed the highest proportion of the genome covered by ROH and the highest mean number of ROH per animal, followed by pigs, which also showed a high ROH genome coverage. Based on the data shown in table 2, further research is called for in species like horse, sheep and goat, since results so far are sparse and do not allow a direct comparison of ROH patterns. In addition, it is necessary to determine whether results reflect population phenomena or are biased (over/under-estimation) due to the set of parameters used to perform ROH analysis.

1.5 ROH and molecular estimates of relatedness and inbreeding

The inbreeding coefficient has traditionally been estimated mainly from the information derived from pedigree data since Wright (1922). The recent development of high-density SNP arrays led to an increasing interest in calculating the inbreeding coefficients from molecular information in livestock (Ferenčaković et al., 2011; Purfield et al., 2012; Bjelland et al., 2013; Ferenčaković et al., 2013a; Silió et al., 2013; Marras et al., 2014; Saura et al., 2015; Zavarez et al., 2015); molecular data are more effective for estimating autozygosity and for detecting the effects of inbreeding than pedigree data (Keller et al., 2011). It is not uncommon for pedigree data to contain errors (Ron et al., 1996). Further, pedigree relatedness is estimated from statistical expectations of the probable proportion of genomic identity by

descent, while genotype-based estimates show the actual relatedness among individuals (Visscher et al. 2006).

As a result, molecular information has introduced significant advances into the analyses of inbreeding coefficients, i.e. inbreeding estimated from observed homozygous genotypes and from ROH (F_{ROH}). Observed homozygosity can be defined as the proportion of homozygous loci at either the individual or population level (Allendorf, 1986; Avise, 1994), and it is an alternative measure of inbreeding (Li et al., 2011).

F_{ROH} estimates present several advantages compared to F_{PED}. F_{ROH} more accurately predict the current autozygotic percentage of the genome and detect autozygosity due to common ancestry even fifty generations ago (Howrigan et al, 2011; Keller et al., 2011). It can be estimated in any genotyped animals, even when the pedigree information is not available; and also, it allows examining genome autozygosity distribution to find the specific locations with high levels of autozygosity, for example, to estimate the F_{ROH} separately for different chromosomes (Keller et al., 2011). Besides, F_{PED} assumes that the entire genome does not undergo selection (Curik et al., 2002) and recombination events, and therefore does not take into account potential bias from these events (Carothers et al., 2006). This variance increases with every meiosis (McQuillan et al., 2008) and relies on recombination events that occur during the formation of gametes (Bjelland et al., 2013).

The moderate to high correlations between F_{PED} and F_{ROH} pointed out that F_{ROH} estimates can be applied as an indicator of inbreeding levels in a number of cattle and swine breeds (Table 3). It is worth mentioning that F_{PED}-F_{ROH} correlation increases with ROH length. According to Marras et al. (2014), this may be explained by considering that ROH reflect both past and recent relatedness, and that F_{PED} estimates are based on pedigree records which may not extend back many generations. When longer ROH reflecting recent relatedness are considered to calculate F_{ROH}, the F_{PED}-F_{ROH} correlation tends to be higher. Saura et al. (2015) reported that the average $F_{ROH>5Mb}$ was very close to F_{PED}, whereas the average $F_{ROH<5Mb}$ was about four times lower than the average F_{PED}.

Scraggs et al. (2014) suggested that the F_{PED} underestimated the true relatedness of Wagyu cattle since differences in inbreeding coefficient levels

between F_{PED} and F_{ROH} were observed. They reported lower estimated F_{PED} compared to $F_{ROH.}$ These results are consistent with the data obtained for cattle (Marras et al., 2014; Kim et al., 2015a) and pigs (Saura et al., 2015), in which F_{ROH} gave a higher estimate, suggesting that F_{PED} may possibly be underestimating inbreeding. In horses, Metzger et al. (2015) estimated F_{ROH}, but did not correlate it to $F_{PED.}$ F_{ROH} estimates for 50 SNP-window ranged from 0.18 to 0.43. Guangul (2014) estimated F_{ROH} in five goat populations at different run lengths (>1 Mb to >16 Mb) and found values ranging from 0.0048 (F_{ROH} >16 Mb) to 0.0500 (F_{ROH} >1 Mb). A more recent study found F_{ROH} values ranging from 0.02 to 0.09 in goats and 0.02 to 0.10 in sheep (Kim et al., 2015b). An overview of F_{ROH} levels and mean F_{ROH} for different species, such as human, cattle and pig is available in Curik et al. (2014).

According to previously described studies we can conclude that F_{ROH} is a better tool for quantifying inbreeding. Our conclusion is based on limitations of F_{PED}. Pedigree-based inbreeding estimates fail to capture the actual proportion of the genome that is IBD, hence, animals in the founder population may be classified as unrelated. Also, the probabilistic approach of F_{PED} does not account for stochastic variations during meiosis. We believe that these shortcomings may be responsible for underestimated/erroneous F_{PED} estimates. The advantage of F_{ROH} goes further in identifying IBD segments with a greater accuracy. F_{ROH} estimates can disclose the age of the inbreeding based on the length of the ROH, which is an important tool in population genetics.

1.6 Deleterious variants and inbreeding depression through detected ROH

In mid-1876, Charles Darwin reported that inbreeding could lead to reduced productivity in plants (Darwin, 1876), and Garrod (1996) realized that some traits in humans, as albinism and alkaptonuria, occurred more frequently in progeny derived from consanguineous marriages. Garrod studied the pattern of recessive inheritance proposed by Mendel (1866) and observed that the high incidence of recessive diseases in inbred individuals resulted from the high probability that they were homozygous for a deleterious recessive allele inherited IBD. The deleterious recessive variants can be identified in inbred individuals by the presence of long

homozygous regions (Lander & Botstein, 1987) or by studying the ROH (Broman & Weber, 1999). Szpiech et al. (2013) and Zhang et al. (2015) observed a strong linear relationship between the frequency of deleterious homozygous variants and the genomic ROH fraction, with values of 0.98 and 0.93, respectively. Szpiech et al. (2013) reported that individuals with a high ROH coverage had a higher fraction of deleterious variants occurring in long ROH, which are in agreement with the hypothesis that recent inbreeding enables rare deleterious variants to exist in homozygous form. Additionally, Zhang et al. (2015) observed that deleterious homozygotes occur more frequently in ROH regions than non-deleterious homozygotes. Therefore, the role played by autozygosity contributed to an ongoing interest in determining whether these segments are correlated to risk factors for simple and complex diseases (Howrigan et al., 2011) and to inbreeding depression (Charlesworth & Willis, 2009).

Biscarini et al. (2014a) studied the distribution of functional bovine variants and used ROH to detect genomic regions associated with susceptibility to infectious, metabolic and reproductive diseases, and the risk of mastitis and locomotion disorders in dairy cattle. They were able to identify the prevalence of ROH in regions that contain important genes that trigger the diseases. One example is the ROH located at 12Mb on BTA12, which contains the *VWA8* gene, whose mutations might be associated with musculoskeletal disorders and coagulation abnormalities. ROH associated with reproductive problems were found on BTA 5, BTA 15 and BTA 18. Susceptibility to infectious diseases was found on BTA 7 and BTA 12. Biscarini et al. (2013) also used ROH to study causal mutations for arthrogryposis and macroglossia in Piedmontese cattle.

Muchadeyi et al. (2015) examined Swakara sheep, looking for extended ROH shared across animals associated with sub-vitality performance. Consensus overlapping ROH (cROH) in sub-vital sheep were observed on chromosomes 3, 4 and 25. These cROH regions carried genes impacting on the nervous system and skeletal and brain development, such as *LRRTM3*, *DPP6*, and *SHH*. Despite the findings, they were unable to support the presence of a recent recessive-lethal mutation causing the sub-vital phenotype.

Scanning the genome for ROH might be an alternative or complementary

strategy to genome-wide association studies (GWAS) for complex disease traits (Biscarini et al., 2014b). Huson et al. (2014) utilized genome-wide association, haplotype analysis, signatures of selection, and ROH analysis to identify a consensus region for the SLICK locus on BTA20 in cattle. Mészáros et al. (2015) utilized ROH and GWAS analysis to identify genomic regions of entropion in Austrian Fleckvieh cattle. Biscarini et al. (2015b) studied different methods to identify QTLs in farm animals instead of traditional GWAS, namely resampled predictive models (Biffani et al., 2015) and a ROH-based approach. Kim et al. (2015a) used genome-wide association testing to explore potential negative correlations between the occurrence of ROH and daughter pregnancy rate and/or somatic cell score in US Jersey cattle.

The genetic basis of inbreeding depression is caused by an increased rate of homozygosity (Charlesworth & Willis, 2009) and frequency of homozygous deleterious alleles (Ouborg et al., 2010; Ku et al., 2011), which leads to reduced individual performance and decreased population variability (Gonzalez-Recio et al., 2007). The performance reduction in plant and animal populations due to inbreeding depression has been described by Charlesworth & Willis (2009). The reduced performance and diversity in populations lead to a reduced selection response of breeding programs (Weigel, 2006). Thus, the study of inbreeding depression and its negative consequences are considered indispensable for conservation genetics studies and for breeding program management (Keller & Waller 2002).

Pryce et al. (2014) estimated inbreeding depression for milk production and fitness traits using pedigree-based inbreeding coefficients. Increasing F_{PED} and F_{GRM} (diagonal of the genomic relationship matrix minus 1) in 1% for Holstein's cows decreased milk yield of 21 and 28 liters/lactation, respectively. For Jersey cows, this reduction was 12 and 27 liters/lactation, respectively. The effect of F_{PED} on fertility was only significant ($P<0.005$) for Holstein cows, where a 1% increase resulted in a calving interval extension of +0.18 days. Also, milk, fat and protein yields decreased for Holstein and Jersey cows, which was attributed to a 1% increase in homozygous SNPs. Kim et al. (2015a) recognized that more than 60 regions displayed increased ROH levels, which can be correlated to F_{PED} increase in US Jersey cattle in the last five decades. A negative association between the occurrence of ROH and daughter

pregnancy rate (DPR) was observed on BTA 3, BTA 7, BTA 8, and BTA 12. The same pattern was observed for somatic cells score (SCS). An increase of ROH levels based on F_{PED} influenced SCS on BTA 1, BTA 3, BTA 4, BTA 5, BTA 13, and BTA 21, suggesting that high autozygosity due to inbreeding may potentially influence fertility and be related to susceptibility to mastitis.

A study to test inbreeding depression on post-weaning growth performance in pigs was performed by Silió et al. (2013). Their estimates of inbreeding depression were expressed by the decreasing performance relative to the mean per a 0.10 increase in inbreeding coefficient, and were -4.40% for daily gain, and -1.52% for weight at 90 days. Saura et al. (2015) observed genomic regions responsible for inbreeding depression for two reproductive traits in a highly inbred line of Iberian pigs (Guadyerbas pigs). A reduction in number of piglets born alive and in the total number of piglets born was also observed per a 10% increase in inbreeding.

These findings suggest that inbreeding increases autozygosity throughout the genome, and might trigger the expression of homozygous recessive alleles that may cause expression of an unfavorable phenotype (Agerholm et al., 2001). Understanding when deleterious mutations arose and why they persist in populations is of interest in breeding and conservation programs (Schubert et al., 2014; Marsden et al., 2016). Marsden et al. (2016) demonstrated that rather than just avoiding inbreeding, a large population size is critical for preventing the accumulation of deleterious variants. Thus, we believe that further efforts need to be made in conservation genetics programs to avoid environmental population segregation. If it is not accomplished in time, the accumulation of deleterious variants and the occurrence of genetic erosion will lead to more severe consequences than the ones triggered by the inbreeding itself.

1.7 Effect of artificial selection on ROH

The identification of recent selection signatures in the genome provides relevant information regarding the response to strong directional selection in domesticated animals. The selection of superior animals reduced phenotypic variability and reshaped the genome, including ROH patterns, when compared to not

intensively selected animal groups (Kim et al., 2013). The search for superior animals via selection has reduced the diversity of haplotypes and increased homozygosity around the target locus, generating high frequency of ROH in regions that house the selection targets (Leocard, 2009; Karimi, 2013; Zhang et al., 2015).

Zhang et al. (2015) observed that ROH patterns were not randomly distributed across the genome. A number of ROH peaks were distributed and shared among individuals, which is likely the result of selection events and not only attributable to demographic history. Genomic regions that are selection targets tend to generate ROH islands, defined as genomic regions with reduced genetic diversity and, consequently, high homozygosity around the selected locus compared to the rest of the genome (Pemberton et al., 2012). According to Sonesson et al. (2010), there is a risk that the genomic selection can also result in long homozygous segments around QTL regions in populations that were selected for any given trait.

Purfield et al. (2012) observed that the genomic regions located on BTA 7, BTA 14, BTA 16, and BTA 18, which had a high incidence of ROH, also contained important cattle genes associated with traits related to immunity, carcass, and dystocia in calving. In particular, BTA 5 and BTA 9 had an increased number of long ROH (> 20Mb) and tended to contain QTL associated with milk fat production and growth traits in cattle. Kim et al. (2013) found that the longest ROH was 87.13 Mb on BTA 8 in a Holstein cow selected from an elite herd. The average number of ROH per individual in the control group, i.e. animals without selection, was significantly lower ($p<0.0001$) compared to the animals that had undergone genetic selection. Animals belonging to groups with higher selection intensity had a high level of autozygosity in their genomes, especially in 13 regions of 11 chromosomes. Differences in patterns of ROH among populations support the possibility that animal selection affects the patterns of autozygosity in their genome. According to Pemberton et al. (2012), recent strong directional selection is expected to have a greater influence on long ROH compared to medium and small ROH because it tends to generate long haplotype segments.

Karimi (2013) studied some European (Angus, Fleckvieh, and Brown Swiss) and zebu (Nellore, Gir, and Brahman) cattle breeds and found a total of six common regions in the genome of the two groups of animals located at regions associated

with QTL for production on BTA 6 (38,268,200:39,451,000), BTA 7 (51,502,500:52,353,000), BTA 10 (24,575,700:25,619,800), BTA 12 (28,434,000:29,628,100), BTA 16 (43,802,200:44,968,700) and BTA 21 (1,360,390:1,853,150). These results suggest that natural and artificial selection may strongly shapes genomic ROH patterns in livestock (Pemberton et al., 2012).

Khanshour (2013a) screened for candidate segments for positive selection performing ROH analysis in distinct Arabian horse populations. Metzger et al. (2015) studied the functional distribution of ROH in selected and non-selected horses, and noticed that genes affecting cellular, metabolic, and developmental processes, as well as the immune system and reproduction, were observed in ROH regions. The Hanoverian breed which has been intensively selected for optimal performance shared 18 ROH regions which harbored six relevant genes for morphology and performance in sport horses.

Intense selection and decreasing the effective population size through the selection of superior animals can endanger the viability and variability of populations in the long term (Saccheri et al., 1996). Selection tends to increase inbreeding rates, resulting in fixed alleles that cause major phenotypic changes, and lead to loss of allelic variation (Muir et al., 2008). The studies reviewed above demonstrate the importance of managing the breeding programs to maintain heterozygosity in the chromosomal regions that house important genes for animal husbandry, thus keeping the genetic diversity around the selection target locus and avoiding the expression of deleterious variants throughout the genome.

1.8 Genetic diversity and ROH

The last decade has witnessed a sharp increase in the intensity of selection in breeding programs, especially in dairy cattle, pigs and poultry. The increased use of elite animals has contributed to increased inbreeding rates, reducing the effective population size. The low effective population size increases the effect of inbreeding and genetic drift and reduces the genetic variability, which may compromise the viability of populations and change the patterns of ROH in the long-term (Frankham & Ralls, 1998). These events are more accentuated in many local breeds with a small

population size, predisposing them to extinction (FAO, 2013).

Genomic tools have been widely used to study and characterize the genetic diversity and population structure of livestock (Rothschild & Plastow, 2014). Therefore, without pedigree data, in many breeds genetic markers can be used to estimate the effective population size, for example, by exploring the extent of LD (de Roos et al., 2008; Corbin et al., 2010; Herrero-Medrano et al., 2013; Uimari & Tapio, 2011; Beynon et al., 2015). The estimated effective population size based on the recombination rate and LD provides useful predictions and consistent comparisons among populations (de Roos et al., 2008; Herrero-Medrano et al., 2013). The individual evaluation of ROH patterns also has practical implications for conservation programs. Animals with high levels of ROH may be excluded or used less frequently in mating populations threatened with extinction or with a small effective population (Herrero-Medrano et al., 2013; Biscarini et al., 2015a).

The mating or cross between outbred individuals or populations is an option for increasing genetic diversity, contributing to the disruption of long ROH in the genome. The effects of hybridization or outbreeding were observed in three African hybrid cattle breeds (Kuri, Sheko, and Borgou) that had the shortest ROH among the studied African breeds (Purfield et al., 2012). In pigs, recent mating between domesticated Manchado de Jabugo animals and other breeds may have resulted in the breakdown of homozygous haplotypes longer than 100 Mb. The Manchado de Jabugo population showed no signs of high inbreeding, probably due to its mixed heritage, despite being highly endangered due to its small population size (Herrero-Medrano et al., 2013). Ai et al. (2013) observed that Chinese Kele pigs had the shortest ROH, leading to a hypothesis that this population has a historical admixture with Western pigs. This hypothesis was based on structural analysis, in which Kele pigs showed close phylogenic relationships and signals of admixture with Western pigs. The White Duroc breed showed the lowest fraction of autozygous segments in Western pigs. It might be attributed to the recent admixture of Duroc and Large White in the population, reducing the length of homozygous segment in the White Duroc genome. The same pattern was observed by Al-Mamun et al. (2015), in which the pure sheep breeds had more ROH across the whole genome than the crosses.

Maintaining the genetic diversity within-breed and across-breed is crucial in

conservation genetics (Ollivier & Foulley, 2005) to sustain livestock production. In this regard, studies using genomic measures of coancestry to minimize the loss of genetic diversity and inbreeding in conservation programs have been proposed (de Cara et al., 2013; Bosse et al., 2015; Gómez-Romano et al., 2016). A measure of coancestry based on IBD segments has been suggested as a strategy to maintain genetic diversity and fitness in conservation programs when the given population has a medium to high inbreeding load (de Cara et al., 2013). As discussed in this review, selection, low effective population size and inbreeding can negatively affect genetic diversity, and breeding programs must monitor genetic variation to prevent an irreversible erosion of genetic diversity in livestock populations, maximizing their capability to adapt to changes (Biscarini et al., 2015a).

1.9 Final Considerations

The objective of this study was to give a comprehensive review of current knowledge and application of ROH in livestock, and also to identify research gaps for future research areas. Studies involving homozygous segments have shown that ROH are common in the genome of livestock species, and have addressed population history and their structure, selection pressure, inbreeding, and the occurrence of deleterious variants throughout the genome. However, most of the studies involved dairy cattle and pig, and fewer were on beef cattle and other livestock species. Studies on a wider range of species are needed to better understand the genetic architecture and the effects of ROH in livestock.

Probably the major difficulty faced by scientists is the lack of consistent criteria among studies regarding the threshold values in each parameter analyzed to define a ROH. A number of software programs have been developed to infer ROH by applying different algorithms and methodologies, but there have been relatively few studies assessing which set of parameters is optimal for detecting ROH so as to better understand their effects on detecting autozygosity (Howrigan et al., 2011; Ferenčaković et al., 2013b; Karimi, 2013; Mastrangelo et al., 2016). Even though some incipient standards can be seen among the studies, such as the number of heterozygous calls allowed being no more than two, further efforts need to be made

to reduce biased values when defining a ROH. In addition, as stated in this review, the frequency and distribution of ROH in livestock is influenced by many factors, of which we highlight the populations' demographic history. However, to date, there have been no studies on the influence of the particularities of a specific population on the ability of these algorithms to infer ROH and to establish consistent criteria.

It has been well demonstrated that ROH is a more effective and accurate alternative for quantifying animal relatedness and inbreeding levels. Most of the ROH analyses have been heavily used for conservation purposes, to manage inbreeding, and in small and isolated populations to make them reservoirs of genetic variation. However, further research is necessary to quantify the impact of animal relatedness measures based on ROH on EBV prediction and genetic evaluation reliability. In this regard, Luan et al. (2014) showed that predicting GEBV based on ROH showed higher or similar accuracy of GEBV prediction for simulated data in Brown Swiss cattle as compared to linkage analysis relationships. It is important to stress that different from dairy cattle populations, in beef cattle and sheep populations the absence of pedigree information or the presence of incomplete pedigree with missing parents is more common. According to Berry et al. (2016), the development of accurate genomic evaluations in beef populations is more difficult than in dairy populations due to the existence of multiple breeds, poor extent of phenotyping, lack of artificial insemination, and beef systems' being generally a lower-margin business and a poorer adopter of technology. Thus, for beef and sheep populations, it is expected that gains in EBV reliability from pedigree information based on ROH are likely to be somewhat higher than those in dairy cattle.

In recent years, the emergence of various next-generation sequencing (NGS) platforms has enabled the sequencing of the full genome at lower cost and in shorter time. NGS platforms bring higher map resolution not only for ROH detection (Pippucci et al., 2014) but also for other structural variations, like the genome copy number variation (Pirooznia et al., 2015). Despite the advances in high-throughput technologies and whole-genome sequencing analysis, tools to integrate the genomic structural variations with whole transcriptome analysis are necessary to better elucidate the genetic mechanisms that determine the genetic and phenotypic differences among animals, causal variation and loss of alleles. These technologies

can assist in better understanding the autozygosity at work in livestock.

Acknowledgements

E. Peripolli received scholarship from Coordination for the Improvement of Higher Education Personnel (CAPES) in cooperation with the Brazilian Agricultural Research Corporation (EMBRAPA).

F. Baldi, D.P Munari, and M.V.G.B. Silva held productivity research fellowship from The Brazilian National Council for Scientific and Technological Development (CNPQ)

Conflict of interest

The authors declare that they have no conflict of interest.

1.10 References

Aberle K.S., Hamann H., Drögemuller C. & Distl O. (2004) Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. *Animal Genetics* 35, 270–7.

Agerholm J.S., Bendixen C., Andersen O. & Arnbjerg J. (2001) Complex vertebral malformation in Holstein calves. *Journal of Veterinary Diagnostic Investigation* 13, 283–9.

Ai H., Huang L. & Ren J. (2013) Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PLoS One* 8, e56001.

Allendorf F.W. (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* 5, 181–90.

Al-Mamun H.A., Clark S.A., Kwan P. & Gondro C. (2015) Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genetics Selection Evolution* 47, 90.

Apollonio M., Randi E. & Toso S. (1988) The systematics of the wild boar (*Sus scrofa L.*) in Italy. *Bollettino di Zoologia* 3, 213–21.

Avise J.C. (1994) Molecular Markers, Natural History and Evolution. Chapman and Hall, New York, NY.

Barrett J.C., Fry B., Maller J. & Daly M.J. (2005) HAPLOVIEW: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–5.

Berry D.P., Garcia J.F. & Garrick D.J. (2016) Development and implementation of genomic predictions in beef cattle. *Animal Frontiers* 6, 32–8.

Beynon S.E., Slavov G.T., Farré M. et al. (2015) Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. *BMC Genetics* 16, 65.

Biffani S., Dimauro C., Macciotta N., Rossoni A., Stella A. & Biscarini F. (2015) Predicting haplotype carriers from SNP genotypes in *Bos taurus* through linear discriminant analysis. *Genetics Selection Evolution* 47, 4.

Biscarini F., Del Corvo M., Stella A., Albera A., Ferenčaković M. & Pollott G. (2013) Busqueda de las mutaciones causales para artrogriposis y macroglosia en vacuno de raza Piemontesa: resultados preliminares. (Proceedings) Actas de las XV Jornadas sobre Produccion Animal-AIDA.

Biscarini F., Biffani S., Nicolazzi E.L., Morandi N. & Stella A. (2014a) Applying runs of homozygosity to the detection of associations between genotype and phenotype in farm animals. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. Vancouver, BC, Canada.

Biscarini F., Biffani S., Morandi N., Nicolazzi E.L. & Stella A. (2014b) Using runs of homozygosity to detect genomic regions associated with susceptibility to infectious and metabolic diseases in dairy cows under intensive farming conditions. *arXiv:1601. 07062.*

Biscarini F., Nicolazzi E.L., Stella A., Boettcher P.J. & Gandini G. (2015a) Challenges and opportunities in genetic improvement of local livestock breeds. *Frontiers in Genetics* 6, 33.

Biscarini F., Biffani S. & Stella A. (2015b) Más allá del GWAS: alternativas para localizar QTLs. q-bio.GN. arXiv:1504.03802v1.

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. *Journal of Dairy Science* 96, 4697–706.

Bosse M., Megens H.-J., Madsen O., Paudel Y., Frantz L.A., Schook L.B., Crooijmans R.P. & Groenen M.A. (2012) Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genetics* 8, e1003100.

Bosse M., Megens H.-J., Madsen O., Crooijmans R.P.M.A., Ryder O.A., Austerlitz F., Groenen M.A.M. & de Cara M.A.R. (2015) Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome Research* 25, 1–12.

Broman K. & Weber J.L. (1999) Long homozygous chromosomal segments in reference families from the Centre d'Étude du Polymorphisme Humain. American Journal of Human Genetics 65, 1493–500.

Browning S.R. & Browning B.L. (2010) High-resolution detection of identity by descent in unrelated individuals. *American Journal of Human Genetics* 86, 526–39.

Carothers A.D., Rudan I., Kolcic I., Polasek O., Hayward C., Wright A.F., Campbell H., Teague P., Hastie N.D. & Weber J.L. (2006) Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Annals of Human Genetics* 70, 666–76.

Charlesworth D. & Willis J.H. (2009) The genetics of inbreeding depression. *Nature Reviews* 10, 783–96.

Clark P.U., Dyke A.S., Shakun J.D., Carlson A.E., Clark J., Wohlfarth B., Mitrovica J.X., Hostetler S.W. & McCabe A.M. (2009) The last glacial maximum. *Science* 325, 710–4.

Corbin L.J., Blott S.C., Swinburne J.E., Vaudin M., Bishop S.C. & Woolliams J.A. (2010) Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics* 41, 8–15.

Curik I., Sölkner J. & Stipic N. (2002) Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. *Journal of Animal Breeding and Genetics* 119, 101–15.

Curik I., Ferenčaković M. & Sölkner J. (2014) Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livestock Science* 166, 26–34.

Darwin C.R. (1876) The Effects of Cross and Self Fertilization in the Vegetable Kingdom. John Murray, London.

De Cara M.A.R., Villanueva B., Toro M.A. & Fernández J. (2013) Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology* 22, 6091–9.

De Roos A.P.W., Hayes B.J., Spelman R.J. & Goddard M.E. (2008) Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–12.

Dong J.T. (2001) Chromosomal deletions and tumor suppressor genes in prostate cancer. *Cancer and Metastasis Reviews* 20, 173–193.

Engel E. (1980) A new genetic concept: uniparental disomy and its potential effect, isodisomy. *American Journal of Medical Genetics* 6, 137–43.

Falconer D.S. & Mackay T.F.C. (1996) Introduction to Quantitative Genetics, 4th edn. Longman, Essex, UK.

FAO (2013) In Vivo Conservation of Animal Genetic Resources (FAO Animal Production and Health Guidelines No. 14). FAO, Rome.

Ferenčaković M., Hamzic E., Gredler B., Curik I. & Sölkner J. (2011) Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. *Agriculturae Conspectus Scientificus* 76, 325–8.

Ferenčaković M., Hamzic E., Gredler B., Solberg T.R., Klemetsdal G., Curik I. & Sölkner J. (2013a) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics* 130, 286–93.

Ferenčaković M., Sölkner J. & Curik I. (2013b) Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution* 45, 42.

Ferreira E., Souto L., Soares A.M.V.M. & Fonseca C. (2009) Genetic structure of the wild boar population in Portugal: evidence of a recent bottleneck. *Mammalian Biology* 74, 274–85.

Frankham R. & Ralls K. (1998) Conservation biology: inbreeding leads to extinction. *Nature* 392, 441–2.

Garrod A.E., Oxon M.D. & Lond F.R.C.P. (1996) The incidence of alkaptonuria: a study of chemical individuality. *Molecular Medicine* 2, 1616–20.

Gibson J., Newton E.M. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* 15, 789–95.

Gomez-Romano F., Villanueva B., Fern_andez J., Woolliams J.A. & Pong-Wong R. (2016) The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Genetics Selection Evolution* 48, 2.

Gonzalez-Recio O., de Maturana E.L. & Gutierrez J.P. (2007) Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *Journal of Dairy Science* 90, 5744–52.

Groenen M.A.M., Archibald A.L., Uenishi H. et al. (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–8.

Gu J., Orr N., Park S.D., Katz L.M., Sulimova G., MacHugh D.E. & Hill E.W. (2009) A genome scan for positive selection in thoroughbred horses. *PLoS One* 4, e5767.

Guangul S.A. (2014) Design of community based breeding programs for two indigenous goat breeds of Ethiopia. Doctoral thesis, University of Natural Resources and Life Sciences, Vienna.

Gusev A., Lowe J.K., Stoffel M., Daly M.J., Altshuler D., Breslow J.L., Friedman J.M. & Peer I. (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome Research* 19, 318–26.

Hall S.J.G. & Hall J.G. (1988) Inbreeding and population dynamics of the Chillingham cattle (Bos taurus). *Journal of Zoology* 216, 479–93.

Hall S.J.G., Fletcher T.J., Gidlow J.R., Ingham B., Shepherd A., Smith A. & Widdows A. (2005) Management of the Chillingham wild cattle. *Government Veterinary Journal* 15, 4–11.

Herrero-Medrano J.M., Megens H.-J., Groenen M.A.M., Ramis G., Bosse M., P_erez-Enciso M. & Crooijmans R.P.M.A. (2013) Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC Genetics* 14, 106.

Howrigan D.P., Simonson M.A. & Keller M.C. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 12, 460.

Huie M.L., Anyane-Yeboa K., Guzman E. & Hirschhorn R. (2002) Homozygosity for multiple contiguous single-nucleotide polymorphisms as an indicator of large heterozygous deletions: identification of a novel heterozygous 8-kb intragenic deletion (IVS7–19 to IVS15–17) in a patient with glycogen storage disease type II. *American Journal of Human Genetics* 70, 1054–7.

Huson H.J., Kim E.-S., Godfrey R.W. et al. (2014) Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle. *Frontiers in Genetics* 5, 1–12.

Karimi S. (2013) Runs of homozygosity patterns in taurine and indicine cattle breeds. Doctoral thesis, University of Natural Resources and Life Sciences, Vienna.

Keller L.F. & Waller M. (2002) Inbreeding effects in wild populations. *Trends in Ecology and Evolution* 17, 230–41.

Keller M., Visscher P. & Goddard M. (2011) Quantification of inbreeding due to distance ancestors and its detection using dense SNP data. *Genetics* 189, 237–49.

Khanshour A.M. (2013a) Genetic diversity and population structure of the Arabian horse populations from Syria and other countries. Doctoral dissertation, Texas A&M University, College Station.

Khanshour A., Conant E., Juras R. & Cothran E.G. (2013b) Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. *Journal of Heredity* 104, 386–98.

Kim E.-S., Cole J.B., Huson H., Wiggans G.R., Van Tassell 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* 8, e80813.

Kim E.-S., Sonstegard T.S., Tassell C.P.V., Wiggans G., Rothschild M.F. et al. (2015a) The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. *PLoS One* 10, e0129967.

Kim E-S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B., Sayre B., Mwacharo J.M. & Rothschild M.F. (2015b) Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* 116, 255–64.

Kirin M., McQuillan R., Franklin C., Campbell H., McKeigue P.M. & Wilson J.F. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS One* 5, e13996.

Koufos A., Hansen M.F., Copeland N.G., Jenkins N.A., Lampkin B.C. & Cavenee W.K. (1985) Loss of heterozygosity in three embryonal tumours suggests a common pathogenetic mechanism. *Nature* 16, 330–4.

Ku C.S., Naidoo N., Teo S.M. & Pawitan Y. (2011) Regions of homozygosity and their impact on complex diseases and traits. *Human Genetics* 129, 1–15.

Lander E.S. & Botstein D. (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 19, 1567–70.

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. *Proceedings of the National Academy of Sciences of the United States of America* 104, 19942–7.

Leocard S. (2009) Selective sweep and the size of the hitchhiking set. *Advances in Applied Probability* 41, 731–64.

Li M.-H., Strandén I., Tiirikka T., Sevón-Aimonen M.L. & Kantanen J. (2011) A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLoS One* 6, e26256.

Luan T., Yu X., Doleza M., Bagnato A., Meuwissen T. (2014) Genomic prediction based on runs of homozygosity. *Genetics Selection Evolution*, 46:64.

MacLeod I.M., Larkin D.M., Lewin H.A., Hayes B.J. & Goddard M.E. (2013) Inferring demography from runs of homozygosity in whole-genome sequence, with correction for sequence errors. *Molecular Biology and Evolution* 30, 2209–23.

Marguerite P., Andersen P., Zara W., Hetrick E.D. & Gottschling D.E. (2008) A genetic screen for increased loss of heterozygosity in Saccharomyces cerevisiae. *Genetics* 179, 1179–95.

Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., Valentini A., Williams J.L. & Macciotta N.P. (2014) Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal Genetics* 46, 110–21.

Marras G., Rossoni A., Schwarzenbacher H., Biffani S., Biscarini F. & Nicolazzi E.L. (2016) ZANARDI: an open-source pipeline for multiple-species genomic analysis of SNP array data. *Animal Genetics* 10.1111/age.12485.

Marsden C.D., Vecchyo D.O-D., O'Brien D.P., Taylor J.F., Ramirez O., Vilá C., Marques-Bonet T., Schnabel R.D., Wayne R.K. & Lohmueller K.E. (2016) Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences of the United States of America* 113, 152–7.

Mastrangelo S., Tolone M., Gerlando R.D., Fontanesi L., Sardina M.T. & Portolano B. (2016) Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal Consortium* 10, 746–54.

McQuillan R., Leutenegger A., Abdel-Rahman R. et al. (2008) Runs of homozygosity in European populations. *American Journal of Human Genetics* 83, 359–72.

Megens H-J., Crooijmans R.P.M.A., Bastiaansen J.W.M. et al. (2009) Comparison of linkage disequilibrium and haplotype diversity on macro- and micro chromosomes in chicken. *BMC Genetics* 10, 86.

Mendel G. (1866) Versuche uber Plflanzenhybriden. Verhandlungen des naturforschenden Vereines in Brunn, Bd. IV fur das Jahr 1865, Abhandlungen, 3–47.

Mészáros G., Stückler M.P., Ferenčaković M. & Sölkner J. (2015) Genomic background of entropion in Fleckvieh cattle. *Poljoprivreda/Agriculture* 21, 48–51.

Metzger J., Karwath M., Tonda R., Beltran S., Águeda L., Gut M., Gut I.G. & Distl O. (2015) Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* 16, 764.

Muchadeyi F.C., Malesa M.T., Soma P. & Dzomba E.F. (2015) Runs of homozygosity in Swakara pelt producing sheep: implications on sub-vital performance. *Proceedings for Association for the Advancement of Animal Breeding and Genetics* 21, 310–3.

Muir W.M., Wong G.K.-S., Zhang Y. et al. (2008) Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds*. Proceedings of the National Academy of Science of the United States of America* 105, 17312–7.

Nalls M., Guerreiro R., Simon-Sanchez J., Bras J.T., Traynor B.J., Gibbs J.R., Launer L., Hardy J. & Singleton A.B. (2009) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics* 10, 183–90.

Neves H.H.R., Desidério J.A., Pimentel E.C.G., Scalez D.C.B. & Queiroz S.A. (2015) Preliminary study to determine extent of linkage disequilibrium and estimates of autozygosity in Brazilian Gyr dairy cattle. *Archivos de Zootecnia* 64, 99–108.

Nothnagel M., Lu T.T., Kayser M. & Krawczak M. (2010) Genomic and geographic distribution of SNP defined runs of homozygosity in Europeans. *Human Molecular Genetics* 19, 2927–35.

Ollivier L. & Foulley J. (2005) Aggregate diversity: new approach combining within and between breed genetic diversity. *Livestock Production Science* 95, 247–54.

Ouborg N.J., Pertoldi C., Loeschcke V., Bijlsma R. & Hedrick P.W. (2010) Conservation genetics in transition to conservation genomics. *Trends in Genetics* 26, 177–87.

Pemberton T., Absher D., Feldman M., Myers R.M., Rosenberg N.A. & Li J.Z. (2012) Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics* 91, 275–92.

Pippucci T., Magi A., Gialluisi A. & Romeo G. (2014) Detection of runs of homozygosity from whole exome sequencing data: state of the art and perspectives for clinical, population and epidemiological studies. *Human Heredity* 77, 63–72.

Pirooznia M., Goes F.S. & Zandi P.P. (2015) Whole-genome CNV analysis: advances in computational approaches. *Frontiers in Genetics* 6, 138.

Polasek O., Hayward C., Bellenguez C. et al. (2010) Comparative assessment of methods for estimating individual genome-wide homozygosity-by-descent from human genomic data. *BMC Genomics* 11, 139.

Pryce J.E., Haile-Mariam M., Goddard M.E. & Hayes B.J. (2014) Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution* 46, 71.

Purcell S., Neale B., Todd-Brown K. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81, 559–75.

Purfield D.C., Berry D., McParland S. & Bradley D.G. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics* 13, 70.

R Development Core Team (2015) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Ron M., Blanc Y., Band M., Ezra E & Weller J.I. (1996) Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. *Journal of Dairy Science* 79, 676–81.

Rothschild M.F. & Plastow G.S. (2014) Applications of genomics to improve livestock in the developing world. *Livestock Science* 166, 76–83.

Saccheri I.J., Brakefield P.M. & Nichols R.A. (1996) Severe inbreeding depression and rapid fitness rebound in the butterfly Bicyclus anynana (Satyridae). *Evolution* 50, 2000–13.

SAS Institute (2012) SAS User Guide. SAS Institute Inc., Cary NC.

Saura M., Fernández A., Varona L., Fernández A.I., de Cara M.A., Barragán C. & Villanueva B. (2015) Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. *Genetics Selection Evolution* 47, 1.

Schubert M., Jonsson H., Chang D. et al. (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Sciences of the United States of America* 111, E5661–9.

Scraggs E., Zanella R., Wojtowicz A., Taylor J.F., Gaskins C.T., Reeves J.J., de Avila J.M. & Neibergs H.L. (2014) Estimation of inbreeding and effective population size of fullblood Wagyu cattle registered with the American Wagyu Cattle Association. *Journal of Breeding and Genetics* 131, 3–10.

Seelow D., Schuelke M., Hildebrandt F. & Nurnberg P. (2009) HOMOZYGOSITY MAPPER - an interactive approach to homozygosity mapping. *Nucleic Acids Research* 37, W593–9.

Silió L., Rodríguez M.C., Fernández A., Barragán C., Benítez R., Óvilo C. & Fernández A.I. (2013) Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP derived metrics. *Journal of Animal Breeding and Genetics* 130, 349–60.

Solkner J., Ferencakovic M., Gredler B. & Curik I. (2010) Genomic metrics of individual autozygosity, applied to a cattle population. Proceedings of the 61st Annual Meeting of the European Association of Animal Production. Heraklion, Greece.

Sonesson A.K., Woolliams J.A. & Meuwissen T.H.E. (2010) Maximizing genetic gain whilst controlling rates of genomic inbreeding using genomic optimum contribution selection. Proceedings of the World Congress of Genetics Applied to Livestock Production, Leipzig, Germany [CD-ROM Communication].

Szpiech Z.A., Xu J., Pemberton T.J., Peng W., Z€ollner S., Rosenberg N.A. & Li J.Z. (2013) Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics* 93, 90– 102.

Toro M.A. & Varona L. (2010) A note on mate allocation for dominance handling in genomic selection. *Genetics Selection Evolution* 42, 33.

Traspov A., Deng W., Kostyunina O., Ji J., Shatokhin K., Lugovoy S., Zinovieva N., Yang B. & Huang L. (2016) Population structure and genome characterization of local pig breeds in Russia, Belorussia, Kazakhstan and Ukraine. *Genetics Selection Evolution* 48, 16.

Uimari P. & Tapio M. (2011) Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *Journal of Animal Science* 89, 609–14.

Vine A.E., McQuillin A., Bass N.J. et al. (2009) No evidence for excess runs of homozygosity in bipolar disorder. *Psychiatric Genetics* 19, 165–70.

Visscher P.M., Medland S.E., Ferreira M.A.R., Morley K.I., Zhu G., Comes B.K., Montgomery G.W. & Martin N.G. (2006) Assumption-free estimation of heritability from genome-wide identity-by descent sharing between full siblings. *PLoS Genetics* 2, e41.

Weigel K. (2006) Controlling inbreeding in modern dairy breeding programs. *Advanced Dairy Science and Technology* 18, 263–74.

Williams J.L., Hall S.J.G, Del Corvo M., Ballingall K.T., Colli L., Marsan P.A. & Biscarini F. (2015) Inbreeding and purging at the genomic level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. *Animal Genetics* 47, 19–27.

Wright S. (1922) Coefficients of inbreeding and relationship. *American Naturalist* 56, 330–8.

Yoder D.M. & Lush J.L. (1937) A genetic history of the Brown Swiss cattle in the United States. *Journal of Heredity* 28, 154–60.

Yokota J., Wada M., Shimosato Y., Terada M. & Sugimura T. (1987) Loss of heterozygosity on chromosomes 3, 13, and 17 in small-cell carcinoma and on chromosome 3 in adenocarcinoma of the lung. *Proceedings of the National Academy of Sciences of the United States of America* 84, 9252–6.

Zavarez L.B., Utsunomiya Y.T., Carmo A.S. et al. (2015) Assessment of autozygosity in Nellore cows (*Bos indicus*) through high density SNP genotypes. *Frontiers in Genetics* 6, 5.

Zhang L., Orloff M.S., Reber S, Li S., Zhao Y. & Eng C. (2013) CGATOH: extended approach for identifying tracts of homozygosity. *PLoS One* 8, e57772.

Zhang Y., Young J.M., Wang C., Sun X., Wolc A. & Dekkers J.C.M. (2014) Inbreeding by pedigree and genomic markers in selection lines of pigs. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. Vancouver, BC, Canada.

Zhang Q., Guldbrandtsen B., Bosse M., Sun X., Wolc A. & Dekkers J.C.M. (2015) Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* 16, 542.

Table 01: Comparison of pre-set parameters for identification and characterization of ROH in different animal species.

Author	Specie	Software	SNP array	Consecutive SNPs/ROH[1]	Density[2] (SNP/kb)	MaximumGap[3] (kb)	Minimum length[4] (kb)	Heterozygous SNP/Sliding window	Missing SNP/Sliding window
Ferenčaković et al. (2011)	Cattle: Dual purpose	Fortran 90 software	Illumina Bovine SNP 50k BeadChip	15	-	-	1000	00	-
Purfield et al. (2012)	Cattle: Beef, dairy and dual purpose[5]	PLINK v1.07	Illumina BovineHD Genotyping BeadChip assay	58	1/50	100	500	01	02
Purfield et al. (2012)	Cattle: Beef, dairy and dual purpose[5]	PLINK v1.07	Illumina Bovine SNP 50k BeadChip	No restriction	1/120	1000	500	01	02
Bjelland et al. (2013)	Cattle: Dairy	PLINK v1.07	Illumina Bovine SNP 50k BeadChip	30	-	-	-	00	01
Ferenčaković et al. (2013a)	Cattle: Dairy and dual purpose[5]	SNP & Variation Suite v7.6.8	Illumina Bovine SNP 50k BeadChip	15	1/1000	1000	1000	00	05
Karimi (2013)	Cattle: Beef, dairy and dual purpose[5]	SNP & Variation Suite v7.6.8 PLINK v1.07 cgaTOH	Illumina BovineHD Genotyping BeadChip assay	30	1/50	250	1000	01	04
Biscarini et al. (2014a)	Cattle: Dairy	PLINK v1.07	Illumina Bovine SNP 50K Beadchip	No restriction	-	1000	-	01	05
Marras et al. (2014)	Cattle: Beef, dairy and dual purpose[5]	SAS 9.2 script (SAS Institute 2012)	Illumina bovine SNP 50K BeadChip	15	-	1000	1000	00	00
Scraggs et al. (2014)	Cattle: Beef	PLINK v1.07	Illumina bovine SNP 50K BeadChip	50	-	-	1000	01	01
Mészáros et al. (2015)	Cattle: Dual purpose	-	Illumina Bovine SNP 50k BeadChip	30	-	-	1000	00	00
Williams et al., (2015)	Cattle: kept for conservation	R Development Core Team, 2008	Illumina BovineHD Genotyping BeadChip assay	-	1/50	100	100	01	02
Zavarez et al. (2015)	Cattle: Beef	SNP & Variation Suite v7.6.8	Illumina BovineHD Genotyping BeadChip assay	30	1/100	500	4000	02 (ROH ≥ 4Mb)	05
Zavarez et al. (2015)	Cattle: Beef	SNP & Variation Suite v7.6.8	Illumina BovineHD Genotyping BeadChip assay	30	1/100	500	500	00 (ROH < 4 Mb)	05

Bosse et al. (2012)	Swine	PLINK v1.07	Porcine SNP60 Beadchip	20	1/1000	-	10	01	-
Ai et al. (2013)	Swine	PLINK v1.07	Porcine SNP60 Beadchip	-	-	-	500	01	05
Herrero-Medrano et al. (2013)	Swine	PLINK v1.07	Porcine SNP60 Beadchip	20	1/1000	1000	10	01	-
Silió et al. (2013)	Swine	SNP & Variation Suite v7.6.8	Porcine SNP60 Beadchip	30	1/100	1000	1000	-	02
Saura et al. (2015)	Swine	Fortran	Porcine SNP60 Beadchip	30	1/100	1000	-	01	02
Zhang et al. (2014)	Swine	PLINK v1.07	Porcine SNP60 Beadchip	10	1/500	1000	5000	-	01
Traspov et al. (2016)	Swine	PLINK v1.09	Porcine SNP60 Beadchip	-	-	-	500	01	05
Khanshour (2013a)	Horse	PLINK v1.07	Equine SNP50 BeadChip	50	1/50	1000	500	-	-
Al-Mamun et al. (2015)	Sheep	PLINK v1.07	Illumina Ovine SNP50 BeadChip	-	-	250	500	01	02
Muchadeyi et al. (2015)	Sheep	PLINK v1.07	Illumina Ovine SNP50 Beadchip	20	1/50	500	-	00	02
Guangul (2014)	Goat	CgaTOH	47K SNPs bead chip	20	-	1000	1000	01	05

ROH, runs of homozygosity; "-", Information not available.
[1] Minimum number of consecutive SNPs needed to define a segment as a ROH.
[2] Minimum allowed density of SNPs inside a run.
[3] Maximum gap between consecutive homozygous SNPs.
[4] Minimum length to define a ROH.
[5] Papers analyzing different breeds, and each breed with a different purpose.

Table 2. Comparison between the mean of the total number of ROH, the mean genome length covered by ROH (Mb), and the mean genome proportion covered by ROH (%) in different livestock species.

Author	Specie	SNP array	Breed/Population	*n*	Average number of ROH	Mean genome length covered by ROH (Mb)	Mean genome proportion covered by ROH (%)	Mean ROH length (Mb)
Purfield et al., 2012	Cattle	Illumina BovineHD Genotyping BeadChip	Angus	39	-	198.60[1]	7.911[2]	5.09[1]
			Hereford	40	-	198.70[1]	7.911[2]	4.96[1]
			Belgian Blue	38	-			2.12 to 2.46[3]
			Charolais	117	-			0.68 to 0.79[3]
			Friesian	98	-			0.89 to 0.95[3]
			Holstein	262	-	80.58 to 93.48[1]	3.20 to 3.721[2]	0.30 to 0.35[3]
			Holstein-Friesian crosses	111	-			0.72 to 0.84[3]
			Limousin	128	-			0.63 to 0.73[3]
			Simmental	58	-			1.39 to 1.61[3]
Ferenčaković et al., 2013a	Cattle	Illumina Bovine SNP 50k BeadChip	Brown Swiss	304	98.9	396.80	15.60[4]	1.30[1]
			Fleckvieh	502	94.5	223.10	8.77[4]	0.44[1]
			Norwegian Red	498	80.0	253.50	9.96[4]	0.51[1]
			Tyrol Grey	117	72.3	221.00	8.68[4]	1.88[1]
Kim et al., 2013	Cattle	Illumina Bovine SNP 50k BeadChip	Unselected line	299	31.1[5]/13.5[6]	-	-	6.61[5]/10.0[6]
			Contemporary Holsteins	1634	43.5[5]/20.1[6]	-	-	
			Selected line for milk production	151	40.4[5]/19.5[6]	-	-	
Marras et al., 2014	Cattle	Illumina bovine SNP 50K BeadChip	Italian Holstein	2093	81.7	297.00	11.61[7]	3.6
			Italian Brown	749	94.6	371.00	14.51[7]	3.9
			Piedmontese	364	54.0	106.00	4.14[7]	1.9
			Marchigiana	410	71.4	210.00	8.21[7]	-
			Italian Simmental	479	94.3	210.00	8.21[7]	2.2
Zavarez et al., 2015	Cattle	Illumina BovineHD Genotyping BeadChip	Nellore	1278	-	120.22[8]	4.79	1.26
Bosse et al., 2012	Swine	Porcine SNP60 Beadchip	European pig breeds European wild boar Asian wild boars Chinese pigs	52	778.8	645.95[9]	23.00	1.11
Herrero-Medrano et al., 2013	Swine	Porcine SNP60 Beadchip	Chato Murciano	25	34	814.47[9]	29.00	23.95[3]
			Bisaro	15	13	280.85[9]	10.00	21.60[3]
			Wild boars	18	30	<561.70[9]	<20.00	<18.72[3]
			Iberian	31	26	<561.70[9]	<20.00	<21.60[3]
			Manchado de Jabugo	08	24	<561.70[9]	<20.00	<23.40
Traspov et al., 2016	Swine	Porcine SNP60 Beadchip	Russia, Belorussia, Kazakhstan and Ukraine breeds	170	-	72.30	2.57[9]	0.42[3]
			Chinese breeds	135	-	106.00	3.77[9]	0.78[3]

			International commercial breeds	153	-	123.28	4.38[9]	0.80[3]
Metzger et al., 2015	Horse	Illumina SNP50 BeadChip.	Sorraia	02	4175[5]	798.63	35.60[10]	0.19[3]
			Dülmen Horse	01	2804[5]	416.55	18.57[10]	0.14[3]
			Arabian	01	3581[5]	565.57	25.21[10]	0.15[3]
			Saxon-Thuringian Heavy Warmblood	01	3138[5]	476.06	21.22[10]	0.15[3]
			Thoroughbred	01	4595[5]	953.19	42.49[10]	0.20[3]
			Hanoverian	04	311[5]	454.02	20.24[10]	0.14[3]
Al-Mamun et al., 2015	Sheep	Illumina OvineSNP50 BeadChip	Border Leicester (BL)	253	49.65	-	-	-
			Merino (MER)	265	7.57	-	-	-
			Poll Dorset (PD)	264	37.89	-	-	-
			MER x BL	260	0.30	-	-	-
			MER x BL x PD	231	1.43	-	-	-
Muchadeyi et al., 2015	Sheep	Illumina OvineSNP50 Beadchip	White-vital	41	214	-	-	1.66
			White subvital	16	84	-	-	1.61
			Black	15	72	-	-	1.56
			Grey	22	109	-	-	1.75
Guangul, 2014	Goat	47K SNPs bead chip	Abergelle	-	33.24	69.11	2.87[11]	2.07[3]
			Western Lowland	-	29.35	63.86	2.65[11]	2.17[3]
			Red Sokoto	-	28.49	115.02	4.78[11]	4.03[3]
			Western African Dwarf	-	42.48	120.17	5.00[11]	2.82[3]
			Sahel	-	21.86	61.68	2.56[11]	2.82[3]

ROH, runs of homozygosity; '–', information not available.
[1] Value defined using the parameters set for the HD panel genotypes;
[2] An estimation considering the overall length of the genome covered by SNPs to be 2510.61 Mb.
[3] Estimation of the mean ROH length (Mb) calculated by dividing the mean genome length covered by ROH (Mb) by the mean of the total number of ROH.
[4] An estimation considering the overall length of the genome covered by SNPs to be 2543.17 Mb.
[5] Values defined using a 50 SNP sliding window.
[6] Values defined using a 100 SNP sliding window.
[7] An estimation considering the overall length of the genome covered by SNPs to be 2556.00 Mb.
[8] An estimation considering the overall length of the genome covered by SNPs to be 2510.00 Mb.
[9] An estimation considering the overall length of the genome covered by SNPs to be 2808.52 Mb, based on refseq database Scrofa 10.2, 2014.
[10] An estimation considering the overall length of the genome covered by SNPs to be 2242.87 Mb.
[11] An estimation considering the overall length of the genome covered by SNPs to be 2402.62 Mb.

Table 3. Correlations between the inbreeding from pedigree data (F_{PED}) and from molecular information through runs of homozygosity (F_{ROH}) for different ROH length and livestock species

Author	Specie	SNP array	Breed/Population	*n*	r (F_{PED}, F_{ROH})	(F_{PED}, $F_{ROH>1Mb}$)	(F_{PED}, $F_{ROH>2Mb}$)	(F_{PED}, $F_{ROH>4Mb}$)	(F_{PED}, $F_{ROH>5Mb}$)	(F_{PED}, $F_{ROH>8Mb}$)	(F_{PED},$F_{ROH>16Mb}$)
Ferenčaković et al., 2011	Cattle	Illumina Bovine SNP 50k BeadChip	Austrian Simmental	500	-	0.64[1]	0.67[1]	0.68[1]	-	0.68[1]	0.63[1]
Purfield et al., 2012	Cattle	Illumina BovineHD Genotyping BeadChip assay	Multiple breeds[2]	891[2]	0.71[3]	-	-	-	-	-	-
Ferenčaković et al., 2013a	Cattle	Illumina Bovine SNP 50k BeadChip	Brown Swiss	304	-	0.66[1]	0.67[1]	-	-	0.60[1]	0.50[1]
			Fleckvieh	502	-	0.66[1]	0.69[1]	-	-	0.70[1]	0.64[1]
			Nowegian Red	498	-	0.61[1]	0.61[1]	-	-	0.62[1]	0.53[1]
			Tyrol Grey	117	-	0.71[1]	0.72[1]	-	-	0.71[1]	0.70[1]
Kim et al., 2013	Cattle	Illumina Bovine SNP 50k BeadChip	Unselected line	299	0.62[4]/0.60[5]	-	-	-	-	-	-
			Contemporary Holsteins	1634	0.68[4]/0.64[5]	-	-	-	-	-	-
			Selected line for milk production	151	0.59[4]/0.58[5]	-	-	-	-	-	-
Marras et al., 2014	Cattle	Illumina Bovine SNP 50k BeadChip	Italian Holstein	2093	-	0.70		0.69	-	0.65	0.56
			Italian Brown	749	-	0.66		0.66	-	0.65	0.58
			Italian Simmental	479	-	0.66		0.74	-	0.76	0.71
Pryce et al., 2014	Cattle	Illumina Bovine SNP 50k BeadChip	Holstein	8853	0.53	-	-	-	-	-	-
			Jersey	4138							
Neves et al., 2015	Cattle	Illumina Bovine SNP 50k BeadChip	Gyr	25	-	-	0.39	0.40	-	0.40	-
Kim et al., 2015	Cattle	Illumina Bovine SNP 50k BeadChip	Jersey	1062	0.70[6]/0.71[7]						
Zhang et al., 2014	Swine	Porcine SNP60 Beadchip	Yorkshire	2358	0.69						
Silió et al., 2013	Swine	Porcine SNP60 Beadchip	Iberian	64	-			0.77[8]		0.81[8]	
Saura et al., 2015	Swine	Porcine SNP60 Beadchip	Guadyerbas	109	0.63			-0.24		0.60	

[1] Pedigree inbreeding coefficients referred to all generation long.
[2] Angus (n=39), Hereford (n=40), Belgian Blue (n=38), Charolais (n=117), Friesian (n=98), Holstein (n=262), Holstein-Friesian crosses (n=111), Limousin (n=128), and Simmental (n=58).
[3] Value defined using the parameters set for the HD panel genotypes in which the sum of ROH per animal was set as ROH >500 kb.
[4] Values defined using a 50 SNP sliding window.
[5] Values defined using a 100 SNP sliding window.

[6] Definition of ROH based on the number of continuous homozygous SNP (100 SNPs).
[7] Definition of ROH based on the number of continuous homozygous SNP (30, 50, and 80 SNPs).
[8] Values for genealogical inbreeding coefficients tracing the pedigree back to the founder animals (F26G) and tracing the pedigree back five generations to common ancestors (F5G).

CAPÍTULO 2 – Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle

Resumo - As corridas de homozigose (ROH) são definidas como segmentos homozigóticos contínuos de uma sequência de DNA. Os ROH têm sido utilizados em estudos para quantificar a autozigosidade individual e como um estimador acurado dos níveis de endogamia em diversas espécies animais. Animais da raça Gir Leiteiro foram genotipados com o BovineHD BeadChip (Illumina Inc., San Diego, CA, USA) que contêm 777.962 SNPs (n=582), BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) contendo 54.609 SNPs (n=1664) e GGP-LD *Indicus* (Geneseek® Genomic Profiler *Indicus* 30K) que contêm 27.533 SNPs (n=662). Todos os genótipos foram imputados para o painel BovineHD BeadChip (Illumina Inc., San Diego, CA, USA) e utilizados para as análises dos padrões genômicos de ROH e para a estimação dos coeficientes de endogamia. Segmentos em homozigose foram identificados em todos os animais, com uma média de 55,12±10,37 segmentos por animal e um comprimento médio de 3,17 Mb. Segmentos pequenos ($ROH_{1\text{-}2Mb}$) foram encontrados abundantes nos genomas dos animais, totalizando aproximadamente 60% de todos os ROH identificados, entretanto, a proporção total do genoma coberto pelos mesmos foi relativamente pequena. Nossos resultados sugerem que em média 7,01% (175,28 Mb) do genoma dessa população é autozigótico. Os coeficientes de endogamia foram estimados por meio da análise da proporção genômica em homozigose (F_{ROH}), pela matrix genômica de parentesco (F_{GRM}) e por estimativas baseadas no pedigree tradicional (F_{PED}). Estimativas do F_{PED} variaram de 0,000 a 0,327 e F_{ROH} de 0,001 a 0,201. Correlações baixas a moderadas foram observadas entre F_{PED}-F_{ROH} e F_{GRM}-F_{ROH}, com valores variando de -0,16 a 0,59. Correlações entre F_{ROH} de diferentes tamanhos e F_{PED} aumentaram com o comprimento do ROH. As baixas correlações ($r<0.44$) entre F_{PED}-F_{ROH} para segmentos pequenos indicam que o F_{PED} não é adequado para capturar eventos remotos de endogamia. Correlações moderadas ($r>0.44$) entre segmentos grandes indicam que os níveis de autozigosidade derivados dos ROH podem ser utilizados como uma estimativa acurada dos níveis de endogamia. Quatorze regiões compartilhando segmentos por mais de 50% dos indivíduos foram identificadas e classificadas como ilhas de ROH. Nessas regiões, genes associados com o conteúdo de gordura no leite (*DGAT1*, *CYP11B*, *EEF1D*, *INSIG2* e *STAT1*), involução da glândula mamária (*IGFBP7*), lactação (*CHR* e *TRAPPC9*) e adaptação. Genes contidos nas ilhas de ROH sugerem uma forte seleção para características leiteiras e adaptação ambiental ao calor (*HSF1*) foram identificados.

Palavras-chave: *Bos indicus*, Características leiteiras, Coeficientes de endogamia, Corridas de homozigose, Ilhas de ROH

Abstract - Runs of homozygosity (ROH) are continuous homozygous segments of the DNA sequence. They have been applied to quantify individual

autozygosity and been used as a potential inbreeding measure in livestock species. Gyr dairy animals were genotyped with the BovineHD BeadChip (Illumina Inc., San Diego, CA, USA), that contains 777,962 bialleleic SNP markers (n=582); the BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), containing 54,609 SNP (n=1664); and with the GGP-LD *Indicus* (Geneseek® Genomic Profiler *Indicus* 30K), that contains 27,533 bialleleic SNP markers (n=662). All genotypes were imputed to the BovineHD BeadChip (Illumina Inc., San Diego, CA, USA) and used for the analysis of genomic ROH patterns and inbreeding coefficients measurements. ROH were identified in all animals, with an average number of 55.12±10.37 segments and a mean length of 3.17 Mb. Short segments ($ROH_{1\text{-}2\ Mb}$) were abundant through the genomes, accounting for 60% of all segments identified, however, the proportion of the genome that was covered by them was relatively small. Our results suggest that on average 7.01% (175.28Mb) of the genome of this population is autozygous. Inbreeding coefficients were estimated by the analysis of the genome portion in runs of homozygosity (F_{ROH}), using the genomic relationship matrix (F_{GRM}), and by pedigree-based estimate coefficient (F_{PED}). F_{PED} estimates ranged from 0.00 to 0.327 and F_{ROH} from 0.001 to 0.201. Low to moderate correlations were observed between F_{PED}-F_{ROH} and F_{GRM}-F_{ROH}, with values ranging from -0.16 to 0.59. Correlations between F_{ROH} from different lengths and F_{PED} increased with ROH length. Low F_{PED}-F_{ROH} correlations for small segments ($r<0.44$) indicate that F_{PED} estimates are not the most suitable method to capture ancient inbreeding. Moderate correlations between larger ROH ($r>0.44$) indicate that the levels of autozygosity derived from ROH can be used as an accurate estimator of recent inbreeding levels. Overlapping ROH were evident across the genomes and 14 regions were identified with ROH frequencies exceeding 50% of the whole population. Several genes associated with milk fat content (*DGAT1*, *CYP11B*, *EEF1D*, *INSIG2*, and *STAT1*), mammary gland involution (*IGFBP7*), lactation (*CHR* and *TRAPPC9*), and heat adaptation (*HSF1*) were identified Genes inside ROH islands suggest a strong selection for dairy traits and enrichment for Gyr cattle environmental adaptation.

Keywords: *Bos indicus*, Dairy traits, Inbreeding coefficients, ROH islands, Runs of homozygosity

2.1 Introduction

Autozygosity occurs when chromosomal segments arising from a common ancestor are identical by descent (IBD) and inherited from both parents on to the offspring genome (BROMAN; WEBER, 1999). This pattern of inheritance gives rise to continuous IBD homozygous segments, characterized as runs of homozygosity (ROH) (GIBSON; MORTON; COLLINS, 2006), which can be a consequence of several population phenomena (FALCONER; MACKAY, 1996). The development of high-density SNP arrays to scan the genome for ROH has been proposed as a useful

method to distinguish non-autozygotic segments that are identical by state (IBS) from those autozygotic and IBD (HOWRIGAN; SIMONSON; KELLER, 2011).

As the occurrence of ROH tend to be revealed in the genome, its identification and characterization can provide an insight into how population structure and demography have evolved over time (BOSSE et al., 2012; PURFIELD et al., 2012; HERRERO-MEDRANO et al., 2013). ROH can disclose the genetic relationships among individuals, estimating with a high accuracy the autozygosity at the individual and population levels (FERENČAKOVIĆ et al., 2011; 2013; KIM et al., 2015; ZAVAREZ et al., 2015), and can elucidate about selection pressure events (KARIMI, 2013; KIM et al., 2013; ZHANG et al., 2015a). As the expected length of the autozygous segment follows an exponential distribution with mean equal to $1/2g$ morgans, where *g* is equal to the number of generations since the common ancestor, the number of generations from the selection events can be inferred from the length and frequency of ROH (HOWRIGAN; SIMONSON; KELLER, 2011).

The autozygosity based on ROH can help to better understand the genetic selection process of quantitative traits, since selection tend to print homozygous stretches on the genome (MARRAS et al., 2014). According to Zhang et al. (2015a), ROH patterns are not randomly distributed across the genome, and ROH islands are seen to be distributed and shared among individuals, which is likely the result of selection events. Therefore, ROH can be used to explore signatures of selection (KARIMI, 2013; MARRAS et al., 2014), since genomic regions sharing ROH potentially contain alleles associated with genetic improvement in livestock (PURFIELD et al., 2012) and are of interest for breeding programs (MARRAS et al., 2014). Runs of homozygosity can also be an accurate estimator of inbreeding, since high levels of inbreeding increases the frequency of homozygous alleles.

Studies have considered pedigree-based estimates of inbreeding (F_{PED}) since Wright (1922), however, the availability of whole-genome marker panels has widespread the use of genomic information in animal breeding (MEUWISSEN; HAYES; GODDARD, 2001). Pedigree-based relatedness is estimated from statistical expectations of the probable proportion of genomic identity by descent, while genotype-based estimates show the current relatedness among individuals (VISSCHER et al., 2006). Molecular approaches based on inbreeding coefficient

estimates derived from ROH (F_{ROH}) and based on genomic relationship matrix (F_{GRM}) (VANRADEN, 2008) are meaningfulness to avoid drawbacks of using pedigrees to analyze inbreeding. F_{ROH} are worth to estimate genome-wide autozygosity since it captures the influence of relatedness among founders from the base population, it takes into account the stochastic nature of recombination and mutations loads (KELLER et al., 2011), and also account for potential bias resulting from selection (CURIK; SÖLKNER; STIPIC, 2002).

The first Gyr (*Bos primigenius indicus*) animals entered in Brazil in 1912, and most of the bulls were imported between 1914 and 1921, and then used in crosses (SANTIAGO, 1986). The imported animals were first used for beef cattle purpose, and some breeders started to use them for milk production. Gyr animals have been intensely used as the basis for crosses with taurine dairy breeds due to its rusticity and greater tolerance to the tropical environment (QUEIROZ; LÔBO, 1993). The mating between imported animals invariably led to an increased inbreeding rate in the population, nevertheless it resulted in genetic gains and fixation of favorable alleles. Over time, the deleterious effects associated with increased homozygosity arising from inbreeding are predisposed to reduce the genetic gains, resulting in a clear loss of genetic variability (reviewed by PERIPOLLI et al., 2016). Hence, the intense use of founders' animals to create the first Gyr dairy lines without an existence of a breeding program at the time (REIS FILHO, 2006), and the limited number of animals imported from India concomitant with the small number of proven sires used to disseminate the breed (WANG, 2015) were presumably the trigger of autozygosity. Therefore, maintaining genetic variability in the Gyr cattle in Brazil is a demanding issue, since Brazil is recognized as a Gyr genetic supplier to some tropical countries that have deficiencies in milk production. Genome-wide autozygosity is an upcoming research area and there is a growing interest in characterizing and comprehending the mechanisms involved in it, to allow a long-term viability and sustainability of Gyr breeding programs.

The aim of this work was to access genome-wide autozygosity in Gyr cattle to identify and characterize the occurrence of ROH, as well as to investigate ROH islands for gene content in segments shared by more than 50% of the animals.

Finally, to compare estimates of molecular inbreeding calculated from ROH (F_{ROH}), GRM approach (F_{GRM}), and pedigree-based coefficient (F_{PED}).

2.2 Material and Methods

2.2.1 Animals and genotyping

The animals used in this study comprise the progeny test program from the National Program for Improvement of Dairy Gir (PNMGL), headed by Embrapa Dairy Cattle (Juiz de Fora, Minas Gerais, Brazil) in cooperation with the Brazilian Association of Dairy Gyr Breeders (ABCGIL) and the Brazilian Association of Zebu Breeders (ABCZ). The objective of the program is to promote the genetic improvement of the Gyr dairy cattle, through the identification and selection of genetically superior bulls for fat, protein and total solids in milk, as well as traits associated with animal conformation and management.

A total of 19 dams and 563 sires born between 1964 and 2013 were genotyped with the BovineHD BeadChip (Illumina Inc., San Diego, CA, USA), containing 777,962 bialleleic SNP markers; 1,664 dams with the BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), that contains 54,609 SNP; and 662 dams with the GGP-LD *Indicus* (GeneSeek® Genomic Profiler *Indicus* 30K), that contains 27,533 bialleleic SNP markers.

Imputation was implemented using the FIMPUTE 2.2 software (SARGOLZAEI; CHESNAIS; SCHENKEL, 2014), and lower density panels were imputed to the HD level. High accuracies of SNP imputation for the Gyr dairy cattle breed were reported by Boison et al. (2015), with values ranging from 0.95 to 0.98 when considering low-density panels with more than 15k markers to impute high-density panel containing 777,962 bialleleic SNP markers (Illumina Inc., San Diego, CA, USA). SNPs unsigned to any chromosome and mapped to sexual chromosomes were removed from the dataset. SNP Markers and samples were also edited for call rate (<90%). The animals genotyped with the BovineHD BeadChip (Illumina Inc., San Diego, CA, USA) were used as reference population for imputation. The missing genotypes were imputed in the reference population and all the markers were retained. After editing

the reference and imputed genotypes, a total of 2,908 animals and 735,236 SNPs were retained for the analyses.

2.2.2 Runs of homozygosity

ROH were identified in every individual using PLINK v1.07 (PURCELL et al., 2007). The PLINK software uses a sliding window of a specified length or number of homozygous SNPs to scan along each individual's genotype at each SNP marker position to detect homozygous segments (HOWRIGAN; SIMONSON; KELLER, 2011). The parameters and thresholds applied to define a ROH were (i) a sliding window of 50 SNPs across the genome; (ii) the proportion of homozygous overlapping windows was 0.05; (iii) the minimum number of consecutive SNPs included in a ROH was 100; (iv) the minimum length of a ROH was set to 1 Mb; (v) the maximum gap between consecutive homozygous SNPs was 500 kb; (vi) a density of one SNP per 50 kb; and (vii) a maximum of five SNPs with missing genotypes and up to one heterozygous genotype were allowed in a ROH. The ROH were defined by a minimum of 1 Mb in length to avoid short and common ROH that occur throughout the genome due to linkage disequilibrium (LD) (PURFIELD et al., 2012). ROH were classified into five length classes: 1-2, 2-4, 4-8, 8-16, and >16 Mb, identified as $ROH_{1\text{-}2\ Mb}$, $ROH_{2\text{-}4\ Mb}$, $ROH_{4\text{-}8\ Mb}$, $ROH_{8\text{-}16\ Mb}$, and $ROH_{>16\ Mb,}$ respectively.

2.2.3 Pedigree and genomic inbreeding coefficients

Three types of inbreeding coefficients (F_{PED}, F_{ROH} and F_{GRM}) were estimated. Pedigree-based inbreeding coefficients (F_{PED}) were estimated for all animals using pedigree records from a dataset containing 101,351 animals born between 1946 and 2015. The pedigree data was provided by Embrapa Dairy Cattle (Juiz de Fora, Minas Gerais, Brazil). The average pedigree depth was approximately three generations ranging from 0 to 7.85. The F_{PED} was estimated through the software INBUPGF90 (AGUILAR; MISZTAL, 2008). Genomic inbreeding coefficients based on ROH (F_{ROH}) were estimated for each animal according to McQuillan et al. (2008):

$$F_{ROH} = \frac{\sum_{j=1}^{n} L_{ROHj}}{L_{total}}$$

where L_{ROHj} is the length of ROH_j, and L_{total} is the total size of the autosomes covered by markers. L_{total} was taken to be 2,510,605,962 bp, based on the consensus map. For each animal F_{ROH} ($F_{ROH1\text{-}2\ Mb}$, $F_{ROH2\text{-}4\ Mb}$, $F_{ROH4\text{-}8\ Mb}$, $F_{ROH8\text{-}16\ Mb}$, and $F_{ROH>16\ Mb}$) was calculated based on ROH distribution of five minimum different lengths (ROH_j): 1-2, 2-4, 4-8, 8-16, and >16 Mb, respectively. A second measure of genomic inbreeding was calculated from a Genomic relationship matrix (G) and was denoted as F_{GRM}. The G matrix was calculated according to the method described by VanRaden et al. (2011) using the following formula:

$$G = \frac{ZZ'}{2\sum_{i=1}^{n} P_i\,(1 - P_i)}$$

where Z is a genotype matrix that contains the 0-2*p* values for homozygotes, 1-2*p* for heterozygotes, and 2-2*p* for opposite homozygotes, where P_i is the reference allele frequency at locus *i*th. The diagonal elements of the matrix G represent the relationship of the animal with itself, thus, it was used to assess the genomic inbreeding coefficient. Pearson's correlation coefficients were calculated between estimates of pedigree-based (F_{PED}) and genomic inbreeding coefficients (F_{ROH} and F_{GRM}).

2.2.4 Gene prospection in shared ROH regions

The homozygous segments shared by more than 50% of the samples were chosen as an indication of possible ROH islands throughout the genome. The --homozyg-group function implemented in PLINK v1.07 (PURCELL et al., 2007) was used to assess ROH islands shared among individuals. The Map Viewer of the bovine genome UMD3.1.1 was used for identification of genes in ROH regions, available at "National Center for Biotechnology Information" (NCBI Map Viewer - https://www.ncbi.nlm.nih.gov/mapview/). Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 tool (HUANG; SHERMAN; LEMPICKI, 2009a;

HUANG; SHERMAN; LEMPICKI, 2009b) was used to identify Gene Ontology (GO) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways using the list of genes from ROH islands and the *Bos taurus* annotation file as background.

2.3 Results and Discussion

2.3.1 Genomic distribution of Runs of homozygosity

ROH were identified in all 2908 individuals, totaling 161,362 homozygous segments among overall samples. On an individual animal basis, the average number of ROH per animal was 55.12±10.37, with values ranging from 17 to 121. The mean ROH length was 3.17 Mb. The shortest ROH was 1.00 Mb (226 SNPs) found on BTA7 and the longest was 108.97 Mb (33,050 SNPs) found on BTA8. Similarly, Kim et al. (2013) found the longest ROH on BTA 8 (87.13 Mb) in a contemporary Holstein cow and Mastrangelo et al. (2015) in Cinisara cattle breed (112.65 Mb on BTA 8). The number of ROH per chromosome was greater for BTA 5 (10,670 segments), and tended to decrease with chromosome length. Contrasting with our results, Purfield et al. (2012), Mastrangelo et al. (2015), and Gurgul et al. (2016) observed the greatest number of ROH per chromosome on BTA 1. The greatest fraction of chromosome residing in ROH was found on BTA 25 (11.98% of chromosomal length in a ROH) (Figure 1).

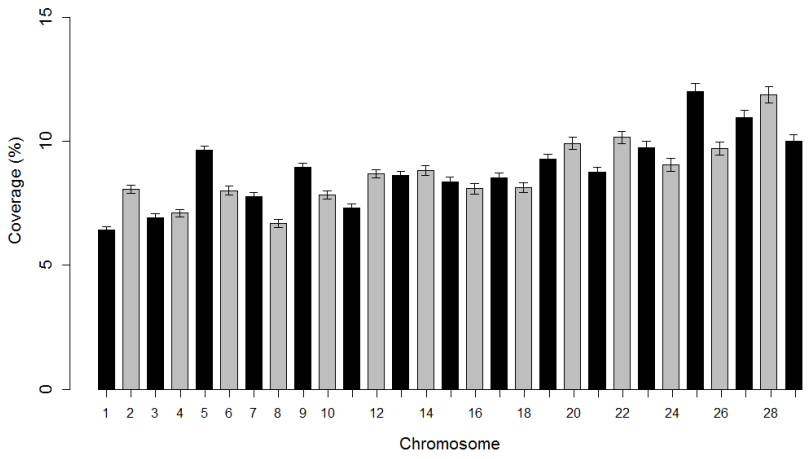

Figure 1. Average percentage of chromosome coverage by runs of homozygosity with a minimum length of 1 Mb. The error bars indicate standard error.

Descriptive statistics of ROH number and length by classes is given in Table 1. Total length of ROH for Gyr is composed mostly of a high number of shorter segments ($ROH_{1\text{-}2\ Mb}$). These segments accounted for approximately 60% of all ROH detected, however contributed less than 25% of the cumulative ROH length. Similar results were observed by Purfield et al. (2012), Ferencaković et al. (2013), Kim et al. (2013), Marras et al. (2014) and Mastrangelo et al. (2015) in which most ROH coverage was in the shorter ROH class and longer ROH were found to be less abundant across the genome.

While shorter ROH were abundant throughout the genome, the proportion of the genome that was covered by them was relatively small. In contrast, larger ROH ($ROH_{>16\ Mb}$) were at least twenty-five fold less abundant than shorter ROH ($ROH_{1\text{-}2\ Mb}$) and still covered a higher proportion of the genome than small and medium ROH (Table 1).

The number of generations of inbreeding can be inferred from the extent of ROH, since their extension is expected to correlate to ancient and recent inbreeding due to recombination events (BROMAN; WEBER, 1999). Therefore, ROH due to recent inbreeding are expected to be longer, since recombination did not have enough time to break up these IBD segments, while short ROH tend to reflect ancient inbreeding because the segments have been broken down by repeated meiosis (KIRIN et al., 2010). The presence of segments larger than 10 Mb is traceable to inbreeding from recent common ancestors that occurred only five generations ago (HOWRIGAN; SIMONSON; KELLER, 2011), and 78% of the animals presented at least one homozygous segment extending over 10 Mb, which is likely a reflection of a recent parental relatedness.

Table 1. Descriptive statistics of runs of homozygosity number (*n*ROH) and length (in Mb) by class (1-2, 2-4, 4-8, 8-16, and >16 Mb).

Class	*n* ROH	(%)	Mean Length	Minimum	Maximum	Standard Deviation	Genome Coverage (%)
$ROH_{1\text{-}2\ Mb}$	95,892	59.42	1.34	1.00	1.99	0.27	1.77
$ROH_{2\text{-}4\ Mb}$	35,395	21.93	2.77	2.00	3.99	0.55	1.34
$ROH_{4\text{-}8\ Mb}$	17,843	11.05	5.54	4.00	7.99	1.12	1.36
$ROH_{8\text{-}16\ Mb}$	8,518	5.27	10.98	8.00	15.99	2.17	1.46
$ROH_{>16\ Mb}$	3,714	2.30	25.23	16.00	108.97	10.06	2.33

The animal displaying the highest autozygosity exhibited a ROH genome coverage encompassing 730.21 Mb of the total autosomal genome extension covered by markers (29.20% of the cattle genome), with 71 ROH ≥ $ROH_{1\text{-}2\ Mb}$, and a mean ROH length of 10.28 Mb. Purfield et al. (2012) observed that dairy breeds were the most autozygous animals among several studied breeds, and had on average 700.3 Mb classified as ROH. Mastrangelo et al. (2015) observed a similar value for the Reggiana dairy breed (725.2 Mb classified as ROH). Besides, Marras et al. (2014) described that dairy breeds had a higher sum of all ROH than beef breeds. The higher autozygosity observed in dairy breeds can be explained by the intense artificial selection and the repeatedly use of superior and proven sires for reproduction by artificial insemination (KIM et al., 2013). The least inbred animal exhibited a ROH genome coverage encompassing 48.81 Mb (1.95% of the cattle genome), with 32 ROH ≥ $ROH_{1\text{-}2\ Mb}$, and a mean ROH length of 1.52 Mb. Figure 2 illustrates the relationship between the total number of ROH per animal and the total length (Mb) of genome in such ROH for each individual. The sum of all ROH per animal allowed the estimation of the percentage of the genome that is autozygous. Overall, the ROH statistics suggest that on average 7.01% (175.28Mb) of the genome of this population is autozygous. A similar value was observed by Marras et al. (2014) (7% for Marchigiana beef purpose breed). Gyr cattle presented a lower genome average autozygosity compared to previous studies reported by Ferenčaković et al. (2011) (9% for Austrian dual purpose Simmental, Brown Swiss, and Tyrol Grey bulls) and Kim et al. (2013) (10% for Holstein cattle), and a higher autozygosity than results obtained by Zavarez et al. (2015) (4.58% for Nelore cattle).

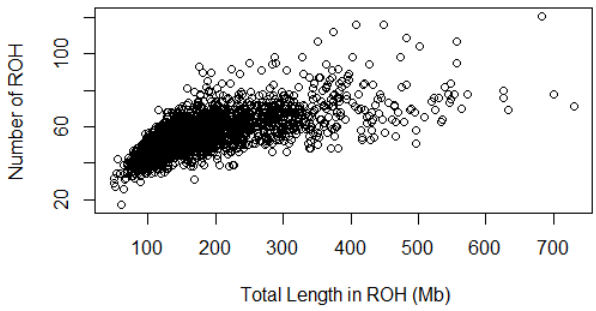

Figure 2. Relationship between the total number of runs of homozygosity (ROH>1 Mb) and the total length (Mb) of the genome in such ROH for each individual. Each circle represents one animal.

2.3.2 Pedigree and genomic inbreeding

Descriptive statistics for F_{PED} and F_{ROH} coefficients are presented in Table 2. Among F_{ROH} estimates, we can observe an increase in variation with ROH length, being evidenced by the coefficient of variation (CV). The higher the CV was, the greater the differences between the mean and median were for each F_{ROH} length. Thus, given the dissimilarity among the CV, we assumed that the mean should not be used as the best measurement of central tendency, indicating that the median should be used instead for F_{PED} and F_{ROH} coefficients.

Table 2. Number of genotyped animals (*n*) and descriptive statistics for pedigree (F_{PED}) coefficient and genomic inbreeding coefficients based on runs of homozygosity (F_{ROH}) for different lengths ($F_{ROH\ 1\text{-}2\ Mb}$, $F_{ROH\ 2\text{-}4\ Mb}$, $F_{ROH\ 4\text{-}8\ Mb}$, $F_{ROH\ 8\text{-}16\ Mb}$, and $F_{ROH\ >16\ Mb}$).

Inbreeding coefficient	Mean	Median	Minimum	Maximum	Coefficient of Variation (%)	*n*
F_{PED}	0.019	0.004	0.000	0.327	3.387	2,758
$F_{ROH1\text{-}2\ Mb}$	0.017	0.017	0.006	0.037	20.70	2,758
$F_{ROH2\text{-}4\ Mb}$	0.013	0.013	0.001	0.039	35.30	2,757
$F_{ROH4\text{-}8\ Mb}$	0.013	0.012	0.001	0.063	55.63	2,740
$F_{ROH8\text{-}16\ Mb}$	0.014	0.012	0.003	0.082	73.04	2,422
$F_{ROH>16\ Mb}$	0.023	0.016	0.006	0.201	97.75	1,533

In the dairy industry, genomic inbreeding coefficients of genotyped animals are commonly calculated from F_{GRM} (BJELLAND et al., 2013). F_{GRM} moderately correlated with F_{PED} (0.37) (Figure 4), concurring with the results obtained by Pryce et al. (2014). VanRaden et al. (2011) reported higher correlations for Holstein (0.59), Jersey (0.68), and Brown Swiss (0.61) animals. Hayes and Goddard (2008) also obtained higher correlations for Australian Angus bulls (0.69). Lower correlations between these estimates were reported by Gurgul et al. (2016), Marras et al. (2014), and Zhang et al. (2015b). Two out of three reasons hypothesized by Pryce et al. (2014) might explain the poor correlation found in our study: (i) individuals from sub-populations for which allele frequencies diverge from those of the entire population are estimated to have high F_{GRM}; and (ii) pedigree completeness.

Low to moderate correlations were observed between F_{PED}-F_{ROH} and F_{GRM}-$F_{ROH.}$ The lowest correlation was observed for $F_{ROH1\text{-}2\ Mb}$- $F_{ROH8\text{-}16\ Mb}$ (-0.04) and the

highest for F_{GRM}-$F_{ROH>16\ Mb}$ (0.59). Correlations between F_{ROH} from different lengths and F_{PED} increased with ROH length. This tendency may be explained by considering that ROH reflect both past and recent relatedness, and that F_{PED} estimates are based on pedigree records which may not extend back many generations (FERENČAKOVIĆ et al., 2013; MARRAS et al., 2014). When longer ROH reflecting recent relatedness are considered to calculate F_{ROH}, the F_{PED}-F_{ROH} correlation tends to be higher (MARRAS et al., 2014; SAURA et al., 2015).

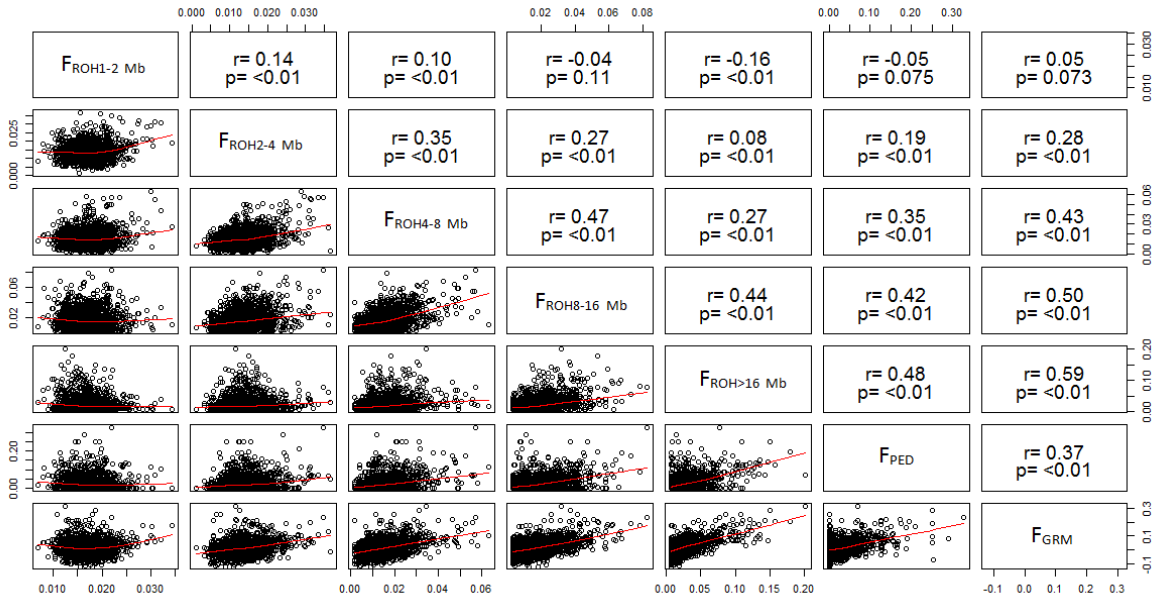

Figure 4. Scatterplots (lower panel) and correlations (upper panel) of genomic inbreeding coefficients F_{ROH} ($F_{ROH\ 1\text{-}2\ Mb}$, $F_{ROH\ 2\text{-}4\ Mb}$, $F_{ROH\ 4\text{-}8\ Mb}$, $F_{ROH\ 8\text{-}16\ Mb}$, and $F_{ROH\ >16\ Mb}$) and F_{GRM}, and pedigree-based inbreeding coefficients (F_{PED}).

Several authors have found a high correlation between F_{PED}-F_{ROH} when a deeper number of described generations are available in the pedigree (FERENČAKOVIĆ et al., 2011; PURFIELD et al., 2012; FERENČAKOVIĆ et al., 2013; MARRAS et al., 2014; GURGUL et al., 2016), suggesting that the correlation between these parameters increases with pedigree deep. Ferenčaković et al. (2011, 2013) observed F_{PED}-F_{ROH} correlations values ranging from 0.61 to 0.67 and 0.50 to 0.72, respectively, for pedigrees with more than five generations. Purfield et al. (2012) used a complete generation equivalents higher than six and obtained F_{PED}-F_{ROH} correlations of 0.73 for ROH>10 Mb and 0.71 for ROH>1Mb, both with the

reduced panel. Marras et al. (2014) observed high F_{PED}-F_{ROH} correlations using pedigree with four, seven and ten generations, with values from 0.56 to 0.74. Gurgul et al. (2016) also reported the highest F_{PED}-F_{ROH} correlation for animals with seven complete generations of pedigree data, with an average value of 0.45. In the present study, a small number of generations were available to estimate F_{PED}, which may have introduced biased F_{PED} values as the pedigree was not able to cover ancient relatedness. Furthermore, incomplete pedigree also reduces F_{PED} estimates (CASSEL et al., 2003).

The correlations between F_{GRM}-F_{ROH} were higher than those between F_{PED}-F_{ROH} for all ROH classes described. Conversely, Marras et al. (2014) observed a higher F_{PED}-F_{ROH} correlation than those reported for F_{GRM}-F_{ROH}. Several studies also have found a low to moderate F_{GRM}-F_{ROH} correlation for dairy breeds (MARRAS et al., 2014; MASTRANGELO et al., 2015). In Holstein cattle, moderate to high correlations were described by Bjelland et al. (2012) (0.81). Pryce et al. (2014) observed a correlation of 0.62 in Holstein and Jersey populations, and Zavarez et al. (2015) correlations ranging from 0.41 to 0.74 in Nellore cattle based on ROH of different minimum lengths.

The inbreeding evolution (F_{PED} and F_{ROH}) for animals born between 1980 and 2012 is shown in Figure 5 and the genotyping sampling of animals per inbreeding coefficient in table 2. The F_{PED} evolution showed a tendency to slightly increase over time (Figure 5a), while F_{ROH} tended to decrease for segments higher than 4 Mb (Figure 5d, 5e, and 5f). It is worth to highlight that the decline in $F_{ROH>8\text{-}16\ Mb}$ and $F_{ROH>16\ Mb}$ reflects an inbreeding up to six and just three generations ago, respectively. The reduction in these coefficients since the 80's happen together with the creation of the Brazilian Dairy Gyr Breeding Program (PNMGL) and the implementation of the Gyr progeny testing, both in 1895 (JUNIOR SANTANA, 2014). Probably, these facts suggested that different proven sires from different lines started to be incorporated into the population and the herds were no longer closed. Additionally, it is expected that better mating decisions were taken by the breeders with the advent of the breeding program, decreasing the genomic inbreeding level in these populations over time.

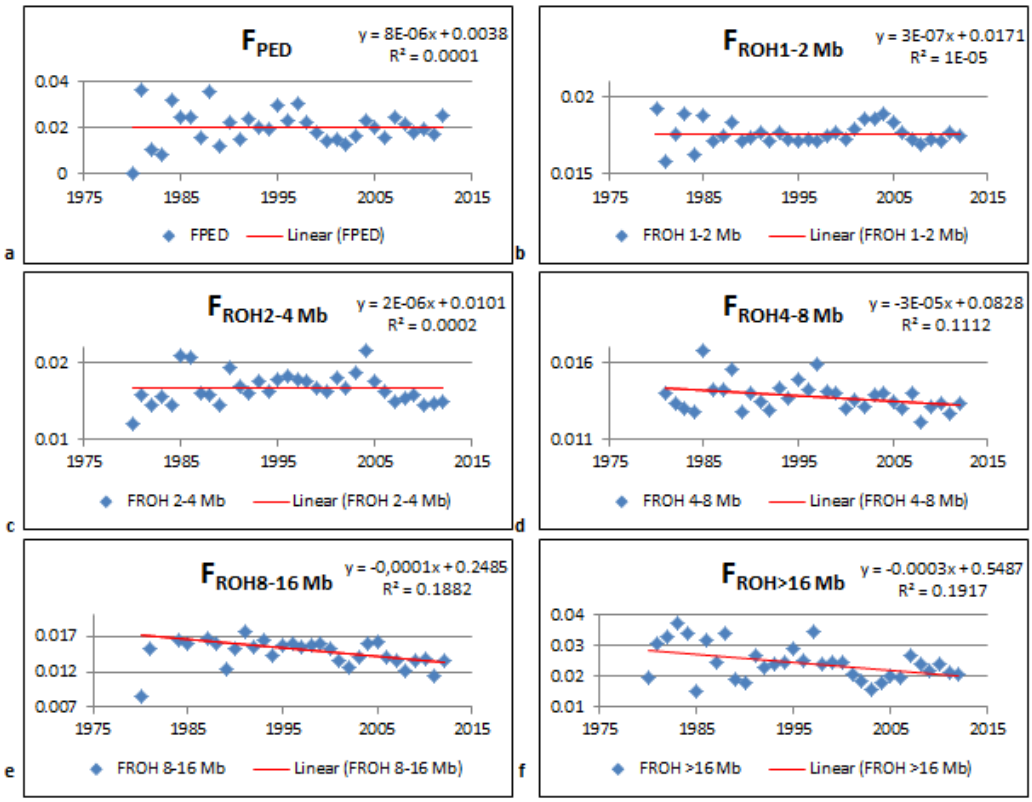

Figure 5. Inbreeding evolution over the past 30 years for pedigree-based inbreeding (F_{PED}) and F_{ROH} ($F_{ROH1\text{-}2\ Mb}$, $F_{ROH2\text{-}4\ Mb}$, $F_{ROH4\text{-}8\ Mb}$, $F_{ROH8\text{-}16\ Mb}$, and $F_{ROH>16\ Mb}$) coefficients. Each blue dot represents the inbreeding average per year.

2.3.3 Gene characterization in ROH islands

Overlapping ROH were evident across the genome, and their genomic distribution was non-uniform both in length and position across chromosomes. A total of 14 regions were identified with ROH frequencies exceeding 50% of the whole population on BTA 2, 6, 10, 12 and 14 (Appendix A), suggesting that these regions are likely a sign of ROH islands shared among animals (FERENČAKOVIĆ et al., 2013). ROH islands can be defined as genomic regions with reduced genetic diversity and, consequently, high homozygosity around the selected locus that might harbor targets of positive selection and are under strong selective pressure (PEMBERTON et al., 2012). The most strong pattern was observed on BTA 2 (78,394,916:87,587,063), with an overlapping ROH region present in 92% of the samples (Figure 3). This region showed an enrichment of genes involved with the immune system (Appendix A). Similarly, Marras et al. (2014) reported a ROH in 90%

of the samples in Piedmontese cattle, however, it was located at the beginning of BTA 2 closest to the myostatin (*MSTN*) locus. Karimi (2013) identified four ROH islands exceeding 40% of individuals for Brahman, Gyr and Nelore breeds, with the most common pattern on BTA 21 exceeding 92% of individuals. The ROH islands on BTA 10 and BTA 12 observed by Karimi (2013) were not found to be located in the same genomic region as in our study. However, the ROH island on BTA 10 (24,575,700:25,619,800) reported by them was located closest (25,895,397:27,374,489). ROH islands on BTA 6 were also seen by Gaspa et al. (2014) in Italian Holstein cattle and Marras et al. (2014) in dairy and beef breeds. ROH islands on BTA 6 may be an indicative of signatures of selection for dairy traits, since this chromosome is well documented to harbour genes that affect milk production traits (COHEN et al., 2004; COHEN-ZINDER et al., 2005; SCHNABEL et al., 2005; SCHOPEN et al., 2011). Other studies with dairy breeds have found different ROH islands among chromosomes (PURFIELD et al., 2012; KIM et al., 2013; HOWARD et al., 2015).

A relevant number of genes (n=282) inside these ROH islands were observed (Appendix A), in which several of them play important role in the mammary gland biology and have a prominent importance in milk, dairy traits, and heat adaptation. Gene ontology (GO) and pathway analysis (KEEG) were performed to obtain a broad functional insight for the set of genes. An enrichment of genes involved in several biological processes, which we highlighted cell differentiation (GO:0030154), regulation of cell death (GO:0010941), positive regulation of cell migration (GO:0030335), synaptic transmission and dopaminergic (GO:0001963), response to unfolded protein (GO:0006986), utero embryonic development (GO:0001701), mRNA transcription (GO:0009299), cellular response to insulin stimulus (GO:0032869), positive regulation of B cell proliferation (GO:0030890), and negative regulation of B cell apoptotic process (GO:0002903), were observed.

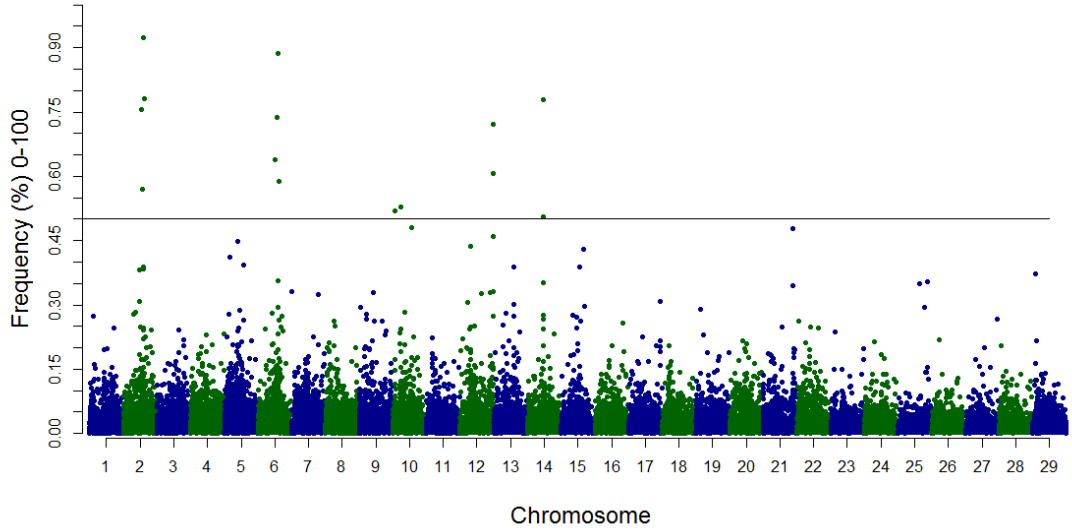

Figure 3. Distribution of overlapping ROH (ROH islands) in Gyr cattle genome. X axis: distribution of ROH across the genome. Y axis: frequency (%) of overlapping ROH among samples.

A total of 10 genes were identified related to cell differentiation process (GO:0030154), in which we highlight the *TRAPPC9* gene on BTA 14. Interestingly, this gene was found to have significant effects on mastitis-related traits in Chinese Holstein populations (WANG et al., 2015). Besides, Jiang et al. (2014) observed a higher TRAPPC9 mRNA expression level in the mammary gland of lactating cows than in the other tissues, such as heart, liver, lung, kidney, ovary, uterus, and muscle.

Three genes associated with regulation of cell death (GO:0010941) were identified and one of them, the *EEF1D* gene on BTA 14, was found to possess one SNP (rs109661298) associated with milk fat percentage in Chinese Holstein population (JIANG et al., 2010). More recently, Xie et al. (2014) identified two alternatively spliced variants of the *EEF1D* gene *(EEF1Da* and *EEF1Db*) in the mammary gland of lactating Holstein cow. The *EEF1Da* (exon 1a + exon 2) was the dominant transcript and expressed in higher level in mammary gland than in other tissues of lactating dairy cows, implying that this gene is likely involved in biological and physiological processess in the mammary gland. Jiang et al. (2014) also observed a high EEF1D mRNA expression level in the mammary gland of lactating cows. This gene was also related with the mRNA transcription (GO:0009299), as also the *HSF1* gene found on BTA 14. This gene is a heat-shock transcription factor, and its transcription is rapidly induced after temperature stress (provided by RefSeq,

Jul 2008). Heat shock transcription factors and heat shock proteins (HSP) play a crucial role in environmental stress adaptation and thermal balance since it allows cells to adapt to gradual environmental changes (SØRENSEN et al., 2003), controlling the balance between survival and an effective immune system in order to adjust to stress (MORANGE, 2006). Chan, Nagaraj and Reverter (2010), in a study comparing *Bos indicus* and *Bos taurus* cattle, observed genes that have been reported to be associated with performance attributes for tropical-adaptation and among their findings, *HSP* (*HSPA14* and *HSPB9*) were associated with heat adaptation. Kumar et al. (2015) observed a higher abundance of *HSP* family genes during summer and winter compared to mid-spring season in *Bos indicus* cattle and Murrah buffaloes, and the magnitude of increase was higher during summer as compared to winter. Among their findings, a significantly higher ($p \leq 0.001$) HSF1 mRNA expression during the summer as compared to mid-spring season was also observed. These findings are consistent with the zebu cattle adaptation traits, in which we highlight its greater ability to tolerate poor feed and hostile climate. Furthermore, *HSP* may be used to access stress response and heat tolerance adaptation in cattle due to its variation in expression under different heat stress conditions (KUMAR et al., 2015).

A total of seven genes were identified in ROH islands related to positive regulation of cell migration (GO:0030335) biological process. Of these, the *IRS2*, *ATP8A1*, *GABRG1*, and *GABRAG2* genes have been previously associated with dairy traits. The *IRS2* gene on BTA 12 encodes the insulin receptor substrate 2, a cytoplasmic signaling molecule that mediates effects of insulin, insulin-like growth factor 1, and other cytokines (provided by RefSeq, Jul 2008). Insulin infusion has been shown to increase milk and protein yields, and reduce milk fat content and yield in lactating goats. It also decreased net uptake of C10:0, C14:0, C16:0, trans-C16:1 and >C18:0 fatty acids, and increased mammary blood flow by 42% (BEQUETTE et al., 2001). The *IRS2 gene* has also been shown to be associated with the positive regulation of B cell proliferation (GO:0030890) and negative regulation of B cell apoptotic process (GO:0002903). The *ATP8A1*, *GABRG1*, and *GABRAG2* genes on BTA 6 laid within the region of the most iHS score as reported by Hayes et al. (2008)

in Norwegian Red cattle, a breed which has been strongly selected for milk production.

The *CRH* gene found on BTA 14 was identified associated with synaptic transmission and dopaminergic (GO:0001963) biological process. This hormone is secreated in response to stress, and binds to corticotropin releasing hormone receptors and stimulates the release of adrenocorticotropic hormone from the pituitary gland (provided by RefSeq, Nov 2015). Since stress disrupts lactation, intracerebroventricularly admnistration of CRH was performed in primiparous rats and a dose-dependent reduction in the amount of milk obtained by the pups was observed (ALMEIDA; YASSOURIDIS; FORGAS-MOYA, 1994).

Three genes were related to cellular response to insulin stimulus (GO:0032869) and among them the *STAT1* gene found on BTA 2 has been reported to be expressed through pregnancy, lactation and involution (WATSON, 2001). *STAT1* is under control of the prolactin hormone, and several events take place when the prolactin binds to its receptor, such as the activation of the STAT1, STAT3, and STAT5 proteins, which regulate the transcription of genes involved in secretion of milk proteins and components (TUCKER, 2000; BOLE-FEYSOT et al., 1998). *STAT1* effects on increased milk fat percentage (0.01%) and milk protein percentage was reported by Cobanoglu et al. (2006).

We also found a number of genes within ROH islands that have been reported to have a prominent importance in milk and dairy traits on BTA 14 (*DGAT1*, *CYP11B1*, and *DERL1)*, BTA 2 (*INSIG2*), and BTA 6 (*IGFBP7*) chromosomes.

The *DGAT1* gene found on BTA 14 catalyzes the last stage in triacylglycerols synthesis (CASES et al., 1980), and a nonconservative lysine to alanine (K232A) substitution in this gene has been shown to have effects on milk production traits in dairy cattle breeds (THALLER et al., 2003; KAUPE et al., 2007), specially for milk fat content (WINTER et al., 2002; GRISART et al., 2004; ARGOV-ARGAMAN et al., 2013) and fatty acid composition (SCHENNINK et al., 2007).

The *CYP11B1* gene on BTA 14 catalyzes the 11β- and 18-hydroxylation of corticosteroids in cattle (MULLER, 1998; LISUREK; BERNHARDT, 2004), and cortisol is one of the principal hormones involved in lipid metabolism (BHATHENA, 2000). Polimorfism on *CYP11B1* (V30A) gene had associated effects on milk fat

content (+0.04%), protein content (+0.01%), and milk yield (+82 kg) (KAUPE et al., 2007). In bufaloes, a novel polymorphism on *CYP11B1* (p.A313T) gene was also found associated with increased milk fat percentage (MARYAM et al., 2015).

It is suggested that the protein coded by *DERL1* gene on BTA 14 has effects associated with tissue remodeling and maintenance of function in reproductive tissues. A greater DERL1 mRNA expression was associated with active ovarian follicular growth and early corpus luteum development and function, suggesting a role of *DERL1* gene in developing follicles (NDIAYE; LUSSIER; PATE, 2010). Our findings are particularly interesting, since fertility and longevity are highly connected (ESSL, 1998) and have a considerable economic impact upon dairy cattle industry. Reduced fertility affects longevity in cattle, decreasing the lifetime reproductive cycles and lactations per female (NDIAYE; LUSSIER; PATE, 2010).

Insulin-induced genes (*INSIG*s) are important mediators in the lipid metabolism (XIAO-YING; SHENG-QIU, 2010). Given the role of the *INSIG*s genes in lipogenesis and cholesterol regulation in animals, the study of variants which might affect the production performance in animals is of substantial interest (DENG et al., 2016). On BTA 2 ROH islands, the *INSIG2* gene was observed, and polimorfism in this gene has been reported to have effects on milk fatty acids composition in Holstein cattle (RINCON et al., 2012). INSIG2 mRNA expression was increased during lactation in cows (BIONAZ; LOOR, 2008) and was highly expressed in the mammary gland in buffalo, suggesting that INSIGs protein might be important in the regulation of fat milk synthesis (WU et al., 2014). Mutations in 88, 436 and 471 nucleotides of the *INSIG2* gene sequence in ewes influenced milk fat content, and the SNP in 1,071 nucleotide was associated with milk yield and milk protein content (LURIDIANA et al., 2014).

The *IGFBP7* gene found on BTA 6 encodes a member of the insulin-like growth factor (IGF)-binding protein (IGFBP) family (provided by RefSeq, Dec 2011). *IGF*s and *IGFBP*s genes play an essential role in mammary gland development by influencing mammary epithelial cell proliferation, differentiation and cell survival (POLLAK; SCHERNHAMMER; HANKINSON, 2004; KLEINBERG; BARCELLOS-HOFF, 2011; BARTELLA et al., 2012). Virgin mice’s lacking the *IGFBP7* gene has been described to have significantly mammary gland development retardation.

Pregnant Igfbp7-null mice exhibited precocious mammary gland involution in the presence of suckling pups, and Igfpb7-null glands contained fewer alveolar structures (CHATTERJEE et al., 2014).

Among the KEGG pathways identified, we highlighted the ones related to environmental information processing, such as neuroactive ligand-receptor interaction (bta04080), PI3K-Akt signaling pathway (bta04151), and AMPK signaling pathway (bta04152) with 11, 10, and 5 genes identified in ROH islands, respectively. PI3K-Akt signaling pathway regulates key cellular functions such as transcription, translation, growth, proliferation, and survival. AMPK signaling pathway acts as a sensor of cellular energy status leading to a concomitant inhibition of energy-consuming biosynthetic pathways and activation of ATP-producing catabolic pathways.

2.4 Final considerations

Despite of the reduced genetic basis and the limited number of animals imported to form the first Gyr dairy lines, the autozygotic proportion of the genome were considerably low in this population. Besides, it can be seen a clearly decay in F_{ROH} for segments higher than 4 Mb, which is consistent with the development of the first Gyr breeding program and the progeny testing program. Low correlations between F_{PED}-F_{ROH} for small segments indicate that F_{PED} estimates are not the most suitable method to capture ancient inbreeding, especially when a deeper number of generations are not described in the pedigree. The existence of a moderate correlation between larger ROH indicates that the levels of autozygosity derived from ROH can be used as an accurate estimator of individual inbreeding levels. Several common ROH islands have been found in the Gyr genome, suggesting that ROH might be used to identify genomic regions under selection. Common islands on BTA 14 are supposed to be a sign of strong selection for dairy traits and environmental adaptation as several genes associated with them were identified. Our findings contribute to the understanding of the effects of inbreeding when accessing genome-wide autozygosity, and how selection can shape the distribution of ROH islands in the cattle genome. Hence, this approach may contribute to comprehend the

evolutionary process of the Gyr breed and provide the basis to overcome future challenges.

2.5 References

AGUILAR I.; MISZTAL I. Technical Note: Recursive Algorithm for Inbreeding Coefficients Assuming Nonzero Inbreeding of Unknown Parents. **Journal of Dairy Science**, v. 91, p. 1669-1972, 2007.

ALMEIDA O.F.; YASSOURIDIS A.; FORGAS-MOYA I. Reduced availability of milk after central injections of corticotropin-releasing hormone in lactating rats. **Neuroendocrinology**, v. 59, supl. 2, p.72-77, 1994.

ARGOV-ARGAMAN N.; MIDA K.; COHEN B-C.; VISKER M.; HETTINGA K. Milk Fat Content and DGAT1 Genotype Determine Lipid Composition of the Milk Fat Globule Membrane. **PLoS ONE**, v. 8, e68707, 2013.

BARTELLA V.; DE MARCO P.; MALAGUARNERA R.; BELFIORE A.; MAGGIOLINI M. New advances on the functional cross-talk between insulin-like growth factor-I and estrogen signaling in cancer. **Cellular Signalling**, v. 24, p. 1515–1521, 2012.

BEQUETTE B.J.; KYLE C.E.; CROMPTON L.A.; BUCHAN V.; HANIGAN M.D. Insulin Regulates Milk Production and Mammary Gland and Hind-Leg Amino Acid Fluxes and Blood Flow in Lactating Goats. **Journal of Dairy Science**, v. 84, p. 241-255, 2001.

BHATHENA S.J. Relationship between fatty acids and the endocrine system. **BioFactors**, v. 13, p. 35–39, 2000.

BIONAZ M.; LOOR J. Gene networks driving bovine milk fat synthesis during the lactation cycle. **BMC Genomics**, v. 9, 2008.

BJELLAND D.W.; WEIGEL K.A.; VUKASINOVIC N.; NKRUMAH J.D. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. **Journal of Dairy Science**, v. 96, p. 4697–4706, 2013.

BOISON S.A.; SANTOS D.J.A.; UTSUNOMIYA A.H.T.; CARVALHEIRO R.; NEVES H.H.R.; O'BRIEN A.M.P.; GARCIA J.F.; SOLKNER J.; SILVA M.V.G.B. 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.**, v. 98, p. 4969-4989, 2015

BOLE-FEYSOT C.; GOFFIN V.; EDERY M.; BINART N; KELLY P.A. Prolactin (PRL) and Its Receptor: Actions, Signal Transduction Pathways and Phenotypes Observed in PRL Receptor Knockout Mice. **Endocrine Reviews**, v. 19, supl. 3, p. 225–268, 1998.

BOSSE M.; MEGENS H-J.; MADSEN O.; PAUDEL Y.; FRANTZ L.A.F.; SCHOOK L.B.; CROOIJMANS R.P.M.A.; GROENEN M.A.M. Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. **PLoS Genetics**, v. 8, e1003100, 2012.

BROMAN K.W.; WEBER J.L. Long Homozygous Chromosomal Segments in Reference Families from the Centre d'E´tude du Polymorphisme Humain. **The American Journal of Human Genetics**, v. 65, p. 1493–1500, 1999.

CASES S.; SMITH S.J.; ZHENG Y-W.; MYERS H.M.; LEAR S.R.; SANDE E.; NOVAK S.; COLLINS C.; WELCHI C.B.; LUSISI A.J.; ERICKSON S.K.; FARESE, R.V.JR. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. **Proceedings of the National Academy of Sciences**, v. 95, p. 13018–13023, 1998.

CASSELL B.G.; ADAMEC V.; PEARSON R.E. Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. **Journal of Dairy Science,** v. 86, p. 2967–2976, 2003.

CHAN, E.K.F.; NAGARAJ S.H.; REVETER A. The evolution of tropical adaptation: comparing taurine and zebu cattle. **Animal Genetics**, v. 41, p. 467-477.

CHATTERJEE S.; BACOPULOS S.; YANG W.; AMEMIYA Y.; SPYROPOULOS D.; RAOUF A.; SETH A. Loss of Igfbp7 Causes Precocious Involution in Lactating Mouse Mammary Gland. **PLoS ONE**, v. 9, e87858, 2014.

COBANOGLU O.; ZAITOUN I.; CHANG Y.M.; SHOOK G.E; KHATIB H. Effects of the Signal Transducer and Activator of Transcription 1 (STAT1) Gene on Milk Production Traits in Holstein Dairy Cattle. **Journal of Dairy Science**, v. 89, p. 4433–4437, 2006.

COHEN M.; REICHENSTEIN M.; WIND A.E-V.D.; HEON-LEE J.; SHANI M.; LEWIN H.A.; WELLER J.I; RON M.; SEROUSSI E. Cloning and characterization of FAM13A1—a gene near a milk protein QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holsteins. **Genomics**, v. 84, p. 374– 383, 2004.

COHEN-ZINDER M.; SEROUSSI E.; LARKIN D.M.; LOOR J.J; WIND A.E-V.D.; LEE J-H.; DRACKLEY J.K.; BAND M.R.; HERNANDEZ A.G.; SHANI M.; LEWIN H.A.; WELLER J.I.; RON M. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. **Genome Research**, v. 15, p. 936–944, 2005.

CURIK I.; SÖLKNER J.; STIPIC N. Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. **Journal of Animal Breeding and Genetics**, v. 119, p. 101–115, 2002.

DENG T.; PANG C.; MA X.; LU X.; DUAN A.; ZHU P.; LIANG X. Four novel polymorphisms of buffalo INSIG2 gene are associated with milk production traits in Chinese buffaloes. **Molecular and Cellular Probes**, v. 5, p. 294-299, 2016.

ESSL A. Longevity in dairy cattle breeding: a review. **Livestock Production Science**, v. 57, p. 79–89, 1998.

FALCONER D.S.; MACKAY T.F.C. **Introduction to quantitative genetics**. 4. ed. Longman, Essex, 1996.

FERENČAKOVIĆ M.; HAMZIC E.; GREDLER B.; CURIK I.; SÖLKNER J. Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. **Agriculturae Conspectus Scientificus**, v. 76, p. 325–328, 2011.

FERENČAKOVIĆ M.; HAMZIĆ E.; GREDLER B.; SOLBERG T.R; KLEMETSDAL G.; CURIK I.; SÖLKNER J. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. **Journal of Animal Breeding and Genetics**, v. 130, p. 286–293, 2013.

GASPA G.; MARRAS G.; SORBOLINI S.; MARSAN P.A.; WILLIAMS J.L.; VALENTINI A.; DIMAURO C.; MACCIOTTA N.P.P. Genome-Wide Homozygosity in Italian Holstein Cattle using HD SNP Panel. In: 10th World Congress of Genetics Applied to Livestock Production, 2014, Vancouver. **Anais eletrônicos**. Disponível em: <https://asas.confex.com/asas/WCGALP14/webprogram/programs.html>. Acesso em 20 set. 2016.

GIBSON J.; MORTON N.E.; COLLINS A. Extended tracts of homozygosity in outbred human populations. **Human Molecular Genetics**, v. 15, p. 789–795, 2006.

GRISART B.; FARNIR F.; KARIM L.; CAMBISANO N.; KIM J-J.; KVASZ A.; MNI M.; SIMON P.; FRERE J-M.; COPPIETERS W.; GEORGES M. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. **Proceedings of the National Academy of Sciences**, v. 101, p. 2398-2403, 2004.

GURGUL A.; SZMATOŁA T.; TOPOLSKI P.; JASIELCZUK I.; ŻUKOWSKI K,; BUGNO-PONIEWIERSKA M. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. **Journal of Applied Genetics**, v. 57, supl. 4, p. 527-530, 2016.

HAYES B.J.; GODDARD M.E. Technical note: Prediction of breeding values using marker-derived relationship matrices. **Journal of Animal Science**, v. 86, p. 2089-2091, 2008.

HAYES B.J.; LIEN S.; NILSEN H.; OLSEN H.G.; BERG P.; MACEACHEM S.; POTTER S.; MEUWISSEN T.H.E. The origin of selection signatures on bovine chromosome 6. **Animal Genetics**, v. 39, p. 105-111, 2008.

HERRERO-MEDRANO J.M.; MEGENS H-J.; GROENEN MAM.; RAMIS G.; BOSSE M.; PÉREZ-ENCISO M.; CROOIJMANS R.P.M.A. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. **BMC Genetics**, v. 14, 2013.

HOWARD J.T.; MALTECCA C.; HAILE-MARIAM M.; HAYES B.J.; PRYCE J.E. Characterizing homozygosity across United States, New Zealand and Australian Jersey cow and bull populations. **BMC Genomics**, v. 16, 2015.

HOWRIGAN D.P.; SIMONSON M.A.; KELLER M.C. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. **BMC Genomics**, v. 12, 2011.

HUANG D.W.; SHERMAN B.T.; LEMPICKI R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. **Nature Protocols**, v. 4, p. 44-57, 2009.

HUANG D.W.; SHERMAN B.T.; LEMPICKI R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. **Nucleic Acids Research**, v. 37, p. 1-13, 2009.

JIANG L.; LIU J.; SUN D.; MA P.; DING X.; YU Y.; ZHANG Q. Genome Wide Association Studies for Milk Production Traits in Chinese Holstein Population. **PLoS ONE**, v. 5, e13661, 2010.

JIANG L.; LIU X.; YANG J.; WANG H.; JIANG J.; LIU L.; HE S.; DING X.; LIU J.; ZHANG Q. Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. **BMC Genomics**, v. 15, 2014.

JUNIOR SANTANA 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**, v. 163, p. 26-33, 2014.

KARIMI, Z. **Runs of Homozygosity patterns in Taurine and Indicine cattle breeds**. 2013. 53 f. Dissertação (Major thesis animal breeding and genetics) - University of Natural Resources and Life Sciences, Vienna, 2013.

KAUPE B.; BRANDT H.; PRINZENBERG E-M.; ERHARDT G. Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. **Journal of Animal Science**, v. 85, p. 11-21, 2007.

KELLER M.C.; VISSCHER P.M.; GODDARD M.E. Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. **Genetics**, v. 189, p. 237-249, 2011.

KIM E-S.; COLE J.B.; HUSON H.; WIGGANS G.R.; VAN TASSELL C.P.; CROOKER B.A.; LIU G.; DA Y.; SONSTEGARD T.D. Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein Cattle. **PLoS ONE**, v. 8, e80813, 2013.

KIM E-S.; SONSTEGARD T.D. VAN TASSELL C.P.; WIGGANS G.; ROTHSCHILD M.F. Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein Cattle. **PLoS ONE**, v. 10, e0129967, 2015.

KIRIN M.; MCQUILLAN R.; FRANKLIN C.S.; CAMPBELL H.; MCKEIGUE P.M.; WILSON J.F. Genomic Runs of Homozygosity Record Population History and Consanguinity. **PLoS ONE**, v. 5, e13996, 2010.

KLEINBERG D.L.; BARCELLOS-HOFF M.H. The Pivotal Role of Insulin-Like Growth Factor I in Normal Mammary Development. **Endocrinology Metabolism Clinics of North America**, v. 40, supl. 3, 2011.

KUMAR A.; ASHRAF S.; GOUD T.S.; GREWAL A.; SINGH S.V.; YADAV B.R.; UPADHYAY R.C. Expression profiling of major heat shock protein genes during different seasons in cattle (*Bos indicus*) and buffalo (*Bubalus bubalis*) under tropical climatic condition. **Journal of Thermal Biology**, v. 51, p. 55-64, 2015.

LISUREK M.; BERNHARDT R. Modulation of aldosterone and cortisol synthesis on the molecular level. **Molecular and Cellular Endocrinology**, v. 215, p. 149–159, 2004.

LURIDIANA S.; MURA M.C.; COSSO G.; DAGA C.; BODANO S.; DIAZ M.L.; BINI P.P.; CARCANGIU V. Ovine insulin induced-gene-2: Molecular characterization, polymorphisms and association with milk traits. **Molecular Biology Reports**, v. 41, p. 4827-4831, 2014.

MARRAS G.; GASPA G.; SORBOLINI S.; DIMAURO C.; AJMONE-MARSAN P.; VALENTINI A.; WILLIAMS J.L.; MACCIOTTA N.P.P. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. **Animal Genetics**, v. 46, p. 110-121, 2014.

MARYAM J.; BABAR M.E.; NADEEMA A.; YAQUB T.; HASHMI A.S. Identification of functional consequence of a novel selection signature in CYP11b1 gene for milk fat content in *Bubalus bubalis*. **Meta Gene**, v. 6, p. 85-90, 2015.

MASTRANGELO S.; TOLONE M.; DI GERLANDO R.; FONTANESI L.; SARDINA M.T.; PORTOLANO B. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. **Animal**, v. 10, supl. 5, p. 746-754, 2016.

MCQUILLAN R.; LEUTENEGGER A-L.; ABDEL-RAHMAN R.; FRANKLIN C.S.; PERICIC M.; BARAC-LAUC L.; SMOLEJ-NARANCIC N.; JANICIJEVIC B.; POLASEK O.; TENESA A.; MACLEOD A.K.; FARRINGTON S.M.; RUDAN P.; HAYWARD C.; VITART V.; RUDAN I.; WILD S.H.; DUNLOP M.G.; WRIGHT A.F.; CAMPBELL H.; WILSON J.F. Runs of Homozygosity in European Populations. **The American Journal of Human Genetics**, v. 83, p. 359–372, 2008.

MEUWISSEN T.H.E.; HAYES B.J.; GODDARD M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, v. 157, p. 1819–1829, 2001.

MORANGE M. **Molecular Chaperones in Health and Disease.** Springer-Verlag Berlin Heidelberg, 2006, p. 153-169

MULLER J. Regulation of aldosterone biosynthesis: the end of the road? **Clinical and Experimental Pharmacology and Physiology**, v. 25, p. 79–85, 1998.

NDIAYE K.; LUSSIER J.G.; PATE J.L. Molecular characterization and expression of DERL1 in bovine ovarian follicles and corpora lutea. **Reproductive Biology and Endocrinology**, v.8, 2010.

PEMBERTON T.J; ABSHER D.; FELDMAN M.W.; MYERS R.M.; ROSENBERG N.A.; LI J.Z. Genomic Patterns of Homozygosity in Worldwide Human Populations. **The American Journal of Human Genetics**, v. 91, p. 275–292, 2012.

PERIPOLLI E.; MUNARI D.P.; SILVA M.V.G.B.; LIMA A.L.F.; IRGANG R.; BALDI F. Runs of homozygosity: Current Knowledge and application in livestock. **Animal Genetics**. Disponível em: < http://dx.doi.org/10.1111/age.12526>.

POLLAK M.N.; SCHERNHAMMER E.S.; HANKINSON S.E. Insulin-like growth factors and neoplasia. **Nature Reviews Cancer**, v. 4, p. 505–518, 2004.

PRYCE J.E.; HAILE-MARIAM M.; GODDARD M.E; HAYES B.J. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. **Genetics Selection Evolution**, v. 46, 2014.

PURCELL S.; NEALE B.; TODD-BROWN K.; THOMAS L., FERREIRA M.A.; BENDER D., MALLER J.; SKLAR P.; DE 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**, v. 81, p. 559-575, 2007.

PURFIELD D.C.; BERRY D.; MCPARLAND S.; BRADLEY D.G. Runs of homozygosity and population history in cattle. **BMC Genetics**, v. 13, 2012.

QUEIROZ S.A.; LÔBO R.B. Genetic relationship, inbreeding and generation interval in registered Gir cattle in Brazil. **Journal of Animal Breeding and Genetics**, p. 228-233, 1993.

REIS FILHO J.C.; VERNEQUE R.S.; TORRES R.A; LOPES P.S.; RAIDAN F.S.S; TORAL F.L.B. Inbreeding on productive and reproductive traits of dairy Gyr cattle. **Revista Brasileira de Zootecnia**, v. 44, supl. 5, p. 174-179, 2015.

RINCON G.; ISLAS-TREJO A.; CASTILLO A.R.; BAUMAN D.E.; GERMAN B.J.; MEDRANO J.F. Polymorphisms in genes in the SREBP1 signalling pathway and SCD are associated with milk fatty acid composition in Holstein cattle. **Journal of Dairy Research**, v. 79, 2012.

SANTIAGO A.A.: **O Zebu na Índia, no Brasil e no mundo**. Instituto Campineiro de ensino agrícola: Campinas, 1986.

SARGOLZAEI M.; CHESNAIS J.P.; SCHENKEL F.F. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, v. 15, 2014.

SAURA M.; FERNÁNDEZ A.; VARONA L.; FERNÁNDEZ A.I.; DE CARA M.A.R.; BARRAGÁN C.; VILLANUEVA B. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. **Genetics Selection Evolution**, v. 47, 2015.

SCHENNINK A.; STOOP W.M.; VISKER M.H.P.W.; HECK J.M.L.; BOVENHUIS H.; VAN DER POEL J.J.; VAN VALENBERG H.J.F.; VAN ARENDONK J.A.M. DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. **Animal Genetics**, v. 38, p. 467-473, 2007.

SCHNABEL R.D.; KIM J-J.; ASHWELL M.S.; SONSTEGARD T.S.; VAN TASSELL C.P.; CONNOR E.E.; TAYLOR J.F. Fine-mapping milk production quantitative trait loci on BTA6: Analysis of the bovine osteopontin gene. **Proceedings of the National Academy of Sciences**, v. 102, 2005.

SCHOPEN G.C.B.; VISKER M.H.P.W.; KOKS P.D.; MULLAART E.; VAN ARENDONK J.A.M.; BOVENHUIS H. Whole-genome association study for milk protein composition in dairy cattle. **Journal of Dairy Science**, v. 94, p. 3148-3158, 2011.

SØRENSEN J.G.;KRISTENSEN T.N.; LOESCHCKE V. The evolutionary and ecological role of heat shock proteins. **Ecology Letters**, v. 6, p. 1025–1037, 2003.

THALLER G.; KRAMER W.; WINTER A.; KAUPE B.; ERHARDT B.; FRIES R. Effects of DGAT1 variants on milk production traits in German cattle breeds. **Journal of Animal Science**, v. 81, p. 1911-1918, 2003.

TUCKER H.A. Hormones, mammary growth, and lactation: A 41-year perspective. **Journal of Dairy Science**, v. 83, p. 874–884, 2000.

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

VANRADEN P.M.; OLSON K.M.; WIGGANS G.R.; COLE J.B; TOOKER M.E. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. **Journal of Dairy Science**, v. 94, p. 5673–5680, 2011.

VISSCHER P.M.; MEDLAND S.E.; FERREIRA M.A.R; MORLEY K.I.; ZHU G.; CORNES B.K.; MONTGOMERY G.W.; MARTIN N.G. Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. **PLoS Genetics**, v. 2, e41, 2006.

WANG Y. **Genetic and Geographic Diversity of Gyr (*Bos Indicus*) Cattle in Brazil**. Dissertação (Major thesis animal breeding and genetics) - University of Natural Resources and Life Sciences, Vienna, 2015.

WANG X.; MA P.; LIU J.; ZHANG Q, ZHANG Y.; DING X.; JIANG L.; WANG Y.; ZHANG Y.; SUN D.; ZHANG S.; SU G.; YU Y. Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. **BMC Genetics**, v. 16, 111, 2015.

WATSON C.J. Stat transcription factors in mammary gland development and tumorigenesis. **Journal of Mammary Gland Biology and Neoplasia**, v. 6, p. 115–127, 2001.

WINTER A.; KRAMER W.; WERNER F.A.O.; KOLLERS S.; KATA S.; DURSTEWITZ G.; BUITKAMP J.; WOMACK J.E.; THALLER G.;FRIES R. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. **Proceedings of the National Academy of Sciences**, v. 99, 2002.

WRIGHT S. Coefficients of inbreeding and relationship. **The American Naturalist**, v. 56, p. 330–338, 1922.

WU C.; LIU L.; HUO J.; LI D.; YUAN Y.; YUAN F.; MIAO Y. Isolation, sequence characterization, and tissue transcription profiles of two novel buffalo genes: INSIG1 and INSIG2. **Tropical Animal Health and Production**, v. 46, p. 33-41, 2014.

XIAO-YING D.; SHENG-QIU T. Insulin-induced gene: A new regulator in lipid metabolism. **Peptides**, v. 31, p. 2145-2150, 2010.

XIE Y.; YANG S.; CUI X.; JIANG L.; ZHANG S.; ZHANG Q.; ZHANG Y.; SUN D. Identification and expression pattern of two novel alternative splicing variants of EEF1D gene of dairy cattle. **Gene**, v. 534, p. 189-196, 2014.

ZAVAREZ L.B.; UTSUNOMIYA Y.T.; CARMO A.S.; NEVES H.H.R.; CARVALHEIRO R.; FERENČAKOVIĆ M.; O'BRIEN A.M.P.; CURIK I.; COLE J.B.; VAN TASSELL C.P.; DA SILVA M.V.G.B.; SONSTEGARD T.S.; SÖLKNER J.; GARCIA J.F. Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. **Frontiers in Genetics**, v. 6, 2015.

ZHANG Q.; GULDBRANDTSEN B.; BOSSE M.; LUND M.S.; SAHANA G. Runs of homozygosity and distribution of functional variants in the cattle genome. **BMC Genomics,** v. 15, 2015a.

ZHANG Q.; CALUS 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**, v. 16, 2015b.

APPENDIX

Appendix A - Gene content inside runs of homozygosity (ROH) overlapping regions

BTA[1]	ROH Frequency	Physical Position (bp)	Length (bp)	Genes content
2	0.755	68,748,659:89,063,900	20,315,241	*CCDC93, INSIG2, DBI, SCTR, CFAP221, PTPN4, EPB41L5, RALB, INHBB, GLI2, TFCP2L1, CLASP1, TSN, CNTNAP5, GYPC, STAT1, STAT4, MYO1B, TMEFF2, SLC39A10, DNAH7, CCDC150, PGAP1, SF3B1, HSPD1, HSPE1, MARS2, BOLL, PLCL1, SATB2*
2	0.569	68,748,659:104,825,968	36,077,309	*DBI, SCTR, INHBB, GLI2, GYPC, STAT1, STAT4, HSPD1, HSPE1, BOLL, AOX1, NDUFB3, CFLAR, CASP8, FZD7, SUMO1, BMPR2, CD28, CTLA4, NDUFS1, CREB1, CRYGD, CRYGC, CRYGB, IDH1, MAP2, MYL1, LANCL1, FN1*
2	0.924	78,394,916:87,587,063	9,192,147	*GYPC, LOC100295717, GLS, STAT1, STAT4, TRNAC-GCA, MYO1B, TRNAE-UUC, NABP1, SDPR, TMEFF2, LOC785710, SLC39A10, DNAH7, STK17B, LOC531691, CCDC150, GTF3C3, C2H2orf66, PGAP1, ANKRD44, SF3B1, COQ10B, HSPD1, HSPE1, MOB4, RFTN2, MARS2, BOLL, PLCL1*
2	0.781	81,983,121:87,587,063	5,603,942	*LOC100138726, LOC781256, SLC39A10, LOC104971271, DNAH7, STK17B, LOC101906937, LOC531691, CCDC150, GTF3C3, C2H2orf66, PGAP1, LOC101907169, ANKRD44, LOC104971272, LOC104971273, LOC782417, LOC504995, SF3B1, COQ10B, HSPD1, HSPE1, MOB4, RFTN2, MARS2, BOLL, PLCL1, LOC104971274, LOC104971275*
6	0.638	58,133,150:59,323,454	1,190,304	*NWD2, LOC101906734, C6H4orf19, TRNAC-GCA, RELL1, LOC783826, PGM2, LOC783708, LOC781379, LOC101906872, TBC1D1, LOC104972736, LOC104972737, LOC104972738, LOC101907152,*
6	0.737	62,281,712:81,603,050	19,321,338	*ATP8A1, KCTD8, GABRG1, GABRA2, GABRA4, GABRB1, ATP10D, NFXL1, CNGA1, ZAR1, OCIAD1, CWH43, SPATA18, RASL11B, FIP1L1, LNX1, CHIC2, KIT, KDR, SRD5A3, CLOCK, CEP135, KIAA1211, PPAT, PAICS, HOPX, POLR2B, IGFBP7, ADGRL3, TECRL*
6	0.887	68,338,834:73,220,200	4,881,366	*CNGA1, NIPAL1, TXK, TEC, SLAIN2, ZAR1, FRYL, OCIAD1, OCIAD2, CWH43, LRRC66, SGCB, SPATA18, USP46, RASL11B, SCFD2, FIP1L1, LNX1 CHIC2, PDGFRA, KIT, KDR, SRD5A3, TMEM165,, CLOCK, PDCL2, NMU, EXOC1, CEP135, KIAA1211*
6	0.587	70,117,799:81,603,050	11,485,251	*RASL11B, SCFD2, FIP1L1, LNX1, CHIC2, GSX2, PDGFRA, KIT, KDR, SRD5A3, TMEM165, CLOCK, PDCL2, EXOC1, NMU, CEP135, KIAA1211, AASDH, PPAT, PAICS, SRP72, ARL9, THEGL, HOPX, SPINK2, NOA1, POLR2B, IGFBP7, ADGRL3, TECRL*
10	0.518	5,133,564:8,452,227	3,318,663	*THOC3, CPLX2, LOC104973021, HRH2, LOC104973023, SFXN1, DRD1, GCNT4, LOC104973024, LOC104973025, LOC104970646, ANKRD31, POLK, TRNAY-*

				GUA, HMGCR, COL4A3BP, ANKDD1B, POC5, SV2C, IQGAP2, F2RL2, F2R, LOC100296562, F2RL1, S100Z, CRHBP, AGGF1, LOC781720, ZBED3, PDE8B
10	0.529	25,895,397:27,374,489	1,479,092	*RPGRIP1, HNRNPC, ZNF219, ARHGEF40, RNASE13, TPPP2, NDRG2, SLC39A2, METTL17, RNASE2, RNASE1, BRB, RNASE1, RNASE6, RNASE4, ANG, ANG2, RNASE12, RNASE11, RNASE10, PNP, TMEM55B, APEX1, OSGEP, KLHL33, CCNB1IP1, TTC5, OR11H4, OR4N4, OR4K14*
12	0.720	86,889,033:89,989,632	3,100,599	*LOC781180, FAM155A, LIG4, ABHD13, TNFSF13B, MYO16, TRNAY-AUA, LOC104973669, LOC104973672, LOC101905776, LOC101905821, LOC104973673, IRS2, LOC104973674, COL4A1, LOC104973675, COL4A2, LOC104973676, LOC104973677, RAB20, NAXD, CARS2, ANKRD10, ING1, LOC784176, LOC104973678, ARHGEF7, TEX29, LOC101906959*
12	0.606	86,889,033:89,992,862	3,103,829	*LOC781180, FAM155A, LIG4, ABHD13, TNFSF13B, MYO16, CARS2, TRNAY-AUA, LOC104973669, LOC104973672, LOC101905776, LOC101905821, LOC104973673, IRS2, LOC104973674, COL4A1, LOC104973675, COL4A2, LOC104973676, LOC104973677, RAB20, NAXD, ANKRD10, ING1, LOC784176, ARHGEF7, TEX29, LOC101906959, LOC104973678*
14	0.778	37,250,059:42,032,707	4,782,648	*EYA1, TRNAC-ACA, MSC, TRPA1, MIR1603, KCNB2, TERF1, SBSPON, LOC101906455, C14H8orf89, RPL7, RDH10, TRNAG-UCC, MIR2284L, STAU2, UBE2W, TCEB1, TMEM70, LY96, TRNAE-UUC, LOC104974051, JPH1, GDAP1, LOC104974053, LOC104974054, PI15, CRISPLD1, HNF4G, LOC104974057, ZFHX4*
14	0.505	39,495,608:41,685,719	2,190,111	*SLC39A4, CPSF1, DGAT1, HSF1, CYC1, EEF1D, CYP11B1, PTK2, TRAPPC9, TG, MYC, DERL1, HAS2, CEBPD, PRKDC, ATP6V1H, RGS20, MOS, PLAG1, PENK, RAB2A, ASPH, CRH, MYBL1, ARFGEF1,*

[1] *Bos Taurus* Autosome (BTA) that presented a frequency of overlapping ROH shared by more than 50% of the samples.