UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

IDENTIFICATION OF STRUCTURAL VARIANTS AND SELECTION SIGNATURES IN CATTLE

Elisa Peripolli

Zootecnista

2021

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL

IDENTIFICATION OF STRUCTURAL VARIANTS AND SELECTION SIGNATURES IN CATTLE

Discente: Elisa Peripolli
Orientador: Prof. Dr. Fernando Sebastián Baldi Rey
Coorientador: Dr. Marcos Vinícius Gualberto Barbosa da Silva

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

2021

P445i

Peripolli, Elisa

Identification of structural variants and selection signatures in cattle / Elisa Peripolli. -- Jaboticabal, 2021

320 p.

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

Orientador: Fernando Sebastián Baldi Rey

Coorientador: Marcos Vinícius Gualberto Barbosa da Silva

1. Adaptação. 2. Bos taurus indicus. 3. Bos taurus taurus. 4. Recursos genéticos. 5. Sequenciamento. I. Título.

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

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: IDENTIFICATION OF STRUCTURAL VARIANTS AND SELECTION SIGNATURES IN CATTLE

AUTORA: ELISA PERIPOLLI
ORIENTADOR: FERNANDO SEBASTIAN BALDI REY
COORIENTADOR: MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA

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

Prof.Dr. FERNANDO SEBASTIAN BALDI REY (Participaçao Virtual)
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Pesquisadora Dra. NEDENIA BONVINO STAFUZZA (Participaçao Virtual)
Instituto de Zootecnia / Sertãozinho/SP

Prof. Dr. DANISIO PRADO MUNARI (Participaçao Virtual)
Departamento de Engenharia e Ciências Exatas (DECEx) / FCAV / Unesp - Jaboticabal

Pós-doutorando RAFAEL ESPIGOLAN (Participaçao Virtual)
FZEA/USP / Pirassununga/SP

Pós-doutoranda MARIANA PIATTO BERTON (Participaçao Virtual)
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Jaboticabal, 02 de fevereiro de 2021

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. Obteve o título de mestre em 17 de fevereiro de 2017. Em março de 2017, ingressou no curso de doutorado no mesmo programa de Pós-graduação da mesma instituição de fomento e sob mesma orientação, bolsista da Fundação de Amparo à Pesquisa do Estado de São Paulo (Processo FAPESP 2016/24084-7). Entre o período de setembro de 2018 a agosto de 2019, fez estágio de pesquisa no exterior na University of Goettingen (Alemanha), com auxílio da Fundação de Amparo à Pesquisa do Estado de São Paulo (Processo FAPESP 2017/27148-9), sob orientação do Prof. Dr. Henner Simianer. Obteve o título de Doutora em 02 de fevereiro de 2021.

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

Santa Madre Paulina

“*Tenha paciência para ver seus galhos se transformarem em flores*”

O pequeno mestre

Dedico

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

Ao meu esposo, Hugo Borges de Quadros, por ser meu melhor amigo e companheiro de vida, te amo!

Aos meus tios, Fernando Christiano Zimmermann (*in memoriam*) e Luiz Carlos de Oliveira Telles, por serem **amor**.

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, por todo o amor e momentos maravilhosos que ficarão para sempre no meu coração. Vocês são e sempre serão meus melhores amigos.

Ao meu esposo, 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!

Ao meu orientador Prof. Dr. Fernando Sebastián Baldi Rey por me receber tão bem no seu grupo de pesquisa e pela confiança em mim depositada na execução desse e de outros trabalhos durante todo o período da pós-graduação. Além de um excelente orientador, é um grande amigo e conselheiro. Obrigada por ouvir meus desabafos pessoais e profissionais e por sempre confiar no meu potencial.

Ao meu co-orientador, Dr. Marcos Vinícius Gualberto Barbosa da Silva, pelo auxílio e prontidão no decorrer do desenvolvimento deste trabalho. À Embrapa Gado de leite (Juiz de Fora - MG) pelo fornecimento dos dados.

Ao programa de melhoramento Montana Composto Tropical® e a Associação Nacional de Criadores e Pesquisadores (ANCP) pelo fornecimento dos bancos de dados.

À Universidade de Göttingen, em especial ao Dr. Henner Simianer e Dr. Christian Reimer, por terem me recebido de portas abertas e por todo o suporte durante o período de doutorado sanduíche na Alemanha.

Ao Washington Luiz Olivato Assagra (Tom), por me ajudar com os problemas computacionais e com os scripts para montar os bancos de dados e rodar as análises. Pela prontidão em sempre me ajudar com os ‘pepinos’ do servidor e por nunca hesitar em compartilhar seu vasto conhecimento em bioinformática e programação comigo.

Aos meus *hermanitos*, Bianca, Fabi, Tonussi, Mari, Sabrina(s), Hermenegildo e Juan Diego por tornarem a salinha divertida e as conversas produtivas quando não estávamos produtivos. À todos os amigos do cafezinho e aos demais alunos do programa de Genética e Melhoramento Animal – UNESP/FCAV.

Às minhas *schatzis* do coração, Géssica, Pri, Karol e Jacke, pela amizade verdadeira e por todas as risadas e momentos que compartilhamos juntas. Por compreenderem minha ausência e por sempre torcerem por mim.

À minha amiga Heloísa Fidelis, por me ouvir e principalmente me aconselhar nos meus últimos meses em Jaboticabal. Amiga, muito obrigada pela nossa amizade, conversas, risadas e gordices.

Aos membros da banca de defesa de tese de doutorado, Dra. Nedenia Bonvino Stafuzza, Dr. Danisio Prado Munari, Dr. Rafael Espigolan e Dra. Mariana Piatto Berton, pelas sugestões que muito contribuíram e acrescentaram a esse trabalho e pela disponibilidade de compor minha banca.

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

À Fundação de Amparo à pesquisa do Estado de São Paulo (FAPESP) pela concessão de bolsa de doutorado (Processo FAPESP 2016/24084-7) e da Bolsa de Estágio e Pesquisa no Exterior (BEPE-DR, Processo FAPESP 2017/27148-9).

À todos que foram importantes em algum momento, mas minha memória não permitiu citar.

SUMÁRIO

Resumo v
Abstract vii
CAPÍTULO 1 – General considerations 9
Introduction 9
Objectives 11
Literature Review 11
Cattle introduction in Brazil and the locally adapted breeds development 11
Pantaneiro 13
Crioulo Lageano 13
Caracu Caldeano 14
Indicine cattle breeds 14
Composite Montana Tropical® beef cattle 15
Runs of homozygosity 15
Selection Signatures 18
Copy number variation 20
References 22
CAPÍTULO 2 – Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits[1] 32
Abstract 32
Introduction 33
Results 35
Genome-wide distribution of Runs of homozygosity 35
Pedigree and genomic inbreeding 36
Autozygosity islands in Nellore lineages 37
Functional annotation of genes 38
Discussion 39
Genome-wide distribution of runs of homozygosity 39
Pedigree and genomic inbreeding 40
Autozygosity islands in Nellore lineages 42
Functional annotation of genes 45
Final considerations 46
Methods 47

Animals and genotyping 47
Runs of homozygosity 48
Pedigree and genomic inbreeding coefficients 49
Identification and gene prospection in autozygosity islands 50
References 50
Figures 62
CAPÍTULO 3 – Genome-wide scan for runs of homozygosity in composite Montana Tropical® beef cattle[1] 66
Abstract 66
Introduction 67
Material and Methods 69
Samples, genotyping and data editing 69
Effective population size 70
Pedigree-based inbreeding coefficient 71
Runs of homozygosity 71
Detection of autozygosity islands 72
Gene searching and functional annotation analysis 73
Results and Discussion 73
Effective population size 73
Distribution of runs of homozygosity 74
Autozygosity islands 77
Functional annotation of genes 77
Final Considerations 81
References 82
Figures 91
CAPÍTULO 4 – Genome-wide detection of signatures of selection in indicine and Brazilian locally adapted taurine cattle breeds using whole-genome re-sequencing data[1] 94
Abstract 94
Introduction 95
Results 96
Data 96
Variant annotation and enrichment 96
Population structure 97
Genomic inbreeding 98

Selective sweeps 98
Selective sweeps and runs of homozygosity 99
Overlap with candidate regions under positive selection in other cattle populations 101
Discussion 101
Population structure 101
Genomic inbreeding 103
Candidate regions under positive selection 104
Overlap with candidate regions under positive selection in other cattle populations 107
Final Considerations 108
Methods 108
Samples, sequencing, and raw data preparation 108
Variant annotation and predicted functional impacts 110
Population differentiation analysis 110
Genomic inbreeding coefficient estimation 111
Selective sweeps detection 111
Selective sweeps and runs of homozygosity 113
Functional annotation of the candidate regions 113
References 114
Figures 127
CAPÍTULO 5 – Assessment of copy number variants in three Brazilian locally adapted cattle breeds using whole-genome re-sequencing data 130
Abstract 130
Introduction 131
Material and methods 133
Samples, sequencing, and raw data preparation 133
CNVs and CNVRs detection 133
Variant annotation and predicted functional impacts 134
Functional annotation 134
Results 134
Data 134
CNV and CNVRs discovery 135
Variant annotation and gene assessment 136
Functional annotation of genes 137

CNVRs and overlapping QTLs in cattle 138
Discussion 139
CNVs and CNVRs discovery 139
Variant and functional annotation of genes 140
CNVRs and overlapping QTLs in cattle 143
Final Considerations 143
References 144
Figures 151
CAPÍTULO 6 – FINAL CONSIDERATIONS 156
APPENDIX A 158
APPENDIX B 185
APPENDIX C 193
APPENDIX D 272

IDENTIFICAÇÃO DE VARIAÇÕES ESTRUTURAIS E ASSINATURAS DE SELEÇÃO EM BOVINOS

RESUMO - Devido aos impactos causados na produção animal recorrentes das mudanças climáticas, é importante caracterizar o genoma bovino para desvendar os mecanismos genéticos envolvidos na variação fenotípica que foram influenciados pelo ambiente e moldados pela seleção natural. O objetivo deste estudo é descrever os principais efeitos da adaptação e seleção em animais zebuínos e taurinos localmente adaptadas através da identificação de variações estruturais e assinaturas de seleção utilizando dados genotípicos e de sequenciamento de genoma inteiro. No capítulo 2, foram utilizados genótipos imputados (*n=735.044* marcadores) de 9,386 animais da raça Nellore e de suas respectivas linhagens a fim de estimar a autozigosidade do genoma baseado nas corridas de homozigose (ROH) por meio do software Plink. Em geral, os coeficientes de endogamia baseados em ROH (F_{ROH}) não foram altos, com valores próximos a 2%. As ilhas de autozigosidade foram evidentes em todo o genoma e sua localização não diferiu em grande número dentro das linhagens. Termos enriquecidos ($p<0,01$) dentro das ilhas de autozigosidade sugeriam uma forte seleção para características relacionadas à resposta imune, podendo explicar uma maior adaptabilidade do gado zebuíno em ambientes severos. O capítulo 3 visou avaliar a autozigosidade de todo o genoma para explorar regiões ricas em ROH que poderiam melhor caracterizar os diferentes tipos biológicos (produtivo ou adaptativo) do gado de corte composto Montana Tropical®. Animais Montana (*n=1.436*) foram genotipados com o GGP-LD BeadChip (*n=30.105* marcadores) e os ROH foram identificados em cada indivíduo usando o software Plink. O número de ilhas de autozigosidade não diferiu consideravelmente entre os tipos biológicos e não foi encontrado nenhum termo enriquecido significativo ($p<0,05$) compartilhado entre eles. Termos enriquecidos associados à resposta imunológica e homeostase foram descritos para o tipo biológico adaptativo, enquanto aqueles ligados ao sistema imunológico, bem como às funções reprodutivas e produtivas, foram identificados para o tipo biológico produtivo. No capítulo 4, quatro métodos estatísticos foram implementados para detectar regiões genômicas sob pressão seletiva usando dados de sequenciamento de genoma inteiro (~12.4 X) de bovinos das raças Gir (GIR, *n=13*), Caracu Caldeano (CAR, *n=12*), Crioulo Lageano (CRL, *n=12*) e Pantaneiro (PAN, *n=12*). As estatísticas dentro de população (CLR e iHS) e entre populações (F_{ST} e XPEHH) foram combinadas separadamente em um único valor por meio do método '*de-correlated composite of multiple signals*' (DCMS). As regiões de varredura seletiva foram identificadas por meio dos valores do limite superior (1%) da distribuição empírica gerada por cada estatística DCMS. As assinaturas de seleção identificadas forneceram uma percepção abrangente de genes candidatos juntamente com QTLs relacionadas a características produtivas e de adaptação ao ambiente hostil no qual estas raças foram expostas. No capítulo 5, o método de leitura baseada em '*read-depth*' implementado no software CNVnator foi utilizado para identificar variações no número de cópias (CNVs) utilizando dados de sequenciamento de genoma inteiro (~14.07 X) de bovinos das raças CAR (*n=12*), CRL (*n=12*) e PAN (*n=12*). Regiões de CNV (CNVRs) foram identificadas sobrepondo as CNVs individuais dentro de cada raça. A anotação

funcional das CNVRs revelou variantes com elevada consequência na sequência proteica abrangendo genes fortemente associados a resiliência ambiental, dentre os quais podemos destacar o *BOLA-DQB*, *BOLA-DQA5*, *CD1A*, *β-defensins*, *PRG3* e *ULBP21*. A análise de enriquecimento funcional utilizando os genes prospectados nas CNVRs também revelou termos significativos ($p<0.01$) fortemente associados à imunidade e resistência do gado a ambientes severos. Nossos resultados elucidaram os mecanismos biológicos inerentes as raças bovinas aqui estudadas, fornecendo informações a respeito de genes candidatos e regiões genômicas que abrangem características adaptativas relevantes, bem como informações úteis para futuras abordagens de conservação, estudos de associação ou seleção.

Key-words: Adaptação, *Bos taurus indicus*, *Bos taurus taurus*, recursos genéticos, sequenciamento de nova geração, varreduras de seleção

IDENTIFICATION OF STRUCTURAL VARIANTS AND SELECTION SIGNATURES IN CATTLE

ABSTRACT – Given the impacts caused by climate change upon livestock production, it is important to characterize the cattle genome to unravel the genetic mechanisms underlying phenotypic variation that were influenced by the environment and shaped by natural selection that allowed them to thrive in distinct ecosystems. Therefore, the objective of this study is to describe the main effects of adaptation and selection in indicine and locally adapted taurine cattle breeds through the identification of structural variants and signatures of selection using genotypic and whole-genome re-sequencing data. In chapter 2, imputed genotypes (*n=735,044* markers) were used to assess genome-wide autozygosity based on runs of homozygosity (ROH) in 9,386 Nellore animals and its lineages using the Plink software. Overall, inbreeding coefficients based on ROH (F_{ROH}) were not high, with values close to 2%. Autozygosity islands were evident across the genome, and their genomic location did not largely differ within lineages. Enriched terms ($p<0.01$) within the autozygosity islands suggested a strong selection for immune response-related traits and might explain the greater adaptability of the indicine cattle in harsh environments. Chapter 3 aimed to assess genome-wide autozygosity to explore ROH hotspot regions which could better characterize the different biological types (productive or adaptive) within the composite Montana Tropical® beef cattle. Montana animals (*n=1,436*) were genotyped with the GGP-LD BeadChip (*n=30,105* markers), and ROH were identified in every individual using the Plink software. The number of autozygosity islands did not differ considerably between biological types, and no significant enriched term ($p<0.05$) was found to be shared between them. Enriched terms associated with the immune response and homeostasis were described for the adaptive biological type, while those linked to the immune system as well as with reproductive and productive functions we identified for the productive biological type. In chapter 4, four statistical methods were implemented to detect genomic regions under selective pressure using whole-genome re-sequencing data from Gir (GIR, *n=13*), Caracu Caldeano (CAR, *n=12*), Crioulo Lageano (CRL, *n=12*), and Pantaneiro (PAN, *n=12*) cattle breeds. Within-population (CLR and iHS) and cross-population statistics (F_{ST} and XPEHH) were combined separately in a single score using the de-correlated composite of multiple signals (DCMS) method, and putative sweep regions were revealed by assessing the top 1% of the empirical distribution generated by each DCMS statistic. The signatures of selection identified herein provided a comprehensive set of putative candidate genes together with QTLs disclosing cattle production traits and adaptation to the challenging environment in which these breeds have been exposed. In chapter 5, the read depth-based method implemented in CNVnator was used for copy number variants (CNV) calling on re-sequenced data (~14.07 X) from CAR (*n=12*), CRL (*n=12*), and PAN (*n=12*) cattle breeds. CNV regions (CNVRs) were identified by overlapping individual CNVs within each breed. The functional annotation of the CNVRs revealed variants with high consequence on protein sequence harboring relevant genes with functions strongly linked to environmental resilience (i.e., *BOLA-DQB, BOLA-DQA5*, *CD1A*, β-defensins, *PRG3*, and *ULBP21*). Enrichment analysis based on the gene list retrieved from the

CNVRs also disclosed over-represented terms ($p<0.01$) greatly associated with immunity and cattle resistance to harsh environments. Our findings improve the knowledge about the genome biology of such cattle breeds and provide candidate genes and genomic regions encompassing relevant traits as well as useful information for future conservation, association, or selection approaches.

Key-words: Adaptation, *Bos taurus indicus*, *Bos taurus taurus*, footprints, genetic resource, next-generation sequencing

CAPÍTULO 1 – GENERAL CONSIDERATIONS

INTRODUCTION

Great changes have occurred in livestock production systems over the last century with the advent of the agricultural industrialization, specialization, mechanization, and globalization. As production systems have evolved, the strong focus on high-yielding breeds have led to a considerable decline in the diversity of many local cattle breeds (MARSONER et al., 2017). High-specialized breeds have become increasingly preffered and largely kept given their production traits, leading to a progressive replacement of traditional multipurpose and/or locally adapated cattle breeds (UGARTE et al., 2001; ZANDER et al., 2013). In recent decades, it is noteworthy that great efforts have been made to improve our knowledge of locally adapted breeds worldwide, and a number of studies related to the economic valuation of these cattle breeds have been carried out in countries where the use of such breeds are particularly important (ZANDER et al., 2013).

Brazil is characterized by a set of ecosystems and biomes, i.e., Amazon rainforest, Cerrado, Mata Atlântica, Caatinga, Pampa and Pantanal, each one with its own particularities. According to Egito et al. (2007), natural selection acting in a remarkably variable set of ecosystems throughout the country together with breed admixture events allowed the development of locally adapted breeds in a wide range of environments, i.e., Curraleiro Pé Duro, Junqueira, Franqueiro, Caracu, Mocho Nacional, Crioulo Lageano, and Pantaneiro. In addition to these locally adapted breeds, the indicine cattle imported from India and the composite beef cattle breeds raised in Brazil are noteworthy to be highlighted given their remarkable adaptation upon tropical and subtropical environments, playing a key role in the Brazilian cattle production systems. These breeds have shown outstanding levels of phenotypic variability and improved fitness to local conditions.

It is important to assure that animal genetic resources will match with the production environments in which they are kept and that the genetic diversity needed to adapt production systems to future changes will be maintained, requiring adjustments to husbandry and production strategies. In this sense, production systems

in which non locally adapted livestock have been introduced may be vulnerable to direct and indirect effects of response to changing conditions. One strategy for adapting production systems to these effects is the introduction of animals better adapted to local conditions, adapting livestock production systems and maintaining the genetic diversity (PILLING; HOFFMANN, 2011). According to Hoffmann (2013), adaptation is necessary to respond adequately to environmental change, and a better characterization of locally adapted breeds will be the key for maintaning genetic resources in these regions. Characterizing locally adapted breeds at a genome-wide level is a powerful tool for creating a germplasm bank and a reservoir of livestock genetic diversity resource upon environmental change and adaptation (PILLING; HOFFMANN, 2011).

Domestication and subsequent natural/artificial selection together with the evolutionary adaptive process in cattle not only have changed the allelic frequencies at causal single nucleotide polymorphism (SNP) over time, but also the surrounding genomic regions due to the hitchhiking effect (SMITH; HAIGH, 1974). The development of large-scale catalogs of genetic variation has stimulated the interest in identifying genomic footprints within the genome of modern cattle, helping us to clarify the roles of selection in the evolutionary processes (BISWAS; AKEY, 2006). A more comprehensive and genomic understanding of how selection has shaped the patterns of genetic variation may provide important insights into the mechanisms of evolutionary change (OTTO, 2000) and facilitate the annotation of significant genomic regions (NIELSEN, 2001). With a better understanding of the genetic difference between breed types, locally adapted breeds can be an important source of genetic information leading to the discovery and validation of genomic regions and DNA variants controlling important traits.

Advances in molecular genetics, genomics, and bioinformatics allowed using high-density arrays and complete DNA sequences for studying the effects of natural and artificial selection in the genome of livestock. Whole-genome sequences are potentially the richest source of genetic data and represent an unprecedented resource to disentangle the genetic architecture of complex traits in cattle. Studies from sequence data can be used to further catalogue large amounts of signatures of selection and genetic variation, allowing to create new datasets to be accurately used

in the discovery of novel single nucleotide polymorphisms (SNPs) (BARRIS et al., 2012; CHOI et al., 2013), in the prospection of genomic key regions related to loss of function (DAS et al., 2015) and adaptation (DAETWYLER et al., 2014; LIAO et al., 2013), and in the identification of structural variations in the bovine genome associated with productive traits (CHOI et al., 2014, 2016; HOU et al., 2012; YUE et al., 2014).

Despite the recent achievements in high-throughput genotyping and re-sequencing, there is still a drastic shortage of studies for less notorious and locally adapted breeds. Very little is known about the genetic composition and the importance of such breeds to a wide range of environments. An understanding of the extent and pattern of genetic variability among breeds may help in the development of more rational breeding programs.

Objectives

The objective of this study is to describe the main effects of selection/adaptation in indicine and Brazilian locally adapted taurine cattle breeds through the identification of structural variants and signatures of selection using genotype and high-throughput sequencing data.

LITERATURE REVIEW

Cattle introduction in Brazil and the locally adapted breeds development

The first cattle heads arrived in Brazil in 1534 (MARTINS et al., 2009; PRIMO, 1992) brought by Spanish and Portuguese conquerors during the Brazilian colonization period (MAZZA et al., 1994). The first cattle landed in the Southeast region through the harbor of São Vicente-São Paulo in the year of 1534 followed by other entries in the Northeast region (Pernambuco and Bahia states) in 1550 (MAZZA et al., 1994; PRIMO, 1992). The animals that arrived in São Vicente irradiated to the Southern fields, Goiás, São Francisco Valley (Minas and Bahia), and also to the Northeast region (Ceará and Piauí), whereas those that arrived in Pernambuco and Bahia states spread to the Northeast region, north of Minas and west of Bahia, and eventually, individuals from both populations may have found themselves (PRIMO, 1993).

The cattle introduced by the European conquerors were exposed to a process of natural selection for several generations (~400 years) in extremely variable environments throughout the country and facing all kinds of difficulties such as scarce food, diseases and parasites and strong weather without any significant selective pressure imposed by man (MARIANTE; CAVALCANTE, 2000a). The natural selection of these herds together with the recurring events of breed miscegenation led to the development of locally adapted cattle breeds with outstanding levels of phenotypic variability and better adapted to local conditions in a wide range of Brazilian environments. Hence, in the Northeast region, the Curraleiro cattle appeared and also spread to the central states of Minas Gerais and Goiás. In the Southeast region, the Junqueira and Franqueiro cattle were developed together with the Caracu and Mocho Nacional breeds. In the South region, the Crioulo Lageano was formed, and in the Pantanal region, the Pantaneiro cattle developed (EGITO, 2007). According to Brito (2013), these locally adapted cattle breeds could be considered isolated populations in a certain ecosystem or region exhibiting their own characteristics of acclimatization, i.e., rusticity and adaptation to adverse conditions and parasites, influenced by the environment and shaped by natural selection.

From the end of the nineteenth century and the beginning of the twentieth century, the search for more productive animals due to the emergent demand for food supply (products of animal origin) led to imports of exotic and more productive breeds of indicine origin (EGITO; MARIANTE; ALBUQUERQUE, 2002; MARIANTE et al., 1999). The animals imported in the last 50 to 100 years, although considered highly productive, lacked the fitness traits found in the local breeds (MARIANTE et al., 2009; SERRANO et al., 2004). Thus, the rapid growth of the commercial populations has occurred at the expense of a second group of locally adapted breeds through the intensive use of absorbent crossbreeds, and gradually they replaced the locally adapted breeds (MARIANTE; EGITO, 2002).

As a consequence of the economic and social changes since the arrival of the first conquerors, a progressive reduction as much in the number as in the geographic distribution of the locally adapted cattle breeds occurred to such an extent that now most of them are now in an advanced state of genetic dilution and threatened with the risk of extinction (EGITO; MARIANTE; ALBUQUERQUE, 2002; FELIX et al., 2013).

Nowadays, four out of five Brazilian locally adapted cattle breeds are in danger of extinction (Curraleiro Pé-Duro, Pantaneiro, Crioulo Lageano, and Mocho Nacional). The Caracu breed is an exception, and it can be considered as already established (FELIX et al., 2013; MARIANTE et al., 2008). Nowadays, the Brazilian Agricultural Research Corporation (EMBRAPA) through the National Research Centre for Genetic Resources (CENARGEN) retains a germplasm bank in order to avoid the genetic dilution and irreplaceable gene losses of the Brazilian locally adapted cattle breeds. The extinction of such breeds may lead to the loss of important traits of interest for production, while their use may mean an important alternative to improve the rusticity of commercial cattle breeds with high productivity, but with low adaptation capacity (EGITO et al., 2007; EGITO; MARIANTE; ALBUQUERQUE, 2002).

Pantaneiro

The Pantanal region is a sedimentary floodplain situated in the upper Paraguay river basin encompassing a large complexity of habitats and biodiversity. This ecosystem is affected by a variable climate and landscape, including the high incidence of solar radiation and thermic amplitude, the prevalence of parasites and predator, and the high-water level fluctuation which seasonally alters the food availability (MAZZA et al., 1992a, 1992b). The large expanses of land together with the lack of fences allowed the cattle brought by the conquers to freely reproduce and adapt to the ecological conditions of the Pantanal region. Hence, these cattle have undergone a natural selection and evolutionary process for more than four centuries, adapting themselves to the diverse ecological conditions of the Pantanal region and giving rise to the Pantaneiro breed (PRIMO, 1992). Through the long process of natural selection, they have acquired rusticity, high fertility rates and the ability to survive under food and water stress conditions (ISSA et al., 2009), playing a greater role in the economy of the Pantanal region until the beginning of the 20th century.

Crioulo Lageano

The Crioulo Lageano cattle have been selected in the Southern region of Brazil (Lages, Santa Catarina) for roughly four hundred years facing several adverse conditions such as acidic and rocky soil, high altitude, harsh winter with extremely low temperatures and frost, and poor vegetation (PRIMO, 1986). Because of all these environmental limitations, agriculture and cultivated pastures were limited in the region, however, the Crioulo Lageano cattle were able to thrive, being perfectly adapted to such ecological conditions of the region. Currently, the total population of the Crioulo Lageano breed is reduced to a herd that should not exceed 500 animals and more than 80% of the population belongs to a single breeder (MARIANTE; CAVALCANTE, 2000b).

Caracu Caldeano

Among the Brazilian locally adapted cattle breeds, the Caracu is the only one no longer in danger of extinction (FELIX et al., 2013; MARIANTE et al., 2008), with more than 85,500 registered animals throughout the country (ABCCaracu, http://www.abccaracu.com.br/NOVOCARACU/). This breed is widely used in crossbreeding, mainly with zebu cows, however, some animals were kept as purebred in the state of São Paulo and Minas Gerais. In the region of Poços de Caldas (Minas Gerais, Brazil), they have been selected for milk production, originating the Caracu Caldeano lineage. In the Experimental Station of Sertãozinho (São Paulo, Brazil), these animals have been the object of a study aiming to evaluate their potential for meat production (MCMANUS et al., 2010). Further, the natural selection has led to anatomical and physiological changes that have given them resistance to tropical environmental conditions, i.e., short coat, resistance to heat and parasites, good uprights and locomotion, resistant hooves for both hard and soaked soils, and the ability to digest rough fibers (KUES et al., 2006).

Indicine cattle breeds

The vast majority of the bovine based population reared for meat production in Brazil is composed mostly of indicine cattle (*Bos taurus indicus*), and among them, the

Nellore cattle have the largest number of animals (SANTIAGO, 1984). The Brazilian Nellore population is the result of less than 7,000 heads of purebred imported animals (BRASIL, 1978), and the major importation took place in 1962, when exceptional bulls were brought over the country and became founders of important lineages that were decisive to the great expansion of the Brazilian herd in the last 30 years (OLIVEIRA; MAGNABOSCO; BORGES, 2002).

Another indicine cattle breed that stood out in the tropics is the Gyr dairy cattle, which were imported to Brazil in 1912, and most of the bulls between 1914 and 1921 (SANTIAGO, 1986). Formerly, these animals were used in crossbreeding schemes for meat production, however, some breeders figured out outstanding animals for dairy-related traits, shifting their breeding objectives towards milk production. Gyr animals have been intensively used in tropical and subtropical regions as a basis for crossbreeding with European dairy breeds to produce a progeny with greater adaptability to hostile environmental conditions (QUEIROZ; LÔBO, 1993).

Composite Montana Tropical® beef cattle

The composite Montana Tropical® beef cattle were developed in 1994 for tropical and sub-tropical beef cattle systems under grazing conditions. The breed is centered on clusters defined by biological types combining physiology, growth, and reproductive and adaptive-related traits from *Bos taurus indicus* and *Bos taurus taurus* populations. Therefore, the base population is centered on four biological types defined as the NABC system, where: **N** are *Bos taurus indicus* cattle breeds; **A** are adapted *Bos taurus taurus* cattle breeds; **B** are *Bos taurus taurus* British breeds; and **C** are European Continental breeds. The composite Montana Tropical® beef cattle are notorious due to the greater carcass yield and meat quality traits, together with important adaptative and robustness traits (FERRAZ et al., 2002).

Runs of homozygosity

Runs of homozygosity (ROH) are continuous homozygous segments of the DNA sequence in diploid genomes (GIBSON; MORTON; COLLINS, 2006) which occurs when parents having a common ancestor pass shared chromosomal segments

identical by descent (IBD) on to their progeny (WRIGHT, 1922). This phenomenon results in inherited continuous IBD homozygous segments in the offspring's genome, characterized as ROH (Figure 1) (BROMAN; WEBER, 1999).

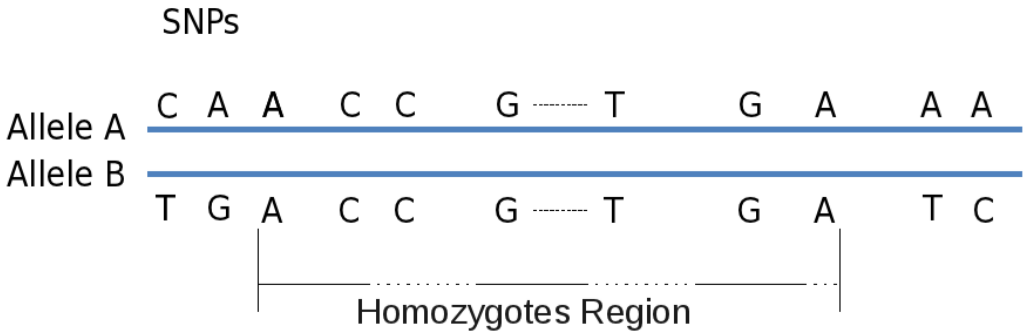


Figure 1: Illustration of a run of homozygosity (Adapted from: http://cancersincommon.herokuapp.com/page/roh)

ROH have been applied to quantifying individual autozygosity in several livestock species such as in chicken (FLEMING et al., 2016; MARCHESI et al., 2018), pig (SILIÓ et al., 2013; ZANELLA et al., 2016), goats (CARDOSO et al., 2018; ONZIMA et al., 2018), sheep (MASTRANGELO et al., 2018; PURFIELD et al., 2017) and cattle (MARRAS et al., 2014; PERIPOLLI et al., 2018a, 2018b; ZAVAREZ et al., 2015), given their high correlation (~0.7) (MCQUILLAN et al., 2008). Demographic events and population phenomena such as genetic drift, population bottleneck, inbreeding, and selection are known to have a strong influence on the occurrence of homozygosity throughout the genome (FALCONER; MACKAY, 1996). In this regard, the genomic footprint of these events at the DNA level enables the investigation of homozygosity patterns in the genome, disclosing how population history, structure, and demography have evolved over time (BERTOLINI et al., 2018; BOSSE et al., 2012; HERRERO-MEDRANO et al., 2013; PURFIELD et al., 2012). ROH can unwrap the genetic relationships among individuals, estimating with high accuracy the true level of autozygosity at the individual and population levels (CURIK; FERENČAKOVIĆ; SÖLKNER, 2014; FERENČAKOVIĆ et al., 2011, 2013).

Studies have considered the inbreeding coefficient traditionally estimated from pedigree data (F_{PED}) since Wright (1922), however, the availability of high-density SNP arrays led to an increasing interest in calculating the inbreeding coefficient from molecular information, such as those derived from ROH (F_{ROH}). As a result, the genomic information has introduced significant advances into the analyses of inbreeding coefficients, and ROH has been widely used as a predictor of whole genome inbreeding levels since they measure more accurately the relatedness among individuals and are not based on statistical expectations of the probable proportion of genomic IBD such as F_{PED} does (VISSCHER et al., 2006). Additionally, F_{ROH} also takes into account the stochastic nature of recombination and mutations loads (KELLER; VISSCHER; GODDARD, 2011). The advantages of F_{ROH} goes further in identifying IBD segments with greater accuracy. It is noteworthy to highpoint that the occurrence of ROH together with its extension can reveal the number of generations since the inbreeding event took place. This approach is possible due to the close correlation between the length of the ROH and the distance with the common ancestor due to recombination events, allowing the detection of autozygosity even 50 generations previously (HOWRIGAN; SIMONSON; KELLER, 2011; KELLER; VISSCHER; GODDARD, 2011).

Regions of the genome that have undergone selection pressure events can be unraveled by the identification and characterization of ROH (BOSSE et al., 2012; MASTRANGELO et al., 2017; PURFIELD et al., 2012; ZHANG et al., 2015) since selection is one of the main forces increasing overall autozygosity and printing continuous lengths of homozygous genotypes across the genomes (MARRAS et al., 2014). ROH patterns are not dispersed through the genome by chance (ZHANG et al., 2015), and genomic regions sharing these segments most likely harbor alleles associated with genetic selection (PURFIELD et al., 2012). The continuous search for elite/superior animals through the intense selection of sires has reduced the heterozygosity and genetic diversity around the target locus, leading to a high frequency of ROH, and consequently, generating autozygotic islands within these regions (LEOCARD, 2009; PEMBERTON et al., 2012). In this regard, the distribution of ROH patterns can be a useful tool to explore signatures of selection by informing the genomic regions that have been undergone to selective pressure over time.

Selection Signatures

The reduction in the genetic variation adjacent to a beneficial mutation is broadly referred as a selective sweep or signatures of selection, and it occurs as the result of natural or human-driven selection pressure altering the frequency of a favorable allele over time (KIM; STEPHAN, 1999; SMITH; HAIGH, 1974). If an allele confers fitness advantage, its carrier is more likely to thrive and leave more offspring than non-carrier, and as a result, the haplotype containing such beneficial allele tends to spread quickly and increases in frequency within the population (SABETI et al., 2002). Variants neighboring such beneficial mutation also tend to increase in frequency in a process called ‘hitchhike’ effect (Figure 2) (FAY; WU, 2000; SMITH; HAIGH, 1974), and extended linkage disequilibrium patterns between the favorable mutation and neighboring SNPs are observed (SABETI et al., 2002; VOIGHT et al., 2006).

As outlined above, if a population undergoes selection pressure events, it leaves distinctive tractable patterns of genetic variation that deviate statistically from that expected purely by chance (KIM; STEPHAN, 2002; OLEKSYK; SMITH; O’BRIEN, 2010). Such unique patterns of genetic variation can be detected as (i) the allele frequency spectrum shifted towards extreme frequencies (skewness); (ii) reduced local variability and excess of homozygous genotypes; (iii) extended haplotype structure, and (iv) extreme local population differentiation (MA et al., 2015; QANBARI et al., 2014). Therefore, detection of signatures of selection can disentangle past responses of the cattle genome to selection as well as increase the understanding of the evolution and biology underlying a particular phenotype (STELLA et al., 2010; UTSUNOMIYA et al., 2015). It can provide a straightforward insight into the mechanisms creating diversity across populations and contribute to mapping loci and meaningful variants underlying adaptive processes and selected traits in the genome (ANDERSSON; GEORGES, 2004; OLEKSYK; SMITH; O’BRIEN, 2010).

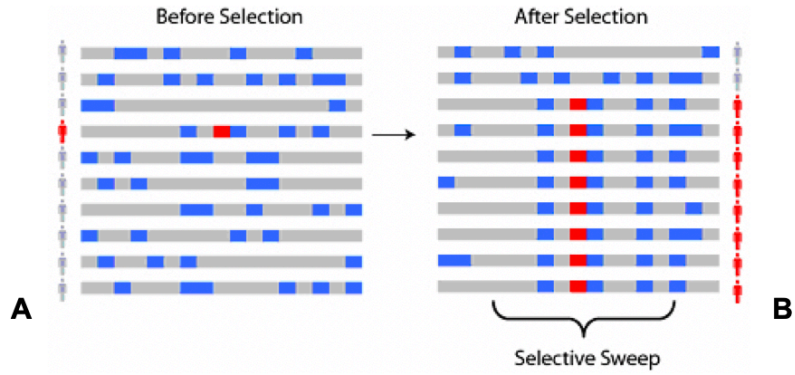


Figure 2: A selective sweep. **A.** Polymorphisms along a chromosome in which the ancestral alleles are shown in grey and the derived alleles in blue. The selected favorable allele is shown in red. **B.** The positively-selected allele (red) rises to high frequency, and the alleles that happen to be close by on the chromosome 'hitchhike' along with it to high frequency, creating a selective sweep (Adapted from: SCHAFFNER; SABETI, 2008).

Genomic regions under selection are currently detectable from SNP data, and the abundance of such markers throughout the genome makes them particularly suitable for the detection of regions where a selective sweep occurred (ANDERSSON; GEORGES, 2004). The availability of whole-genome SNP arrays has considerably improved the accuracy of signatures of selection studies (GIBBS et al., 2009), especially when considering the reconstruction of haplotypes and identification of linkage disequilibrium at high resolution (FRAZER et al., 2007). With the availability of large-scale SNP arrays and full genome sequencing, the ability to detect signatures of selection has made a breakthrough, and multiple statistical tests have been developed based on different demographic or selection models (VITTI; GROSSMAN; SABETI, 2013). In this regard, popular statistics to capture signatures of selection among populations include the composite likelihood ratio (CLR, NIELSEN et al., 2005) allele frequency spectrum-based method, long-range haplotype based methods such as the integrated haplotype score (iHS, VOIGHT et al., 2006) and the cross population extended haplotype homozygosity (XP-EHH, SABETI et al., 2007), and population

differentiation-based methods including the fixation index (FST, WEIR; COCKERHAM, 2006; WRIGHT, 1950).

Copy number variation

Chromosomal rearrangements can lead to a significant modification in the order (inversions and translocations) or number (duplications and deletions) of genomic regions, contributing to phenotypic diversity and evolutionary adaptation in several animals and plants species (CLOP; VIDAL; AMILLS, 2012). The identification of such chromosomal rearrangements has been a major focus on genomic studies. Over time, researchers shifted from microsatellites to SNP as the central measure of genetic variation in cattle, however, substantial improvement has been made in understanding additional forms of genetic variation, such as genomic structural variation (LIU et al., 2010).

Structural variants encompassing changes in DNA structure and content together with phenotypic variation are a significant source of genetic and phenotypic variation among individuals (BECKMANN; ESTIVILL; ANTONARAKIS, 2007; CONRAD; ANTONARAKIS, 2007; FEUK; CARSON; SCHERER, 2006a). In this regard, copy number variations (CNVs) are structural variants that alters the number of copies of a genomic region in comparison with a reference genome, which can range from one kb to numerous mega base pairs (Mbp) in length (FEUK; CARSON; SCHERER, 2006b).

CNVs are widespread among humans and often comprise a large proportion of the genome (~12 to 15%) (BAILEY; EICHLER, 2006; REDON et al., 2006; STANKIEWICZ; LUPSKI, 2010). In cattle, estimates suggest that 0.68% (FADISTA et al., 2010) and 1.07% (LIU et al., 2010) of the bovine genome is covered by CNVs. Hou et al. (2011) described a higher proportion of the genome covered by CNVs, with a value close to 4.60%. It is worth highlighting that the latter used the BovineSNP50 genotyping array while the others the array-based comparative genomic hybridization (aCGH), which has a limited resolution. A considerable proportion of CNVs is likely to have functional consequences by influencing gene expression since most of them

overlap with protein-coding regions (SEBAT et al., 2004) and may potentially alter gene dosage/regulation and transcript structure (LI; OLIVIER, 2013).

Three major mutational mechanisms have been implicated in genomic rearrangements and the formation of CNVs: (i) non-allelic homologous recombination (NAHR), (ii) non-homologous end-joining (NHEJ), and (iii) fork stalling and template switching (FoSTeS) (GU; ZHANG; LUPSKI, 2008). A schematic representation of those process is shown in Figure 3.

NAHR often occurs during meiosis and it is caused by the alignment of and subsequent crossover between two nonallelic DNA sequence repeats sharing high sequence homology. CNVs are frequently found close to low copy repeat regions (>10 kb in length with 95–97% similarity) in the genome, suggesting an increased predisposition to NAHR events in such regions, and consequently, CNVs formation (KIM et al., 2008; SHAW, 2004). NHEJ is a DNA repair mechanism throughout the cell cycle initiated in response to double-strand breaks (DSBs) in DNA sequence. NHEJ proceeds in four main steps: (i) detection of DSBs, (ii) molecular bridging of both broken DNA ends, (iii) modification of the ends to make them compatible, and (iv) the ligation step (WETERINGS; VAN GENT, 2004). This process can leave ‘information scars’ at the rejoining sites as the editing of the ends includes cleavage or addition of several nucleotides (LIEBER, 2008). Further, NHEJ mediated repair is not dependent on the presence of segmental duplications and low copy repeat regions (GU; ZHANG; LUPSKI, 2008). FoSTeS occurs when the DNA replication machinery pauses, the lagging strand dissociates from the polymerase holoenzyme from the original template and switches to another replication fork and restarts DNA synthesis on the new fork by priming it via the microhomology between the switched template site and the original fork (LEE; CARVALHO; LUPSKI, 2007).

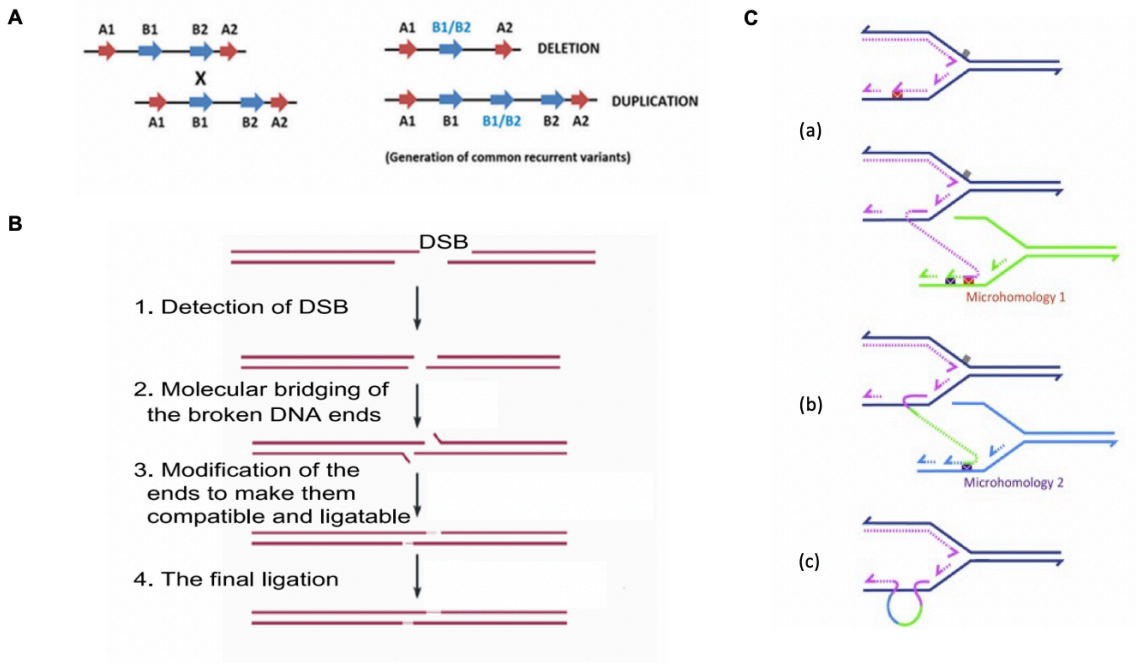


Figure 3. Genomic rearrangement mechanisms implicated in the formation of copy number variations (CNVs). **A.** Low copy repeat regions (red and blue) with counter features (homology, size, and distance) lead to the formation of CNVs by non-allelic homologous recombination (NAHR) events (adapted from Cardoso et al. (2016). **B.** Non-homologous end-joining (NHEJ) mechanism (adapted from (GU; ZHANG; LUPSKI, 2008). **C.** Fork stalling and template switching (FoSTeS). (a) The lagging strand disengages, invades an adjacent active replication fork and anneals to a region with microhomology, which primes DNA synthesis. (b) The lagging strand disengages once again and invades a further adjacent active replication fork, where it anneals to another micro homologous region to restart synthesis. (c) Eventually, the lagging strand returns to the original replication fork to continue replication to the end of the chromosome (adapted from Ottaviani; Lecain; Sheer (2014).

REFERENCES

ANDERSSON, L.; GEORGES, M. Domestic-animal genomics: Deciphering the genetics of complex traits. **Nature Reviews Genetics**, v. 5, n. 3, p. 202–212, 2004.

BAILEY, J. A.; EICHLER, E. E. Primate segmental duplications: crucibles of evolution, diversity and disease. **Nature Reviews Genetics**, v. 7, p. 552–564, 2006.

BARRIS, W. et al. Next generation sequencing of African and Indicine cattle to identify single nucleotide polymorphisms. **Animal Production Science**, v. 52, p. 133–142, 2012.

BECKMANN, J. S.; ESTIVILL, X.; ANTONARAKIS, S. E. Copy number variants and genetic traits: Closer to the resolution of phenotypic to genotypic variability. **Nature Reviews Genetics**, v. 8, n. 8, p. 639–646, 2007.

BERTOLINI, F. et al. Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. **Genetics Selection Evolution**, v. 50, n. 1, p. 59, 2018.

BISWAS, S.; AKEY, J. M. Genomic insights into positive selection. **Trends in Genetics**, v. 22, n. 8, p. 437–446, 2006.

BOSSE, M. et al. Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. **PLoS Genetics**, v. 8, n. 11, 2012.

BRASIL. **Projeto de melhoramento genético da zebuinocultura: PROZEBU: 1978-1984**. Belo Horizonte: Associação Brasileira dos Criadores de Zebu, 1978.

BRITO, C. M. C. **Citologenética do gado pé-duro**. vol. 9 ed. Goiânia: Centro Científico Conhecer, 2013.

BROMAN, K. W.; WEBER, J. L. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. **American Journal of human genetics**, v. 65, n. 6, p. 1493–500, 1999.

CARDOSO, A. R. et al. Major influence of repetitive elements on disease-associated copy number variants (CNVs). **Human Genomics**, v. 10:30, 2016.

CARDOSO, T. F. et al. Patterns of homozygosity in insular and continental goat breeds. **Genetics Selection Evolution**, v. 50:56, 2018.

CHOI, J.-W. et al. Massively parallel sequencing of Chikso (Korean brindle cattle) to discover genome-wide SNPs and InDels. **Molecules and Cells**, v. 36, p. 203–211, 2013.

CHOI, J.-W. et al. Whole-genome analyses of Korean native and Holstein cattle breeds by massively parallel sequencing. **PLoS ONE**, v. 9:e101127, n. 7, 2014.

CHOI, J.-W. et al. Copy number variations in Hanwoo and Yanbian cattle genomes

using the massively parallel sequencing data. **Gene**, v. 589, n. 1, p. 36–42, 2016.

CLOP, A.; VIDAL, O.; AMILLS, M. Copy number variation in the genomes of domestic animals. **Animal Genetics**, v. 43, n. 5, p. 503–517, 2012.

CONRAD, B.; ANTONARAKIS, S. E. Gene Duplication: A drive for phenotypic diversity and cause of human disease. **Annual Review of Genomics and Human Genetics**, v. 8, p. 17–35, 2007.

CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. **Livestock Science**, v. 166, n. 1, p. 26–34, 2014.

DAETWYLER, H. D. et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. **Nature Genetics**, v. 46, n. 8, p. 858–865, 2014.

DAS, A. et al. Deep sequencing of Danish Holstein dairy cattle for variant detection and insight into potential loss-of-function variants in protein coding genes. **BMC genomics**, v. 16:1043, 2015.

EGITO, A. et al. Microsatellite based genetic diversity and relationships among ten Creole and commercial cattle breeds raised in Brazil. **BMC Genetics**, v. 8:83, 2007.

EGITO, A. A.; MARIANTE, A. S.; ALBUQUERQUE, M. S. M. Programa brasileiro de conservação de recursos genéticos animais. **Archivos de zootecnia**, v. 51, n. 193, p. 7, 2002.

EGITO, A. A. DO. **Diversidade genética, ancestralidade individual e miscigenação nas raças bovinas no Brasil com base em Microssatélites e Haplótipos de DNA Mitocandrial: subsídios para a conservação**. [s.l.] Universidade Brasília, 2007.

FADISTA, J. et al. Copy number variation in the bovine genome. **BMC Genomics**, v. 11:284, 2010.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. Essex: Pearson, 1996.

FAY, J. C.; WU, C. I. Hitchhiking under positive Darwinian selection. **Genetics**, v. 155, n. 3, p. 1405–1413, 2000.

FELIX, G. et al. Potencial de uso de raças bovinas locais brasileiras: Curraleiro Pé-duro e Pantaneiro. **Enciclopédia Biosfera**, v. 9, n. 16, p. 1715–1741, 2013.

FERENČAKOVIĆ, M. et al. Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle. **Agriculturae Conspectus Scientificus**, v. 76, n. 4,

p. 325–329, 2011.

FERENČAKOVIĆ, M. et al. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. **Journal of Animal Breeding and Genetics**, v. 130, n. 4, p. 286–293, 2013.

FERRAZ, J. B. S. et al. **(CO)variance component estimation for growth weights of Montana Tropical®, a Brazilian beef composite**. 7th World Congress on Genetics Applied to Livestock Production. **Anais**...Montpellier: 2002

FEUK, L.; CARSON, A. R.; SCHERER, S. W. Structural variation in the human genome. **Nature Reviews Genetics**, v. 7, n. 2, p. 85–97, 2006a.

FEUK, L.; CARSON, A. R.; SCHERER, S. W. **Structural variation in the human genomeNature Reviews Genetics**, 2006b.

FLEMING, D. S. et al. Genomic analysis of Ugandan and Rwandan chicken ecotypes using a 600 k genotyping array. **BMC Genomics**, v. 17, n. 1, p. 407, 2016.

FRAZER, K. A. et al. A second generation human haplotype map of over 3.1 million SNPs. **Nature**, v. 449, p. 851–861, 2007.

GIBBS, R. A. et al. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. **Science**, v. 324, n. 5926, p. 528–532, 2009.

GIBSON, J.; MORTON, N. E.; COLLINS, A. Extended tracts of homozygosity in outbred human populations. **Human Molecular Genetics**, v. 15, n. 5, p. 789–795, 2006.

GU, W.; ZHANG, F.; LUPSKI, J. R. Mechanisms for human genomic rearrangements. **PathoGenetics**, v. 1, n. 1, 2008.

HERRERO-MEDRANO, J. M. et al. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. **BMC genetics**, v. 14:106, 2013.

HOFFMANN, I. Adaptation to climate change--exploring the potential of locally adapted breeds. **Animal**, v. 7, n. 2, p. 346–362, 2013.

HOU, Y. et al. Genomic characteristics of cattle copy number variations. **BMC Genomics**, v. 12:127, 2011.

HOU, Y. et al. Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. **BMC Genomics**, v. 13:376, 2012.

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. 12:460, 2011.

ISSA, É. C. et al. Cytogenetic analysis of the Y chromosome of native brazilian bovine breeds: preliminary data. **Archivos de Zootecnia**, v. 58, n. 221, p. 93–101, 2009.

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, n. 1, p. 237–249, 2011.

KIM, P. M. et al. Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. **Genome Research**, v. 18, p. 1865–1874, 2008.

KIM, Y.; STEPHAN, W. Allele frequency changes in artificial selection experiments: Statistical power and precision of QTL mapping. **Genetical Research**, v. 73, n. 2, p. 177–184, abr. 1999.

KIM, Y.; STEPHAN, W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. **Genetics**, v. 160, p. 765–777, 2002.

KUES, W. A. et al. High incidence of single nucleotide polymorphisms in the prion protein gene of native Brazilian Caracu cattle. **Journal of Animal Breeding and Genetics**, v. 123, p. 326–330, 2006.

LEE, J. A.; CARVALHO, C. M. B.; LUPSKI, J. R. A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders. **Cell**, v. 131, n. 7, p. 1235–1247, 2007.

LEOCARD, S. Selective sweep and the size of the hitchhiking set. **Advances in Applied Probability**, v. 41, p. 731–764, 2009.

LI, W.; OLIVIER, M. Current analysis platforms and methods for detecting copy number variation. **Physiological Genomics**, v. 45, p. 1–16, 2013.

LIAO, X. et al. Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection. **Genome**, v. 56, n. 10, p. 592–598, 2013.

LIEBER, M. R. The mechanism of human nonhomologous DNA End joining. **Journal of Biological Chemistry**, v. 283, n. 1, p. 1–5, 2008.

LIU, G. E. et al. Analysis of copy number variations among diverse cattle breeds. **Genome Research**, v. 20, n. 5, p. 693–703, 2010.

MA, Y. et al. Properties of different selection signature statistics and a new strategy for combining them. **Heredity**, v. 115, p. 426–436, 2015.

MARCHESI, J. A. P. et al. Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. **Animal**, v. 12, n. 6, p. 1126–1134, 2018.

MARIANTE, A. .; EGITO, A. . Animal genetic resources in Brazil: Result of five centuries of natural selection. **Theriogenology**, v. 57, n. 1, p. 223–235, 2002.

MARIANTE, A.; CAVALCANTE, N. **Animais do descobrimento: raças domésticas da história do Brasil.** [s.l.] Empresa Brasileira de Pesquisa Agropcuária, Centro de Pesquisa Agropecuária do Pantanal, 2000a.

MARIANTE, A. DA S. et al. Advances in the Brazilian animal genetic resources conservation programme. **Animal Genetic Resources Information**, v. 25, p. 107–121, 1999.

MARIANTE, A. S. et al. Managing genetic diversity and society needs. **Revista Brasileira de Zootecnia**, v. 37, p. 127–136, 2008.

MARIANTE, A. S. et al. Present status of the conservation of livestock genetic resources in Brazil. **Livestock Science**, v. 120, n. 3, p. 204–212, 2009.

MARIANTE, A. S.; CAVALCANTE, N. **Animais do Descobrimento: raças domésticas da história do Brasil**. Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2000b.

MARRAS, G. et al. 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.

MARSONER, T. et al. Indigenous livestock breeds as indicators for cultural ecosystem services: A spatial analysis within the Alpine Space. **Ecological Indicators**, v. 94, p. 55–63, 2017.

MARTINS, V. M. V. et al. **Raça Crioula Lageana. O esteio do ontem, o labor do hoje e a oportunidade do amanhã**. [s.l.] Editora Associação Brasileira dos Criadores da Raça Crioula Lageana (ABCCL), 2009.

MASTRANGELO, S. et al. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. **Genetics Selection Evolution**, v. 49:84, 2017.

MASTRANGELO, S. et al. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. **Animal Genetics**, v. 49, n. 1, p. 71–81, 2018.

MAZZA, M. et al. **Etnobiologia e conservação do bovino Pantaneiro**. [s.l.] Empresa

Brasileira de Pesquisa Agropcuária, Centro de Pesquisa Agropecuária do Pantanal, 1994.

MAZZA, M. C. M. et al. Phenotypical characterization of Pantaneiro cattle in Brazil. **Archivos de Zootecnia**, v. 41, p. 477–484, 1992a.

MAZZA, M. C. M. et al. Conservation of Pantaneiro cattle in Brazil: Historical origin. **Archivos de zootecnia**, v. 41, p. 443–453, 1992b.

MCMANUS, C. et al. **A Raça Caracu**. [s.l: s.n.]. Disponível em: <animal.unb.br>.

MCQUILLAN, R. et al. Runs of Homozygosity in European Populations. **American Journal of Human Genetics**, v. 83, n. 3, p. 359–372, 2008.

NIELSEN, R. Statistical tests of selective neutrality in the age of genomics. **Heredity**, v. 86, p. 641–647, 2001.

NIELSEN, R. et al. Genomic scans for selective sweeps using SNP data. **Genome research**, v. 15, n. 11, p. 1566–75, 2005.

OLEKSYK, T. K.; SMITH, M. W.; O'BRIEN, S. J. Genome-wide scans for footprints of natural selection. **Philosophical Transactions of the Royal Society B**, v. 365, p. 185–205, 2010.

OLIVEIRA, J. H. F.; MAGNABOSCO, C. DE U.; BORGES, A. M. S. **Nelore: base genética e evolução seletiva no Brasil**. Distrito Federal: Embrapa Cerrados, 2002.

ONZIMA, R. B. et al. Genome-Wide Characterization of Selection Signatures and Runs of Homozygosity in Ugandan Goat Breeds. **Frontiers in genetics**, v. 9, n. 318, 2018.

OTTAVIANI, D.; LECAIN, M.; SHEER, D. The role of microhomology in genomic structural variation. **Trends in Genetics**, v. 30, n. 3, p. 85–94, 2014.

OTTO, S. P. Detecting the form of selection from DNA sequence data. **Trends in Genetics**, v. 16, n. 12, p. 526–529, 2000.

PEMBERTON, T. J. et al. Genomic patterns of homozygosity in worldwide human populations. **American Journal of Human Genetics**, v. 91, n. 2, p. 275–292, 2012.

PERIPOLLI, E. et al. Autozygosity islands and ROH patterns in Nellore lineages: Evidence of selection for functionally important traits. **BMC Genomics**, v. 19:680, 2018a.

PERIPOLLI, E. et al. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. **BMC Genomics**, v. 19:34, 2018b.

PILLING, D.; HOFFMANN, I. Biodiversity for Food and Agriculture Biodiversity for Food

and Agriculture. **Commission on genetic resources for food and agriculture**, n. 53, 2011.

PRIMO, A. El ganado bovino ibérico en las Américas: 500 años después. **Archivos de zootecnia**, v. 41, n. 154, p. 421–432, 1992.

PRIMO, A. T. Conservación de Recursos Genéticos Animales em el Brasil. In: **Ganado Bovino Criollo**. Buenos Aires: [s.n.]. p. 224.

PRIMO, A. T. **Os bovinos ibéricos nas Américas**. Reunião anual da Sociedade Brasileira de Zootecnia. **Anais**...Rio de Janeiro: 1993

PURFIELD, D. C. et al. Runs of homozygosity and population history in cattle. **BMC genetics**, v. 13:70, 2012.

PURFIELD, D. C. et al. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. **PLoS ONE**, v. 12:e017678, n. 5, 2017.

QANBARI, S. et al. Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. **PLoS Genetics**, v. 10:e100414, n. 2, 2014.

QUEIROZ, S. A.; LÔBO, R. B. Genetic relationship, inbreeding and generation interval in registered Gir cattle in Brazil. **Journal of animal breeding and genetics**, v. 110, p. 228–233, 1993.

REDON, R. et al. Global variation in copy number in the human genome. **Nature**, v. 444, p. 444–454, 2006.

SABETI, P. C. et al. Detecting recent positive selection in the human genome from haplotype structure. **Nature**, v. 419, p. 832–837, 2002.

SABETI, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. **Nature**, v. 449, n. 7164, p. 913–918, 2007.

SANTIAGO, A. A. **O Guzerá**. Recife: Tropical, 1984.

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

SCHAFFNER, S.; SABETI, P. Evolutionary adaptation in the human lineage. **Nature Education**, v. 1:14, n. 1, 2008.

SEBAT, J. et al. Large-scale copy number polymorphism in the human genome. **Science**, v. 305, n. 5683, p. 525–528, 2004.

SERRANO, G. et al. Genetic diversity and population structure of Brazilian native

bovine breeds. **Pesquisa Agropecuaria Brasileira**, v. 39, n. 6, p. 543–549, 2004.

SHAW, C. J. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. **Human Molecular Genetics**, v. 13, n. 1, p. 57R – 64, 2004.

SILIÓ, L. et al. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. **Journal of Animal Breeding and Genetics**, v. 130, n. 5, p. 349–360, 2013.

SMITH, B. Y. J. M.; HAIGH, J. The hitch-hiking effect of a favourable gene. **Genetic Research**, v. 23, p. 23–35, 1974.

STANKIEWICZ, P.; LUPSKI, J. R. Structural variation in the human genome and its role in disease. **Annual Review of Medicine**, v. 61, p. 437–455, 2010.

STELLA, A. et al. Identification of selection signatures in cattle breeds selected for dairy production. **Genetics**, v. 185, n. 4, p. 1451–1461, 2010.

UGARTE, E. et al. Impact of high-yielding foreign breeds on the Spanish dairy sheep industry. **Livestock Production Science**, v. 71, p. 3–10, 2001.

UTSUNOMIYA, Y. T. et al. Genomic data as the “hitchhiker’s guide” to cattle adaptation: Tracking the milestones of past selection in the bovine genome. **Frontiers in Genetics**, v. 6, n. 36, 2015.

VISSCHER, P. M. et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. **PLoS Genetics**, v. 2, n. 3, p. 0316–0325, 2006.

VITTI, J. J.; GROSSMAN, S. R.; SABETI, P. C. Detecting Natural Selection in Genomic Data. **Annual Review of Genetics**, v. 47, p. 97–120, 2013.

VOIGHT, B. F. et al. A map of recent positive selection in the human genome. **PLoS Biology**, v. 4:e72, n. 3, 2006.

WEIR, B. S.; COCKERHAM, C. C. Estimating F-Statistics for the Analysis of Population Structure. **Evolution**, v. 38, n. 6, p. 1358–1370, 2006.

WETERINGS, E.; VAN GENT, D. C. The mechanism of non-homologous end-joining: A synopsis of synapsis. **DNA Repair**, v. 3, n. 11, p. 1425–1435, 2004.

WRIGHT, S. Coefficients of inbreeding and relationship. **The American Naturalist**, v. 56, n. 645, p. 330–338, 1922.

WRIGHT, S. The Genetical Structure of populations. **Nature**, v. 166, n. 4215, p. 247–

249, 1950.

YUE, X. P. et al. Copy number variations of the extensively amplified Y-linked genes, HSFY and ZNF280BY, in cattle and their association with male reproductive traits in Holstein bulls. **BMC Genomics**, v. 15:113, 2014.

ZANDER, K. K. et al. Assessing the total economic value of threatened livestock breeds in Italy : Implications for conservation policy. **Ecological Economics**, v. 93, p. 219–229, 2013.

ZANELLA, R. et al. Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. **Genetics Selection Evolution**, v. 48:24, 2016.

ZAVAREZ, L. B. et al. Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. **Frontiers in Genetics**, v. 6, n. 5, p. 1–8, 2015.

ZHANG, Q. et al. Runs of homozygosity and distribution of functional variants in the cattle genome. **BMC Genomics**, v. 16:542, 2015.

CAPÍTULO 2 – AUTOZYGOSITY ISLANDS AND ROH PATTERNS IN NELLORE LINEAGES: EVIDENCE OF SELECTION FOR FUNCTIONALLY IMPORTANT TRAITS[1]

ABSTRACT

The aim of this study was to assess genome-wide autozygosity in a Nellore cattle population, and to characterize ROH patterns and autozygosity islands that may have occurred due to selection within its lineages. It attempts also to compare estimates of inbreeding calculated from ROH (F_{ROH}), genomic relationship matrix (F_{GRM}), and pedigree-based coefficient (F_{PED}). The average number of ROH per animal was 55.15±13.01 with an average size of 3.24 Mb. The Nellore genome is composed mostly by a high number of shorter segments accounting for 78% of all ROH, although the proportion of the genome covered by them was relatively small. The genome autozygosity proportion indicates moderate to high inbreeding levels for classical standards, with an average value of 7.15% (178.70 Mb). The average of F_{PED} and F_{ROH}, and their correlations (-0.05 to 0.26) were low. Estimates of correlation between F_{GRM}-F_{PED} was zero, while the correlation (-0.01 to -0.07) between F_{GRM}-F_{ROH} decreased as a function of ROH length, except for $F_{ROH>8Mb}$ (-0.03). Overall, inbreeding coefficients were not high for the genotyped animals. Autozygosity islands were evident across the genome (*n=62*) and their genomic location did not largely differ within lineages. Enriched terms (p<0.01) associated with defense response to bacteria (GO:0042742), immune complex reaction (GO:0045647), pregnancy-associated glycoproteins genes (GO:0030163), and organism growth (GO:0040014) were described within the autozygotic islands. Low F_{PED}-F_{ROH} correlation estimates indicate that F_{PED} is not the most suitable method for capturing ancient inbreeding when the pedigree does not extend back many generations and F_{ROH} should be used instead. Enriched terms (p<0.01) suggest a strong selection for immune response. Non-overlapping islands within the lineages greatly explain the mechanism underlying selection for functionally important traits in Nellore cattle.

Key-words: *Bos taurus indicus*, Indicine, Genomic Inbreeding, Gene Ontology

[1] Este capítulo corresponde ao artigo científico publicado na revista BMC Genomics. 2018; 19:680.

INTRODUCTION

Brazilian livestock and agriculture production have a prominent impact upon the world's food commerce. Brazilian beef production is one of the largest players in the world and produced roughly 9.56 million tons of carcass weight equivalents in 2015 [1]. The vast majority of the bovine based population reared for meat production in Brazil is composed mostly of indicine cattle (*Bos taurus indicus*). According to the Brazilian Zebu Breeders Association (ABCZ, www.abcz.com.br) such population is around 80% of the total cattle. Given the physical and physiological characteristics that they possess which greatly explain their better adaptation towards grazing systems in tropical environments [2–4], it is not surprisingly that much use of the indicine cattle has been made in these regions.

The Nellore breed has the largest number of animals (horned and polled) among the indicine cattle raised in Brazil, followed by Guzerat and Gyr. Most of Nellore importation was from India during the last century and lasted up to the seventies when the importation was banned [5]. The Nellore population in Brazil is the result of less than 7,000 heads of purebred imported animals [6]. The major importation took place in 1962, when exceptional bulls were brought over the country standing out as progenitors of the main Nellore lineages [7]. Magnabosco et al. [8] reported the existence of six predominant lineages of Nellore breed (Karvadi Imp; Taj Mahal Imp; Kurupathy Imp; Golias Imp; Godhavari Imp, and Rastã Imp) that contributed to the development of the current Brazilian Nellore population. These lineages were derived from outstanding bulls named Karvadi, Taj Mahal, Kurupathy, Golias, Godhavari and Rastã which gained fame as breeders given their high rates of productive and reproductive performance [7]. Although the selection criteria used to improve the Nellore cattle among Brazilian breeding programs are closely linked and mainly associated with reproductive and carcass quality traits, there is evidence of different genetic patterns among the lineages based on the selection criterion used to improve each of them over time [9,10]. In this manner, a question can be raised whether the genetic progress is going or not towards the same direction within the lineages raised in Brazil.

Genetic evaluations of Nellore cattle using BLUP (Best Linear Unbiased Prediction) methodology have established significant progress since the eighties, when several genetic evaluation programs started to expand in Brazil [11]. Despite the reduced number of animals imported from India, Santana et al. [12] has reported an average inbreeding coefficient of 3% in a Nellore population, indicating that these animals have been under relative control for at least three decades. Therefore, breeding programs are always seeking for strategies to preserve populations, and there is a growing interest in characterizing and monitoring genome-wide autozygosity to maintain the genetic diversity [13,14], allowing a long-term conservation of genetic resources and sustainability in animal breeding programs.

Runs of homozygosis (ROH) have been widely applied to quantify individual autozygosity in livestock [15–20] given their high correlation (~0.7) [21]. A small number of studies have described the autozygosity in Nellore cattle and most of them do not make use of a large sample size. Karimi [22] identified region patterns with a high prevalence of ROH in taurine and indicine breeds and made use of merely 134 Nellore samples. Additionally, Zavarez et al. [19] reported the distribution of genome-wide autozygosity levels based on ROH in only 1,278 Nellore cows genotyped for over 777,000 markers.

Since homozygous stretches printed on the genome may have arisen as a result of artificial selection, autozygosity based on ROH can strongly disclose the understanding of genetic selection [18]. ROH patterns are not seen to be randomly distributed across the genomes [23] and genomic regions sharing ROH patterns potentially contain alleles associated with genetic improvement in livestock [24]. The correlation of ROH and selection for productivity was first identified by Kim et al [25]. Furthermore, ROH has been successfully utilized as a measure of inbreeding by estimating the level of autozygosity in the genome [15,16,25–28].

Up to date, studies characterizing genome-wide autozygosity in the main Nellore lineages are incipient. Hence, this study was carried out to assess genome-wide autozygosity in a Nellore cattle population to identify and characterize ROH patterns as well as to identify autozygosity islands that may have occurred due to selection for functionally important traits in different Nellore lineages and verify whether these lineages differ or not from one another. It attempts also to compare estimates of

molecular inbreeding calculated from ROH (F_{ROH}), genomic relationship matrix (F_{GRM}), and from pedigree-based coefficient (F_{PED}).

RESULTS

Genome-wide distribution of Runs of homozygosity

On individual animal basis, the average number of ROH per animal, considering the genotyped animals (*n=9,386*), was 55.15±13.01 with an average size of 3.24 Mb. The longest ROH was 99.30 Mb in length (28,778 SNPs) on *Bos taurus* autosome (BTA) 5. The number of ROH per chromosome was also greater for BTA5 (33,492 segments) (Figure 1a) and the greatest fraction of chromosome covered with ROH was found on BTA28 (15.06% of chromosomal length within an ROH) (Figure 1b).

ROH analysis for the different length classes for the genotyped animals (*n=9,386*) revealed that the Nellore genome is composed mostly of a high number of short segments ($ROH_{1\text{-}2\ Mb}$ and $ROH_{2\text{-}4\ Mb}$), which accounted for approximately 78% of all ROH detected and roughly contributed to 43% of the cumulative ROH length (Table 1). Short and medium ($ROH_{4\text{-}8\ Mb}$) ROH displayed a similar genome coverage as well as a cumulative ROH length, with values varying from 20.53 to 22.88%. Despite the total length of ROH being composed mostly of a high number of short segments, the proportion of the genome covered by them was relatively small when compared to larger ROH ($ROH_{>8\ Mb}$).

Table 1. Descriptive statistics of runs of homozygosity number (*n*ROH) and length (in Mb) for four different length classes ($ROH_{1\text{-}2\ Mb}$, $ROH_{2\text{-}4\ Mb}$, $ROH_{4\text{-}8\ Mb}$, and $ROH_{>8\ Mb}$)

Class	*n* ROH	(%)	Mean Length	Standard Deviation	Genome Coverage (%)	Cumulative ROH Length (%)
$ROH_{1\text{-}2\ Mb}$	285,085	55.07	1.34	0.27	1.63	22.88
$ROH_{2\text{-}4\ Mb}$	123,254	23.81	2.79	0.56	1.47	20.53
$ROH_{4\text{-}8\ Mb}$	68,407	13.21	5.53	1.11	1.63	22.59
$ROH_{>8\ Mb}$	40,925	7.91	13.93	7.18	2.58	34.00

The most autozygous animal exhibited a ROH genome coverage encompassing 718.96 Mb of the total autosomal genome extension (UMD3.1) covered by markers (28.75% of the cattle genome), totaling 92 ROH ≥ $ROH_{1\text{-}2\ Mb}$. On average, 7.15% (178.70 Mb) of the genome was considered to be a region of homozygosity.

Pedigree and genomic inbreeding

Descriptive statistics for F_{PED} and F_{ROH} coefficients for the genotyped animals (*n=9,386*) are presented in Table 2. The average F_{PED} and F_{ROH} were low in the studied population, and it is noteworthy to highpoint that 94.20% of the genotyped animals exhibited a F_{PED} below 5%. Low correlations were observed between F_{PED}-F_{ROH}, and it gradually increased as a function of ROH length (Figure 2). No estimates of correlation were found between F_{GRM}-F_{PED} and those between F_{GRM}-F_{ROH} decreased as a function of ROH length. The inbreeding evolution (Figure 3) demonstrates a significant ($p<0.01$) decay in $F_{ROH>8\ Mb}$.

Table 2. Number of genotyped animals (*n*) and descriptive statistics of the pedigree-based inbreeding coefficient (F_{PED}) and runs of homozygosity-based inbreeding coefficient (F_{ROH}) for different lenghts ($F_{ROH1\text{-}2}$, $F_{ROH2\text{-}4}$, $F_{ROH4\text{-}8}$, and $F_{ROH>8\ Mb}$)

Coefficient	Mean	Median	Minimum	Maximum	Coefficient of Variation (%)	*n*
F_{PED}	0.017	0.013	0.000	0.258	3.387	8502
$F_{ROH1\text{-}2\ Mb}$	0.016	0.016	0.000	0.199	27.14	9387
$F_{ROH2\text{-}4\ Mb}$	0.014	0.014	0.000	0.100	37.71	9352
$F_{ROH4\text{-}8\ Mb}$	0.016	0.015	0.001	0.059	47.81	9281
$F_{ROH>8\ Mb}$	0.025	0.021	0.003	0.222	77.03	8836

F_{PED} and F_{ROH} averages for each Nellore lineage (*n=8,646*) are presented in Table 3. The highest F_{PED} ($p<0.05$) values were observed for Karvadi, Golias, and Godhavari lineages. F_{ROH} estimates were close to F_{PED}, and they did not differ ($p<0.05$) for Karvadi and Godhavari lineages.

Table 3. Average mean (number of observations) of pedigree-based inbreeding coefficient (F_{PED}) and runs of homozygosity-based inbreeding coefficient (F_{ROH}) for different lenghts (F_{ROH1-2}, F_{ROH2-4}, F_{ROH4-8}, and $F_{ROH>8\ Mb}$) for six Nellore lineages

Coefficient	Karvadi	Golias	Godhavari	Taj Mahal	Akasamu	Nagpur
F_{PED}[1]	0.020[a] (7,282)	0.019[a] (178)	0.020[a] (90)	0.016[ab] (103)	0.011[b] (42)	-
$F_{ROH1-2\ Mb}$	0.016[a] (7,853)	0.014[c] (288)	0.015[ab] (205)	0.014[bc] (149)	0.014[c] (79)	0.014[c] (50)
$F_{ROH2-4\ Mb}$	0.014[a] (7,810)	0.012[b] (284)	0.014[a] (198)	0.012[b] (144)	0.011[b] (73)	0.012[b] (44)
$F_{ROH4-8\ Mb}$	0.015[a] (7,664)	0.014[b] (266)	0.016[a] (185)	0.014[b] (136)	0.014[b] (70)	0.012[b] (40)
$F_{ROH>8\ Mb}$	0.025[a] (7,443)	0.022[bc] (245)	0.024[ab] (171)	0.018[c] (130)	0.022[bc] (70)	0.017[c] (34)

F_{PED} was not available for the Nagpur lineage. Means sharing a common letter within a row were not significantly different ($p<0.05$) from one another.

Autozygosity islands in Nellore lineages

Autozygosity islands were evident across the genome, and their distributions along the genome vary in length and position across chromosomes. A total of 62 regions with 100 outlying consecutive SNPs were identified for the genotyped animals (*n=9,386*) in almost all autosomes, with the exception of BTA2, BTA11, BTA18, BTA25, and BTA28 (Appendix 1A). Overall, the mean length was 1.40±0.85 Mb, and the longest island was observed on BTA7 (107,000,000:111,700,000 bp) encompassing 4.70 Mb of length. Interestingly, BTA7 also contained the highest number of islands (*n=8*) followed by BTA1, BTA12 and BTA20, all-encompassing five islands each.

To verify if the autozygosity islands possess genes related to environmental adaptation processes, those 62 autozygosity islands were overlapped with 9,803 CNVRs strongly associated with adaptation for the Nellore cattle described by Lemos et al. [29]. Only 338 CNVRs were observed within the autozygosity islands, and the overlapping regions harbored 484 genes with described functions.

When analyzing the autozygosity islands within the lineages (*n=8,646*), the Karvadi lineage showed the highest number of islands (*n=54*), followed by Godhavari (*n=31*), Golias (*n=26*), Taj Mahal (*n=18*), Akasamu (*n=13*), and Nagpur (*n=6*). It should be noted that overlapping islands were observed in between the lineages (Appendix 2A and 3A). Interestingly, the region on BTA7 encompassing 51,610,000 to 52,930,000 bp in length was found to be described in all lineages. Non-overlapping autozygosity islands were also observed in some lineages in specific genomic regions and were

screened for gene content (Appendix 4A). These regions could be an indicative of selection signatures or it may reflect inbreeding events within a lineage [26].

Functional annotation of genes

As most of autozygosity islands identified for the genotyped animals (*n=9,386*) overlapped with those described for the Nellore lineages (Appendix 5A), the analysis performed using the DAVID v.6.8 [30,31] comprised 946 genes identified for the genotyped animals (Table 4). Appendix 6A describes the set of genes involved in each GO term and KEGG pathway.

Table 4. Gene Ontology (GO) terms and KEGG pathways annotation analysis enriched (P<0.01) based on autozygosity islands set of genes

Terms	**Genes**	**P-value**
GO Biological Process		
(GO:0042742) Defense response to bacteria	14	7.07E-5
(GO:0030163) Protein catabolic process	9	6.33E-4
(GO:0070200) Establishment of protein localization to telomere	4	1.70E-3
(GO:0040014) Regulation of multicellular organism growth	6	2.68E-3
(GO:0045647) Negative regulation of erythrocyte differentiation	4	4.46E-3
(GO:0030901) Midbrain development	6	4.84E-3
GO Molecular Function		
(GO:0008289) Lipid binding	13	2.07E-4
(GO:0004190) Aspartic-type endopeptidase activity	9	3.24E-4
GO Cellular Component		
(GO:0005776) Autophagosome	8	3.07E-3
(GO:0005634) Nucleus	155	6.11E-3
(GO:0005815) Microtubule organizing center	10	8.36E-3
(GO:0005730) Nucleolus	41	8.50E-3
KEGG pathway		
(bta01100) Metabolic pathways	72	4.21E-4

To obtain a broad functional insight into the set of genes (*n=484*) observed within the autozygosity islands and the CNVRs overlapping regions, an enrichment analysis was also performed. An enhancement of genes involved in several GO terms

(four biological processes, one molecular function, and one cellular component process) was significant (p≤0.01) and one for KEEG (Appendix 7A). Despite the large number of overlapping regions, and consequently, the large number of genes found within these regions, no significant GO term and KEGG pathway was found commonly associated in both studies and neither associated in some way with environmental adaptation processes.

DISCUSSION

Genome-wide distribution of runs of homozygosity

The longest ROH was described on BTA5, however, results in taurine and indicine cattle [20,25,32] have reported the longest on BTA8. Corroborating with the results, Peripolli et al. [20] observed the greatest number of ROH on BTA5 in indicine cattle, however, studies have described the greatest number on BTA1 [24,32,33]. BTA5, which presented the longest and the greater number of ROH, has been reported to harbor QTL related to weight [34,35], reproduction [36,37], and milk fat yield traits [37,38] in cattle.

Dissimilarity among animals was observed between the number of ROH and the length of the genome covered by them (Figure 4). Animals exhibiting the same homozygous genome length displayed a variable number of ROH. This pattern was also described by Mészáros et al. [39], who attributed this event as a consequence of the distinct distances from the common ancestor. Therefore, when considering animals with the same homozygous genome length, we can infer that those displaying more ROH have an increased distance with the common ancestor since these segments are expected to be shorter due to repeated meiosis events that break up ROH through recombination [40].

The highest autozygosity value per animal was similar to those reported in the literature for dairy breeds [20,24,32,41]. Conversely, Marras et al. [18] described that dairy breeds had a higher sum of all ROH than did beef breeds, and Purfield et al. [24] observed that dairy breeds were the most autozygous animals among several studied breeds. In addition, the autozygotic proportion of the genome described for this population seems to indicate moderate to high inbreeding levels for classical

standards. Similar results were described by Marras et al. [18] for Marchigiana beef cattle (7%) and Peripolli et al. [20] for Gyr dairy cattle (7.10%). Compared to Zavarez et al. [19] study on a Nellore population whose findings showed a value of 4.58%, this sample of Nellore animals presented a higher average autosomal coverage. The high autozygosity value per animal and homozygous proportion of the genome observed for this population might be a result of the small number of imported progenitors to speed up the genetic progress and develop the first Nellore lineages during the major importation in the sixties. Furthermore, the formation of lineages can be made by the use of consanguinity in which the same breeder is mated with its descendants along the generations aiming to fix genes related to important traits [8].

Pedigree and genomic inbreeding

F_{PED} was lower than results reported by Barbosa et al. [42] and higher than those described by Santana et al. [43], with values of 8.32% and 1.42% for inbred Nellore populations, respectively.

F_{ROH} can disclose the age of the inbreeding given the approximate correlation between the length of the ROH and the distance with the common ancestor due to recombination events over time. Therefore, calculated F_{ROH} are expected to correspond to the reference ancestral population dating 50 ($F_{ROH1\text{-}2\ Mb}$), 20 ($F_{ROH2\text{-}4\ Mb}$), 12.5 ($_{FROH4\text{-}8\ Mb}$), and 6 ($F_{ROH>8Mb}$) generations ago by considering that 1 cM equals to 1 Mb [44]. According to Zavarez et al. [19], incomplete pedigree cannot account for inbreeding caused by distant ancestors and estimates based on F_{PED} are only comparable with F_{ROH} calculated over large ROH. F_{PED} estimate was then compared with $F_{ROH>8\ Mb}$, and the genome autozygotic proportion from $F_{ROH>8\ Mb}$ exceeded F_{PED}. This variation can be attributed to the fact that the pedigree might not have been deep enough to allow F_{PED} to capture the relatedness since its average depth is close to four generations, whereas $F_{ROH>8\ Mb}$ reflects an inbreeding that occurred nearly six generations ago. Furthermore, F_{PED} does not take into account the stochastic events of recombination during meiosis [26] and pedigree relatedness does not show the actual relatedness among individuals since it is estimated from statistical expectations of the probable identical by descendent (IBD) genomic proportion [45].

F_{PED}-F_{ROH} correlations were seen to be higher when longer ROH reflecting recent relatedness were included in F_{ROH} estimates. It is noticeable to highlight that most of the pedigree records did not extend back many generations, therefore, correlations with shorter ROH reflecting ancient relatedness tended to be lower and those with longer ROH reflecting recent relatedness had a tendency to be higher [18,46]. Additionally, several authors have reported a high correlation between F_{PED}-F_{ROH} when a deeper number of described generations are available in the pedigree [15,16,18,24,33].

No estimates of correlation between F_{GRM}-F_{PED} may be explained by considering that individuals from sub-populations for which allele frequencies diverge from the entire population may have been estimated to have high F_{GRM} [47], which may have led to biased correlation. According to Zhang et al. [48], inbreeding coefficients based on methods using allele frequency are sensitive compared to ROH-based methods, especially for populations with divergent allele frequencies. Correlations between F_{GRM}-F_{ROH} decreased as a function of ROH length, and Zavarez et al. [19] associated it with the properties of the G matrix, which is based on individual loci, whereas F_{ROH} is based on chromosomal segments.

The inbreeding evolution stress out a significant ($p<0.01$) decline in $F_{ROH>8\ Mb}$, and it is worth highlighting that it reflects inbreeding up to six generations prior (~30 years). The reduction in this coefficient since the 1990's happened together with the foundation of the Nellore Brazil Breeding program in 1988 (ANCP, http://www.ancp.org.br). These results pointed out, that mating decisions were taken since this time by the breeders to avoid mating between relatives, decreasing the genomic inbreeding level in this population over time. The $F_{ROH\ 4\text{-}8\ Mb}$ reflects inbreeding up to 12.5 generations prior (~60 years) and the slight reduction in this coefficient since the 1960's happened together with the beginning of bull evaluation for weight gain in test stations. The results obtained for $F_{ROH1\text{-}2\ Mb}$ and $F_{ROH2\text{-}4\ Mb}$ showed that mating decisions before the major importations might have favored the increasing of inbreeding.

Inbreeding coefficients were not high for the genotyped animals with lineages records (*n=8,646*), with values around to 2%. According to Pereira [49], the lineage diversification within a breed can provide substantial gains for selection by reducing

inbreeding rates and restoring the genetic variability. The use of Karvadi and Godhavari lineages can be evidenced by the high inbreeding rates described for them when compared to other lineages. According to Oliveira et al. [7], when considering a small number of progenitors in a studied breed, the prevalence use of some ancestors can be explained by their marginal contribution in the reference population. Hence, when assessing the marginal contribution of each lineage to the ANCP Nellore cattle population, an eminent contribution of Karvadi and Godhavari lineages can be observed (10.44 and 1.48%, respectively), agreeing with F_{ROH} estimates. Lineages such as Golias, Taj Mahal, Akasamu, and Nagpur did not show an expressive marginal contribution, and interestingly, displayed lower inbreeding averages ($p<0.05$) for $F_{ROH1\text{-}2\ Mb}$, $F_{ROH2\text{-}4\ Mb}$, and $F_{ROH4\text{-}8\ Mb}$.

Autozygosity islands in Nellore lineages

Autozygosity islands in the genotyped animals (*n=9,386*) were seen overlapping with previous studies on several cattle breeds (Appendix 8A). Within these studies, islands were not reported overlapping only with those described for Nellore cattle. Remarkably, Sölkner et al. [50] and Szmatoła et al. [41] displayed islands in common on BTA7 encompassing the same chromosomal region around 51-53 Mb, and Szmatoła et al. [41] also described islands located on the same chromosomal region on BTA7 (42-44 Mb) in Holstein, Red Polish, Simmental and Limousin cattle breeds. Sölkner et al. [50] and Gaspa et al. [51] exhibited overlapping islands around 1.3-1.9 Mb on BTA21. Overlapping islands between these studies and the current one (43,510,000:43,592,173 – BTA7; 51,574,295:52,353,000 - BTA7, and 1,360,390: 1,829,761 – BTA21) were inspected in detail. These islands are suggested to harbor targets of positive selection in cattle [52] and may be used to identify regions of the genome under selection, and to map genes that affect traits of interest [18]. Further, ROH islands were found overlapping in cattle breeds selected for different purposes, suggesting that selection pressure can also be undergoing on traits other than those specific to dairy or beef traits.

When examining in detail, the region encompassing 51-52 Mb on BTA7 harbored relevant genes for beef cattle production. Among them, we highpoint the

CTNNA1 gene which has been associated with myostatin expression level in skeletal muscle of Holstein-Friesian bulls [53]. Myostatin is a key protein that plays an essential role in regulating skeletal muscle growth, and it is considered to be one of the most important factors responsible for meat productivity traits in cattle [54]. The *MATR3* gene was also described within the overlapping region and it has been related to fat deposition in cattle [55,56]. It is also worth highlighting the *ECSCR* gene. This gene regulates insulin sensitivity and predisposition to obesity [57]. Besides, the protein encoded by this gene is primarily found in endothelial cells and blood vessels (provided by RefSeq, Jun 2014). Endothelial cells are the important players in angiogenesis, a physiological process by which new blood vessels develop from pre-existing vasculature [58]. Blood vessels dilate to dissipate heat to external environment by a process denominated vasodilation. In this regard, the *ECSCR* gene might be a key role in elucidating the better tolerance of some cattle breeds to heat stress, i.e., *Bos taurus indicus*. The increased number of blood vessels through the angiogenic process allows more blood to be dissipated, decreasing the body temperature.

Overlapping islands within the lineages (*n=8,646*) were described in this study and two reasons might have led to this result. First, the Nellore cattle sampled in Brazil is derived from the Ongole cattle imported from the Indian district of Andhra Pradesh [59]. Prior such importations, the Ongole cattle was already notorious in India due to their greater adaptation upon high temperatures, ability to carry lower burdens of cattle tick and tolerate poor feed management [60]. Therefore, these overlapping regions might reflect the acquired adaptedness of zebu cattle in tropical environments due to natural selection over the time [61]. Second, these findings support the concept that despite having different lineages within the Nellore breed, the genetic progress of economically important traits goes toward the same direction and IBD genomic regions harboring traits of interest are being conserved over time.

The region on BTA7 described to be overlapping in all lineages (51610000:52930000 bp) harbored five genes (*CTNNA1*, *LRRTM2*, *SIL1*, *MATR3*, *and PAIP2*). Among them, the *CTNNA1* (Catenin Alpha 1) gene has been described associated with myostatin expression level and molecular function in skeletal muscle in Holstein-Friesian bulls [53], as previously mentioned. Furthermore, the *LRRTM2*

(Leucine Rich Repeat Transmembrane Neuronal 2) gene was found related to maturation of male germ cells and male fertility [62,63].

Non-overlapping islands within the lineages were explored for gene content, and among the genes identified within these regions, we can highpoint those described in Table 5. Remarkably, six genes were also reported in Nellore-specific studies associated with carcass traits (*PPM1*) [64], age at first calving (*NPBWR1, OPRK1,* and *MRPL1*) [65], and birth weight (*RPS20* and *TGS1*) [66].

Table 5. Gene content of non-overlapping ROH islands within the Nellore lineages highlighted according to their function

Lineage	Gene	Function	Author
Godhavari	*LAMB4*	Immune System	[67]
Karvadi	*RFX4*	Immune System	[68]
Godhavari	*IFRD, PPM1B, DTX4, MTMR7*	Productive traits	[64,68–71]
Taj Mahal	*CAPZA2*	Productive traits	[72]
Karvadi	*ZBTB20, RPS20, STAC3, STAT6, RIC8B, LYPLA1, XKR4, TMEM68, TGS1*	Productive traits	[66,68,73–78]
Godhavari	*NAMPT*	Reproductive traits	[79,80]
Godhavari	*PPM1B, JMJD1C*	Reproductive traits	[81,82]
Karvadi	*RFX4, NPBWR1, OPRK1, MRPL15*	Reproductive traits	[65,83]
Karvadi	*DRD3, ZBTB20*	Reproductive traits	[84,85]
Karvadi	*CSNK1A1, TBC1D12*	Thermotolerance	[86,87]

Despite having non-overlapping autozygosity islands within the lineages, several genes have been found described associated with productive and reproductive traits within the lineages. Productive related genes were mainly associated with average daily gain (*IFRD1*), muscle (*PPM1B* and *STAC3*), fat (*DTX4* and *XKR4*), body and birth weight *(MTMR7*, *RPS20*, and *TGS1*), meat and carcass quality traits (*MTMR7*, *CAPZA2*, *STAT6*, and *RIC8B*), and feed intake (*LYPLA1* and *TMEM68*). Reproductive related genes largely encompassed those linked to heifer's fertility (*RFX4*), age at first calving (*NPBWR1*, *OPRK1*, and *MRPL15*), and oocyte maturation and expression (*NAMPT* and *JMJD1C*).

Although they were not located in the same genomic regions, these autozygosity islands showed an enrichment of genes involved in cattle growth, meat and carcass quality traits, immune system, and thermotolerance functions. These findings help to reinforce the concept that the genetic progress goes towards the same direction within the lineages and different genetic patterns among the lineages based on the selection criterion used to improve each of them could not be identified in this study.

Functional annotation of genes

The analyses performed on DAVID revealed only the metabolic pathways (bta01100) KEGG pathway as significant ($p<0.01$), while the Gene Ontology analyses showed several enriched terms for the ROH gene list. The defense response to bacteria (GO:0042742) on biological process encompasses several reactions triggered in response to the presence of a bacteria that act to protect the cell or organism. We highlighted the beta-defensin genes (*DEFB1, DEFB4A, DEFB5, DEFB6, DEFB7, DEFB10,* and *DEFB13*) that encode host defense peptides that are critical to protection against bacterial, viral and fungal infections, and acts as an important link between innate and adaptive immune responses [88]. In addition to their antimicrobial properties, beta-defensins have an important role in several functions including regulation of the immune response, fertility, reproduction, and embryo development [88,89].

The negative regulation of erythrocyte differentiation (GO:0045647) on biological process is defined as any process that stops, prevents, or reduces the frequency, rate or extent of erythrocyte differentiation. Erythrocytes were described by Nelson [90] as belonging to the immune complex reaction (bacteria, complement, and antibody). In fish and chickens, erythrocytes have been shown to facilitate the clearance of pathogens by macrophages [91], and could produce specific signaling molecules such as cytokines in response to binding [92,93].

The protein catabolic process (GO:0030163) includes chemical reactions and pathways resulting in the breakdown of mature proteins, which play an important role in the immune and inflammatory response. Khansefid et al. [94] identified the protein

catabolic process enriched in genes significantly associated with residual feed intake in Angus and Holstein cattle breeds. Regarding the genes related to protein catabolic process identified in our study, most of them are pregnancy-associated glycoproteins genes (PAG) (Appendix 6A) mapped on BTA29. Goszczynski et al. [95] identified eight genes belonging to the PAG gene family within ROH islands in Retinta cattle breed, while Szmatoła et al. [41] identified sixteen PAG genes in Holstein cattle breed. PAG glycoproteins are one major group of the proteins secreted from trophoblast cells of the placenta into the maternal blood shortly after implantation and are detectable throughout gestation [56]. These proteins have been used to monitor embryonic viability as biochemical pregnancy markers in the cow's blood or milk [96] as well as placental functions in cattle [97,98]. Significant reductions in PAG concentrations during the late embryonic/early fetal period are associated with pregnancy failures in cattle [97,99]. PAG proteins also play an important role in implantation, placentogenesis, fetal antigen sequestering, and fetal–maternal interactions [97,100–102]. Modifications in circulating PAG concentrations also were associated with several parameters linked to pregnancy loss in cattle, including parity, artificial insemination service number, milk yield, and metabolic diseases [103].

The regulation of multicellular organism growth (GO:0040014) biological process encompasses any process that modulates the frequency, rate or extent of growth of the body of an organism so that it reaches its usual body size, while the midbrain development (GO:0030901) biological process encompass the process whose specific outcome is the progression of the midbrain over time, from its formation to the mature structure.

FINAL CONSIDERATIONS

This study is the first of its kind to bring out results characterizing genome-wide autozygosity in the main Nellore lineages. The average F_{PED} and F_{ROH} of different lengths were low in the studied population, however, the autozygotic proportion in the genome indicates moderate to high inbreeding levels. Low correlations between F_{PED}-F_{ROH} may be partly due to the relatively superficial depth of the pedigree, emphasizing

the concept that autozygosity based on ROH should be used as an accurate estimator of ancient individual inbreeding levels (BJELLAND et al., 2013; FERENČAKOVIĆ et al., 2011; GURGUL et al., 2016; PURFIELD et al., 2012). Overall, inbreeding coefficients were not high within the lineages, and the findings obtained in this study suggest that lineages displaying an eminent marginal contribution in the reference population also display the highest F_{ROH} values, i.e., Karvadi and Godhavari.

Genomic regions that are selection targets tend to generate autozygosity islands, and several of them have been described in the Nellore genome. Most remarkable is the clear evidence of autozygosity islands patterns within the lineages, suggesting that IBD genomic regions have been selected for the same traits over time. Autozygosity islands harbored enriched terms in which we highlight the defense response to bacteria (GO:0042742) and the negative regulation of erythrocyte differentiation (GO:0045647), which might help to better elucidate the better adaptation of indicine cattle in host environment given its association with immune responses mechanisms. Additionally, non-overlapping autozygosity islands within the lineages were found to contain genes related to cattle growth, reproduction, and meat and carcass quality traits. The results of this study give a comprehensive insight about the autozygosity patterns in the main Nellore lineages and their potential role in explaining selection for functionally important traits in cattle. Despite having different lineages within the Nellore breed, it has clearly shown that selection is going towards the same direction and different genetic patterns could not be described.

METHODS

Animals and genotyping

The animals used in this study comprise a dataset and progeny test program from the National Association of Breeders and Researchers (ANCP – Ribeirão Preto-SP, Brazil). The progeny test program headed by ANCP aims to disseminate semen of genetically superior Nellore young bulls evaluated for sexual precocity, growth, morphologic composition, feed efficiency, and carcass quality traits.

Nellore animals were genotyped with the low-density panel (CLARIFIDE® Nelore 2.0) containing over 20,000 markers (n=*7,729 animals*); GGP-LD BeadChip

(GeneSeek® Genomic Profiler 30K) that contains 30,106 markers (*n=201* animals); Illumina BovineSNP50® Beadchip (Illumina Inc., San Diego, CA, USA) containing 54,001 markers (*n=58* animals); GGPi BeadChip (GeneSeek® Genomic Profiler Indicus) that contains 74,153 markers (*n=487* animals); and with Illumina BovineHD BeadChip (Illumina Inc., San Diego, CA, USA) containing 777,962 markers (*n=911* animals). Imputation was implemented using the FIMPUTE 2.2 software [105] and all genotypes were imputed to a panel containing 735,044 markers. A reference population with 963 sires and dams genotyped with the Illumina BovineHD BeadChip (Illumina Inc., San Diego, CA, USA) was used. Prior imputation, markers were edited for call rate (<90%) for the genotyped and the reference populations. SNPs unsigned to any chromosome and those assigned to sexual chromosomes were removed from the dataset. After editing, a total of 9,386 animals and 735,044 SNP markers were retained for the analyses. Genotyped animals with lineages records (*n=8,646*) were categorized as follows: Karvadi Imp (*n=7,860*), Golias Imp (*n=290*), Godhavari Imp (*n=210*), Taj Mahal Imp (*n=150*), Akasamu Imp (*n=81*), and Nagpur Imp (*n=55*). Lineages were classified using the PEDIG package [106], which estimates the average consanguinity between a set of individuals and a reference group. The reference group encompassed founder's animals from the Nellore base population in which the Nellore lineages were derived from.

Runs of homozygosity

Individual ROH was identified using PLINK v1.90 software [107], which uses a sliding window approach to scan each individual's genotype at each marker position to detect homozygous segments [44]. The parameters and thresholds applied to define ROH were set as follows: a sliding window of 50 SNPs across the genome, a minimum number of 100 consecutive SNPs included in a ROH, a minimum ROH length of 1 Mb, a maximum gap between consecutive homozygous SNPs of 0.5 Mb, one SNP per 50 kb, and a maximum of five SNPs with missing genotypes and up to one heterozygous genotype in a ROH. ROH were classified into four length classes: 1-2, 2-4, 4-8, and >8 Mb, identified as $ROH_{1\text{-}2\ Mb}$, $ROH_{2\text{-}4\ Mb}$ $ROH_{4\text{-}8\ Mb}$, and $ROH_{>8\ Mb,}$ respectively. ROH

were performed separately for all genotyped animals (*n=9,386*) and for each Nellore lineage (*n=8,646*).

Pedigree and genomic inbreeding coefficients

Pedigree-based inbreeding coefficients (F_{PED}) were estimated using pedigree records from a dataset containing 45,917 animals born between 1934 and 2017. The pedigree dataset was provided by the National Association of Breeders and Researchers (ANCP – Ribeirão Preto-SP, Brazil). The average pedigree depth was approximately four generations, with a maximum depth value of nine. The F_{PED} was estimated for both datasets (*n=9,386* and *n=8,646*) through the software INBUPGF90 [108]. Genomic inbreeding coefficients based on ROH (F_{ROH}) were estimated for each animal and both datasets, according to the genome autozygotic proportion described by McQuillan et al. [21]:

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

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} ($F_{ROH1\text{-}2\ Mb}$, $F_{ROH2\text{-}4\ Mb}$, $F_{ROH4\text{-}8\ Mb}$, and $F_{ROH>8\ Mb}$) was calculated based on ROH distribution of four minimum different lengths (ROH_j): 1-2, 2-4, 4-8, and >8 Mb, respectively. A second measure of genomic inbreeding was calculated just for the whole dataset (*n=9,386*) using the Genomic relationship matrix (G) (F_{GRM}). The G matrix was calculated according to VanRaden et al. [109] as follows:

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

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. Spearman method was used to estimate correlations between the inbreeding measures.

Identification and gene prospection in autozygosity islands

Autozygosity islands were defined as regions where SNPs were outliers according to boxplot distribution for each autosome (Appendix 9A and 10A). A file generated by PLINK v1.90 software [107] which specifies how many times each SNP appeared in an ROH was used and regions displaying at least 100 consecutive outlier SNPs were then classified as an autozygosity island. Raw data regarding how many times each SNP appeared in an ROH was log-transformed (Log_{10}). Autozygosity islands were identified separately for all genotyped animals (*n=9,386*) and for each Nellore lineage (*n=8,646*).

The gene content of the autozygosity islands was identified using the UMD3.1 bovine genome assembly from the Ensembl BioMart tool [110]. Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 tool [30,31] was used to identify significant ($p \leq 0.01$) Gene Ontology (GO) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways using the list of genes from autozygosity islands and the *Bos taurus taurus* annotation file as background.

Autozygosity islands previously identified for the genotyped animals were overlapped with copy number variation regions (CNVRs) described for Nellore cattle by Lemos et al. [29]. Overlap analysis was carried out using the Bioconductor package *GenomicRanges* [111].

REFERENCES

1. ABIEC. Associação Brasileira das Indústrias Exportadoras de Carnes [Internet]. 2016 [cited 2017 Jun 16]. Available from: http://www.abiec.com.br/estatisticas
2. Turner JW. Genetic and biological aspects of Zebu adaptability. J. Anim. Sci. 1980;50:1201–5.
3. Hansen PJ. Physiological and cellular adaptations of zebu cattle to thermal stress. Anim. Reprod. Sci. 2004;82–83:349–60.

4. Jonsson NN. The productivity effects of cattle tick (Boophilus microplus) infestation on cattle, with particular reference to Bos indicus cattle and their crosses. Vet. Parasitol. 2006;137:1–10.
5. Santiago AA. O Guzerá. Recife: Tropical; 1984.
6. Brasil. Projeto de melhoramento genético da zebuinocultura: PROZEBU: 1978-1984. Belo Horizonte: Associação Brasileira dos Criadores de Zebu; 1978.
7. Oliveira JHF, Magnabosco C de U, Borges AMS. Nelore: base genética e evolução seletiva no Brasil. Distrito Federal: Embrapa Cerrados; 2002.
8. Magnabosco C de U, Cordeiro CMT, Trovo JB de F, Mariante A da S, Lôbo RB, Josahkian LA. Catálogo de linhagens do germoplasma zebuíno: raça Nelore. Brasília: Embrapa Recursos Genéticos e Biotecnologia; 1997.
9. Lôbo RB, Marcondes CR, Vozzi PA, Lima FP, Bezerra LAF, Zambianchi AR. A tecnologia da informação e a sustentabilidade da raça Nelore. 40 Reun. Anu. da Soc. Bras. Zootec. Santa Maria; 2003.
10. Bonin M de N, Ferraz JBS, Eler JP, Silva S da L, Rezende FM, Córdova D, et al. Características de carcaça e qualidade de carne em linhagens da raça Nelore Carcass and meat quality traits in lineages of Nellore breed. Ciência Rural. 2014;44:1860–6.
11. Ferraz JBS, Felício PE de. Production systems - An example from Brazil. Meat Sci. 2010;84:238–43.
12. Pereira RJ, Santana ML, Ayres DR, Bignardi AB, Menezes GRO, Silva LOC, et al. Inbreeding depression in Zebu cattle traits. J. Anim. Breed. Genet. 2016;133:523–33.
13. De Cara MÁR, Villanueva B, Toro MÁ, Fernández J. Using genomic tools to maintain diversity and fitness in conservation programmes. Mol. Ecol. 2013;22:6091–9.
14. Bosse M, Megens HJ, Madsen O, Crooijmans RPMA, Ryder OA, Austerlitz F, et al. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. Genome Res. 2015;25:970–81.
15. Ferenčaković M, Hamzic E, Gredler B, Curik I, Sölkner J. Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle. Agric. Conspec. Sci. 2011;76:325–9.

16. Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, et al. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. J. Anim. Breed. Genet. 2013;130:286–93.
17. Silió L, Rodríguez MC, Fernández A, Barragán C, Benítez R, Óvilo C, et al. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. J. Anim. Breed. Genet. 2013;130:349–60.
18. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet. 2014;46:110–21.
19. Zavarez LB, Utsunomiya YT, Carmo AS, Neves HHR, Carvalheiro R, Ferencakovic M, et al. Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. Front. Genet. 2015;6:1–8.
20. Peripolli E, Baldi F, da Silva MVGB, Irgang R, Lima ALF, R. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. BMC Genomics. 2018;19:34.
21. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. Am. J. Hum. Genet. 2008;83:359–72.
22. Karimi Z. Runs of Homozygosity patterns in Taurine and Indicine cattle breeds (Master thesis). Vienna: BOKU - University of Natural Resources and Life Sciences; 2013.
23. Zhang Q, Guldbrandtsen B, Bosse M, Lund MS, Sahana G. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genomics. 2015;16:542.
24. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. BMC Genet. 2012;13:70.
25. Kim ES, Cole JB, Huson H, Wiggans GR, Van Tassel CP, Crooker BA, et al. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. PLoS One. 2013;8:e80813.
26. Curik I, Ferenčaković M, Sölkner J. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci. 2014;166:26–34.
27. Kim ES, Sonstegard TS, Van Tassell CP, Wiggans G, Rothschild MF. The

relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. PLoS One. 2015;10:e0129967.

28. Scraggs E, Zanella R, Wojtowicz A, Taylor JF, Gaskins CT, Reeves JJ, et al. Estimation of inbreeding and effective population size of full-blood wagyu cattle registered with the American Wagyu Cattle Association. J. Anim. Breed. Genet. 2014;131:3–10.
29. Lemos MVA de, Piatto Berton M, Ferreira de Camargo GM, Peripolli E, de Oliveira Silva RM, Ferreira Olivieri B, et al. Copy number variation regions in Nellore cattle: evidences of environment adaptation. Livest. Sci. Elsevier B.V.; 2018;207:51–8.
30. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13.
31. Huang DW, Sherman BT, Lempicki R a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 2009;4:44–57.
32. Mastrangelo S, Tolone M, Di Gerlando R, Fontanesi L, Sardina MT, Portolano B. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. Animal. 2016;10:746–54.
33. 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. J. Appl. Genet. 2016;57:527–30.
34. Mcclure MC, Morsci NS, Schnabel RD, Kim JW, Yao P, Rolf MM, et al. A genome scan for quantitative trait loci influencing carcass , post-natal growth and reproductive traits in commercial Angus cattle. 2010;41:597–607.
35. Li C, Basarab J, Snelling WM, Benkel B, Murdoch B, Moore SS, et al. The identification of common haplotypes on bovine chromosome 5 within commercial lines of Bos taurus and their associations with growth traits. J. Anim. Sci. 2002;80:1187–94.
36. Kirkpatrick BW, Byla BM, Gregory KE. Mapping quantitative trait loci for bovine ovulation rate. Mamm. Genome. 2000;11:136–9.
37. Lien S, Karlsen A, Klemetsdal G, Våge DI, Olsaker I, Klungland H, et al. A primary

screen of the bovine genome for quantitative trait loci affecting twinning rate. Mamm. Genome. 2000;11:877–82.

38. Wiener P, Maclean I, Williams JL, Woolliams JA. Testing for the presence of previously identified QTL for milk production traits in new populations. Anim. Genet. 2000;31:385–95.
39. Mészáros G, Boison AS, Pérez O'Brien AM, Ferenčaković M, Curik I, da Silva MVGB, et al. Genomic analysis for managing small and endangered populations : a case study in Tyrol Grey cattle. Front. Genet. 2015;6.
40. Kirin M, McQuillan R, Franklin CS, Campbell H, Mckeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. PLoS One. 2010;5:e13996.
41. Szmatoła T, Gurgul A, Ropka-molik K, Jasielczuk I, Tomasz Z, Bugno-poniewierska M. Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. Livest. Sci. 2016;188:72–80.
42. Barbosa ACB, Malhado CHM, Carneiro PLS, Muniz LMS, Ambrosini DP, Carrillo JA, et al. Population structure of nellore cattle in northeastern brazil. Rev. Bras. Zootec. 2013;42:639–44.
43. Santana ML, Oliveira PS, Pedrosa VB, Eler JP, Groeneveld E, Ferraz JBS. Effect of inbreeding on growth and reproductive traits of Nellore cattle in Brazil. Livest. Sci. 2010;131:212–7.
44. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics. 2011;12:460.
45. Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet. 2006;2:0316–25.
46. Saura M, Fernández A, Varona L, Fernández AI, de Cara MÁR, Barragán C, et al. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. Genet. Sel. Evol. 2015;47:1.
47. Pryce JE, Haile-Mariam M, Goddard ME, Hayes BJ. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genet. Sel. Evol. 2014;46:71.

48. Zhang Q, Calus MPL, Guldbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet. 2015;16:88.
49. Pereira JCC. Melhoramento genético aplicado à produção animal. Belo Horizonte: Fundação de Estudo e Pesquisa em Medicina Veterinária e Zootecnia; 1996.
50. Sölkner J, Karimi Z, Pérez O'Brien AM, Mészáros G, Eaglen S, Boison SA, et al. Extremely Non-uniform : Patterns of Runs of Homozygosity in Bovine Populations. 10th World Congr. Genet. Appl. to Livest. Prod. Vancouver; 2014.
51. Gaspa G, Marras G, Sorbolini S, Marsan PA, Willians JL, Valentini A, et al. Genome-Wide Homozygosity in Italian Holstein Cattle using HD SNP Panel. 10th World Congr. Genet. Appl. to Livest. Prod. Vancouver; 2014.
52. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet. 2012;91:275–92.
53. Sadkowski T, Jank M, Zwierzchowski L, Siadkowska E, Oprzdek J, Motyl T. Gene expression profiling in skeletal muscle of Holstein-Friesian bulls with single-nucleotide polymorphism in the myostatin gene 5'-flanking region. J. Appl. Genet. 2008;49:237–50.
54. McPherron AC, Lawler AM, Lee S-J. Regulation of skeletal muscle mass in mice by a new TGF-p superfamily member. Nature. 1997;387:83–90.
55. Lehnert SA, Reverter A, Byrne KA, Wang Y, Nattrass GS, Hudson NJ, et al. Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. BMC Dev. Biol. 2007;7:95.
56. D'Andre CH, Wallace P, Shen X, Nie Q, Yang G, Zhang X. Genes Related to Economically Important Traits in Beef Cattle. Asian J. Anim. Sci. 2011;5:34–45.
57. Akakabe Y, Koide M, Kitamura Y, Matsuo K, Ueyama T, Matoba S, et al. Ecscr regulates insulin sensitivity and predisposition to obesity by modulating endothelial cell functions. Nat. Commun. Nature Publishing Group; 2013;4:2389.
58. Risau W. Mechanisms of angiogenesis. Nature. 1997;386:671–4.
59. Felius M. Cattle Breeds: an Encyclopaedia. First. Doetinchem, Netherlands: Misset; 1995.
60. Karthickeyan SMK, Kumarasamy P, Sivaselvam SN, Saravanan R, Thangaraju P.

Analysis of microsatellite markers in Ongole breed of cattle. Indian J. Biotechnol. 2008;7:113–6.

61. Sanders JO. History and development of zebu cattle in the United States. J. Anim. Sci. New Orleans; 1980;50:1188–200.
62. Delbes G, Yanagiya A, Sonenberg N, Robaire B. PABP Interacting Protein 2 (Paip2) Is a Major Translational Regulator Involved in the Maturation of Male Germ Cells and Male Fertility. Biol. Reprod. Oxford University Press; 2009;81:167.
63. Guan D, Luo N, Tan X, Zhao Z, Huang Y, Na R, et al. Scanning of selection signature provides a glimpse into important economic traits in goats (Capra hircus). Sci. Rep. 2016;6:36372.
64. Silva-Vignato B, Coutinho LL, Cesar ASM, Poleti MD, Regitano LCA, Balieiro JCC. Comparative muscle transcriptome associated with carcass traits of Nellore cattle. BMC Genomics. 2017;18:506.
65. Mota RR, Guimarães SEF, Fortes MRS, Hayes B, Silva FF, Verardo LL, et al. Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nellore cattle. J. Anim. Breed. Genet. 2017;134:484–92.
66. Utsunomiya YT, do Carmo AS, Carvalheiro R, Neves HH, Matos MC, Zavarez LB, et al. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. BMC Genet. 2013;14:52.
67. van Hulzen KJE, Schopen GCB, van Arendonk JAM, Nielen M, Koets AP, Schrooten C, et al. Genome-wide association study to identify chromosomal regions associated with antibody response to Mycobacterium avium subspecies paratuberculosis in milk of Dutch Holstein-Friesians. J. Dairy Sci. Elsevier; 2012;95:2740–8.
68. Taye M, Lee W, Jeon S, Yoon J, Dessie T, Hanotte O, et al. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. Mamm. Genome. 2017;28:528–41.
69. Sorbolini S, Bongiorni S, Cellesi M, Gaspa G, Dimauro C, Valentini A, et al. Genome wide association study on beef production traits in Marchigiana cattle

breed. J. Anim. Breed. Genet. 2017;134:43–8.

70. Guo Y, Zhang X, Huang W, Miao X. Identification and characterization of differentially expressed miRNAs in subcutaneous adipose between Wagyu and Holstein cattle. Sci. Rep. 2017;7:44026.
71. Borowska A, Reyer H, Wimmers K, Varley PF, Szwaczkowski T. Detection of pig genome regions determining production traits using an information theory approach. Livest. Sci. 2017;205:31–5.
72. Schellander K. Identifying genes associated with quantitative traits in pigs: integrating quantitative and molecular approaches for meat quality. Ital. J. Anim. Sci. 2009;8:19–25.
73. Utsunomiya YT, Pérez O'Brien AM, Sonstegard TS, Van Tassell CP, do Carmo AS, Mészáros G, et al. Detecting Loci under Recent Positive Selection in Dairy and Beef Cattle by Combining Different Genome-Wide Scan Methods. PLoS One. 2013;8:e64280.
74. Cong X, Doering J, Mazala DAG, Chin ER, Grange RW, Jiang H. The SH3 and cysteine-rich domain 3 (Stac3) gene is important to growth, fiber composition, and calcium release from the sarcoplasmic reticulum in postnatal skeletal muscle. Skelet. Muscle. 2016;6:17.
75. Rincon G, Farber EA, Farber CR, Nkrumah JD, Medrano JF. Polymorphisms in the STAT6 gene and their association with carcass traits in feedlot cattle. Anim. Genet. 2009;40:878–82.
76. Kawaguchi F, Kigoshi H, Nakajima A, Matsumoto Y, Uemoto Y, Fukushima M, et al. Pool-based genome-wide association study identified novel candidate regions on BTA9 and 14 for oleic acid percentage in Japanese Black cattle. Anim. Sci. J. 2018;
77. Lindholm-Perry AK, Kuehn LA, Smith TPL, Ferrell CL, Jenkins TG, Freetly HC, et al. A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle. Anim. Genet. 2012;43:216–9.
78. Porto Neto LR, Bunch RJ, Harrison BE, Barendse W. Variation in the XKR4 gene was significantly associated with subcutaneous rump fat thickness in indicine and composite cattle. Anim. Genet. 2012;43:785–9.

79. Brisard D, Desmarchais A, Touzé J-L, Lardic L, Freret S, Elis S, et al. Alteration of energy metabolism gene expression in cumulus cells affects oocyte maturation via MOS–mitogen-activated protein kinase pathway in dairy cows with an unfavorable “Fertil−” haplotype of one female fertility quantitative trait locus. Theriogenology. 2014;81:599–612.

80. Reverchon M, Rame C, Bunel A, Chen W, Froment P, Dupont J. VISFATIN (NAMPT) Improves in Vitro IGF1-Induced Steroidogenesis and IGF1 Receptor Signaling Through SIRT1 in Bovine Granulosa Cells. Biol. Reprod. 2016;94:54.

81. Höglund JK, Sahana G, Guldbrandtsen B, Lund MS. Validation of associations for female fertility traits in Nordic Holstein, Nordic Red and Jersey dairy cattle. BMC Genet. BioMed Central; 2014;15:8.

82. Li CH, Gao Y, Wang S, Xu FF, Dai LS, Jiang H, et al. Expression pattern of JMJD1C in oocytes and its impact on early embryonic development. Genet. Mol. Res. 2015;14:18249–58.

83. Gao Y, Gautier M, Ding X, Zhang H, Wang Y, Wang X, et al. Species composition and environmental adaptation of indigenous Chinese cattle. Sci. Rep. 2017;7:16196.

84. dos Santos FC, Peixoto MGCD, Fonseca PA de S, Pires M de FÁ, Ventura RV, Rosse I da C, et al. Identification of Candidate Genes for Reactivity in Guzerat (Bos indicus) Cattle: A Genome-Wide Association Study. Davoli R, editor. PLoS One. Public Library of Science; 2017;12:e0169163.

85. Garza-Brenner E, Sifuentes-Rincón AM, Randel RD, Paredes-Sánchez FA, Parra-Bracamonte GM, Arellano Vera W, et al. Association of SNPs in dopamine and serotonin pathway genes and their interacting genes with temperament traits in Charolais cows. J. Appl. Genet. 2017;58:363–71.

86. Taye M, Lee W, Caetano-Anolles K, Dessie T, Hanotte O, Mwai OA, et al. Whole genome detection of signature of positive selection in African cattle reveals selection for thermotolerance. Anim. Sci. J. 2017;88:1889–901.

87. Kim E-S, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity. 2016;116:255–64.

88. Meade KG, Cormican P, Narciandi F, Lloyd A, O’Farrelly C. Bovine β-defensin

gene family: opportunities to improve animal health? Physiol. Genomics. 2014;46:17–28.

89. Dorin JR, Barratt CLR. Importance of β-defensins in sperm function. Mol. Hum. Reprod. 2014;20:821–6.
90. Nelson RA. The immune-adherence phenomenon; an immunologically specific reaction between microorganisms and erythrocytes leading to enhanced phagocytosis. Science. 1953;118:733–7.
91. Passantino L, Altamura M, Cianciotta A, Patruno R, Tafaro A, Jirillo E, et al. Fish immunology. I. Binding and engulfment of candida albicans by erythrocytes of rainbow trout (Salmo gairdneri richardson). Immunopharmacol. Immunotoxicol. 2002;24:665–78.
92. Passantino L, Massaro MA, Jirillo F, Di Modugno D, Ribaud MR, Di Modugno G, et al. Antigenically Activated Avian Erythrocytes Release Cytokine-Like Factors: A Conserved Phylogenetic Function Discovered in Fish. Immunopharmacol. Immunotoxicol. 2007;29:141–52.
93. Passantino L, Altamura M, Cianciotta A, Jirillo F, Ribaud MR, Jirillo E, et al. Maturation of Fish Erythrocytes Coincides with Changes in their Morphology, Enhanced Ability to Interact with Candida albicans and Release of Cytokine-Like Factors Active Upon Autologous Macrophages. Immunopharmacol. Immunotoxicol. 2004;26:573–85.
94. Khansefid M, Millen CA, Chen Y, Pryce JE, Chamberlain AJ, Vander Jagt CJ, et al. Gene expression analysis of blood, liver, and muscle in cattle divergently selected for high and low residual feed intake. J. Anim. Sci. 2017;95:4764–75.
95. Goszczynski D, Molina A, Terán E, Morales-Durand H, Ross P, Cheng H, et al. Runs of homozygosity in a selected cattle population with extremely inbred bulls: Descriptive and functional analyses revealed highly variable patterns. PLoS One. 2018;13:e0200069.
96. Commun L, Velek K, Barbry J-B, Pun S, Rice A, Mestek A, et al. Detection of pregnancy-associated glycoproteins in milk and blood as a test for early pregnancy in dairy cows. J. Vet. Diagnostic Investig. 2016;28:207–13.
97. Pohler KG, Geary TW, Johnson CL, Atkins JA, Jinks EM, Busch DC, et al. Circulating bovine pregnancy associated glycoproteins are associated with late

embryonic/fetal survival but not ovulatory follicle size in suckled beef cows1. J. Anim. Sci. 2013;91:4158–67.

98. Pohler KG, Pereira MHC, Lopes FR, Lawrence JC, Keisler DH, Smith MF, et al. Circulating concentrations of bovine pregnancy-associated glycoproteins and late embryonic mortality in lactating dairy herds. J. Dairy Sci. 2016;99:1584–94.
99. Pohler KG, Peres RFG, Green JA, Graff H, Martins T, Vasconcelos JLM, et al. Use of bovine pregnancy-associated glycoproteins to predict late embryonic mortality in postpartum Nelore beef cows. Theriogenology. 2016;85:1652–9.
100. Wooding FBP, Roberts RM, Green JA. Light and electron microscope immunocytochemical studies of the distribution of pregnancy associated glycoproteins (PAGs) throughout pregnancy in the cow: possible functional implications. Placenta. 2005;26:807–27.
101. Breukelman SP, Perényi Z, Taverne MAM, Jonker H, van der Weijden GC, Vos PLAM, et al. Characterisation of pregnancy losses after embryo transfer by measuring plasma progesterone and bovine pregnancy-associated glycoprotein-1 concentrations. Vet. J. 2012;194:71–6.
102. Wallace RM, Pohler KG, Smith MF, Green JA. Placental PAGs: gene origins, expression patterns, and use as markers of pregnancy. Reproduction. 2015;149:R115–26.
103. Mercadante PM, Ribeiro ES, Risco C, Ealy AD. Associations between pregnancy-associated glycoproteins and pregnancy outcomes, milk yield, parity, and clinical diseases in high-producing dairy cows. J. Dairy Sci. 2016;99:3031–40.
104. Bjelland DW, Weigel K a, Vukasinovic N, Nkrumah JD. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J. Dairy Sci. Elsevier; 2013;96:4697–706.
105. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014;15:478.
106. Boichard D. PEDIG : A Fortran package for pedigree analysis suited for large populations. 7th World Congr. Genet. Appl. to Livest. Prod. Montpellier; 2002.
107. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007;81:559–75.

108. Aguilar I, Misztal I. Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. J. Dairy Sci. 2008;91:1669–72.

109. VanRaden PM, Olson KM, Wiggans GR, Cole JB, Tooker ME. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 2011;94:5673–82.

110. Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. BioMart central portal - Unified access to biological data. Nucleic Acids Res. 2009;37:23–7.

111. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. 2013;9:e1003118.

FIGURES

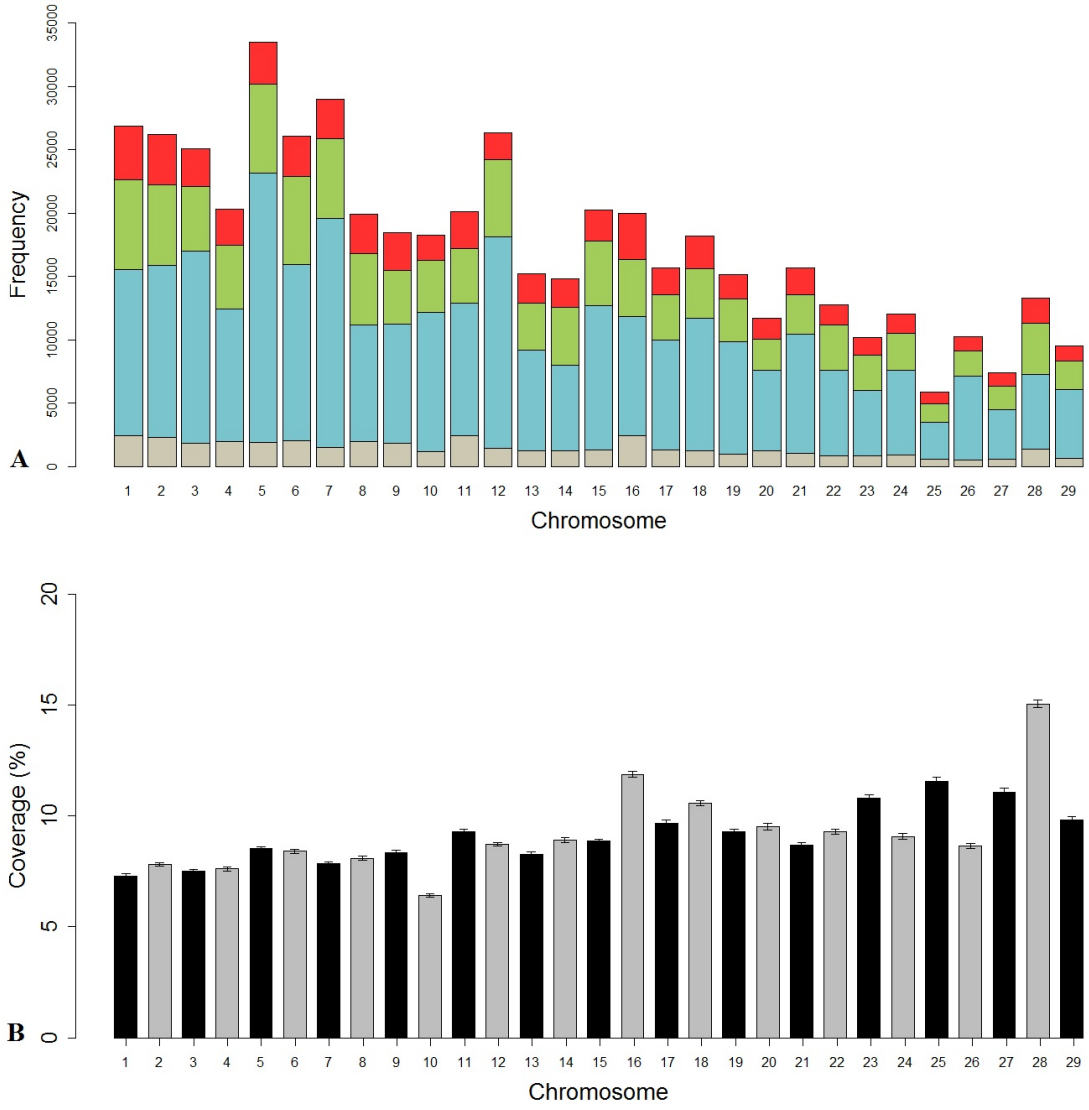


Figure 1. Runs of homozygosity distribution and coverage for each chromosome in Nellore cattle. 1A. Frequency distribution of the number of ROH in different length classes: blue ($ROH_{1\text{-}2\ Mb}$), green ($ROH_{2\text{-}4\ Mb}$), red ($ROH_{4\text{-}8Mb}$), and grey ($ROH_{>8\ Mb}$). 1B. Average percentage of chromosome coverage by runs of homozygosity of minimum length of 1 Mb. The error bars indicate standard error.

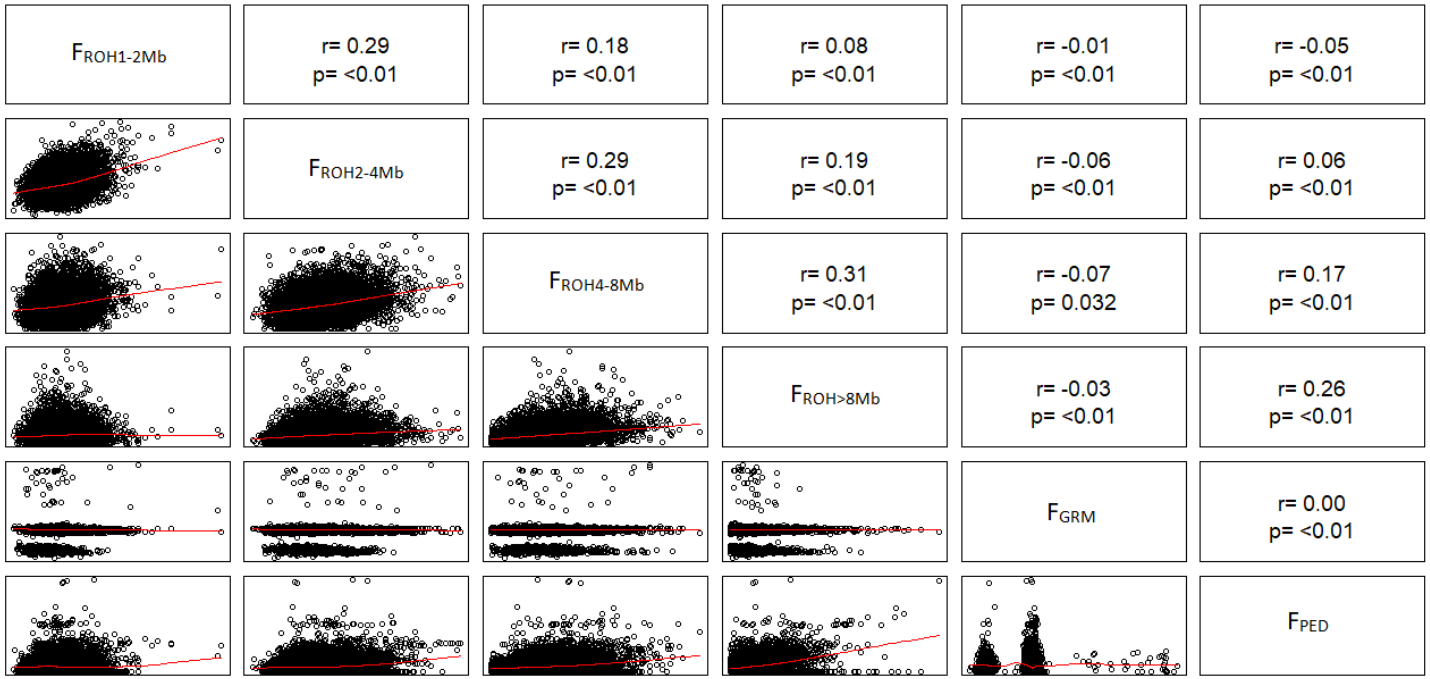


Figure 2. Scatterplots (lower panel) and Spearmann's correlations (upper panel) of genomic inbreeding coefficients F_{ROH} ($F_{ROH\ 1\text{-}2\ Mb}$, $F_{ROH\ 2\text{-}4\ Mb}$, $F_{ROH\ 4\text{-}8\ Mb}$, and $F_{ROH\ >8\ Mb}$) and F_{GRM}, and pedigree-based inbreeding coefficients (F_{PED}).

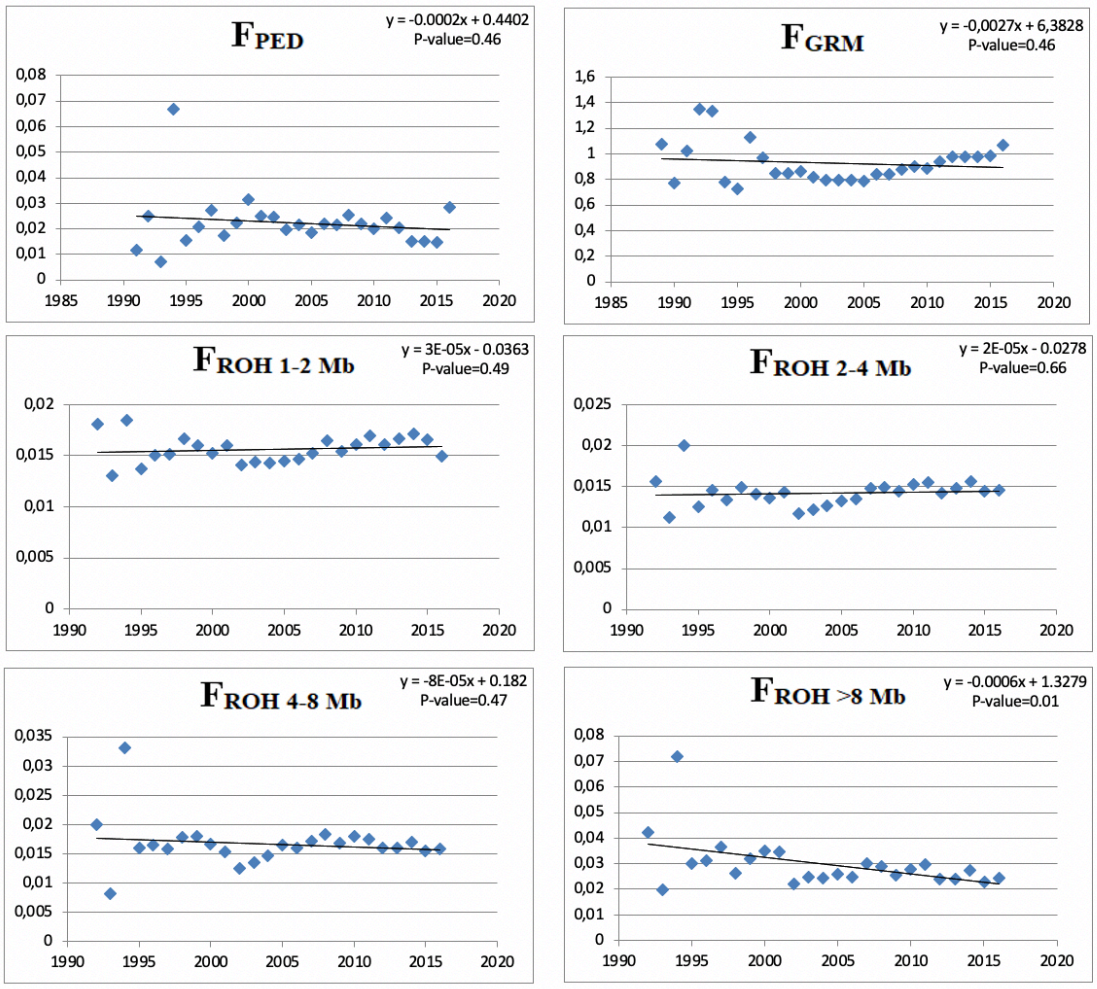


Figure 3. Inbreeding evolution over the past 30 years for pedigree-based inbreeding (F_{PED}), genomic relationship matrix approach (F_{GRM}), and F_{ROH} ($F_{ROH1–2\ Mb}$, $F_{ROH2–4\ Mb}$, $F_{ROH4–8\ Mb}$, and $F_{ROH>8\ Mb}$) coefficients and their respective regression equations and p-values. The X-axis represents the year and the Y-axis shows the inbreeding coefficients. Each blue dot represents the inbreeding average per year.

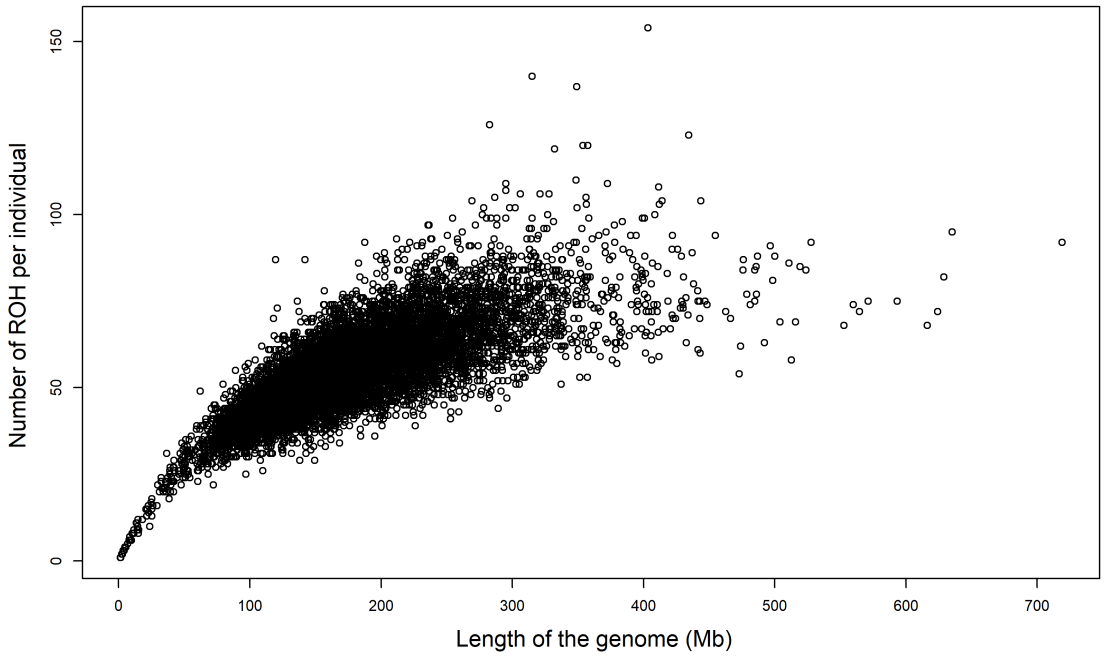


Figure 4. Relationship between the number of runs of homozygosity (ROH) per individual and the total length of the genome covered by them. Each hollow circle stands for one animal.

CAPÍTULO 3 – GENOME-WIDE SCAN FOR RUNS OF HOMOZYGOSITY IN COMPOSITE MONTANA TROPICAL® BEEF CATTLE[1]

ABSTRACT

The aim of this study was to assess the distribution of runs of homozygosity (ROH) and autozygosity islands in the composite Montana Tropical® beef cattle to explore hotspot regions which could better characterize the different biological types within the composite breed. Montana animals (*n=1,436*) were genotyped with the GGP-LD BeadChip (~30,000 markers). ROH were identified in every individual using PLINK v1.90 software. Medium and long ROH prevailed in the genome, which accounted for approximately 74% of all ROH detected. On average, 2.0% of the genome was within ROH, agreeing with the pedigree-based inbreeding coefficient. Montana cattle with a higher proportion of productive breed types showed the highest number of autozygosity islands *(n=17*), followed by those with a higher proportion of breeds adapted to tropical environments (*n=15*). Enriched terms ($p<0.05$) associated with the immune and inflammatory response, homeostasis, reproduction, mineral absorption, and lipid metabolism were described within the autozygosity islands. In this regard, over-represented GO terms and KEGG pathways described in this population may play a key role in providing information to explore the genetic and biological mechanisms together with the genomic regions underlying each biological type that favored their optimal performance ability in tropical and subtropical regions.

Key-words: Autozygosity, Brazilian cattle, *Bos taurus indicus*, *Bos taurus taurus*, crossbreed, heterosis

[1] Este capítulo corresponde ao artigo científico publicado na revista Journal of Animal Breeding and Genetics 2019, doi: 10.1111/jbg.12428

INTRODUCTION

Most livestock production in the world occurs in tropical and subtropical areas, in a wide range of heterogeneous production systems that can vary from grassland-based to feedlot systems. Animal husbandry faces many conflicting challenges since several environmental factors can affect livestock production, especially in tropical regions where the air temperature and relative humidity directly influence the animal's production potential (Marino et al., 2016). Given the variable climates and landscapes, it is essential to match the animal biological type to the environment of which it will be raised, increasing its optimal performance ability to the challenging environment. Climatic adaptation in cattle is a complex issue, and there are strong differences between breeds regarding heat tolerance (Beatty et al., 2006; Cartwright, 1955; Renaudeau et al., 2012; Ribeiro et al., 2009) and other efficiency and adaptive-related traits (Prayaga et al., 2009; Wolcott, Johnston, & Barwick, 2014).

The Montana Tropical® is a composite breed developed for tropical and sub-tropical beef cattle systems under grazing conditions. The composite system of the Montana Tropical® beef cattle propose the formation of clusters defined by biological types according to likeness, physiology, growth and reproduction traits, combining both *Bos taurus indicus* and *Bos taurus taurus* individuals. The base population is mainly centered on four different biological types defined as the NABC system, where: **N** are *Bos taurus indicus* cattle breeds already adapted to tropical conditions (heat tolerance, resistance to parasites, and poor feeding management); **A** are *Bos taurus taurus* cattle breeds known by their fertility and adaptive traits under tropical conditions; **B** are *Bos taurus taurus* British breeds notorious for sexual precocity, carcass quality traits, and high growth rate; and **C** are European Continental breeds recognized by their high growth rates and carcass quality traits.

The composite Montana Tropical® beef cattle can be classified into sixteenths of the breed proportion from the NABC system. In this regard, the traditional cattle have the same proportions of NABC biological types (4:4:4:4, N=4, A=4, B=4, and C=4), always summing up a value of 16 in the total composition. However, the composition of these cattle may vary due to regional climates and breeder's preference, and as a result, they can be empirically classified into two main biological

types (adaptive and productive) given the proportion of the NABC biological types that make them up. The adaptive group has a high proportion (≥ 50%) of adapted (**A**) biological types breeds (i.e., 4:8:2:2 and 4:8:4:0), whereas animals that present <50% of **A** together with a high proportion of **B** and **C** productive biological type breeds (i.e., 4:6:2:4 and 4:6:4:2) are classified as productive. At the beginning of the breed establishment, several breeds have been used to make up the genetic basis of the Montana Tropical® beef cattle, however, fewer breeds are predominant within the composite breed nowadays (Nellore, Senepol, Bonsmara, Limousin, and Hereford). It is noteworthy to highpoint that Montana animals are well establish and now they can be used as a purebred without the need of any ongoing crossbreeding program.

The great limiting factor of newly composite programs, such as the composite Montana Tropical® beef cattle which started in 1994 (Ferraz, Eller, Dias, & Golden, 2002), is the effective population size when compared to ancient breeds and the availability of genomic information. In this regard, it is essential to define mating strategies to preserve the genetic diversity and avoid high inbreeding rates (Zhang, Calus, Guldbrandtsen, Lund, & Sahana, 2015) so as to maintain long-term viability and sustainability of breeding programs. One of the main advantages of a composite breed is that it maintains heterosis over time we normally associate with continuous crossbreeding and it also explores the genetic differences among breeds such as complementarity to achieve an optimum additive genetic composition (Gregory, Cundiff, & Koch, 1993, 1999).

With the widespread use of whole-genome marker panels, an increasing interest in identifying autozygosity from molecular information has aroused. Autozygosity occurs when chromosomal segments identical by descent (IBD) arising from a common ancestor are inherited from both parents on to the offspring genome (Broman & Weber, 1999), resulting in continuous IBD homozygous segments characterized as runs of homozygosity (ROH) (Gibson, Morton, & Collins, 2006). The autozygosity based on ROH can disclose the genetic relationships among individuals, being an accurate estimator for detecting the effects of inbreeding (Ferenčaković, Hamzić, et al., 2013; Ferenčaković, Hamzić, Gredler, Curik, & Sölkner, 2011). Besides, it can reveal selection pressure events (Kim et al., 2013; Zhang, Guldbrandtsen, Bosse, Lund, & Sahana, 2015) since selection is one of the main forces triggering

homozygous stretches on the genome (Marras et al., 2015). The selection also tends to generate autozygosity islands, which can be defined as ROH shared regions among individuals 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). ROH have not been widely applied in crossbred or composite populations, however, Howard et al. (2016) characterized the frequency of ROH in a swine population within purebred breeds and its persistence within the crossbred progeny.

The aim of this study was to assess the distribution of ROH in the composite Montana Tropical® beef cattle to describe the genome-wide autozygosity. It attempts also to investigate ROH hotspot regions for traces of selection and gene content which could better characterize the different biological types contributing to the composite Montana Tropical® beef cattle raised in tropical and subtropical regions.

MATERIAL AND METHODS

Samples, genotyping and data editing

The animals used in this study comprise a dataset from the composite Montana Tropical® cattle breeding program. Montana animals (*n=1,436*) were genotyped with the GeneSeek® Genomic Profile Low-Density BeadChip containing over 30,105 markers. Animals were sampled from 14 farms located in Brazil (South, Southeast, and Midwest regions) and one in Uruguay. The biological type composition, according to the NABC system, for the animals sampled in this study is described in table 1. For all samples, markers unsigned to any chromosome and those assigned to sexual chromosomes were removed from the dataset. Additionally, markers and samples were edited for call rate frequency higher than 0.90.

Table 1. The biological type composition according to the NABC system for the composite Montana Tropical® beef cattle sampled in this study.

Number of samples	Biological type	Biological type proportion[1]			
		N	**A**	**B**	**C**
155	Productive/Adaptive	4	4	4	4
40	Productive	4	6	2	4
228	Productive	4	6	4	2
769	Adaptive	4	8	2	2
244	Adaptive	4	8	4	0

[1] Comprises the NABC system classification based on pedigree records: **N** are *Bos taurus indicus* cattle breeds already adapted to tropical conditions, **A** are *Bos taurus taurus* cattle breeds known by their adaptive traits under tropical conditions, **B** are *Bos taurus taurus* British breeds, and **C** are European Continental breeds.

Effective population size

The effective population size (*Ne*) was estimated using the SNP1101 v1.0 software (Sargolzaei, 2014). The analysis was based on the extent of linkage disequilibrium (LD) using the r^2 statistic (Sved, 1971), represented as follows:

$$N_e = \left[\left(\frac{1}{E(r^2)}\right) - 1\right]\frac{1}{4c}$$

where *c* is the distance in Morgans between two markers estimated for each chromosome in the LD. The $E(r^2)$ is the expected r^2 at distance *c,* calculated as follows:

$$E(r^2) = \frac{1}{1 + 4N_e c}$$

Each genetic distance (*c*) corresponds to a value of *t* generations in the past (Hayes, Visscher, McPartlan, & Goddard, 2003), obtained as follows:

$$t = \frac{1}{2c}$$

The *Ne* was investigated at four time points: 5, 10, 20, and 50 generations ago. Studies have shown that including markers with low minor allele frequencies (MAF) can bias LD estimates (Espigolan et al., 2013; Goddard, Hopkins, Hall, & Witte, 2000; Qanbari et al., 2010), therefore a MAF threshold of 0.01 was applied on the data for this analysis. After quality control, a total of 27,560 markers and 1,391 samples were left for *Ne* analysis.

Pedigree-based inbreeding coefficient

Pedigree-based inbreeding coefficients (F_{PED}) were estimated using pedigree records from a dataset containing information from 6,169 sires and 366,353 dams. The pedigree data was provided by the Animal Breeding and Biotechnology Group of the College of Animal Science and Food Engineering (Pirassununga, São Paulo, Brazil). The pedigree ranged from one to nine generations. The F_{PED} was estimated through the software INBUPGF90 (Aguilar & Misztal, 2008).

Runs of homozygosity

Runs of homozygosity (ROH) were estimated in every individual using PLINK v1.90 software (Purcell et al., 2007) and no pruning was performed based on MAF. High LD estimates lead to short and common ROH throughout the genome (Purfield, Berry, McParland, & Bradley, 2012), whereas a low LD value permits the identification of short segments that are more likely to be IBD rather than derived from LD. In this regard, the average LD estimate (0.13) for all autosomes was used to determine the minimum length of a ROH, allowing us to lower down the minimum length of an autozygous segment to 0.5 Mb. The criterion and threshold used to define ROH are described in table 2.

Table 2. Preset parameters and criterion to define runs of homozygosity (ROH) in the composite Montana Tropical® beef cattle.

Parameters	Threshold
Sliding window (number of SNPs)	40
Minimum number of consecutive SNPs	15
Minimum length of a ROH	0.5 Mb
Maximum gap between consecutive homozygous SNPs	1 Mb
Density (SNP/Kb)	1/120
Missing genotypes	2
Heterozygous genotype	0

ROH were classified into four length classes: 0.5-2, 2-4, 4-8, and >8 Mb, identified as $ROH_{0.5\text{-}2\ Mb}$, $ROH_{2\text{-}4\ Mb}$ $ROH_{4\text{-}8\ Mb}$, and $ROH_{>8\ Mb,}$ respectively. The average level of autozygosity per animal was calculated as the ratio of the total length of genome covered by ROH to the total length of the genome covered by autosomes markers, as proposed by McQuillan et al. (2008). After filtering, Montana animals held 27,929 markers and 1,391 samples for ROH analysis.

Detection of autozygosity islands

As described in the introduction, the composite Montana Tropical® beef cattle can be classified into sixteenths of the breed proportion from the NABC system. This categorization was based on pedigree records from 680,552 animals containing the breed composition. The animals were classified into two main biological types (adaptive and productive) according to their NABC system (Table 1). The first group comprised animals with a high proportion of adapted biological type breeds (**A**) (4:8:2:2 and 4:8:4:0). The second one also encompassed animals with a considerable proportion of adapted cattle, however, with a high proportion of British (**B**) and Continental (**C**) biological type breeds (4:6:2:4 and 4:6:4:2). The traditional composite Montana Tropical® beef cattle (4:4:4:4) was included in both biological types analysis as they have the same proportion of NABC biological types.

Autozygosity islands were identified using an outlier approach. The boxplot distribution for each autosome displaying the number of time each SNP fell within a ROH was used to define the regions where SNPs were outliers in the upper quartile

(Appendix 1B). A file generated by PLINK v1.90 software (Purcell et al., 2007) which specifies how many times each SNP appeared in an ROH was used and regions displaying at least 15 consecutive outlier SNPs were then classified as an autozygosity island. Autozygosity islands were identified separately for the adaptive and productive biological types groups.

Gene searching and functional annotation analysis

The gene content of the autozygosity islands for each biological type (adaptive and productive) was identified using the Ensembl Biomart tool (Haider et al., 2009) (Genes 94, *Bos taurus* UMD3.1). Database for Annotation, Visualization and Integrated Discovery (DAVID) v. 6.8 tool (Huang, Sherman, & Lempicki, 2009a, 2009b) was used to identify significant ($p<0.05$) Gene Ontology (GO) terms and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways using the list of genes from the autozygosity islands from each biological type and the *Bos taurus taurus* annotation file as background.

RESULTS AND DISCUSSION

Effective population size

The *Ne* obtained in this population was estimated from five to 50 generations ago (Figure 1), and its decay over time indicates that the ancestral population based on 50 past generations had a much larger *Ne* (*n=528* animals) compared to the most current generations. The *Ne* for the last five generations showed a value of 128 animals, falling within the minimum value of 50 individuals for any livestock species to ensure the viability and genetic improvement in breeding programs (FAO, 2004). Furthermore, the maintenance of a sufficiently large *Ne* is essential for retention of heterozygosity and heterosis in composite breeds (Gregory et al., 1999).

The average r^2 in all autosomes was 0.13 by considering a maximum distance of 100 kb between adjacent SNPs. Since there were no previous results from the composite Montana Tropical® beef cattle regarding LD analysis, our results were compared to those described for other cattle and composite breeds. Studies have

described a r^2 value of 0.17 for a distance of 100 kb in Nellore cattle (Espigolan et al., 2013), and values varying between 0.20 and 0.22 in Angus, 0.13 to 0.16 in Brahman and 0.15 to 0.26 in Limousin cattle breeds when considering a physical distance close or equal to 100 kb (McKay et al., 2007; Porto-Neto, Kijas, & Reverter, 2014). Additionally, a r^2 varying from 0.13 to 0.16 has been reported in composite cattle (Tropical Composite, Santa Gertrudis, and Belmont Red) within a distance of 70 kb between adjacent SNPs (Porto-Neto, Kijas, & Reverter, 2014).

Distribution of runs of homozygosity

ROH were identified in almost all Montana individuals with the exception of 60 samples. A total of 7,530 ROH were identified distributed among 1,331 Montana individuals with an average value of 5.65 ROH per animal. An average ROH length of 7.73 Mb was estimated across all the autosomes with a maximum value of 73.18 Mb in length (708 SNPs) on *Bos taurus* autosome (BTA) 11. Similar results regarding the average and maximum ROH length were reported by Mastrangelo et al. (2017) study in sheep using medium-density SNP array. According to the authors, values for total ROH length and number might have been underestimated since many ROH remain undetected when using low- and medium-density SNP array. Therefore, our results may be slightly biased since a low-density array was used to characterize ROH, not accurately identifying the total ROH number per animal due to the lack of power to detect these segments when using a shallow density panel.

The number of ROH per chromosome was greatest for BTA5 (452 segments) and the greatest fraction of chromosome covered with ROH was found on BTA25 (16.91% of chromosomal length within an ROH) (Figure 2). Our previous studies in indicine cattle (Peripolli, Metzger, et al., 2018; Peripolli, Stafuzza, et al., 2018) also have described the greatest number of ROH on BTA5, whereas others have found on BTA1 (Gurgul et al., 2016; Mastrangelo et al., 2016; Purfield, Berry, McParland, & Bradley, 2012).

ROH analysis for the different length classes revealed that medium ($ROH_{4\text{-}8\ Mb}$) and long ($ROH_{>8\ Mb}$) segments prevailed in the genome of the composite Montana Tropical® beef cattle, which accounted for approximately 74% of all ROH detected and

greatly contributed to 90% of the cumulative ROH length (Table 3). The high proportion of medium and long ROH described in our study might reflect the reduced power of low-density arrays in identifying ROH between 0.5 to 2 Mb in length (*n=327* segments), as discussed by Purfield and colleagues (Purfield, Berry, McParland, & Bradley, 2012). Additionally, by not allowing any heterozygous call within a ROH, long ROH might not have been overestimated. In fact, these results contradict those reported in cattle (Ferenčaković, Hamzić, et al., 2013; Ferenčaković, Hamzić, Gredler, Curik, & Sölkner, 2011; Marras et al., 2015; Peripolli, Metzger, et al., 2018; Peripolli, Stafuzza, et al., 2018; Szmatoła et al., 2016; Zhang, Calus, et al., 2015), sheep (Purfield, McParland, Wall, & Berry, 2017) and pigs (Saura et al., 2015), in which the total length of ROH was composed mostly of high number of shorter ROH. It is noteworthy to highlight that the inconsistency among the criteria for defining ROH make the comparison of ROH studies not straightforward. The lack of consensus allows different thresholds across studies (Howrigan, Simonson, & Keller, 2011; Ku, Naidoo, Teo, & Pawitan, 2011), and it may be responsible for bias in ROH-based estimates of autozygosity (Ferenčaković, Sölkner, & Curik, 2013). The studies described above reported a high number of shorter ROH, and most of them made use of medium-density arrays (50K). According to Ferenčaković, Hamzić et al. (2013), the 50K array tends to reveal an abundance of small segments, however, it overestimates the numbers of segments between 1 to 4 Mb, suggesting that it is not sensitive enough for its accurate determination. In this regard, a strict comparison has to be made when assessing different studies, taking into account the parameters used to define ROH since they may cause biased estimation.

It should be noted that it is unclear how frequently ROH persist in a crossbred population and whether longer ROH exist. In this context, the persistence of ROH in crossbred and composite population likely result in decreased heterozygosity for that region, which reduces the degree of heterosis. Furthermore, long ROH reduces the probability of creating new favorable haplotype combinations by recombination, then, managing these populations to maintain genetic diversity and reduce the length and frequency of ROH is a desirable effect regarding genetic diversity (Howard, Tiezzi, Huang, Gray, & Maltecca, 2016).

Table 3. Descriptive statistics of runs of homozygosity number (*n*ROH) and mean length (in Mb) for four different length classes ($ROH0_{0.5\text{-}2\ Mb}$, $ROH_{2\text{-}4\ Mb}$, $ROH_{4\text{-}8\ Mb}$, and $ROH_{>8\ Mb}$) in the composite Montana Tropical® beef cattle.

Class	***n* ROH**	**(%)**	**Mean Length**	**Cumulative ROH Length (%)**
$ROH_{0.5\text{-}2\ Mb}$	327	4.34	1.44	0.81
$ROH_{2\text{-}4\ Mb}$	1,655	21.98	3.15	8.97
$ROH_{4\text{-}8\ Mb}$	3,307	43.92	5.62	31.96
$ROH_{>8\ Mb}$	2,241	29.76	15.14	58.26

The extension and frequency of ROH can disclose the number of generations of inbreeding given the approximate correlation between the length of the ROH and the distance with the common ancestor due to recombination events. By considering that 1 cM equals to 1 Mb, the expected length of autozygous segments follows an exponential distribution with mean equal to 1/2g Morgans, where g is the number of generations since the common ancestor (Howrigan, Simonson, & Keller, 2011). Therefore, considering that $ROH_{>8\ Mb}$ are expected to correspond to the reference ancestral population dating six generations ago or less together with the higher frequencies of ROH in this length category, we can disclose that recent inbreeding was observed in the studied population. Additionally, the ROH pattern in this population is consistent with the recent development of the composite breed just in 1994 (Ferraz, Eller, Dias, & Golden, 2002), reinforcing the idea of not long past inbreeding events in such population. The small number of proven sires mated to disseminate the breed presumably triggered the autozygosity in this population, however, when assessing the proportion of the genome under autozygosity, an average value close to 2% was observed. Concurring with this result, F_{PED} estimates were low in this population, with a mean value of 0.6%. These results might reflect the recent establishment of the breed together with the introduction of new genes through genic combinations to explore the complementarity among the breeds within each biological type, resulting in decreased inbreeding rates. However, it should be taken into consideration that the average level of autozygosity described in here might not reflect the true level of autozygosity since

many ROH remain undetected when using a low-density panel, as discussed previously.

Animals exhibiting the same homozygous genome length displayed a variable number of ROH (Figure 3), and this pattern can be attributed as a consequence of the distinct distances from the common ancestor (Mészáros et al., 2015). Hence, when considering animals with the same homozygous genome length, we can infer that those displaying a lower number of ROH have a higher proportion of longer segments and then a decreased distance with the common ancestor than those exhibiting a higher number of ROH. The most extreme animal exhibited a ROH genome coverage encompassing 786.84 Mb of the total autosomal genome extension (UMD3.1) covered by markers (31.47% of the cattle genome). Similar results were described in several cattle breeds, whose findings reported a coverage varying from 25 to 29.20% of the cattle genome (Marras et al., 2015; Mastrangelo et al., 2016; Peripolli, Metzger, et al., 2018; Peripolli, Stafuzza, et al., 2018; Purfield, Berry, McParland, & Bradley, 2012; Szmatoła et al., 2016).

Autozygosity islands

Autozygosity islands were evident across the genome and their distributions varied in length and position across chromosomes for both biological types (Appendix 2B). The number of islands did not differ considerably between biological types, resulting in 15 islands identified for the adaptive type and 17 for the productive. Additionally, the longest island found on the adaptive biological type encompassed 5.95 Mb (199,195:6,154,638 bp) in length on BTA1. This region was screened for gene content and no genes with described functions were identified. For the productive biological type, the longest island was found covering 4.34 Mb (32861744:37203531 bp) in length on BTA22 and harbored five genes with described functions (*FAM19A4*, *FAM19A1*, *SUCLG2*, *KBTBD8*, and *LRIG1*).

Functional annotation of genes

A total of 487 protein-coding genes (adaptive: *n=273* and productive: *n=217*) were identified within the autozygosity islands regions using the bovine reference

genome assembly UMD3.1. Only three genes (*XKR4*, *MT1E*, and *CSMD3*) were identified in both biological types, and the first two are noteworthy to highlight given their role in cattle productive traits. The *XKR4* (XK, Kell blood group complex subunit-related family, member 4) gene has been associated with several economically important traits in beef cattle such as intramuscular fat (Ramayo-Caldas et al., 2014) and subcutaneous rump fat thickness (Bolormaa et al., 2011; Porto Neto, Bunch, Harrison, & Barendse, 2012). This gene has also been described to have functions associated with serum prolactin concentrations in Angus-Simmental-Charolais crossbred (Bastin et al., 2014), feed intake in crossbred steers (Lindholm-Perry et al., 2012), age at puberty in Brahman (Fortes et al., 2012), and backfat thickness (Silva et al., 2017), birth weight (Terakado et al., 2018) and meat tenderness (Magalhães et al., 2016) in Nellore cattle. The second gene (*MT1EI*, metallothionein 1E) encodes a protein that exhibits antioxidant activity (Chung, Hogstrand, & Lee, 2006) and displayed a significant negative correlation with dry matter intake in beef steers (Sun, Zhao, Zhou, Chen, & Guan, 2019).

The analyzes set to study the functional enrichment using the DAVID tool revealed significant ($p<0.05$) GO terms and KEGG pathways for each biological type (Appendix 3B and 4B), and it was used to give an insight about the predicted gene networks. No significant GO term neither KEGG pathway was found to be shared between biological types. For the adaptive biological type, the analysis showed 20 GO terms and six KEGG pathways as significant ($p<0.05$, Appendix 3B) for the gene list. Among them, we highlight terms involved in the immune system activation in response to pathogens and those associated with adaptive traits related to homeostasis, briefly described below.

The type I interferon receptor activity (GO:0004905) and type I interferon signaling pathway (GO:0060337) terms have functions linked to molecular signals that act to initiate changes in the cell activity to promote the first line of defense against viral infection, i.e., foot-and-mouth disease virus (Ma et al., 2018), bovine herpesvirus 1 (Jones, 2019), and bovine viral diarrhea virus (Van Wyk, Snider, Scruten, van Drunen Littel-van den Hurk, & Napper, 2016). The natural killer cell mediated cytotoxicity (bta04650) was identified in the adaptive biological type associated with the immune system activities since natural killer cells are lymphocytes of the innate immune system

involved in early defense against both allogeneic and autologous cells undergoing infection with bacteria, viruses or parasites. The Jak-STAT signaling pathway (bta04630) is one pleiotropic cascade used to transduce several signals for the development and homeostasis in animals, acting as a central pathway for the improvement and function of the immune system and playing important roles in other biological systems (Liongue, O'Sullivan, Trengove, & Ward, 2012).

The blood coagulation, fibrin clot formation (GO:0072378), platelet activation (GO:0030168), complement and coagulation cascades (bta04610), respiratory chain (GO:0070469), oxidoreductase activity (GO:0016491), and cAMP signaling pathway (bta04024) were identified as overrepresented in the adaptive biological type which functions related to several physiological process in order to maintain homeostasis. Homeostasis is the state of equilibrium that the body reaches after responding to a foreign antigen, and the immune system plays a remarkable role by providing several functions to maintain homeostasis to respond effectively to a new antigenic challenge (Taniguchi et al., 2009; Van Parijs & Abbas, 1998).

The functional enrichment analysis for the productive biological type gene list covered a total of 17 GO terms and four KEGG pathways ($p<0.05$, Appendix 4B), in which we highlight those related to the immune system, reproductive, and productive functions. The GO terms related to inflammatory immune response included innate immune response (GO:0045087), lymphocyte chemotaxis (GO:0048247), monocyte chemotaxis (GO:0002548), CCR chemokine receptor binding (GO:0048020), and chemokine signaling pathway (bta04062). Chemokines are a family of small signaling peptides that have a crucial role in the development and maintenance of the innate and adaptive immune response against pathogens, showing vital roles in inflammation, disease modulation, and homeostasis (Widdison & Coffey, 2011). During the inflammation process, chemokines and adhesion molecules work together to promote differential leukocyte trafficking between circulation and the tissue through chemotaxis (Raman, Sobolik-Delmaire, & Richmond, 2011; Thelen, 2001). Chemokines are also involved in embryo implantation, development, and growth (Raman, Sobolik-Delmaire, & Richmond, 2011).

The endodermal cell differentiation (GO:0035987) is a biological process related to reproduction, in which relatively unspecialized cells acquire the specialized features

of endoderm cells, one of the three germ layers of the embryo. Platelet activation (bta04611) pathway plays a key role for primary homeostasis on the disruption of the integrity of vessel wall, and it has been associated with the establishment of pregnancy in cows through maternal platelet activation during early pregnancy (Kojima, Akagi, Zeniya, Shimizu, & Tomizuka, 1996).

Regarding the mineral absorption (bta04978) pathway, the animal's tissues need moderate quantities of some minerals (Ca, P, K, Na, Mg, S, and Cl) and smaller amounts of others (Mn, Fe, I, Co, Cr, Cu, Zn, and Se). Minerals in the diet must be absorbed by either passive or active transport systems across the gastrointestinal mucosa to enter into the blood flow for maintenance, growth, and reproduction. Among the minerals, Mg is vital to bone mineral formation, nerve, and muscle functions; Na plays a crucial role in the absorption of dietary sugars, amino acids, and water; Cl is the main anion related to the regulation of osmotic pressure, responsible for the low pH in the lumen of the abomasum; while Ca plays several roles in the animal's body, acting as a main component of bone and as an intracellular messenger in muscle contraction/relaxation allowing normal muscle and nerve functions (Goff, 2018).

Propanoate (propionic acid) metabolism (bta00640) is an essential metabolic pathway since propionate, a byproduct of ruminal fermentation, is the main precursor for glucose synthesis through gluconeogenesis in the liver of ruminants (Hocquette & Bauchart, 1999). In ruminants, glucose is one of the main forces triggering lipogenesis and marbling, and it also plays a key role in providing fuel for cellular and tissue functions. In this regard, mechanisms involved in the glucose absorption in the small intestine, liver gluconeogenesis, and glucose retention by the tissues are essential to produce high marbling meat and to increase meat quality traits in ruminants (Ladeira et al., 2018). Lee, Park, Kim, Yoon, & Seo (2014), studying metabolic differences between muscle and intramuscular adipose tissues in the *Longissimus dorsi* of Hanwoo beef cattle, identified the propanoate metabolism downregulated in the intramuscular adipose tissue. Nguyen, Zacchi, Schulz, Moore, & Fortes (2018) identified the propanoate metabolism pathway working together with other pathways influencing the adipose tissue in Brahman heifers.

Enriched terms associated with the immune response and homeostasis described for the adaptive biological type can help to better elucidate the mechanisms

underlying the cattle adaptation in hostile environments since the survivability benefit could be achieved with the evolutionary success of the immune system (Lemos et al., 2018; Stothard et al., 2011). Although it described terms related to the immune response, the productive biological type displayed terms associated with reproduction, glucose synthesis, and lipid functions as well, most likely reflecting the fixation of genomic regions harboring genes related to higher productive potential in those specialized breeds that compose the **B** and **C** biological types. According to Frisch & Vercoe (1979), there is an antagonism between some components of adaptation and production potential, which preclude the possibility to create an animal which has both high production potential coupled with a high level of adaptation.

FINAL CONSIDERATIONS

This study describes, for the first time, ROH patterns and autozygosity islands in composite Montana Tropical® beef cattle so as to better characterize the composite breed and the biological types within the NABC system. The ROH patterns described in this population suggested not long past inbreeding events, agreeing with such recent development of the composite Montana Tropical® beef cattle. Despite our results indicate recent inbreeding, autozygosity levels in such population were considered low, agreeing with F_{PED} estimate.

Autozygosity islands were assessed to better identify regions of the genome that have undergone directional selection and how they differ between biological types selected for different objectives within the NABC system. Over-represented GO terms and KEGG pathways provided important genomic information to explore the genetic mechanisms underlying the biological types and the environment that favored their optimal performance ability. The challenge to increase productivity in tropical environments is to combine in one breed several desirable traits, i.e., sexual precocity, resistance to parasites, heat tolerance and growth traits, adapted to pasture-based systems. In this regard, composite breeds such as the Montana Tropical® beef cattle can be an alternative for production systems in challenging environments as a unique genetic resource since it is possible for the breeder to choose what biological type should better adapt to the environmental conditions where the animals will be raised.

REFERENCES

Aguilar, I., & Misztal, I. (2008). Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. *Journal of Dairy Science*, *91*(4), 1669–1672. https://doi.org/10.3168/jds.2007-0575

Bastin, B. C., Houser, A., Bagley, C. P., Ely, K. M., Payton, R. R., Saxton, A. M., … Kojima, C. J. (2014). A polymorphism in XKR4 is significantly associated with serum prolactin concentrations in beef cows grazing tall fescue. *Animal Genetics*, *45*(3), 439–441. https://doi.org/10.1111/age.12134

Beatty, D. T., Barnes, A., Taylor, E., Pethick, D., McCarthy, M., & Maloney, S. K. (2006). Physiological responses of Bos taurus and Bos indicus cattle to prolonged, continuous heat and humidity. *Journal of Animal Science*, *84*(4), 972–985.

Bolormaa, S., Porto Neto, L. R., Zhang, Y. D., Bunch, R. J., Harrison, B. E., Goddard, M. E., & Barendse, W. (2011). A genome-wide association study of meat and carcass traits in australian cattle. *Journal of Animal Science*, *89*(8), 2297–2309. https://doi.org/10.2527/jas.2010-3138

Broman, K. W., & Weber, J. L. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *American Journal of Human Genetics*, *65*(6), 1493–1500. https://doi.org/10.1086/302661

Cartwright, T. C. (1955). Responses of beef cattle to high ambient temperatures. *Journal of Animal Science*, *14*(2), 350–362.

Chung, M. J., Hogstrand, C., & Lee, S. J. (2006). Cytotoxicity of nitric oxide is alleviated by zinc-mediated expression of antioxidant genes. In *Experimental Biology and Medicine* (Vol. 231, pp. 1555–1563). https://doi.org/10.1177/153537020623100916

Espigolan, R., Baldi, F., Boligon, A. A., Souza, F. R. P., Gordo, D. G. M., Tonussi, R. L., … Albuquerque, L. G. (2013). Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics*, *14:305*. https://doi.org/doi.org/10.1186/1471-2164-14-305

FAO. (2004). *Secondary guidelines for development of national farm animal genetic*

resources management plans. Rome, Italy.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I., & Sölkner, J. (2013). Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics*, *130*(4), 286–293. https://doi.org/10.1111/jbg.12012

Ferenčaković, M., Sölkner, J., Curik, I., & Curik, I. (2013). Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genetics Selection Evolution*, *45*(1), 42. https://doi.org/10.1186/1297-9686-45-42

Ferenčaković, M, Hamzić, E., Gredler, B., Curik, I., & Sölkner, J. (2011). Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle. *Agriculturae Conspectus Scientificus*, *76*(4), 325–329.

Ferraz, J. B. S., Eller, J. P., Dias, F., & Golden, B. L. (2002). (CO)variance component estimation for growth weights of Montana Tropical®, a Brazilian beef composite. In *7th World Congress on Genetics Applied to Livestock Production*. Montpellier.

Fortes, M. R. S., Lehnert, S. A., Bolormaa, S., Reich, C., Fordyce, G., Corbet, N. J., … Reverter, A. (2012). Finding genes for economically important traits: Brahman cattle puberty. *Animal Production Science*, *52*, 143–150. https://doi.org/10.1071/an11165

Frisch, J. E., & Vercoe, J. E. (1979). Adaptive and productive features of cattle growth in the tropics: their relevance to buffalo production. *Tropical Animal Health and Production*, *4*(3), 214–222.

Gibson, J., Morton, N. E., & Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics*, *15*(5), 789–795.

Goddard, K. A. B., Hopkins, P. J., Hall, J. M., & Witte, J. S. (2000). Linkage Disequilibrium and Allele-Frequency Distributions for 114 Single-Nucleotide Polymorphisms in Five Populations. *Am. J. Hum. Genet*, *66*(1), 216–234.

Goff, J. P. (2018). Invited review: Mineral absorption mechanisms, mineral interactions that affect acid–base and antioxidant status, and diet considerations to improve mineral status. *Journal of Dairy Science*, *101*(4), 2763–2813. https://doi.org/10.3168/jds.2017-13112

Gregory, K., Cundiff, L. V, & Koch, R. M. (1993). Composite Breeds - What Does the Research Tell Us? In *Range Beef Cow Symposium*, 207. Lincoln.

Gregory, K. E., Cundiff, L. V., & Koch, R. M. (1999). Composite breeds to use heterosis and breed differences to improve efficiency of beef production. U.S. Department of Agriculture Technical Bulletin No. 1875, 81 pp.

Gurgul, A., Szmatoła, T., Topolski, P., Jasielczuk, I., Żukowski, K., & Bugno-Poniewierska, M. (2016). The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. *Journal of Applied Genetics*, *57*(4), 527–530. https://doi.org/10.1007/s13353-016-0337-6

Haider, S., Ballester, B., Smedley, D., Zhang, J., Rice, P., & Kasprzyk, A. (2009). BioMart central portal - Unified access to biological data. *Nucleic Acids Research*, *37*, 23–27. https://doi.org/10.1093/nar/gkp265

Hayes, B. J., Visscher, P. M., McPartlan, H. C., & Goddard, M. E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, *13*(4), 635–643.

Hocquette, J.-F., & Bauchart, D. (1999). Intestinal absorption, blood transport and hepatic and muscle metabolism of fatty acids in preruminant and ruminant animals. *Reproduction Nutrition Development*, *39*(1), 27–48. https://doi.org/10.1051/rnd:19990102

Howard, J. T., Tiezzi, F., Huang, Y., Gray, K. A., & Maltecca, C. (2016). Characterization and management of long runs of homozygosity in parental nucleus lines and their associated crossbred progeny. *Genetics Selection Evolution*, *48*(1), 91.

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*. https://doi.org/10.1186/1471-2164-12-460

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, *37*(1), 1–13. https://doi.org/10.1093/nar/gkn923

Huang, D. W., Sherman, B. T., & Lempicki, R. a. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Jones, C. (2019). Bovine Herpesvirus 1 Counteracts Immune Responses and Immune-Surveillance to Enhance Pathogenesis and Virus Transmission. *Frontiers in*

Immunology, *10*, 1008. https://doi.org/10.3389/fimmu.2019.01008

Kim, E. S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassel, C. P., Crooker, B. A., … Sonstegard, T. S. (2013). Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. *PLoS ONE*, *8*(11), e80813. https://doi.org/10.1371/journal.pone.0080813

Kojima, T., Akagi, S., Zeniya, Y., Shimizu, M., & Tomizuka, T. (1996). Evidence of Platelet Activation Associated with Establishment of Pregnancy in Cows with Transferred Embryos. *Journal of Reproduction and Development*, *42*(4), 225–235. https://doi.org/10.1262/jrd.42.225

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), 1–15. https://doi.org/10.1007/s00439-010-0920-6

Ladeira, M. M., Schoonmaker, J. P., Swanson, K. C., Duckett, S. K., Gionbelli, M. P., Rodrigues, L. M., & Teixeira, P. D. (2018). Review: Nutrigenomics of marbling and fatty acid profile in ruminant meat. *Animal*, *12*(2), 282–294. https://doi.org/10.1017/s1751731118001933

Lee, H. J., Park, H. S., Kim, W., Yoon, D., & Seo, S. (2014). Comparison of metabolic network between muscle and intramuscular adipose tissues in Hanwoo beef cattle using a systems biology approach. *International Journal of Genomics*, 2014, Article ID 679437. https://doi.org/10.1155/2014/679437

Lemos, M. V. A., Berton, M. P., Camargo, G. M. ., Peripolli, E., Silva, R. M. ., Ferreira Olivieri, B., … Baldi, F. (2018). Copy number variation regions in Nellore cattle: Evidences of environment adaptation. *Livestock Science*, *207*. https://doi.org/10.1016/j.livsci.2017.11.008

Lindholm-Perry, A. K., Kuehn, L. A., Smith, T. P. L., Ferrell, C. L., Jenkins, T. G., Freetly, H. C., & Snelling, W. M. (2012). A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle. *Animal Genetics*, *43*(2), 216–219. https://doi.org/10.1111/j.1365-2052.2011.02232.x

Liongue, C., O'Sullivan, L. A., Trengove, M. C., & Ward, A. C. (2012). Evolution of JAK-STAT pathway components: Mechanisms and role in immune system development. *PLoS ONE*, *7*(3), e32777.

https://doi.org/10.1371/journal.pone.0032777

Ma, X. X., Ma, L. N., Chang, Q. Y., Ma, P., Li, L. J., Wang, Y. Y., … Cao, X. (2018). Type I interferon induced and antagonized by foot-and-mouth disease virus. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2018.01862

Magalhães, A. F. B., De Camargo, G. M. F., Junior Fernandes, G. A., Gordo, D. G. M., Tonussi, R. L., Costa, R. B., … De Albuquerque, L. G. (2016). Genome-Wide Association Study of Meat Quality Traits in Nellore Cattle. *PLoS ONE*, *11*(6), e0157845. https://doi.org/10.1371/journal.pone.0157845

Marino, R., Atzori, A. S., D'Andrea, M., Iovane, G., Trabalza-Marinucci, M., & Rinaldi, L. (2016). Climate change: Production performance, health issues, greenhouse gas emissions and mitigation strategies in sheep and goat farming. *Small Ruminant Research*, *135*, 50–59. https://doi.org/10.1016/J.SMALLRUMRES.2015.12.012

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., … MacCiotta, N. P. P. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Animal Genetics*, *46*, 110–121. https://doi.org/10.1111/age.12259

Mastrangelo, S., Tolone, M., Di Gerlando, R., 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*, *10*, 746–754. https://doi.org/10.1017/S1751731115002943

Mastrangelo, Salvatore, Tolone, M., Sardina, M. T., Sottile, G., Sutera, A. M., Di Gerlando, R., & Portolano, B. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genetics Selection Evolution*, *49:84*. https://doi.org/10.1186/s12711-017-0360-z

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., … Moore, S. S. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genetics*, *8:74*.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., … Wilson, J. F. (2008). Runs of Homozygosity in European Populations. *American Journal of Human Genetics*, *83*(3), 359–372.

https://doi.org/10.1016/j.ajhg.2008.08.007

Mészáros, G., Boison, A. S., Pérez O'Brien, A. M., Ferenčaković, M., Curik, I., da Silva, M. V. G. B., … Sölkner, J. (2015). Genomic analysis for managing small and endangered populations : a case study in Tyrol Grey cattle. *Frontiers in Genetics*, *6*(173). https://doi.org/10.3389/fgene.2015.00173

Nguyen, L. T., Zacchi, L. F., Schulz, B. L., Moore, S. S., & Fortes, M. R. S. (2018). Adipose tissue proteomic analyses to study puberty in brahman heifers. *Journal of Animal Science*, *96*(6), 2392–2398. https://doi.org/10.1093/jas/sky128

Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., & Li, J. Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics*, *91*(2), 275–292. https://doi.org/10.1016/j.ajhg.2012.06.014

Peripolli, E., Metzger, J., Vinícius, M., De Lemos, A., Stafuzza, N. B., Kluska, S., … Baldi, F. (2018). Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits. *BMC Genomics*, *19:680*.

Peripolli, E., Stafuzza, N. B., Munari, D. P., Lima, A. L. F., Irgang, R., Machado, M. A., … da Silva, M. V. G. B. (2018). Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. *BMC Genomics*, *19:34*.

Porto-Neto, L. R., Kijas, J. W., & Reverter, A. (2014). The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genetics Selection Evolution*, *46:22*.

Porto Neto, L. R., Bunch, R. J., Harrison, B. E., & Barendse, W. (2012). Variation in the XKR4 gene was significantly associated with subcutaneous rump fat thickness in indicine and composite cattle. *Animal Genetics*, *43*(6), 785–789. https://doi.org/10.1111/j.1365-2052.2012.02330.x

Prayaga, K. C., Corbet A, B, N. J., Johnston, D. J., Wolcott, M. L., Fordyce, G., & Burrow, H. M. (2009). Genetics of adaptive traits in heifers and their relationship to growth, pubertal and carcass traits in two tropical beef cattle genotypes. *Animal Production Science*, *49*, 413–425. https://doi.org/10.1071/EA08247

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and

population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

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

Purfield, D. C., McParland, S., Wall, E., & Berry, D. P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS ONE*, *12*(5), e0176780.

Qanbari, S., Pimentel, E. C. G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A. R., & Simianer, H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics*, *41*(4), 346–356.

Raman, D., Sobolik-Delmaire, T., & Richmond, A. (2011). Chemokines in health and disease. *Experimental Cell Research, 317(5)*, 575-589 https://doi.org/10.1016/j.yexcr.2011.01.005

Ramayo-Caldas, Y., Fortes, M. R. S., Hudson, N. J., Porto-Neto, L. R., Bolormaa, S., Barendse, W., … Reverter, A. (2014). A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G, and FOXP3 in intramuscular fat deposition of beef cattle. In *Journal of Animal Science, 92(7)*, 2832–2845). https://doi.org/10.2527/jas.2013-7484

Renaudeau, D., Collin, A., Yahav, S., De Basilio, V., Gourdine, J. L., & Collier, R. J. (2012). Adaptation to hot climate and strategies to alleviate heat stress in livestock production. In *Animal, 6,* 707–728. https://doi.org/10.1017/S1751731111002448

Ribeiro, A. R. ., Alencar, M. M., Freitas, A. R., Regitano, L. C. A., Oliveira, M. C. S., & Ibelli, A. M. G. (2009). Heat tolerance of Nelore, Senepol x Nelore and Angus x Nelore heifers in the southeast region of Brazi. *South African Journal of Animal Science*, *29*(1), 263–265. https://doi.org/10.4314/sajas.v39i1.61261

Sargolzaei, M. (2014). SNP1101 User Guide., *1.0*.

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

Silva, R. M. D. O., Stafuzza, N. B., Fragomeni, B. D. O., Ferreira De Camargo, G. M.,

Ceacero, T. M., Cyrillo, J. N. D. S. G., … De Albuquerque, L. G. (2017). Genome-wide association study for carcass traits in an experimental Nelore cattle population. *PLoS ONE*, *12*, e0169860. https://doi.org/10.1371/journal.pone.0169860

Stothard, P., Choi, J.-W., Basu, U., Sumner-Thomson, J. M., Meng, Y., Liao, X., & Moore, S. S. (2011). Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics*, *12:559*. https://doi.org/10.1186/1471-2164-12-559

Sun, H.-Z., Zhao, K., Zhou, M., Chen, Y., & Guan, L. L. (2019). Landscape of multi-tissue global gene expression reveals the regulatory signatures of feed efficiency in beef cattle. *Bioinformatics*, *35*(10), 1712–1719. https://doi.org/10.1093/bioinformatics/bty883

Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, *2*(2), 125–141.

Szmatoła, T., Gurgul, A., Ropka-molik, K., Jasielczuk, I., Tomasz, Z., & Bugno-poniewierska, M. (2016). Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. *Livestock Science*, *188*, 72–80.

Taniguchi, Y., Yoshioka, N., Nakata, K., Nishizawa, T., Inagawa, H., Kohchi, C., & Soma, G.-I. (2009). Mechanism for maintaining homeostasis in the immune system of the intestine. *Anticancer Research*, *29*(11), 4855–4860.

Terakado, A. P. N., Costa, R. B., de Camargo, G. M. F., Irano, N., Bresolin, T., Takada, L., … de Albuquerque, L. G. (2018). Genome-wide association study for growth traits in Nelore cattle. *Animal*, *12*(7), 1358–1362. https://doi.org/10.1017/s1751731117003068

Thelen, M. (2001). Dancing to the tune of chemokines. *Nature Immunology, 2*, 129-134. https://doi.org/10.1038/84224

Van Parijs, L., & Abbas, A. K. (1998). Homeostasis and self-tolerance in the immune system: Turning lymphocytes off. *Science*, *280*(5361), 243–248. https://doi.org/10.1126/science.280.5361.243

Van Wyk, B., Snider, M., Scruten, E., van Drunen Littel-van den Hurk, S., & Napper, S. (2016). Induction of functional interferon alpha and gamma responses during acute infection of cattle with non-cytopathic bovine viral diarrhea virus. *Veterinary*

Microbiology, *195*, 104–114. https://doi.org/10.1016/j.vetmic.2016.09.015

Widdison, S., & Coffey, T. J. (2011). Cattle and chemokines: Evidence for species-specific evolution of the bovine chemokine system. *Animal Genetics, 42(4)*, 341-353. https://doi.org/10.1111/j.1365-2052.2011.02200.x

Wolcott, M. L., Johnston, D. J., & Barwick, S. A. (2014). Genetic relationships of female reproduction with growth, body composition, maternal weaning weight and tropical adaptation in two tropical beef genotypes. *Animal Production Science*, *54*(1), 60–73. https://doi.org/10.1071/AN13012

Zhang, Q., Calus, M. P. L., Guldbrandtsen, B., Lund, M. S., & Sahana, G. (2015). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genetics*, *16*, 88.

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. S., & Sahana, G. (2015). Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics*, *16*(1), 542. https://doi.org/10.1186/s12864-015-1715-x

FIGURES

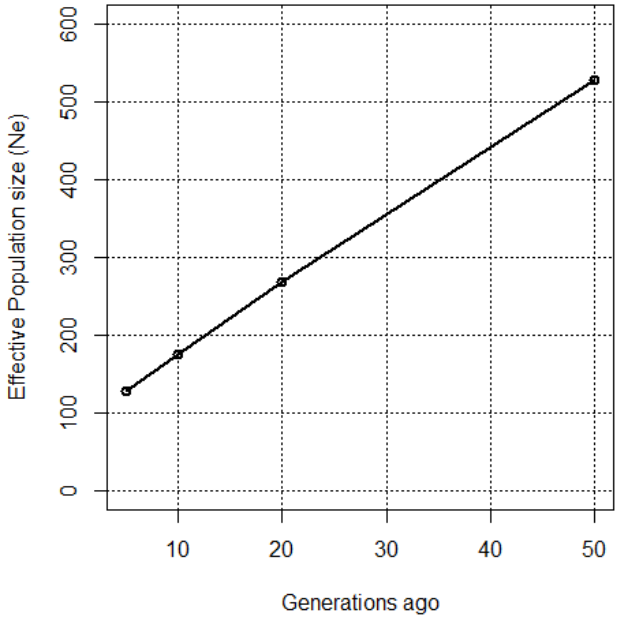


Figure 1. Estimated effective population size (Ne) over time for the composite Montana Tropical® beef cattle.

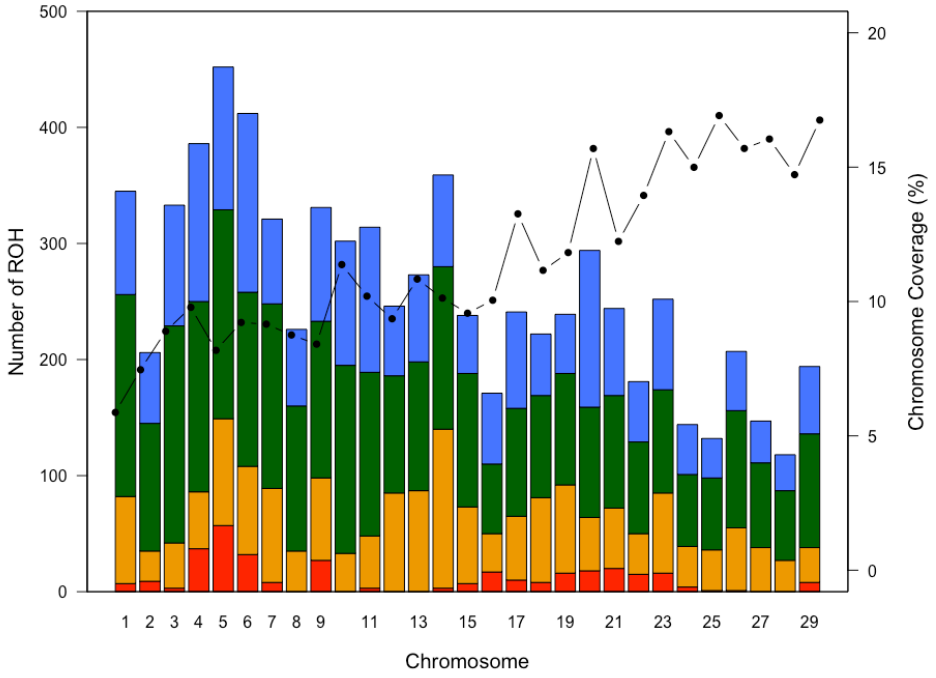


Figure 2. Runs of homozygosity distribution and coverage for each autosome in composite Montana Tropical® beef cattle. **Barplot.** Frequency distribution of the number of runs of homozygosity in different length classes: red ($ROH_{0.5\text{-}2\ Mb}$), orange ($ROH_{2\text{-}4\ Mb}$), green ($ROH_{4\text{-}8\ Mb}$), and blue ($ROH_{>8\ Mb}$). **Lines.** Average percentage of chromosome coverage by runs of homozygosity of minimum length of 0.5 Mb.

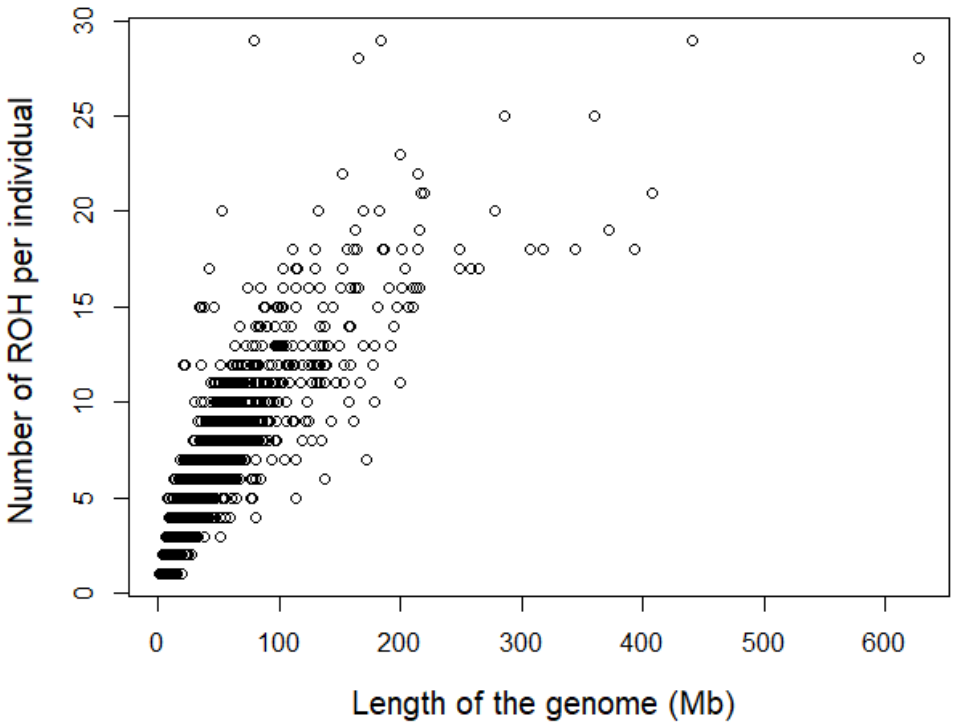


Figure 3. Number of runs of homozygosity (ROH) per individual and the total length of the genome covered by ROH.

CAPÍTULO 4 – GENOME-WIDE DETECTION OF SIGNATURES OF SELECTION IN INDICINE AND BRAZILIAN LOCALLY ADAPTED TAURINE CATTLE BREEDS USING WHOLE-GENOME RE-SEQUENCING DATA[1]

ABSTRACT

The cattle introduced by European conquerors during the Brazilian colonization period were exposed to a process of natural selection in different types of biomes throughout the country, leading to the development of locally adapted cattle breeds. In this study, whole-genome re-sequencing data from indicine and Brazilian locally adapted taurine cattle breeds were used to detect genomic regions under selective pressure. Within-population and cross-population statistics were combined separately in a single score using the de-correlated composite of multiple signals (DCMS) method. Putative sweep regions were revealed by assessing the top 1% of the empirical distribution generated by the DCMS statistics. A total of 33,328,447 biallelic SNPs with an average read depth of 12.4X passed the hard filtering process and were used to access putative sweep regions. Admixture has occurred in some locally adapted taurine populations due to the introgression of exotic breeds. The genomic inbreeding coefficient based on runs of homozygosity (ROH) concurred with the populations' historical background. Signatures of selection retrieved from the DCMS statistics provided a comprehensive set of putative candidate genes and revealed QTLs disclosing cattle production traits and adaptation to the challenging environments. Additionally, several candidate regions overlapped with previous regions under selection described in the literature for other cattle breeds. The current study reported putative sweep regions that can provide important insights to better understand the selective forces shaping the genome of the indicine and Brazilian locally adapted taurine cattle breeds. Such regions likely harbor traces of natural selection pressures by which these populations have been exposed and may elucidate footprints for adaptation to the challenging climatic conditions.

Key-words: *Bos taurus indicus, Bos taurus taurus*, signatures of selection, local adaptation, Next-generation sequencing

[1] Este capítulo corresponde ao artigo científico publicado na revista BMC Genomics 2020, 21:624.

INTRODUCTION

The first cattle herds were brought to Brazil by Portuguese conquerors in 1534 during the Brazilian colonization period [1]. These cattle have undergone to a process of natural selection for more than 450 years in a wide range of ecosystems throughout the country [2]. Natural selection in a remarkably diverse set of environments together with recurring events of breed admixture led to the development of locally adapted cattle breeds, i.e., Curraleiro Pé-Duro, Pantaneiro, Crioulo Lageano, Caracu, and Mocho Nacional [3]. By the end of the nineteenth century, the increasing demand for food supply triggered the imports of exotic and more productive breeds of indicine origin [3, 4]. As a consequence, a reduction in locally adapted cattle breed populations has occurred to such an extent that nowadays, most of them are threatened with extinction [3, 5].

Brazilian locally adapted cattle breeds have been subjected to strong environmental pressures and faced several difficulties including hot, dry or humid tropical climate conditions, scarce food availability, diseases, and parasite infestations without any significant selective pressure imposed by man [2]. Influenced by the environment and shaped by natural selection, these animals acquired very particular traits to thrive in distinct ecosystems, which has presumably left detectable signatures of selection within their genomes. In this regard, Brazilian locally adapted cattle breeds represent an important genetic resource for the understanding of the role of natural selection in diverse environments, providing new insights into the genetic mechanisms inherent to adaptation and survivorship [6]. Although their productivity is much lower compared to highly-specialized breeds under intensive production systems [7, 8], great efforts have been made to improve our knowledge of locally adapted breeds [5, 9, 10] and their use in crossbred schemes.

According to Utsunomiya et al. [11], signatures of selection studies should strongly focus on small local breeds given their endangered status and the putative importance of their genomes in unraveling footprints of selection by elucidating genes and structural variants underlying phenotypic variation. Advances in molecular genetics and statistical methodologies together with the availability of whole-genome re-sequencing has notably improved the accuracy to disentangle the effects of natural

and artificial selection in the genome of livestock [12–14]. However, despite the recent achievements in high-throughput sequencing, studies to detect positive selection in endangered Brazilian locally adapted cattle breeds are incipient. Previous studies on such breeds have mainly focused on population structure and genetic diversity using Random Amplified Polymorphic DNA (RAPD), pedigree data, microsatellite, and Single-Nucleotide Polymorphism (SNP) arrays [15–19].

In this study, we report for the first-time signatures of selection derived from whole-genome re-sequencing data in three Brazilian locally adapted taurine cattle breeds as well as in one indicine breed. Potential biological functions of the genes screened within the putative candidate regions were also examined to better elucidate the phenotypic variation related to adaptation shaped by natural selection.

RESULTS

Data

DNA samples from 13 Gir (GIR), 12 Caracu Caldeano (CAR), 12 Crioulo Lageano (CRL), and 12 Pantaneiro (PAN) re-sequenced to 15X genome coverage were used. An average alignment rate of 99.59% was obtained. After SNP calling and filtering, a total of 33,328,447 SNPs distributed across all 29 autosomes were retained for subsequent analyses with an average read depth of 12.37X (9.57 ~17.52X).

Variant annotation and enrichment

Of the total SNPs identified (*n=33,328,447 SNPs*), most of them were located in intergenic (67.17%) and intronic (25.85%) regions (Appendix 1C). A total of 1,065,515 (3.19%) variants were located in the 5-kb regions upstream from genes, and 928,061 (2.78%) in the 5-kb regions downstream from genes. Several variants with high consequence on protein sequence were identified, including splice acceptor variant (*n=471*), splice donor variant (*n=481*), stop gained (*n=1,111*, stop lost (*n=58*), and start lost (*n=208*). According to SIFT scores, 24,159 variants (23,428 missense, 578 splice region, and 143 start lost) were classified as deleterious.

Following variant annotation, we further investigated the gene content within the predicted variants to cause relevant biological functions. A total of 1,189 genes were described within variants with high consequence on protein sequence and 7,373 genes within those causing a deleterious mutation based on the SIFT score. Functional enrichment analysis revealed several gene ontology (GO) terms and one Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway overrepresented ($p<0.01$) for the set of genes previously described (Appendix 2C and 3C), however, none of them have been associated with the traits/phenotypes that could be affected by the natural selection which those breeds have been subjected to.

Population structure

The population structure among breeds was dissected by analyzing the first two principal components, which accounted for roughly 20% of the genetic variability and divided the populations into three clusters (Figure 1a). A clear separation could be observed between indicine (*Bos taurus indicus*) and locally adapted taurine (*Bos taurus taurus*) populations. Within the taurine populations, the greatest overlap of genetic variation was observed between CRL and PAN breeds. Despite clustering together, the analysis of molecular variance (AMOVA) revealed genetic differentiation between those two breeds ($p<0.001$, Appendix 4C), indicating that all four breeds could be considered as genetically independent entities. Further, when analyzing the first two principal components encompassing the locally adapted taurine cattle breeds (Figure 1b), an evident separation could be observed between CAR and the remaining two populations. The analysis also distinguished CRL from PAN, agreeing with the AMOVA results.

Admixture analysis was performed to further estimate the proportions of ancestry (K) in each population (Figure 2). The lowest cross-validation error (0.387) was observed for $K=2$, revealing the presence of two main clusters differentiating the locally adapted taurine populations from the indicine population. Within the taurine populations, the CAR breed did not show admixed ancestry while CRL and PAN breeds showed 77% of taurine and 23% of indicine ancestry on average. When $K=3$ was assumed, CRL samples revealed evidence of admixed ancestry from other

breeds, whereas PAN samples were quite homogeneous, with little indication of introgression from other breeds. CAR and GIR breeds displayed a greater uniformity and did not reveal major signs of admixture of other breeds, being consistent with *K=2*.

Genomic inbreeding

Descriptive statistics for runs of homozygosity-based inbreeding coefficients (F_{ROH}) are shown in Table 1. The average inbreeding coefficients did not differ significantly ($p<0.05$) among breeds, with the exception of CAR animals. It is worth to highlight that these animals also displayed the smallest inbreeding variability among all breeds, supported by the lowest coefficient of variation.

Table 1. Descriptive statistics of runs of homozygosity-based inbreeding coefficient (F_{ROH}) for Gir (GIR), Crioulo Lageano (CRL), Caracu Caldeano (CAR), and Pantaneiro (PAN) cattle breeds

Breed	**Mean**	**Median**	**Minimum**	**Maximum**	**Coefficient of variation (%)**
Gir	0.040[b]	0.038	0.020	0.060	29.37
Crioulo Lageano	0.036[b]	0.028	0.017	0.082	53.69
Caracu Caldeano	0.138[a]	0.140	0.121	0.153	8.63
Pantaneiro	0.045[b]	0.042	0.022	0.096	43.56

Means sharing a common letter within a column were not significantly different ($p<0.05$) from one another.

Selective sweeps

A total of 499 putative sweep regions encompassing 221 genes were identified from the top 1% of the empirical distribution generated by the within-population de-correlated composite of multiple signals (DCMS) statistic [20] (Figure 3, Appendix 5C). For the cross-population DCMS statistic, the top 1% of the empirical distribution revealed 503 putative sweep regions comprehending 242 genes (Appendix 6C). The *Bos taurus* autosome (BTA) 3 displayed the highest number of putative sweep regions for the within-population DCMS statistic (*n=33*), while BTA11 did for the cross-

population DCMS statistic (*n=67*). The functional importance of the annotated genes was assessed by performing GO and KEGG pathway enrichment analysis separately for each DCMS statistic and its respective retrieved gene list. No overall significant enrichment of any particular GO nor KEGG was found after adjusting the p-values for False Discovery Rate [21].

Five genomic regions overlapped between the candidate sweep regions of the within-population and cross-population DCMS statistics (BTA4:101600000-101650000, BTA5:3700000-3750000, BTA9:98650000-98700000, BTA11:22300000-22350000, and BTA11:53900000-53950000). When inspecting in detail, the region on BTA4:101600000-101650000 harbored two quantitative trait locus (QTLs) with functions related to the bovine respiratory disease [22] and body condition score [23]. The remaining four regions have not been associated with any QTL in cattle so far, however, they were found to be in close vicinity (~15 to 237 kb) with specific QTLs for beef cattle production traits. Such QTLs included body weight at yearling, calving ease, body weight gain, and marbling score [24–26]. Further, among the five overlapping candidates sweep regions, only the one on BTA9 was found to harbor a gene, the *PRKN*.

Selective sweeps and runs of homozygosity

Shared genomic regions harboring several protein-coding genes were identified between runs of homozygosity (ROH) hotspots and the putative sweep regions retrieved from the DCMS statistics (Table 2). ROH hotspots for each breed are described in Appendix 7C. For the shared regions disclosed when considering the within-population DCMS statistic, the ones located on BTA1:8300000-8350000 and BTA1:41600000-41650000 coincided with a QTL for somatic cell score [27] and maturity rate [28], respectively. It is noteworthy to underscore that despite not displaying any overlapping QTL, the region on BTA8:15700224-15700228 was described nearby (~99 kb) a QTL for tick resistance [29], and those on BTA21:6550000-6600000 and BTA21:63250000-63300000 were very close (<14 kb) to QTLs for reproductive-related traits [30, 31]. When considering the cross-population DCMS statistic, the candidate regions overlapped previously identified QTLs formerly

implicated in dairy-related [32–35] and body-related traits (weight [24], energy content [36], and conformation [32]). Further, several QTLs associated with body conformation and growth [23, 24, 37], reproductive-related traits [28, 38], and coat texture [39] were described to be in very close proximity (~18.98 to 88.38kb).

Table 2. Gene annotation and reported QTLs for the shared genomic regions between runs of homozygosity (ROH) hotspots and the putative sweep regions retrieved from the within-population and cross-populations DCMS statistics.

BTA[1]	Start	End	Genes	QTL[2]
		Within-population DCMS statistic x ROH		
1	8,300,000	8,350,000	-	**Somatic cell score** [27]
1	41,600,000	41,650,000	*EPHA6*, *ARL6*	**Maturity rate** [28]
1	112,250,000	112,300,000	*KCNAB1*	-
8	15,800,000	15,850,000	-	Tick resistance [29]
15	35,365,655	35,399,999	*OTOG*	-
15	35,400,001	35,450,000	-	-
18	34,718,675	34,750,000	*CDH16*, *RRAD*	-
21	6,550,000	6,600,000	*ADAMTS17*	Calving ease [30]
21	63,250,000	63,300,000	*VRK1*	Interval to first estrus after calving [31]
		Cross-population DCMS statistic x ROH		
3	77,250,000	77,300,000	-	Body condition score [23]
5	31,800,000	31,850,000	-	Body weight (yearling) [24], Conception rate [38]
5	38,761,637	38,761,745	*YAF2*	
7	57,050,000	57,100,000	-	Rump angle [37]
11	67,450,000	67,500,000	*ANTXR1*, *GFPT1*	**Body weight (yearling)** [24]**, Body energy content** [36]
11	67,700,000	67,749,999	-	-
11	67,750,001	67,800,000	*NFU1*	-
11	68,550,000	68,600,000	*PCYOX1*	-
14	52,900,000	52,914,848	-	Maturity rate [28]
15	10,150,000	10,200,000	-	-
15	10,900,000	10,950,000	-	**Calving ease (maternal)** [32]**, Daughter pregnancy rate** [32]**, Foot angle** [32]**, Milk fat percentage** [32] **, Milk fat yield** [32]**, Net merit** [32]**, Length of productive life** [32]**, Milk**

				protein percentage [32]**, Milk protein yield** [32]**, Calving ease** [32]**, Somatic cell score** [32]
20	38,000,000	38,050,000	*RANBP3L*, *NADK2*	**Milk protein percentage** [33]**, Milk protein yield** [34]**, Milk yield** [34]**,** Coat texture [39]
21	200,000	250,000	-	-
25	1,345,564	1,350,000	*NME3*, *MRPS34*	**Milk fat yield** [35]

[1] BTA: *Bos taurus* autosome; [2] QTLs within the candidate genomic regions are highlighted in bold. Non-bold QTLs were the closest and most suitable candidate QTL for the given candidate region.

Overlap with candidate regions under positive selection in other cattle populations

Several putative sweep regions identified from the top 1% of the empirical distribution generated by the within-population and cross-population DCMS statistics were in agreement with previous research on signatures of selection in cattle (Appendix 8C and 9C, respectively). Such studies included indigenous African and Spanish [6, 40–43], native [44–46], tropical-adapted [47–50], Chinese [50, 51], and commercial beef and dairy [13, 41, 50, 52–55] cattle breeds. For the five genomic regions identified overlapping in between the DCMS statistics, the one on BTA9:98650000-98700000 matched with a previous study on cattle breeds selected for dairy production [55]. Besides, common signals found between ROH hotspots and the within-population and cross-population DCMS statistics were also supported by previously published data on signatures of selection [41, 43, 44, 46, 49, 51, 54] (Appendix 10C and 11C, respectively).

DISCUSSION

Population structure

The segregation between indicine and taurine cattle populations described in both principal component and admixture analysis (*K=2*) reflects the divergence and evolutionary process started roughly two million years ago [56, 57]. As a result of the domestication process and selective breeding over time, the cattle can be classified into temperate (*Bos taurus taurus* or taurine) and tropical (*Bos taurus indicus* or

indicine) based on the common adaptive and evolutionary traits they have acquired [58]. Within the Brazilian locally adapted taurine breeds, the principal component analysis (PCA) indicates the highest relatedness between CRL and PAN breeds and their divergence from the CAR breed may be explained by the European cattle type introduced in Brazil during the colonization period [59]. These results were similar to those obtained using RAPD [17] and microsatellites [19]. Portuguese purebred cattle brought to Brazil belonged to three different bloodlines: *Bos taurus aquitanicus*, *Bos taurus batavicus*, and *Bos taurus ibericus*. In this regard, CRL and PAN breeds descended from a common ancestral pool and have their origin in breeds from *Bos taurus ibericus* cattle, while the CAR cattle is derived from the *Bos taurus aquitanicus* cattle [17]. Further, the divergence within the locally adapted cattle breeds may be a result of artificial selection events over time since the CAR cattle have been selected for milk production for the past 100 years, while CRL and PAN started recently to be artificially selected.

Levels of introgression of indicine genes in taurine breeds described herein are consistent with previous studies on Brazilian locally adapted taurine breeds [16, 17, 19]. This gene flow reinforces the concept that the import of exotic breeds at the beginning of the 20th century [3] led to the miscegenation of the locally adapted breeds due to crossbreeding practices, resulting nearly in their extinction [4]. In this regard, the CRL breed experienced some introduction of Nellore (*Bos taurus indicus*) genes for a short period in the eighties [17], which can be visualized when assuming *K=2* and *K=3*. Concurring with our findings, Egito et al. [19] also revealed that CRL and PAN animals were the closest to the indicine cattle among four Brazilian locally adapted cattle breeds, displaying the highest frequency of indicine gene introgression. A cytogenetic analysis study on the PAN cattle also revealed absorbing crosses with the indicine cattle [60]. In addition, the absence of admixture patterns in CAR individuals has been previously described by Campos et al. [16] and Egito et al. [21]. The homogeneity of such population most likely reflects its formation process and the objective of selection for dairy traits since 1893 [61], which may have distinguished them from other locally adapted taurine breeds when taking into consideration the genetic structure integrity.

Genomic inbreeding

As already stated, the Brazilian locally adapted cattle breeds nearly disappeared between the late 19th and beginning of the 20th century, and most of them are nowadays threatened with extinction [3, 5]. It is worth to stress out that the CAR cattle are an exception, and they can be considered as an established breed [5, 62]. In this regard, animals comprising our dual purpose cattle populations, which were exploited for meat production in former times [63], are nowadays mainly used in animal genetic resources conservation programs (*in situ* and *ex situ*) and as a germplasm reservoir to preserve the genetic variability [4, 64]. Different from the dual-purpose cattle populations, the dairy populations are no longer considered endangered, and such animals have been selected for milk production traits in the southeastern region of Brazil since 1893 (CAR, [61]) and the early nineties (GIR, [65]).

Most of the locally adapted cattle breeds in Brazil developed from a narrow genetic base, and in such cases, inbreeding can increase over generations and reduce genetic variability [66]. Despite their population background, CRL and PAN animals displayed low F_{ROH} estimates, concurring with heterozygosity estimates (Results not shown). Decreased levels of inbreeding and high genetic variability have been previously described for both breeds, probably resulting from a slight selection pressure and herd management focused on maintaining genetic diversity by using a male:female relationship larger than usual [19]. Egito et al. [15] attributed such results to the formation of new PAN herds from 2009 onwards while Pezzini et al. [18] associated it with the diversification in the use of CRL sires. Further, Egito et al. [19] stated that CRL and PAN cattle were the most diverse population with the highest mean allelic richness among four locally adapted cattle breeds investigated. Such results are consistent with F_{ROH} estimates found in this current work, reflecting mild selection pressure in our dual-purpose cattle populations together with rationale mating decisions and herd management taken by the breeders and associations.

The highest F_{ROH} found for the CAR population most likely reflects its history of selective breeding for milk-related traits from a limited genetic base and the occurrence of a population decrease in the sixties, as discussed by Egito et al. [19]. According to Marras et al. [67], it is not unusual to disclose a higher sum of ROH in dairy than in

beef populations. In this regard, the reduction of genetic variability through the increase of autozygosity in dairy breeds can be explained by the intense artificial selection with the use of a relatively small number of proven sires [68]. Despite being also specialized for milk-related traits, it is not surprising that the GIR population did not show as high F_{ROH} levels as did CAR. Previous studies have also shown low inbreeding rates for the GIR cattle considering pedigree-based inbreeding coefficient [69, 70] and F_{ROH} [71, 72]. A trend in the decrease of inbreeding has been previously described [69, 71], and it happens along with the establishment of the Brazilian Dairy Gir Breeding Program (PNMGL) and the Gir progeny testing. Presumptively, these two concomitant events led to the dissemination of the breed, allowing formerly closed herds to start using semen of proven sires, increasing the overall genetic exchange and reducing the average inbreeding over time.

Candidate regions under positive selection

After combining the top 1% putative sweep regions retrieved from the within-population and cross-population DCMS statistics, five candidate regions harboring two QTLs and only one protein-coding gene were identified. Such results allowed us to highlight the body condition score QTL [23] on BTA4:101600000-101650000, which can be defined as the amount of metabolized energy stored in fat and muscle of a live animal [73]. During periods of energy shortage, key hormones expression and tissue responsiveness adjust to increase lipolysis to meet energy requirements and maintain physiological equilibrium [74, 75]. Regulation and coordination of energy partitioning and homeostasis is a challenge to sustainable intensification of cattle productivity in the tropics. The variation in the animal's nutritional and energetic balance may explain the observed variability in performance between animals in different environments [76]. Negative energy balance most likely reduce energy expenditure, impairing reproductive performance [77], and increasing the susceptibility to infections [78]. As formerly described, the Brazilian locally adapted cattle breeds faced several environmental pressures to thrive in the tropics under harsh environmental conditions, suggesting that animals that were able to minimize the mobilization of adipose tissue

reserves in response to the energy deficit might have conferred fitness advantage than the average individual in the given population.

The *PRKN* (also known as *PARK2*) was the only annotated gene identified in between the DCMS statistics, and its functions have been associated with adipose metabolism and adipogenesis [79]. Remarkably, it is considered a strong positional candidate for adiposity regulation in chicken [80].

We also explored common signals between ROH hotspots and the top 1% putative sweep regions retrieved from both DCMS statistics to increase the power of signals. Among the genes identified when considering the within-population DCMS statistic, we revealed the presence of two interesting genes that have been described to have effects on temperament (*EPHA6*) [81] and body size (*ADAMTS17*) [82] in cattle. Further, a gene associated with temperament (*ANTXR1*) [83] was also highlighted when considering the cross-population DCMS statistic.

In tropical and subtropical regions, cattle productivity depends not only on the inherent ability of animals to grow and reproduce but also on their ability to overcome environmental stressors that impact several aspects of cattle production [84]. In cattle, stress responsiveness has been associated with cattle behavior, more specifically, temperament. Temperament can adversely affect key physiological processes involved in cattle growth, reproduction, and immune functions [85]. Studies have shown that non-temperamental cattle tend to gain weight faster [86–88], spend more time eating [88], and have a higher dry matter intake and average daily gain [86, 89] than temperamental cattle. Further, studies have discussed the negative impacts of temperamental animals on immune-related functions (reviewed by [85]). Two reasons might explain those genes associated with temperament located on ROH hotspots overlapping regions on BTA1:41600000-41650000 and BTA11:67450000-67500000. The first reason is that such genes likely reflect levels of introgression of indicine genes in taurine locally adapted cattle breeds, as confirmed by admixture analysis. *Bos indicus* and their crosses have been reported to be more temperamental than *Bos taurus* cattle when reared under similar conditions [90]. The second reason is that the taurine locally adapted cattle breeds were able to overcome environmental stressors through natural selection over time and could prosper in such harsh tropical environment.

The *ADAMTS17* gene, described enclosing a ROH hotspot overlapping region on BTA21:6550000-6600000, is a well-known candidate gene with a major impact on body size [82, 91, 92]. Much has been discussed about the relationship between body size and environmental adaptation. Variations in body size may be explained as an adaptive response to climate and/or can be driven by changes in feed resources and seasonal influences [93, 94]. In this regard, large body size animals can better tolerate austere conditions, having advantages under cold stress as well as in the use of abundant forage resources [95]. On the other hand, smaller animals exhibit better adaptation to warmer and dry climates [96–98] and are more efficient for grazing under seasonal and scarce forage resources [99]. Based on morphological measurements, it should be noted that the indicine and Brazilian locally adapted taurine cattle breeds are small to medium-sized breeds. Both GIR, CRL, and PAN have reduced body size and lightweight, in which females exhibit an average adult live weight of 418 kg [100], 430 kg [101], and 298 kg [102], respectively. CAR animals have a greater body size among the locally adapted cattle breeds, with females displaying an average live weight of 650 kg [103].

Two intersecting QTLs associated with productivity traits usually favored in commercial breeds (somatic cell score and maturity rate QTLs) were found in ROH hotspots overlapping regions when considering the within-population DCMS statistic. Among the QTLs identified when considering the cross-population DCMS statistic, the one associated with body energy content [36] must be highlighted given its importance in energy partitioning and homeostasis, as previously discussed. Additionally, several remarkably QTLs neighboring the candidate regions intervals were identified. These QTLs have been associated with different biological functions linked to local environment adaptation, such as parasite vector resistance (tick resistance QTL), reproductive-related traits (calving ease, interval to first estrus after calving, conception and maturity rate QTLs), body conformation and morphology traits (body condition score, body weight at yearling, rump angle QTLs), and coat color (coat texture QTL).

The genes and QTLs identified within the candidate regions provide a hint about the selective forces shaping the genome of the indicine and Brazilian locally adapted taurine cattle breeds. Such selective forces were described to be likely associated with adaptation to a challenging environment and environmental stressors. Further, several

QTLs identified nearby the candidate regions intervals were also associated to a lesser extent with beef cattle production traits, while others with various biological functions presumably linked to selection to environmental resilience as well.

Overlap with candidate regions under positive selection in other cattle populations

The greatest number of the putative sweep regions identified from the top 1% of the within-population DCMS statistic overlapped with candidate regions under positive selection previously reported in five cattle breeds selected for dairy production [55], comprehending roughly 22% (*n=52*) of the overlapping regions. For the top 1% of the cross-population DCMS statistic, the greatest number was described for native cattle breeds from Siberia, eastern and northern Europe [46], totaling nearly 17% (*n=50*) of the overlapping regions. Remarkably, in both statistics, the majority of the shared signals within those reported in the literature was found associated with specialized cattle breeds (i.e., dairy and beef). We also identified signatures of selection within those reported in the literature shared by breeds showing different production selection within the same candidate region. According to Gutiérrez-Gil et al. [104], such genomic regions may reflect selection for general traits such as metabolic homeostasis, or they might disclose the pleiotropic effects of genes on relevant traits underlying specialized cattle breeds.

The greater number (seven out of 11) of the putative sweep regions shared between ROH hotspots and the top 1% putative sweep regions retrieved from both DCMS statistics overlapped with regions previously described on local and native cattle breeds [41, 43, 44, 46]. Such results allow us to assume that the same selective forces are most likely acting across these populations, and such regions might have been shaped by selection events rather than genetic drift or admixture events.

It is noteworthy to underscore that the regions under positive selection for other cattle populations reported herein were mainly obtained through medium and high-density SNP arrays. SNP genotyping arrays suffer from SNP ascertainment bias, and it strongly influences population genetic inferences (reviewed by Lachance and Tishkoff [105]). Besides, some scan methodologies based on site frequency spectrum and population differentiation may be more likely to ascertainment bias than others

[106, 107], compromising the power of the tests and may yielding to flawed results [108] when compared to those obtained from whole-genome re-sequencing data.

FINAL CONSIDERATIONS

By using whole-genome re-sequencing data, we identified candidate sweep regions in indicine and Brazilian locally adapted taurine cattle breeds, of which the latter have been exposed to a process of natural selection for several generations in extremely variable environments. The signatures of selection across the genome could provide important insights for the understanding of the adaptive process and the differences in the breeding history underlying such breeds. Our findings suggest that admixture has occurred in some locally adapted taurine populations due to the introgression of exotic breeds, and the stratification results revealed the genetic structure integrity of the dairy populations sampled in this study. Candidates sweep regions, most of which overlapped with or were nearby reported QTLs and candidate genes closely linked to cattle production traits and environmental adaptation. Putative sweep regions together with ROH hotspots also provided valuable shreds of evidence of footprints for adaptation to the challenging climatic conditions faced by the breeds. The candidate sweeps regions and the gene list retrieved from them can improve our understanding of the biological mechanisms underlying important phenotypic variation related to adaptation to hostile environments and selective pressures events to which these breeds have undergone. Furthermore, the study provides complementary information which could be used in the implementation of breeding programs for the conservation of such breeds.

METHODS

Samples, sequencing, and raw data preparation

Sequencing analysis was based on data from 13 Gir (*Bos taurus indicus*, dairy production use), 12 Caracu Caldeano (*Bos taurus taurus*, dairy production use), 12 Crioulo Lageano (*Bos taurus taurus*, dual purpose use), and 12 Pantaneiro (*Bos taurus taurus*, dual purpose use) animals. The studied breeds can be classified into two

groups: (i) indicine breeds represented by the Gir (GIR) cattle; and (ii) locally adapted taurine cattle breeds encompassing Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle. Animals were sampled from three Brazilian geographical regions, including the south (CRL), southeast (GIR and CAR), and mid-west (PAN) (Additional file 12).

DNA was extracted from semen samples that were collected from GIR bulls and blood samples from the remaining breeds. The semen straws were acquired from three commercial artificial insemination centers (American Breeders Service (ABS), Cooperatie Rundvee Verbetering (CRV), and Alta Genetics) and the DNA samples from the Animal Genetics Laboratory (AGL) at EMBRAPA Genetic Resources and Biotechnology (Cenargen, Brasília-DF, Brazil). Paired-end whole-genome re-sequencing with 2x100 bp reads (CRL) and 2x125 bp reads (GIR, CAR, and PAN) was performed on the Illumina HiSeq2500 platform with an aimed average sequencing depth of 15X.

Pair-end reads were aligned to the *Bos taurus taurus* genome assembly UMD3.1 using Burrows-Wheeler Alignment MEM (BWA-MEM) tool v.0.7.17 [109] and converted into a binary format using SAMtools v.1.8 [110]. Polymerase chain reaction (PCR) duplicates were marked using Picard tools (http://picard.sourceforge.net, v.2.18.2). For downstream processing, GATK v.4.0.10.1 [111–113] software was used. Base quality score recalibration was performed using a SNP database (dbSNP Build 150) retrieved from the NCBI [114] followed by SNP calling using the HaplotypeCaller algorithm. To remove unreliable SNP calls and reduce the false discovery rate, hard filtering steps were applied on the variant call. Insertions and deletions polymorphism (Indels) and multi-allelic SNPs were filtered out, and then hard filtering was applied for clustered SNPs (>5 SNPs) in a window size of 20 bp. An outlier approach was used and values above 14.44 (highest 5%) for Fisher strand test were removed. The same was applied for the highest and lowest 2.5% values for base quality rank sum test (-2.26 and 3.04), mapping quality rank sum test (-2.46 and 1.58), read position rank sum test (-1.64 and 2.18), and read depth (267 and 883). Variants with a mapping quality value lower than 30 (0.1% error probability) were also removed from the call set. SNPs that passed the filtering process and located on autosomal chromosomes were retained for subsequent analysis.

Variant annotation and predicted functional impacts

A functional annotation analysis of the called variants was performed to assess their possible biological impact using the Variant Effect Predictor (VEP, [115]) together with the Ensembl cow gene set 94 release. Variants are categorized according to their consequence impact on protein sequence as high, moderate, low, or modifier (more severe to less severe). Variants with high consequence on protein sequence (i.e., splice acceptor variant, splice donor variant, stop gained, frameshift variant, stop lost, and start lost) were selected for further assessment. The impact of amino acid substitutions on protein function were predicted using the sorting intolerant from tolerant (SIFT) scores implemented on VEP tool, and variants with SIFT scores lower than 0.05 were considered as deleterious to protein function.

Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 tool [116, 117] was used to identify overrepresented GO terms and KEGG pathways using the list of genes retrieved from the variants classified with high consequence on protein sequence and as deleterious, and the *Bos taurus taurus* annotation file as a background. The p-values were adjusted by False Discovery Rate [21], and significant terms and pathways were considered when $p<0.01$.

Population differentiation analysis

A PCA implemented with a custom R script was used to examine the genetic structure of the four breeds. AMOVA [118] was also implemented to test for genetic differentiation among breeds. Such method consists in assessing population differentiation using molecular markers together with a pairwise distance matrix, and it can easily incorporate additional hierarchical levels of population structure. AMOVA computations were conducted using the ‘amova’ function in R package pegas [119]. The analyses were based on pairwise squared Euclidean distances using the ‘dist’ function implemented in R [120] and the statistical significances were tested by permutations (n = 1,000). Additionally, the software ADMIXTURE v1.3 [121] was used to reveal admixture patterns among breeds by measuring the proportion of individual

ancestry from different numbers of hypothetical ancestral populations (K). Linkage disequilibrium (LD) pruning for admixture analysis was performed on PLINK v1.90 software [122] to remove SNP with a R^2 value greater than 0.1 with any other SNP within a 50-SNP sliding window. The optimal number of *K* was defined based on the cross-validation error value (*K*=1 to 5) implemented in ADMIXTURE.

Genomic inbreeding coefficient estimation

Genomic inbreeding coefficients based on runs of homozygosity (F_{ROH}) were estimated for every animal according to the genome autozygotic proportion described by McQuillan et al. [123]:

$$F_{ROH}^{i} = \frac{S_{ROH}^{i}}{L_{GEN}}$$

where S_{ROH}^{i} is the sum of ROH across the genome for the *i*th animals and L_{GEN} is the total length of the autosomes covered by SNPs. L_{GEN} was taken to be 2511.4 Mb based on the *Bos taurus taurus* genome assembly UMD3.1. ROH were identified in every individual using PLINK v1.90 [122] software in non-overlapping sliding windows of 50 SNPs. The minimum length of a ROH was set to 500 kb. A maximum of three SNPs with missing genotypes and three heterozygous SNPs were admitted in each window, as discussed by Ceballos et al. [124]. Tukey's post-hoc test [125] was used to identify significant pairwise comparisons ($p<0.05$).

Selective sweeps detection

Four statistical methods were implemented to detect genomic regions under selective pressure. Cross-population methods encompassed the Wright's fixation index (F_{ST}) and the Cross-Population Extended Haplotype Homozygosity (XPEHH). Within-population methods included the Composite Likelihood Ratio (CLR) statistic and the integrated Haplotype Score (iHS).

F_{ST} [126] was calculated between all six pairwise combinations of the four breeds with custom R scripts as follows:

$$FST = \frac{\bar{p}(1-\bar{p}) - \sum c_i p_i (1-p_i)}{\bar{p}(1-\bar{p})}$$

where $\bar{p}$ is is the average frequency of an allele in the total population, p_i is the allele frequency in the ith population, and c_i is the relative number of SNPs in the ith population. F_{ST} scores were then averaged in non-overlapping sliding windows of 50 kb.

SweepFinder2 software [127] was used to calculate the CLR statistic [128] within each breed in non-overlapping sliding windows of 50 kb across the genome. The ancestral allele information was assessed from a cattle reference allele list retrieved from Rocha et al. [129]. The CLR analysis was performed considering only SNPs containing the ancestral allele information (*n*=11,260,629 SNPs).

The iHS [130] and XP-EHH [131] statistics were calculated using the program selscan v1.2.0a [132] with default parameters. Within each population, haplotype phasing was performed using Beagle 5.0 [133] and the genetic distances were determined by assuming that 1 Mb ≈ 1 centiMorgan (cM). The iHS scores were calculated within each breed and XP-EHH between all six pairwise combinations of the four breeds. The unstandardized iHS and XP-EHH scores were standard normalized using the script norm with default parameters, as provided by selscan. Absolute iHS and XP-EHH values were averaged in non-overlapping sliding windows of 50 kb. To compute the iHS statistic, the same subset of SNPs (*n*=11,260,629 SNPs) applied in the CLR statistic was used, however, without considering any ancestral allele information. Independent results for each statistical method and population implemented herein are presented in Additional file 13.

Selective sweeps detection can be enhanced by combining multiple genome-wide scan methodologies, benefiting from advantageous complementarities among them together with the increase in the statistical power [20, 134–137]. Further, combining within-population statistics from multiple breeds may decrease false-positive signals that arise due to population stratification (reviewed by Hellwege et al.

[138]). Accordingly, within-population and cross-population statistics were combined separately in a single score using the DCMS statistic [20]. The DCMS statistic was calculated for each 50 kb window using the MINOTAUR package [139] and the empirical p-values of each statistic were derived from a skewness normal distribution with an appropriate one-tailed test (Additional file 14). Candidates sweep regions under selection were revealed by assessing the top 1% of the empirical distribution generated by the DCMS statistics.

Candidate regions identified herein were compared with previous regions under selection described in the literature for other cattle breeds. Overlap analysis was carried out using the Bioconductor package *GenomicRanges* [140].

Selective sweeps and runs of homozygosity

Candidate sweep regions revealed from the top 1% of the empirical distribution generated by the DCMS statistics were intersected with ROH hotspots to identify common signals between both methodologies. ROH formerly identified to estimate F_{ROH} were applied, and ROH hotspots were determined by selecting segments shared by more than 50% of the samples within each breed.

Overlap analysis was performed separately for each DCMS statistic using the Bioconductor package *GenomicRanges* [140].

Functional annotation of the candidate regions

Genes were annotated within the candidate sweep regions using the cow gene set Ensembl release 94 fetched from the Biomart tool [141]. BEDTools [142] was used to identify overlaps between the retrieved gene set list and the putative sweep regions. DAVID v6.8 tool [116, 117] was used to identify overrepresented GO terms and KEGG pathways using the list of genes from the putative sweep regions and the *Bos taurus taurus* annotation file as a background. The p-values were adjusted by False Discovery Rate [21], and significant terms and pathways were considered when $p<0.01$. QTL retrieved from the CattleQTL database [143] were overlapped with the candidate sweep regions using BEDtools [142].

REFERENCES

1. Primo A. El ganado bovino ibérico en las Américas: 500 años después. Arch Zootec. 1992;41:421–32.
2. Mariante A, Cavalcante N. Animais do descobrimento: raças domésticas da história do Brasil. Empresa Brasileira de Pesquisa Agropcuária, Centro de Pesquisa Agropecuária do Pantanal; 2000.
3. Egito AA, Mariante AS, Albuquerque MSM. Programa brasileiro de conservação de recursos genéticos animais. Arch Zootec. 2002;51:7.
4. Mariante A da S, Albuquerque M do SM, do Egito AA, McManus C. Advances in the Brazilian animal genetic resources conservation programme. Anim Genet Resour Inf. 1999;25:107–21.
5. Felix G, Piovezan U, Juliano R, Silva M, Fioravanti M. Potencial de uso de raças bovinas locais brasileiras: Curraleiro Pé-duro e Pantaneiro. Enciclopédia Biosf. 2013;9:1715–41.
6. Kim J, Hanotte O, Mwai OA, Dessie T, Salim B, Diallo B, et al. The genome landscape of indigenous African cattle. Genome Biol. 2017;18:34.
7. Zander KK, Signorello G, Salvo M De, Gandini G, Drucker AG. Assessing the total economic value of threatened livestock breeds in Italy : Implications for conservation policy. Ecol Econ. 2013;93:219–29.
8. Ugarte E, Ruiz R, Gabia D, Beltrán de Heredia I. Impact of high-yielding foreign breeds on the Spanish dairy sheep industry. Livest Prod Sci. 2001;71:3–10.
9. Carvalho GMC, Fé Da Silva LR;, Almeida MJO;, Lima Neto AF;, Beffa LM. Phenotypic evaluation of Curraleiro Pé-duro breed of cattle from semiarid areas of Brazil. Arch Zootec. 2013;62:23–5.
10. Cardoso CC, Lima FG, Fioravanti MCS, Egito AA, Paula e Silva FC, Tanure CB, et al. Heat tolerance in curraleiro pe-duro, pantaneiro and nelore cattle using thermographic images. Animals. 2016;6.
11. Utsunomiya YT, Pérez O'Brien AMP, Sonstegard TS, Sölkner J, Garcia JF. Genomic data as the "hitchhiker's guide" to cattle adaptation: Tracking the milestones of past selection in the bovine genome. Front Genet. 2015;6.

12. Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet. 2014;46.

13. Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. PLoS Genet. 2014;10:e100414.

14. Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, et al. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. Sci Rep. 2016;6:38932.

15. Egito AA, Martinez AM, Juliano RS, Landi V, Moura MI, Silva MC, et al. Population study of Pantaneiro cattle herds aiming the management and genetic handling of the breed. Actas Iberoam en Conserv Anim. 2016;7:59–63.

16. Campos BM, Carmo AS, Egito AA, Mariante AS, Albuquerque MSM, Gouveia JJS, et al. Genetic diversity, population structure, and correlations between locally adapted zebu and taurine breeds in Brazil using SNP markers. Trop Anim Health Prod. 2017;49:1677–84.

17. Serrano G, Egito A, McManus C, Mariante A. Genetic diversity and population structure of Brazilian native bovine breeds. Pesqui Agropecu Bras. 2004;39:543–9.

18. Pezzini T, Mariante AS, Martins E, Paiva S, Seixas L, Costa JBG, et al. Population structure of Brazilian Crioula lageana cattle (Bos taurus) breed. Rev Colomb Ciencias Pecu. 2018;31:93–102.

19. Egito A, Paiva S, Albuquerque M do S, Mariante A, Almeida L, Castro S, et al. Microsatellite based genetic diversity and relationships among ten Creole and commercial cattle breeds raised in Brazil. BMC Genet. 2007;8:83.

20. Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. Properties of different selection signature statistics and a new strategy for combining them. Heredity (Edinb). 2015;115:426–36.

21. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc. 1995;57:289–300.

22. Kiser JN, Lawrence TE, Neupane M, Seabury CM, Taylor JF, Womack JE, et al. Rapid communication: Subclinical bovine respiratory disease - loci and pathogens associated with lung lesions in feedlot cattle. J Anim Sci. 2017;95:2726–31.

23. Veerkamp RF, Coffey MP, Berry DP, De Haas Y, Strandberg E, Bovenhuis H, et al. Genome-wide associations for feed utilisation complex in primiparous Holstein-Friesian dairy cows from experimental research herds in four European countries. Animal. 2012;6:1738–49.

24. Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TPL, et al. Genome-wide association study of growth in crossbred beef cattle. J Anim Sci. 2010;88:837–48.

25. Purfield DC, Bradley DG, Evans RD, Kearney FJ, Berry DP. Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. Genet Sel Evol. 2015;47:47.

26. Mateescu RG, Garrick DJ, Reecy JM. Network analysis reveals putative genes affecting meat quality in Angus cattle. Front Genet. 2017;8.

27. Strillacci MG, Frigo E, Schiavini F, Samoré AB, Canavesi F, Vevey M, et al. Genome-wide association study for somatic cell score in Valdostana Red Pied cattle breed using pooled DNA. BMC Genet. 2014;15:106.

28. Crispim AC, Kelly MJ, Guimarães SEF, E Silva FF, Fortes MRS, Wenceslau RR, et al. Multi-trait GWAS and new candidate genes annotation for growth curve parameters in brahman cattle. PLoS One. 2015;10:e0139906.

29. Mapholi NO, Maiwashe A, Matika O, Riggio V, Bishop SC, MacNeil MD, et al. Genome-wide association study of tick resistance in South African Nguni cattle. Ticks Tick Borne Dis. 2016;7:487–97.

30. Frischknecht M, Bapst B, Seefried FR, Signer-Hasler H, Garrick D, Stricker C, et al. Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. BMC Genomics. 2017;18.

31. Hawken RJ, Zhang YD, Fortes MRS, Collis E, Barris WC, Corbet NJ, et al. Genome-wide association studies of female reproduction in tropically adapted beef cattle. J Anim Sci. 2012;90:1398–410.

32. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 2011;12:408.

33. Nayeri S, Sargolzaei M, Abo-Ismail MK, May N, Miller SP, Schenkel F, et al.

Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 2016;17:75.

34. Meredith BK, Kearney FJ, Finlay EK, Bradley DG, Fahey AG, Berry DP, et al. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet. 2012;13:21.

35. Iso-Touru T, Sahana G, Guldbrandtsen B, Lund MS, Vilkki J. Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. BMC Genet. 2016;17:55.

36. Tetens J, Seidenspinner T, Buttchereit N, Thaller G. Whole-genome association study for energy balance and fat/protein ratio in German Holstein bull dams. Anim Genet. 2013;44:1–8.

37. Wu X, Fang M, Liu L, Wang S, Liu J, Ding X, et al. Genome wide association studies for body conformation traits in the Chinese Holstein cattle population. BMC Genomics. :897.

38. Parker Gaddis KL, Null DJ, Cole JB. Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits. J Dairy Sci. 2016;99:6420–35.

39. Huson HJ, Kim E-S, Godfrey RW, Olson TA, McClure MC, Chase CC, et al. Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle. Front Genet. 2014;5.

40. Bahbahani H, Clifford H, Wragg D, Mbole-Kariuki MN, Van Tassell C, Sonstegard T, et al. Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. Sci Rep. 2015;5:11729.

41. Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Van Tassell CP, et al. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. Mol Biol Evol. 2015;32:711–25.

42. Makina SO, Muchadeyi FC, Van Marle-Köster E, Taylor JF, Makgahlela ML, Maiwashe A. Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genet Sel Evol. 2015;47:92.

43. González-Rodríguez A, Munilla S, Mouresan EF, Cañas-Álvarez JJ, Díaz C, Piedrafita J, et al. On the performance of tests for the detection of signatures of selection: a case study with the Spanish autochthonous beef cattle populations. Genet

Sel Evol. 2016;48:81.
44. Rothammer S, Seichter D, Förster M, Medugorac I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. BMC Genomics. 2013;14:908.
45. Pitt D, Bruford MW, Barbato M, Orozco-terWengel P, Martínez R, Sevane N. Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. Evol Appl. 2019;12:105–22.
46. Iso-Touru T, Tapio M, Vilkki J, Kiseleva T, Ammosov I, Ivanova Z, et al. Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. Anim Genet. 2016;47:647–57.
47. Somavilla AL, Sonstegard TS, Higa RH, Rosa AN, Siqueira F, Silva LOC, et al. A genome-wide scan for selection signatures in Nellore cattle. Anim Genet. 2014;45:771–81.
48. Liao X, Peng F, Forni S, McLaren D, Plastow G, Stothard P. Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection. Genome. 2013;56:592–8.
49. Kim J, Hanotte O, Mwai OA, Dessie T, Salim B, Diallo B, et al. The genome landscape of indigenous African cattle. Genome Biol. 2017;18:34.
50. Mei C, Wang H, Liao Q, Wang L, Cheng G, Wang H, et al. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. Mol Biol Evol. 2018;35:688–99.
51. Wang Z, Ma H, Xu L, Zhu B, Liu Y, Bordbar F, et al. Genome-Wide Scan Identifies Selection Signatures in Chinese Wagyu Cattle Using a High-Density SNP Array. Animals. 2019;9.
52. Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet Sel Evol. 2015;47:49.
53. Pérez O'Brien AM, Utsunomiya YT, Mészáros G, Bickhart DM, Liu GE, Van Tassell CP, et al. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. Genet Sel Evol. 2014;46:19.
54. Boitard S, Boussaha M, Capitan A, Rocha D, Servin B. Uncovering adaptation from sequence data: Lessons from genome resequencing of four cattle breeds. Genetics.

2016;203:433–50.
55. Stella A, Ajmone-Marsan P, Lazzari B, Boettcher P. Identification of selection signatures in cattle breeds selected for dairy production. Genetics. 2010;185:1451–61.
56. Machugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG. Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (Bos Taurus and Bos Indicus). Genetics. 1997;146:1071–86.
57. Hiendleder S, Lewalski H, Janke A. Complete mitochondrial genomes of Bos taurus and Bos indicus provide new insights into intra-species variation, taxonomy and domestication. Cytogenet Genome Res. 2008;120:150–6.
58. Chan EKF, Nagaraj SH, Reverter A. The evolution of tropical adaptation: Comparing taurine and zebu cattle. Anim Genet. 2010;41:467–77.
59. Mazza M, Mazza C, Sereno J, Santos S, Pellegrin A. Etnobiologia e conservação do bovino Pantaneiro. Empresa Brasileira de Pesquisa Agropcuária, Centro de Pesquisa Agropecuária do Pantanal; 1994.
60. Issa ÉC, Jorge W, Sereno JRB. Cytogenetic and molecular analysis of the Pantaneiro cattle breed. Pesqui Agropecu Bras. 2006;41:1609–15.
61. Queiroz SA, Pelicioni LC, Silva BF, Sesana JC, Martins MIEG, Sanches A. Selection indices for a dual purpose breed Caracu. Rev Bras Zootec. 2005;34:827–37.
62. Mariante AS, Egito AA, Albuquerque M do SM, Paiva SR, Ramos AF. Managing genetic diversity and society needs. Rev Bras Zootec. 2008;37:127–36.
63. Mazza MCM, Mazza CA, Sereno JRB, Santos SAL, Mariante AS. Conservation of Pantaneiro cattle in Brazil: Historical origin. Arch Zootec. 1992;41:443–53.
64. Mariante AS, Albuquerque M do SM, Egito AA, McManus C, Lopes MA, Paiva SR. Present status of the conservation of livestock genetic resources in Brazil. Livest Sci. 2009;120:204–12.
65. Queiroz SA, Lôbo RB. Genetic relationship, inbreeding and generation interval in registered Gir cattle in Brazil. J Anim Breed Genet. 1993;110:228–33.
66. Wright S. Coefficients of Inbreeding and Relationship. Am Nat. 1922;56:330–8.
67. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim Genet. 2014;46:110–21.

68. Kim ES, Cole JB, Huson H, Wiggans GR, Van Tassel CP, Crooker BA, et al. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. PLoS One. 2013;8:e80813.
69. Reis Filho JC, Lopes PS, Verneque R da S, Torres R de A, Teodoro RL, Carneiro PLS. Population structure of Brazilian Gyr dairy cattle. Rev Bras Zootec. 2010;39:2640–5.
70. Santana Junior ML, Pereira RJ, Bignardi AB, El Faro L, Tonhati H, Albuquerque LG. History, structure, and genetic diversity of Brazilian Gir cattle. Livest Sci. 2014;163:26–33.
71. Peripolli E, Baldi F, da Silva MVGB, Irgang R, Lima ALF, R. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. BMC Genomics. 2018;19:34.
72. Neves HHR, Scalez DCB, Queiroz SA, Desidério JA, Pimentel ECG. Preliminary study to determine extent of linkage disequilibrium and estimates of autozygosity in Brazilian Gyr dairy cattle. Arch Zootec. 2015;64:99–108.
73. Ferguson JD, Galligan DT, Thomsen N. Principal Descriptors of Body Condition Score in Holstein Cows. J Dairy Sci. 1994;77:2695–703.
74. Bauman DE, Bruce Currie W. Partitioning of Nutrients During Pregnancy and Lactation: A Review of Mechanisms Involving Homeostasis and Homeorhesis. J Dairy Sci. 1980;63:1514–29.
75. Bell AW. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. J Anim Sci. 1995;73:2804–19.
76. Whitaker DA, Goodger WJ, Garcia M, Perera BMAO, Wittwer F. Use of metabolic profiles in dairy cattle in tropical and subtropical countries on smallholder dairy farms. Prev Vet Med. 1999;38:119–31.
77. Stockdale CR. Body condition at calving and the performance of dairy cows in early lactation under Australian conditions: A review. Aust J Exp Agric. 2001;41:823–39.
78. Collard BL, Boettcher PJ, Dekkers JCM, Petitclerc D, Schaeffer LR. Relationships between energy balance and health traits of dairy cattle in early lactation. J Dairy Sci. 2000;83:2683–90.
79. Taye M, Kim J, Yoon SH, Lee W, Hanotte O, Dessie T, et al. Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher

quality beef. BMC Genet. 2017;18:11.
80. Roux PF, Boitard S, Blum Y, Parks B, Montagner A, Mouisel E, et al. Combined QTL and selective sweep mappings with coding SNP annotation and cis-eQTL analysis revealed PARK2 and JAG2 as new candidate genes for adiposity regulation. G3 Genes, Genomes, Genet. 2015;5:517–29.
81. dos Santos FC, Peixoto MGCD, Fonseca PA de S, Pires M de FÁ, Ventura RV, Rosse I da C, et al. Identification of Candidate Genes for Reactivity in Guzerat (Bos indicus) Cattle: A Genome-Wide Association Study. PLoS One. 2017;12:e0169163.
82. Lee YL, Bosse M, Mullaart E, Groenen MAM, Veerkamp RF, Bouwman AC. Functional and population genetic features of copy number variations in two dairy cattle populations. BMC Genomics. 2020;21:89.
83. Valente TS, Baldi F, Sant'Anna AC, Albuquerque LG, Costa MJRP Da. Genome-wide association study between single nucleotide polymorphisms and flight speed in Nellore cattle. PLoS One. 2016;11:e0156956.
84. Burrow HM, Prayaga KC. Correlated responses in productive and adaptive traits and temperament following selection for growth and heat resistance in tropical beef cattle. Livest Prod Sci. 2004;86:143–61.
85. Burdick NC, Randel RD, Carroll JA, Welsh TH. Interactions between temperament, stress, and immune function in cattle. Int J Zool. 2011;2011.
86. Voisinet BD, Grandin T, Tatum JD, O'Connor SF, Struthers JJ. Feedlot cattle with calm temperaments have higher average daily gains than cattle with excitable temperaments. J Anim Sci. 1997;75:892–6.
87. Silveira IDB, Fischer V, Farinatti LHE, Restle J, Filho DCA, de Menezes LFG. Relationship between temperament with performance and meat quality of feedlot steers with predominantly Charolais or Nellore breed. Rev Bras Zootec. 2012;41:1468–76.
88. Cafe LM, Robinson DL, Ferguson DM, Mcintyre BL, Geesink GH, Greenwood PL. Cattle temperament: Persistence of assessments and associations with productivity, efficiency, carcass and meat quality traits. J Anim Sci. 2011;89:1452–65.
89. Petherick JC, Holroyd RG, Swain AJ. Performance of lot-fed Bos indicus steers exposed to aspects of a feedlot environment before lot-feeding. Aust J Exp Agric. 2003;43:1181–91. doi:10.1071/EA02118.

90. Burrow HM. Measurement of temperament and their relationship with performance traits of beef cattle. Anim Breed Abstr. 1997;65:478–495.
91. Frischknecht M, Flury C, Leeb T, Rieder S, Neuditschko M. Selection signatures in Shetland ponies. Anim Genet. 2016;47:370–2.
92. Avila F, Mickelson JR, Schaefer RJ, McCue ME. Genome-wide signatures of selection reveal genes associated with performance in American Quarter Horse subpopulations. Front Genet. 2018;9.
93. Gardner JL, Peters A, Kearney MR, Joseph L, Heinsohn R. Declining body size: A third universal response to warming? Trends Ecol Evol. 2011;26:285–91.
94. Martin JM, Mead JI, Barboza PS. Bison body size and climate change. Ecol Evol. 2018;8:4564–74.
95. Dickerson GE. Animal size and efficiency: Basic concepts. Anim Prod. 1978;27:367–79.
96. Mccain CM, King SRB. Body size and activity times mediate mammalian responses to climate change. Glob Chang Biol. 2014;20:1760–9.
97. Pacifici M, Visconti P, Butchart SHM, Watson JEM, Cassola FM, Rondinini C. Species' traits influenced their response to recent climate change. Nat Clim Chang. 2017;7:205–8.
98. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14:807–20.
99. Taylor CR, Caldwell SL, Rowntree VJ. Running up and down hills: Some consequences of size. Science (80-). 1972;178:1096–7.
100. Araújo Teixeira RM, Lana R de P, Fernandes L de O, de Oliveira AS, de Queiroz AC, de Oliveira Pimentel JJ. Desempenho produtivo de vacas da raça Gir leiteira em confinamento alimentadas com níveis de concentrado e proteína bruta nas dietas. Rev Bras Zootec. 2010;39:2527–34.
101. Mcmanus C, Seixas L. A Raça Crioula Lageana. 2010. www.animal.unb.br.
102. Issa ÉC, Jorge W, Egito AA, Sereno JRB. Cytogenetic analysis of the Y chromosome of native brazilian bovine breeds: preliminary data. Arch Zootec. 2009;58:93–101.
103. Araujo AM de, Ramos AF, Egito AA do, Mariante A da S, Varela ES, Figueiredo EAP de, et al. Núcleos de conservação de Bovinos. In: Albuquerque M do SM, Ianella

P, editors. Inventário de Recursos Genéticos Animais da Embrapa. Brasília: Empresa Brasileira de Pesquisa Agropecuária; 2016. p. 17–23.
104. Gutiérrez-Gil B, Arranz JJ, Wiener P. An interpretive review of selective sweep studies in Bos taurus cattle populations: Identification of unique and shared selection signals across breeds. Front Genet. 2015;6:167.
105. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. BioEssays. 2013;35:780–6.
106. Jakobsson M, Edge MD, Rosenberg NA. The relationship between FST and the frequency of the most frequent allele. Genetics. 2013;193:515–28.
107. Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. 2005;15:1496–502.
108. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. Livest Sci. 2014;166:133–43.
109. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv. 2013;1303.
110. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
111. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
112. DePristo MA, Rivas MA, McKenna A, Hartl C, del Angel G, Sivachenko AY, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.
113. Garimella K V, Levy-Moonshine A, Jordan T, Van der Auwera GA, Hartl C, del Angel G, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinforma. 2013;11:11.10.1-11.10.33.
114. Sherry ST. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–11.
115. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17:122.

116. Huang DW, Sherman BT, Lempicki R a. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
117. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1–13.
118. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131:479–91.
119. Paradis E. pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics. 2010;26:419–20.
120. R Core Team R. R: A Language and Environment for Statistical Computing. Available online at https://www.R-project.org/.; 2015.
121. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
122. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.
123. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of Homozygosity in European Populations. Am J Hum Genet. 2008;83:359–72.
124. Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of Homozygosity: A comparison of SNP Array and whole genome sequence low coverage data. BMC Genomics. 2018;19:106.
125. Tukey JW. Comparing Individual Means in the Analysis of Variance. Biometrics. 1949;5:99–114.
126. Wright S. The Genetical Structure of populations. Nature. 1950;166:247–9.
127. DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. SweepFinder2: increased sensitivity, robustness and flexibility. Bioinformatics. 2016;32:1895–7.
128. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Res. 2005;15:1566–75.
129. Rocha D, Billerey C, Samson F, Boichard D, Boussaha M. Identification of the putative ancestral allele of bovine single-nucleotide polymorphisms. J Anim Breed

Genet. 2014;131:483–6.
130. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006;4:e72.
131. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449:913–8.
132. Szpiech ZA, Hernandez RD. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014;31:2824–7.
133. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am J Hum Genet. 2018;103:338–48.
134. Utsunomiya YT, Pérez O'Brien AM, Sonstegard TS, Van Tassell CP, do Carmo AS, Mészáros G, et al. Detecting Loci under Recent Positive Selection in Dairy and Beef Cattle by Combining Different Genome-Wide Scan Methods. PLoS One. 2013;8:e64280.
135. Randhawa IAS, Khatkar MS, Thomson PC, Raadsma HW. Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. BMC Genet. 2014;15:34. doi:10.1186/1471-2156-15-34.
136. Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, et al. A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection. Science (80-). 2010;327:883–6.
137. Lin K, Li H, Schlötterer C, Futschik A. Distinguishing positive selection from neutral evolution: Boosting the performance of summary statistics. Genetics. 2011;187:229–44.
138. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. Curr Protoc Hum Genet. 2017;95:1.22.1-1.22.23.
139. Verity R, Collins C, Card DC, Schaal SM, Wang L, Lotterhos KE. MINOTAUR: A platform for the analysis and visualization of multivariate results from genome scans with R Shiny. Mol Ecol Resour. 2017;17:33–43.
140. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol. 2013;9:e1003118.

141. Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. BioMart Central Portal--unified access to biological data. Nucleic Acids Res. 2009;37:W23–7.
142. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
143. Hu ZL, Park CA, Reecy JM. Developmental progress and current status of the Animal QTLdb. Nucleic Acids Res. 2016;44:D827–33.

FIGURES

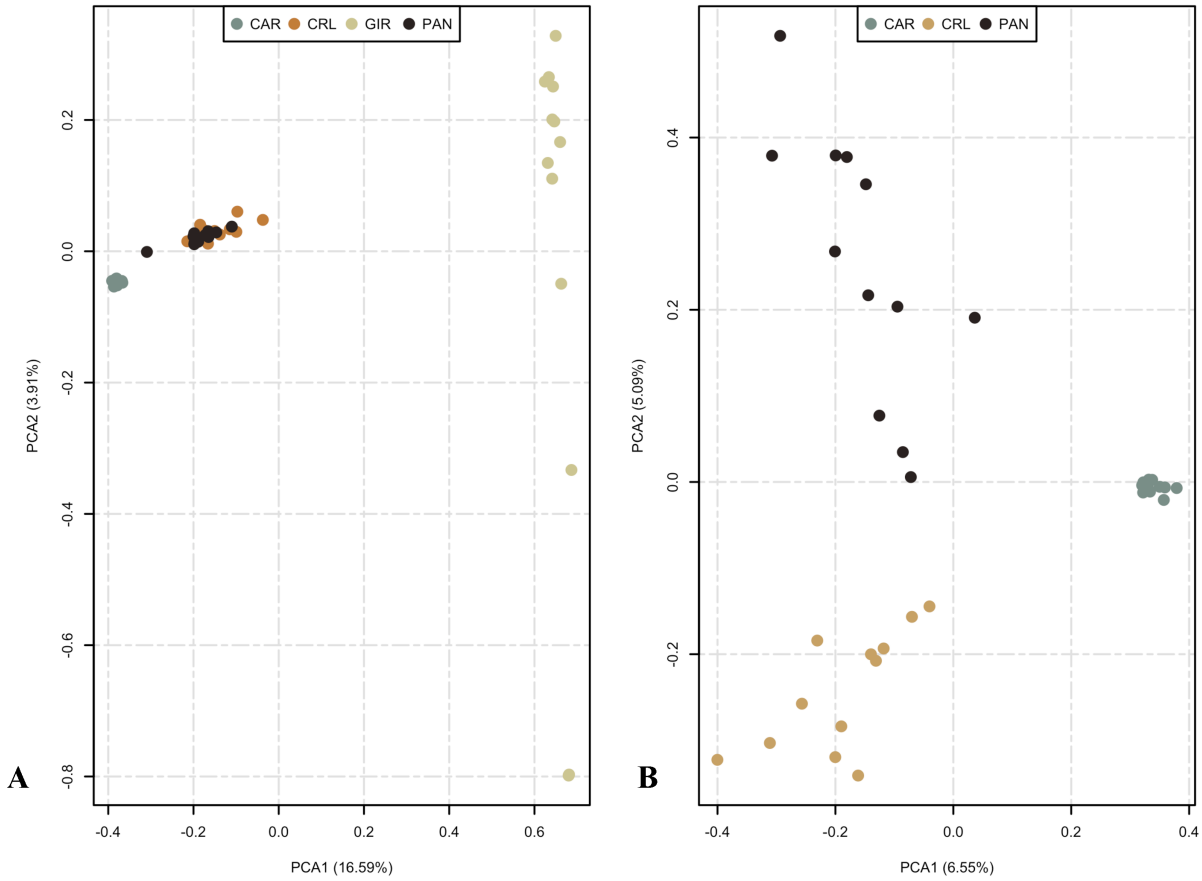


Figure 1. Principal components analysis (PCA) scores plot with variance explained by the first two principal components in brackets. (**A)** PCA scores for the four breeds (Caracu Caldeano – CAR, Crioulo Lageano – CRL, Gir – GIR, and Pantaneiro - PAN. (**B)** PCA scores for the locally adapted taurine cattle breeds (Caracu Caldeano – CAR, Crioulo Lageano – CRL, and Pantaneiro – PAN).

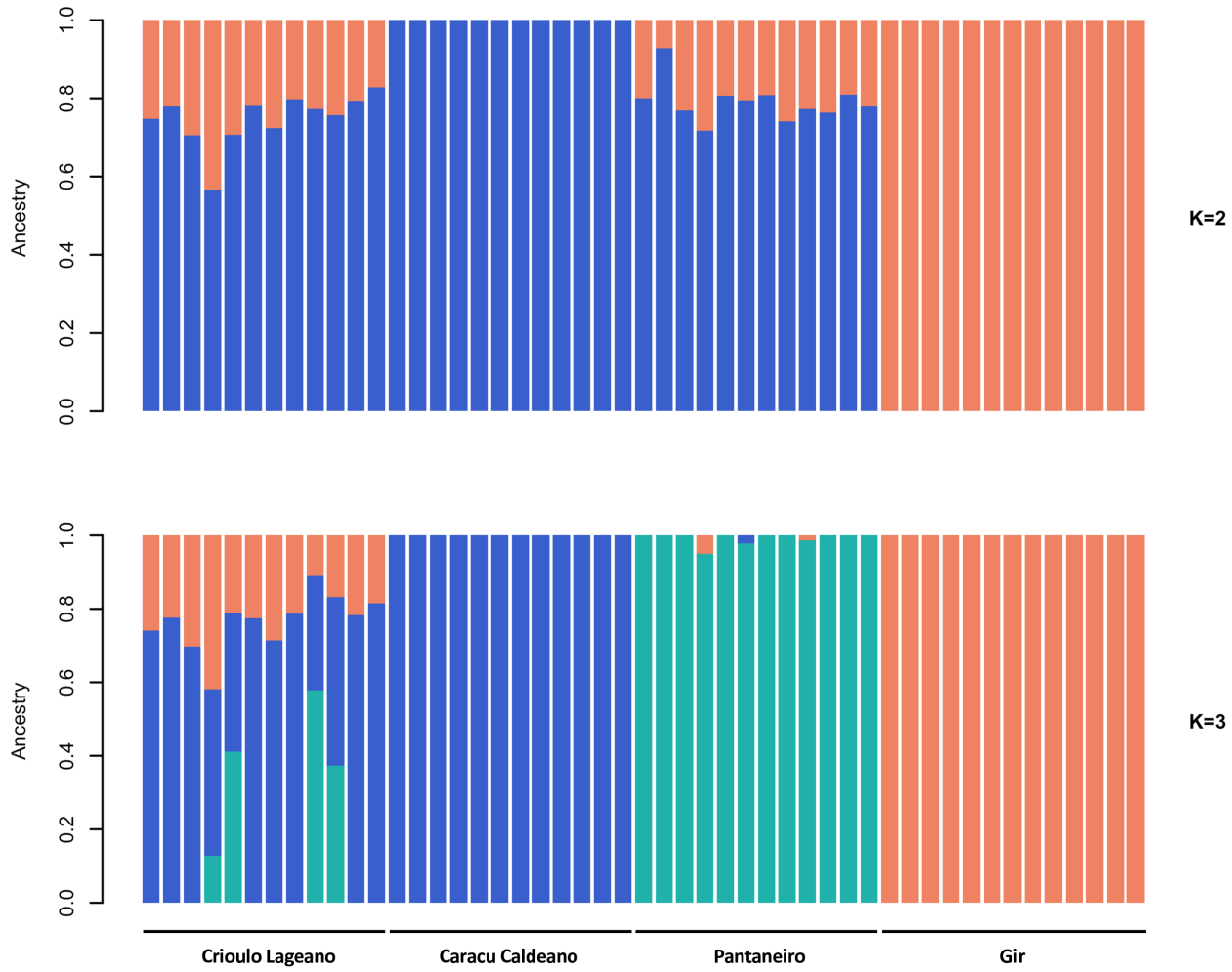


Figure 2. Population structure inferred by using the ADMIXTURE software. Each sample is denoted by a single vertical bar partitioned into *K* colors according to its proportion of ancestry in each of the clusters. Ancestral contributions for *K=2* and *K=3* are graphically represented.

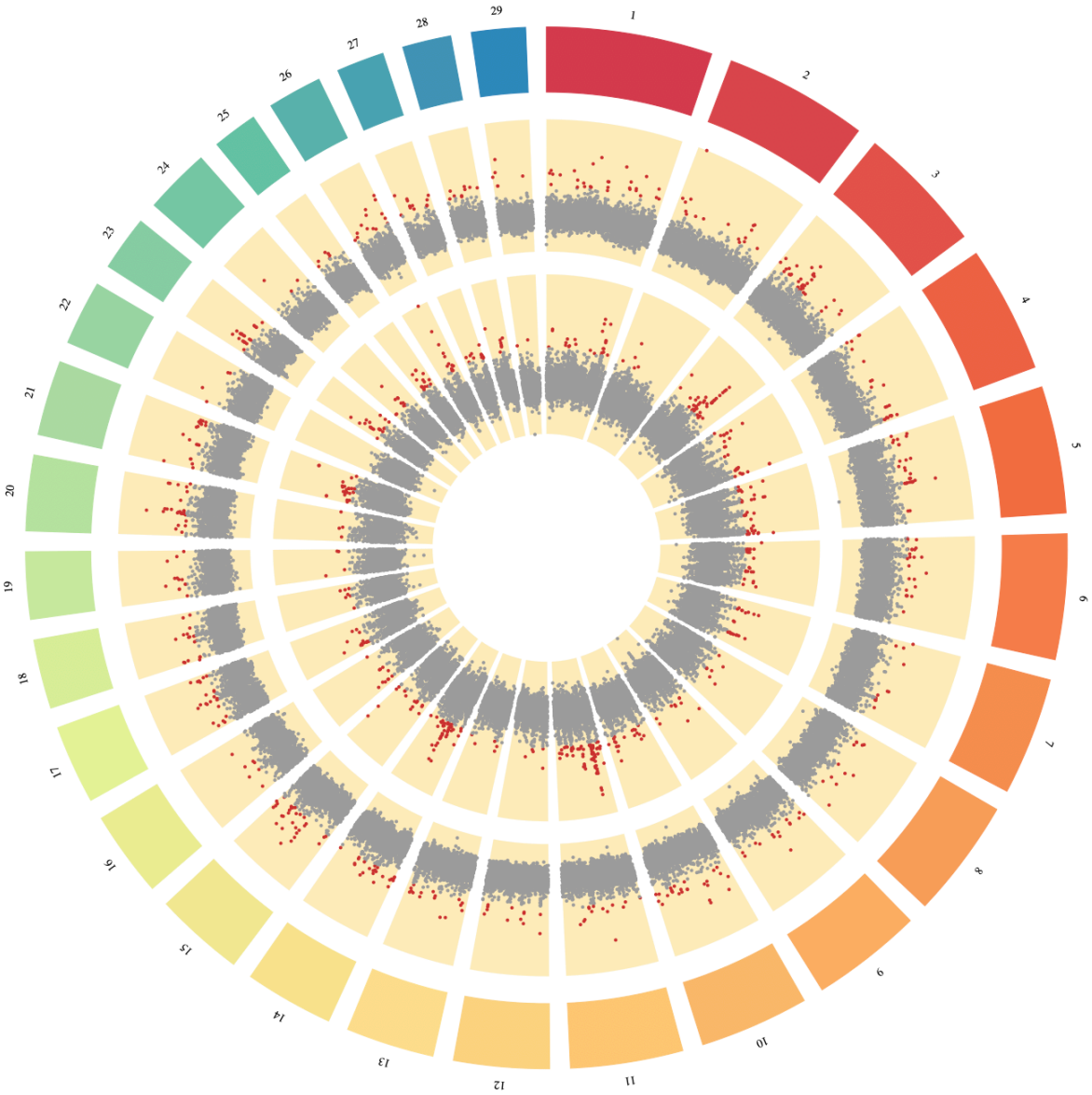


Figure 3. Whole-genome signatures of selection for the within-population DCMS statistic (outer circle) and cross-population DCMS statistic (inner circle). The x-axis shows the window position along the chromosome, and the y-axis the DCMS value associated with such window. Reds dots correspond to the top 1% of the empirical distribution generated by the DCMS statistics.

CAPÍTULO 5 – ASSESSMENT OF COPY NUMBER VARIANTS IN THREE BRAZILIAN LOCALLY ADAPTED CATTLE BREEDS USING WHOLE-GENOME RE-SEQUENCING DATA

ABSTRACT

Further characterization of genetic structural variations should strongly focus on small and endangered local breeds given their role in unraveling genes and structural variants underlying selective pressures and phenotype variation. Therefore, a comprehensive genome-wide assessment of copy number variants (CNVs) based on whole-genome re-sequencing data was performed on three Brazilian locally adapted cattle breeds (Caracu Caldeano - CAR, Crioulo Lageano - CRL, and Pantaneiro - PAN) using the ARS-UCD1.2 genome assembly. Data from 36 individuals with an average coverage depth of 14.07X per individual was used. A total of 24,945 CNVs were identified distributed among the three breeds (CAR=7,285, CRL=7,297, and PAN=10,363). Deletion events were 1.75 to 2.07-fold higher than duplications, and the total length of CNVs is composed mostly of a high number of segments between 10 and 30 kb. CNVs regions (CNVRs) are not uniformly scattered throughout the genomes (*n=463*), and 105 CNVRs were found overlapping among the studied breeds. Functional annotation of the CNVRs revealed variants with high consequence on protein sequence harboring relevant genes, in which we can highlight the *BOLA-DQB, BOLA-DQA5*, *CD1A*, β-defensins, *PRG3*, and *ULBP21* genes. Enrichment analysis based on the gene list retrieved from the CNVRs disclosed over-represented terms ($p<0.01$) strongly associated with immunity and cattle resilience to harsh environments. Additionally, QTLs associated with body conformation and dairy-related traits were also unveiled within the CNVRs. These results can provide important understandings to better receipt the selective forces shaping the genome of such cattle breeds and identify traces of natural selection pressures by which these populations have been exposed to challenging environmental conditions.

Key-words: *Bos taurus taurus*, CNV, local breeds, next-generation sequencing, structural variants

INTRODUCTION

Copy number variants (CNVs) are chromosomal rearrangements (≥1 kilobase) triggered by changes in DNA content and structure (FEUK; CARSON; SCHERER, 2006) leading to a change in the order (inversions and translocations) and the number of copies (duplications and deletions) of a genomic region (HENRICHSEN; CHAIGNAT; REYMOND, 2009). CNVs represent an important source of genetic and phenotypic variability among individuals and populations (BECKMANN; ESTIVILL; ANTONARAKIS, 2007; CONRAD; ANTONARAKIS, 2007; LOW et al., 2019; ZHOU et al., 2016), exerting a significant evolutionary impact by generating the required variation in the population through the change in gene structure and dosage as well as by regulating gene expression and function (ZHANG et al., 2009). Hence, this source of variation may account for more differences among individuals due to the cumulative number of nucleotides affected than do single-nucleotide polymorphisms (SNP) (CONRAD et al., 2010; MCCARROLL; ALTSHULER, 2007; ZHANG et al., 2009). Furthermore, a significant proportion of CNVs encompass genomic regions not well covered by SNP arrays such as segmental duplications regions, and consequently, were not properly genotyped (ESTIVILL; ARMENGOL, 2007). Therefore, CNVs may provide genomic structural information complementary to SNP data (SCHERER et al., 2007).

Different methodologies have been applied to identify CNVs at a genome-wide scale, including comparative genomic hybridization arrays, SNP-genotyping microarrays, and high-throughput sequencing (CLOP; VIDAL; AMILLS, 2012; GERLANDO et al., 2019). Although the first two array platforms may be affected by low probe density (BICKHART et al., 2012), they have been widely used for CNVs detection in several livestock species, particularly in cattle (BAE et al., 2010; FADISTA et al., 2010; GERLANDO et al., 2019; HOU et al., 2011; KIJAS et al., 2011; LIU et al., 2010). Advances in high-throughput genome scan technologies combined with appropriate algorithms have provided better approaches to systematically identify genome-wide CNVs at a higher effective resolution, frequency and sensitivity, allowing the identification of a vast number of structural variants, especially those that have

been previously undetectable due to their small sizes (ALKAN et al., 2009; BICKHART et al., 2012; CLOP; VIDAL; AMILLS, 2012).

CNVs have been associated with heritable complex traits in several species, and lately, the interest in CNVs discovery has extended into livestock species (DUPUIS et al., 2013; FONTANESI et al., 2010, 2011; RAMAYO-CALDAS et al., 2010). Interestingly, genome-wide CNVs studies in local and less notorious breeds have been addressed in the literature (GERLANDO et al., 2019; MOLNÁR et al., 2014; TIAN et al., 2013; WANG et al., 2016; YANG et al., 2017; ZHANG et al., 2015; ZHOU et al., 2014), however, despite the importance of such breeds to a wide range of challenging environments, studies deciphering their genetic structure are still a minority when compared to those accomplished in highly-specialized commercial breeds. Brazilian locally adapted taurine cattle breeds originated from the cattle brought by Portuguese conquerors in 1534 during the Brazilian colonization period (MARIANTE et al., 1999; MARTINS et al., 2009; MAZZA et al., 1994; PRIMO, 1992). These cattle have undergone to a process of natural selection in a remarkably set of ecosystems throughout the country for more than 450 years facing hot, dry or humid tropical climate conditions, scarce food availability, diseases, and parasite infestations (MARIANTE; CAVALCANTE, 2000). Strong environmental pressures, natural selection, and recurring events of breed admixture led to the development of the Brazilian locally adapted cattle breeds, which have acquired very particular traits over time to thrive in distinct ecosystems (MARIANTE et al., 1999).

Further characterization of genetic structural variations, particularly in local breeds, is an important step towards deciphering the molecular mechanisms underlying trait variation, survivorship, and breed adaptation. Therefore, this study reports for the first-time a genome-wide characterization of CNVs derived from whole-genome re-sequencing data in Caracu Caldeano, Crioulo Lageano and Pantaneiro, three Brazilian locally adapted taurine cattle breeds. The breeds examined herein have evolved under challenging environments and might harbor important phenotypic traits and evidence of positive selection that will help secure cattle production in a changing environment.

MATERIAL AND METHODS

Samples, sequencing, and raw data preparation

Sequencing analysis was based on data from one dairy (12 Caracu Caldeano; CAR) and two dual-purpose (12 Crioulo Lageano; CRL and 12 Pantaneiro; PAN) cattle breeds from Embrapa Dairy Cattle (Juiz de Fora, Minas Gerais, Brazil). Animals were sampled from three Brazilian geographical regions, including the south (CRL), southeast (CAR), and mid-west (PAN). The population structure among the breeds together with their history and breed development can be further assessed in Peripolli et al. (2020)

DNA was extracted from blood samples and paired-end whole-genome re-sequencing with 2x100 base pair reads (CRL) and 2x125 base pair reads (CAR and PAN) was performed on the Illumina HiSeq2500 platform with an aimed average sequencing depth of 15X. Pair-end reads were aligned to the *Bos taurus taurus* genome assembly ARS-UCD1.2 using Burrows-Wheeler Alignment MEM (BWA-MEM) tool v.0.7.17 (LI, 2013) and converted into a binary format using SAMtools v.1.8 (LI et al., 2009). PCR duplicates were marked using Picard tools (http://picard.sourceforge.net, v.2.18.2).

CNVs and CNVRs detection

The read depth-based method implemented in CNVnator v0.4.1 (ABYZOV et al., 2011) software was used to call CNVs for each sample relative to the *Bos taurus taurus* genome assembly ARS-UCD1.2. The bin size was set to 500 (CAR and CRL) and 600 bp (PAN) based on the ratio of the average read depth signal to its standard deviation. Quality control was undertaken to remove unreliable raw CNVs and reduce the false discovery rate. CNVs calls with a p-value lower than 0.01 for the t-test statistics (e-val1) together with the fraction of mapped reads with zero quality (q0) lower than 0.5 and CNVs smaller than 1 kb in length were filtered out. Only autosomal chromosomes were included in the analysis.

CNV regions (CNVRs) were identified by overlapping individual CNVs within each breed (REDON et al., 2006), and only those found overlapping in all individuals

within a breed by at least 1 bp were used for downstream analysis. Shared CNVRs among the studied breeds were also identified by overlapping the CNVRs identified within each breed, and only those described overlapping in all three breeds were used for further analysis. Overlapping analyses were carried out using the Bioconductor package *GenomicRanges* (LAWRENCE et al., 2013).

Variant annotation and predicted functional impacts

A functional annotation analysis of the called variants (CNVRs) was performed to assess their possible biological impact using the Variant Effect Predictor (VEP) (MCLAREN et al., 2016) together with the Ensembl genes release 100, version April 2020 (assembly ARS-UCD1.2). Variants with a high consequence on protein sequence (i.e., splice acceptor variant, splice donor variant, stop gained, frameshift variant, stop lost, and start lost) were selected for further assessment.

Functional annotation

Genes were annotated within the CNVRs using the cow gene set Ensembl genes release 100 (ARS-UCD1.2) fetched from the Biomart tool (HAIDER et al., 2009). Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 tool (HUANG; SHERMAN; LEMPICKI, 2009a, 2009b) was used to identify overrepresented ($p<0.01$) gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the list of genes from CNVRs and the *Bos taurus taurus* annotation file as a background. Quantitative trait locus (QTL) retrieved from the CattleQTL database (HU; PARK; REECY, 2016) were overlapped with the CNVRs using Bedtools (QUINLAN; HALL, 2010).

RESULTS

Data

With Illumina paired-end sequencing technology, we obtained re-sequencing data from 36 individuals from three different Brazilian locally adapted taurine cattle

breeds. After mapping the reads to the genome assembly ARS-UCD1.2, an average coverage depth of 14.07X was obtained. As disclosed in the literature, an average coverage depth between 4 to 8X allows sufficient power for CNVs detection using the read depth-based method (BICKHART et al., 2012; SUDMANT et al., 2010).

CNV and CNVRs discovery

Four outlier samples (one for CRL and three for PAN) were filtered out from the dataset after CNV calling due to the discrepant number of CNVs identified.

A total of 7,285 CNVs (4,640 deletions and 2,645 duplications) was identified in the CAR breed. On an individual animal basis, the average number of CNVs per animal was 607.08, with an average length of 28.30 kb and encompassing approximately 0.63% (17.18 Mb) of the total autosomal genome extension (ARS-UCD1.2). In the CRL breed, the total number of CNVs was 7,297 (4,726 deletions and 2,571 duplications), displaying an average number of 663.36 CNVs per animal together with an average length of 27.60 kb and covering roughly 0.67% (18.31 Mb) of the total autosomal genome extension. For the PAN breed, 10,363 CNVs (6,998 deletions and 3,365 duplications) were identified, with an average number of 1151.44 CNVs per animal and an average length of 34.06 kb, encompassing nearly 1.44% (39.22 Mb) of the total autosomal genome extension.

The longest CNVs within each breed were very close in size among the studied breeds and were all events of deletion, with values of 1004.99 kb in length on BTA10:23775501-24780500 bp (CRL), 1006.99 kb in length on BTA10:23773501-24780500 bp (CAR), and 1007.39 kb in length on BTA9:104447401-105454800 bp and BTA10:23773201-24780600 bp (PAN). Remarkably, the genomic region on BTA10:23775501-24780500 bp was found overlapping in all three breeds within the longest CNVs described. When inspecting in detail, such genomic region did not harbor any gene nor QTL. The number of CNVs per chromosome was greater on BTA1 for the PAN (*n=662*) cattle and on BTA15 for the CRL (*n=518*) and CAR (*n=496*) cattle breeds (Appendix 1D to 3D). The total length of CNVs for the studied breeds is composed mostly of a high number of segments between 10 and 30 kb, which

accounted for approximately 47% (CAR; *n=3,443* and CRL; *n=3,422*) and 55% (PAN; *n=5,737*) of all CNVs detected (Figure 1A).

The CNVRs were not evenly distributed throughout the genomes, with some chromosomes missing CNVRs and others containing several such regions (Figure 2, Appendix 4D to 6D). The total length of CNVRs is also composed mostly of a high number of segments between 10 and 30 kb in length (Figure 1B). A total of 153 CNVRs were identified in the CAR breed, including 49 deletions, 102 duplications, and 2 mixed (deletion and duplication within the same region) events. Such CNVRs covered roughly 0.09% (2.45 Mb) of the autosomal genome extension (ARS-UCD1.2), with an average length size of 16.05 kb and values ranging from 1.00 to 79.50 kb. In the CRL breed, the total number of CNVRs was 140 (46 deletions, 86 duplications, and 8 mixed events), covering approximately 0.08% (2.17 Mb) of the autosomal genome extension with an average length size of 15.53 kb and values ranging from 0.50 to 114.50 kb. For the PAN breed, a total of 170 CNVRs were described, encompassing 61 deletions, 99 duplications, and 10 mixed events. The CNVRs covered nearly 0.13% (3.60 Mb) of the autosomal genome extension, with an average length size of 21.122 kb and values ranging from 0.50 to 200.50 kb.

The number of CNVRs per chromosome was greater on BTA1 for the CAR (*n=17*) and CRL (*n=13*) cattle breeds (Figure 3A and B, respectively), and BTA12 showed the greatest enrichment for the PAN (*n=29*) cattle (Figure 3C). It is worth highlighting that the number of CNVRs duplication events was higher (~1.85-fold) than did the deletions. Shared CNVRs (*n=105*) were observed in between the studied breeds, with a length size varying from 1.00 to 52.00 kb and a mean size of 14.34 kb (Appendix 7D).

Variant annotation and gene assessment

Functional classification showed that most of the variants identified within the CNVRs were located in intergenic and intronic regions (Appendix 8D), and several variants with a high consequence on protein sequence were identified (CAR *n=43*; CRL *n=37*; PAN *n=57*; and shared CNVRs *n=53*; Appendix 9D to 12D). Following variant annotation, we further investigated the gene content within the predicted

variants to cause relevant biological functions. A total of 30, 22, 42, and 26 protein-coding genes were described within variants with a high consequence on protein sequence for CAR, CRL, PAN, and shared CNVRs, respectively. Among them, it is worth to underscore the *BOLA-DQB, BOLA-DQA5*, *CD1A*, β-defensins, *PRG3* and *ULBP21* genes (Figure 4), which functions have been strongly linked to cattle environmental resilience, including immune response and ectoparasite resistance.

Functional annotation of genes

Enrichment analysis was performed to obtain a broad functional insight into the set of genes (Appendix 13D) observed in CNVRs described in each breed, as well as in shared CNVRs observed in between the three studied breeds. GO enrichment analysis revealed five biological processes, three molecular functions and four cellular component processes enriched ($p<0.01$, Table 1), and suggested that several of the CNVRs genes are mainly enhanced in functions related to the immune response. Some overrepresented terms were described in more than one breed, and those in shared CNVRs have been previously identified when analyzing the breeds individually.

Table 1. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis enriched (p<0.01) based on copy number variation regions (CNVRs) identified within each breed (Caracu Caldeano, Crioulo Lageano, and Pantaneiro) and based on shared CNVRs observed in between the three studied breeds.

Category[1]	**Term**	***n* genes**	**p-value**	**Genes**
Caracu Caldeano				
MF	GO:0047961~glycine N-acyltransferase activity	3	5.24E-06	*GLYAT, GAT, GLYATL2*
BP	GO:0042742~defense response to bacterium	4	8.38E-05	*DEFB7, EBD, DEFB13, DEFB4A*
BP	GO:0006955~immune response	4	1.47E-03	*PRG3, BOLA-DQA5, BOLA-DQB*
CC	GO:0005576~extracellular region	5	3.59E-03	*DEFB7, EBD, PRG3, DEFB13, DEFB4A*

MF	GO:0046703~natural killer cell lectin-like receptor binding	2	4.57E-03	*ULBP21, RAET1G*
Crioulo Lageano				
BP	GO:0002504~antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	2	7.79E-03	*BOLA-DQA5, BOLA-DQB*
CC	GO:0042613~MHC class II protein complex	2	9.79E-03	*BOLA-DQA5, BOLA-DQB*
Pantaneiro				
MF	GO:0005044~scavenger receptor activity	3	2.04E-04	*WC1, CD163L1, WC1.3*
BP	GO:0042742~defense response to bacterium	3	1.48E-03	*DEFB7, DEFB13, DEFB10*
CC	GO:0005576~extracellular region	4	7.76E-03	*DEFB7, CD163L1, DEFB13, DEFB10*
Shared CNVRs				
BP	GO:0002504~antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	2	5.20E-03	*BOLA-DQA5, BOLA-DQB*
CC	GO:0042613~MHC class II protein complex	2	6.13E-03	*BOLA-DQA5, BOLA-DQB*

[1]MF: Molecular function; BP: Biological process; CC: Cellular component.

CNVRs and overlapping QTLs in cattle

CNVRs were disclosed in genomic regions containing QTLs in cattle formerly implicated in body conformation (*n=2*) and dairy-related traits (*n=10*) (Appendix 14D). It is noteworthy to underscore that most of the QTLs described herein were found within the shared CNVRs in between the studied breeds. The CAR and CRL cattle did not display any further QTL besides those described in the shared CNVRs. Further, the PAN cattle displayed QTLs related to milk protein percentage and fatty acid content on BTA3 and BTA29, respectively, in addition to those identified within the shared CNVRs. It should be noted that the majority of the QTLs harbored duplication events and just one on BTA17:68058001-68079500 bp (Non-return rate QTL) (FRISCHKNECHT et al., 2017) was found encompassing a deletion event.

DISCUSSION

CNVs and CNVRs discovery

The widespread availability of array-based methods has led to much interest in the discovery and mapping of CNVs and their association with phenotypes (YAU; HOLMES, 2008). Previous studies assessing CNVs in several cattle breeds have been mainly based on array comparative genomic hybridization (aCGH) (FADISTA et al., 2010; LIU et al., 2010, 2019) and SNP arrays (BAE et al., 2010; CICCONARDI et al., 2013; HOU et al., 2012; JIANG et al., 2013; YANG et al., 2017; ZHANG et al., 2015). Although they promoted the progress of CNV studies, much has been discussed about the limitations of such methodologies associated with the power to detect CNVs (LAI et al., 2005; PINTO et al., 2011; WINCHESTER; YAU; RAGOUSSIS, 2009). SNP arrays were not specifically designed for CNVs detection since they do not well cover the whole genome, restraining their application and leading to biased results (HOU et al., 2011; JIANG et al., 2013). Studies have reported that coverage bias and platform resolution resulted in differences regarding the number and sizes between CNVs when using next-generation sequencing (NGS) and array-based methods (BEN SASSI et al., 2016; DA SILVA et al., 2016; JIANG et al., 2013; ZHAN et al., 2011). In this regard, CNV studies based on NGS data have been shown to overcome the sensitivity limits of array-based methods and to detect more precisely CNVs' boundaries (ALKAN; COE; EICHLER, 2011). Hence, differences in CNV calls from different platforms make the comparison among studies not straightforward and emphasize the importance of a careful assessment when contrasting studies.

Current studies on local and endangered cattle breeds using whole-genome resequencing data are very minimal when compared to specialized breeds (i.e., dairy and beef) (BEN SASSI et al., 2016; BICKHART et al., 2012; BOUSSAHA et al., 2015; GAO et al., 2017; STOTHARD et al., 2011; ZHAN et al., 2011). Accordingly, we investigated structural variations in three Brazilian locally adapted cattle breeds using a read depth approach based on whole-genome re-sequencing data. Our results revealed that CNVs are non-uniformly scattered across the genomes and represent a small proportion of the reference assembly used for mapping (~0.63 to 1.44%), as also reported for other cattle populations (BICKHART et al., 2012; STOTHARD et al., 2011;

ZHAN et al., 2011; ZHANG et al., 2015). The number of autosomal CNVs identified in each breed is consistent with previous reports based on NGS data (STOTHARD et al., 2011; ZHAN et al., 2011), and higher than those described by Bickhart et al. (2012) and Ben Sassi et al. (2016). Further, deletions events were approximately 1.75 to 2.07-fold more recurrent than did duplications, concurring with former NGS studies for taurine cattle breeds: ~1.72-fold (GAO et al., 2017) and 1.15-fold (BOUSSAHA et al., 2015). The increased number of deletions described herein might be associated with the mechanism by which CNVs are formed within the genome. Studies have shown that non-homologous end-joining (NHEJ) formation mechanism is the major mean responsible for deletion and translocations (SHAW; LUPSKI, 2005; TOFFOLATTI et al., 2002). NHEJ is a repair mechanism frequently initiated in response to double-strand breaks after DNA processing (VAN GENT; VAN DER BURG, 2007), and it can occasionally error-prone, leading to loss or small insertion of nucleotides at the lesion site (LABHART, 1999).

The sizes of the identified CNVs mostly ranged from 10 to 30 kb for all breeds, with a few outliers having a size higher than 500 kb. Such results are consistent with those based on SNP array (BAE et al., 2010; LEMOS et al., 2018; WU et al., 2015; ZHANG et al., 2015) on diverse cattle breeds, however, it differed from NGS data (DA SILVA et al., 2016) in which CNVs were most frequent between 100 to 200 kb. Nevertheless, it is worth to underscore that the CNVRs size range distribution concurred with those described in the literature for both SNP array-based and NGS data (BEN SASSI et al., 2016; DA SILVA et al., 2016; GAO et al., 2017; LEMOS et al., 2018).

Variant and functional annotation of genes

Genome-wide characterization of CNVs and the comprehensive assessment of CNVRs are a powerful strategy to ascertain potential key genes and biological mechanisms encompassing traits of interest in several livestock species. In this regard, CNVRs identified herein were better assessed to predict the impact of variants on protein sequence and determine their likely biological effects. Further, the gene content

within those regions were inspected in detail to disentangle their roles in shaping particular characteristics and phenotypes of the studied populations.

When further investigating the gene content harboring variants with a high consequence on protein sequence, the majority of them were described to be closely linked to adaptation and immune response functions. Among them, the *BOLA-DQA5* and *BOLA-DQB* genes were found located within the major histocompatibility complex (MHC) region. In cattle, the MHC region is known as the bovine leukocyte antigen (BoLA), which is encoded on BTA23 (FRIES; EGGEN; WOMACK, 1993). BoLA plays a crucial role in determining immune responsiveness, and genetic variations in such region has been greatly associated with disease susceptibility and resistance (reviewed by Takeshima & Aida (2006)). Additionally, several cattle studies have described CNVs adjacent to the BoLA region (HOU et al., 2011; LIU et al., 2010; PORTO-NETO et al., 2013; PRINSEN et al., 2017; ZHOU et al., 2016).

An enrichment of β-defensins genes (*DEFB10*, *DEFB13*, *DEFB4A*, *DEFB7*, and *EBD*) have also been identified harboring high impact variants within the CNVRs. β-defensins are antimicrobial peptides (AMPs) acting against many Gram-positive and negative bacteria, fungi, enveloped viruses, and other unicellular parasites (BROGDEN, 2005; LEHRER; LICHTENSTEIN; GANZ, 1993; NICOLAS; MOR, 1995). AMPs are among the most evolutionarily ancient molecules of the immune system and are present in a variety of vertebrates, insects, and plants (SELSTED; OUELLETTE, 2005). Besides their antimicrobial activity, β-defensins have chemoattractant activity for immature dendritic and T cells (YANG et al., 1999), playing a critical role in the immediate reaction to a broad spectrum of pathogens by inducing primary immunological responsiveness (BANCHEREAU et al., 2000; SAKAGUCHI et al., 2008). Further, bovine β-defensins located within the bovine cluster D are mainly expressed in the mammary gland, and therefore, contribute to local host defense and impart resistance against intramammary infections (GURAO; KASHYAP; SINGH, 2017).

Besides those genes previously described with immune system-related functions, three annotated genes encompassing variants with high impact on protein sequence are also worth to be highlighted given their role in cattle adaptation. The first one is the *ULBP2* gene, and it is hypothesized that cattle ULBP gene family evolved

under adaptive diversifying selection in response to selective pressure exerted by a viral pathogen (LARSON et al., 2006). The remaining two genes (*CD1A* and *PRG3*) have been associated with tick resistance. The *CD1A* gene has been described to be highly expressed at the tick attachment site from Holstein-Friesian animals (PIPER et al., 2008), and a study on Angus cattle (HOU et al., 2012) revealed that parasite resistance animals with high estimated breeding values (EBV) for eggs per gram displayed such gene within regulatory networks linked to gastrointestinal nematodes. The second gene linked to tick resistance is the *PRG3*. It forms a protective barrier by stimulating the histamine biosynthetic process and activating basophils, which are important effectors of tick rejection and a major component of the acquired resistance of the host (FALCONE; PRITCHARD; GIBBS, 2001; WIKEL, 1996). Such mechanism leads to an unfriendly environment for tick attachment and feeding (KONGSUWAN et al., 2008).

All of the previously discussed genes have been described within the significant GO terms, strongly supporting their enriched functions associated with immunity and cattle resilience to harsh environments. It should be noted that only one over-represented term (GO:0047961~glycine N-acyltransferase activity) has not been directly associated somehow with immune-related functions. Several other CNVs cattle studies displayed an enrichment of genes linked to immune response and environmental interaction, including sensory response and chemical stimuli (BICKHART et al., 2012; LIU et al., 2010; STOTHARD et al., 2011; UPADHYAY et al., 2017; WANG et al., 2015; YANG et al., 2017). Immune-related genes seem to be evolved under positive selection (SACKTON et al., 2007), reflecting a coevolutionary process between infectious pathogenic exposure and the host's defense system to acquire a broad range of antimicrobial defense (LUENSER; LUDWIG, 2005; MCTAGGART et al., 2012). Therefore, it has been hypothesized that the increased dosage of such genes may offer survivability and adaptive benefits (LIU et al., 2010; NGUYEN et al., 2008), suggesting that adaptation to diverse pathogenic environments most likely have exerted important selective forces in the cattle genome.

It is not surprising that an abundance of genes and over-represented terms were found described to be involved in processes closely associated with immune functions and parasite resistance. The Brazilian locally adapted cattle breeds studied herein

exhibit distinguishing levels of phenotypic variability and enhanced fitness to local conditions due to a long process of natural selection in extremely variable and harsh environments (MARIANTE; CAVALCANTE, 2000). Such breeds have undergone strong environmental pressures for more than 450 years without any significant selective pressure imposed by man, facing adverse tropical climate conditions (heat, dryness, and humidity), limited food availability, disease's susceptibility, and parasite infestations (MARIANTE; CAVALCANTE, 2000). Hence, these limitations led them to acquire very particular traits over time to thrive in such distinct ecosystems (MARIANTE et al., 1999) and may have left footprints of selection within their genome.

CNVRs and overlapping QTLs in cattle

Most of the CNVRs overlapped with previously reported regions harboring QTLs that mostly affect dairy-related traits, and two reasons might have led to this result. First, when examining in detail the QTLs associations by trait classes in the CattleQTL database (HU; PARK; REECY, 2016), the greatest number of reported QTLs (~36%) has been associated with milk-related traits (*n=50,208*), followed by reproductive (*n=44,369*), and productive (*n=22,519*) traits. The second reason relies on the fact that the CAR breed has been selected for milk production traits in the southeastern region of Brazil since 1893 (QUEIROZ et al., 2005). Further, the remaining two breeds despite not being considered high-specialized cattle breeds are classified as dual-purpose and might have undergone mild selection for dairy-related traits (LARA et al., 2002; OLIVEIRA-BROCHADO et al., 2018).

FINAL CONSIDERATIONS

By using whole-genome re-sequencing data, we reported for the first time a genome-wide characterization of CNVs in three Brazilian locally adapted taurine cattle breeds. Our results provide substantial information about the potential use of CNVs to identify putative regions that have been functionally relevant and have played a substantial role in shaping the genome of such cattle breeds based on the environmental conditions in which they have been raised. Enrichment analysis, variant

annotation, and QTL identification retrieved from the CNVRs revealed a large proportion of genes associated with immune system functioning, parasite resistance, and some production-related traits. These results provide evidence of positive selection for traits linked to cattle resilience to challenging environments.

The cattle populations studied herein represent an important model for understanding the role of environmental stressors and the effect of different selective forces acting on the genome diversity of the Brazilian locally adapted taurine cattle breeds. These findings are of particular interest since it is important to assure that animal genetic resources will match with the production environments in which they are raised. The identification of genomic regions harboring structural variations plays an important role in the introgression of locally adapted breeds in crossbreeding schemes. Hence, production systems may benefit from the introduction of crossbred animals, taking advantage of animals better adapted to local conditions displaying key adaptative traits for survival in challenging environments together with production traits from high-specialized cattle breeds.

REFERENCES

ABYZOV, A. et al. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. **Genome Research**, v. 21, n. 6, p. 974–984, 2011.

ALKAN, C. et al. Personalized copy number and segmental duplication maps using next-generation sequencing. **Nature Genetics**, v. 41, n. 10, p. 1061–1067, 2009.

ALKAN, C.; COE, B. P.; EICHLER, E. E. Genome structural variation discovery and genotyping. **Nature Reviews Genetics**, v. 12, n. 5, p. 363–376, maio 2011.

BAE, J. S. et al. Identification of copy number variations and common deletion polymorphisms in cattle. **BMC Genomics**, v. 11:232, 2010.

BANCHEREAU, J. et al. Immunobiology of Dendritic Cells - Annual Review of Immunology, 18(1):767. **Annual review of immunology**, v. 18, p. 767–811, 2000.

BECKMANN, J. S.; ESTIVILL, X.; ANTONARAKIS, S. E. Copy number variants and genetic traits: Closer to the resolution of phenotypic to genotypic variability. **Nature Reviews Genetics**, v. 8, n. 8, p. 639–646, 2007.

BEN SASSI, N. et al. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. **Journal of Dairy Science**, v. 99, n. 8, p. 6371–6380, 2016.

BICKHART, D. M. et al. Copy number variation of individual cattle genomes using next-generation sequencing. v. 22, p. 778–790, 2012.

BOUSSAHA, M. et al. Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. **PLoS ONE**, v. 10, n. 8, p. e0135931, 2015.

BROGDEN, K. A. Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? **Nature Reviews Microbiology**, v. 3, p. 238–250, 2005.

CICCONARDI, F. et al. Massive screening of copy number population-scale variation in Bos taurus genome. **BMC Genomics**, v. 14:124, 2013.

CLOP, A.; VIDAL, O.; AMILLS, M. Copy number variation in the genomes of domestic animals. **Animal Genetics**, v. 43, n. 5, p. 503–517, 2012.

CONRAD, B.; ANTONARAKIS, S. E. Gene Duplication: A drive for phenotypic diversity and cause of human disease. **Annual Review of Genomics and Human Genetics**, v. 8, p. 17–35, 2007.

CONRAD, D. F. et al. Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. **Nature Genetics**, v. 42, n. 5, p. 385–391, 2010.

DA SILVA, J. M. et al. Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. **BMC Genomics**, v. 17:454, 2016.

DUPUIS, M. C. et al. Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. **Animal Genetics**, v. 44, n. 2, p. 206–208, 2013.

ESTIVILL, X.; ARMENGOL, L. Copy number variants and common disorders: Filling the gaps and exploring complexity in genome-wide association studies. **PLoS Genetics**, v. 3, n. 10, p. 1787–1799, 2007.

FADISTA, J. et al. Copy number variation in the bovine genome. **BMC Genomics**, v. 11:284, 2010.

FALCONE, F. H.; PRITCHARD, D. I.; GIBBS, B. F. Do basophils play a role in immunity against parasites? **Trends in Parasitology**, v. 17, n. 3, p. 126–129, 2001.

FEUK, L.; CARSON, A. R.; SCHERER, S. W. **Structural variation in the human genomeNature Reviews Genetics**, 2006.

FONTANESI, L. et al. An initial comparative map of copy number variations in the goat (Capra hircus) genome. **BMC Genomics**, v. 11:639, 2010.

FONTANESI, L. et al. A first comparative map of copy number variations in the sheep genome. **Genomics**, v. 97, p. 158–165, 2011.

FRIES, R.; EGGEN, A.; WOMACK, J. E. The bovine genome map. **Mammalian Genome**, v. 4, p. 405–428, 1993.

FRISCHKNECHT, M. et al. Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. **BMC Genomics**, v. 18, n. 910, 2017.

GAO, Y. et al. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. **BMC Genomics**, v. 18:265, 2017.

GERLANDO, R. DI et al. Genome-wide detection of copy-number variations in local cattle breeds. **Animal Production Science**, v. 59, n. 5, p. 815–822, 2019.

GURAO, A.; KASHYAP, S. K.; SINGH, R. **β-defensins: An innate defense for bovine mastitisVeterinary World**, 2017.

HAIDER, S. et al. BioMart Central Portal--unified access to biological data. **Nucleic acids research**, v. 37, p. W23–W27, 2009.

HENRICHSEN, C. N.; CHAIGNAT, E.; REYMOND, A. Copy number variants, diseases and gene expression. **Human Molecular Genetics**, v. 18, n. R1, p. R1–R8, 2009.

HOU, Y. et al. Genomic characteristics of cattle copy number variations. **BMC Genomics**, v. 12:127, 2011.

HOU, Y. et al. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. **Functional and Integrative Genomics**, v. 12, n. 1, p. 81–92, 2012.

HU, Z. L.; PARK, C. A.; REECY, J. M. Developmental progress and current status of the Animal QTLdb. **Nucleic Acids Research**, v. 44, p. D827–D833, 2016.

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, n. 1, p. 1–13, 2009a.

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, n. 1, p. 44–57, 2009b.

JIANG, L. et al. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. **BMC Genomics**, v. 14:131, 2013.

KIJAS, J. W. et al. Analysis of copy number variants in the cattle genome. **Gene**, v. 482, p. 73–77, ago. 2011.

KONGSUWAN, K. et al. **Identification of genes involved with tick infestation in Bos taurus and Bos indicus**. (M.-H. Pinard et al., Eds.)Animal Genomics for Animal Health. **Anais**...Basel: 2008

LABHART, P. **Nonhomologous DNA end joining in cell-free systemsEuropean Journal of Biochemistry**, 1999.

LAI, W. R. et al. Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. **Bioinformatics**, v. 21, n. 19, p. 3763–3770, 2005.

LARA, M. et al. Genetic polymorphisms at the k-casein locus in Pantaneiro cattle. **Archivos de zootecnia**, v. 51, n. 193, p. 11, 2002.

LARSON, J. H. et al. Genomic organization and evolution of the ULBP genes in cattle. **BMC Genomics**, v. 7:227, 2006.

LAWRENCE, M. et al. Software for Computing and Annotating Genomic Ranges. **PLoS Computational Biology**, v. 9, n. 8, p. e1003118, 2013.

LEHRER, R. I.; LICHTENSTEIN, A. K.; GANZ, T. Defensins: Antimicrobial and Cytotoxic Peptides of Mammalian Cells. **Annual Review of Immunology**, v. 11, p. 105–128, 1993.

LEMOS, M. V. A. et al. Copy number variation regions in Nellore cattle: Evidences of environment adaptation. **Livestock Science**, v. 207, 2018.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009.

LI, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **ArXiv**, v. 1303, 2013.

LIU, G. E. et al. Analysis of copy number variations among diverse cattle breeds. **Genome Research**, v. 20, n. 5, p. 693–703, 2010.

LIU, M. et al. Array CGH-based detection of CNV regions and their potential

association with reproduction and other economic traits in Holsteins. **BMC Genomics**, v. 20:181, mar. 2019.

LOW, W. Y. et al. Haplotype-Resolved Cattle Genomes Provide Insights Into Structural Variation and Adaptation. **bioRxiv**, p. 720797, 2019.

LUENSER, K.; LUDWIG, A. Variability and evolution of bovine β-defensin genes. **Genes and Immunity**, v. 6, p. 115–122, 2005.

MARIANTE, A.; CAVALCANTE, N. **Animais do descobrimento: raças domésticas da história do Brasil.** [s.l.] Empresa Brasileira de Pesquisa Agropcuária, Centro de Pesquisa Agropecuária do Pantanal, 2000.

MARIANTE, A. DA S. et al. Advances in the Brazilian animal genetic resources conservation programme. **Animal Genetic Resources Information**, v. 25, p. 107–121, 1999.

MARTINS, V. M. V. et al. **Raça Crioula Lageana. O esteio do ontem, o labor do hoje e a oportunidade do amanhã**. [s.l.] Editora Associação Brasileira dos Criadores da Raça Crioula Lageana (ABCCL), 2009.

MAZZA, M. et al. **Etnobiologia e conservação do bovino Pantaneiro**. [s.l.] Empresa Brasileira de Pesquisa Agropcuária, Centro de Pesquisa Agropecuária do Pantanal, 1994.

MCCARROLL, S. A.; ALTSHULER, D. M. Copy-number variation and association studies of human disease. **Nature Genetics**, v. 39, n. 7, p. S37–S42, 2007.

MCLAREN, W. et al. The Ensembl Variant Effect Predictor. **Genome Biology**, v. 17, p. 122, 2016.

MCTAGGART, S. J. et al. Immune genes undergo more adaptive evolution than non-immune system genes in Daphnia pulex. **BMC Evolutionary Biology**, v. 12:63, 2012.

MOLNÁR, J. et al. Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. **BMC Genomics**, v. 15:761, 2014.

NGUYEN, D. Q. et al. Reduced purifying selection prevails over positive selection in human copy number variant evolution. **Genome Research**, v. 18, p. 1711–1723, 2008.

NICOLAS, P.; MOR, A. Peptides as weapons against microorganisms in the chemical defense system of vertebrates. **Annual Review of Microbiology**, v. 49, n.

1, p. 277–304, 1995.

OLIVEIRA-BROCHADO, N. C. et al. Dairy potential of Pantaneira breed upper region of pantanal of Mato Grosso Sul. **Arquivo Brasileiro de Medicina Veterinaria e Zootecnia**, v. 70, n. 2, p. 644–648, 2018.

PERIPOLLI, E. et al. Genome-wide detection of signatures of selection in indicine and Brazilian locally adapted taurine cattle breeds using whole-genome re-sequencing data. **BMC Genomics**, v. 21:624, 2020.

PINTO, D. et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. **Nature Biotechnology**, v. 29, n. 6, p. 512–520, 2011.

PIPER, E. K. et al. Gene expression in the skin of Bos taurus and Bos indicus cattle infested with the cattle tick, Rhipicephalus (Boophilus) microplus. **Veterinary Immunology and Immunopathology**, v. 126, p. 110–119, 2008.

PORTO-NETO, L. R. et al. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. **BMC Genomics**, v. 14:876, 2013.

PRIMO, A. El ganado bovino ibérico en las Américas: 500 años después. **Archivos de zootecnia**, v. 41, n. 154, p. 421–432, 1992.

PRINSEN, R. T. M. M. et al. A genome wide association study between CNVs and quantitative traits in Brown Swiss cattle. **Livestock Science**, v. 202, p. 7–12, 2017.

QUEIROZ, S. A. et al. Selection indices for a dual purpose breed Caracu. **Revista Brasileira de Zootecnia**, v. 34, n. 3, p. 827–837, 2005.

QUINLAN, A. R.; HALL, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. **Bioinformatics**, v. 26, n. 6, p. 841–842, 2010.

RAMAYO-CALDAS, Y. et al. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. **BMC Genomics**, v. 11:593, 2010.

REDON, R. et al. Global variation in copy number in the human genome. **Nature**, v. 444, p. 444–454, 2006.

SACKTON, T. B. et al. Dynamic evolution of the innate immune system in Drosophila. **Nature Genetics**, v. 39, p. 1461–1468, 2007.

SAKAGUCHI, S. et al. Regulatory T Cells and Immune Tolerance. **Cell**, v. 133, p. 775–787, 2008.

SCHERER, S. W. et al. Challenges and standards in integrating surveys of structural variation. **Nature Genetics**, v. 39, n. 7, p. S7–S15, 2007.

SELSTED, M. E.; OUELLETTE, A. J. Mammalian defensins in the antimicrobial immune response. **Nature Immunology**, v. 6, n. 6, p. 551–557, 2005.

SHAW, C. J.; LUPSKI, J. R. Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. **Human Genetics**, v. 116, n. 1–2, p. 1–7, 22 jan. 2005.

STOTHARD, P. et al. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. **BMC Genomics**, v. 12:559, 2011.

SUDMANT, P. H. et al. Diversity of human copy number variation and multicopy genes. **Science**, v. 330, n. 6004, p. 641–646, 2010.

TAKESHIMA, S. N.; AIDA, Y. Structure, function and disease susceptibility of the bovine major histocompatibility complex. **Animal Science Journal**, v. 77, p. 138–150, 2006.

TIAN, M. et al. Copy number variants in locally raised Chinese chicken genomes determined using array comparative genomic hybridization. **BMC Genomics**, v. 14:262, 2013.

TOFFOLATTI, L. et al. Investigating the Mechanism of Chromosomal Deletion: Characterization of 39 Deletion Breakpoints in Introns 47 and 48 of the Human Dystrophin Gene. **Genomics**, v. 80, n. 5, p. 523–530, 2002.

UPADHYAY, M. et al. Distribution and functionality of copy number variation across European cattle populations. **Frontiers in Genetics**, v. 8:108, 2017.

VAN GENT, D. C.; VAN DER BURG, M. **Non-homologous end-joining, a sticky affairOncogene**, 2007.

WANG, M. D. et al. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. **BMC Genomics**, v. 16:894, 2015.

WANG, M. D. et al. Tropically adapted cattle of Africa: Perspectives on potential role of copy number variations. **Animal Genetics**, v. 47, p. 154–164, 2016.

WIKEL, S. K. Host immunity to ticks. **Annual review of entomology. Vol. 41**, v. 41, p. 1–22, 1996.

WINCHESTER, L.; YAU, C.; RAGOUSSIS, J. Comparing CNV detection methods for SNP arrays. **Briefings in Functional Genomics and Proteomics**, v. 8,

n. 5, p. 353–366, 2009.

WU, Y. et al. A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. **Animal Genetics**, v. 46, n. 3, p. 289–298, 2015.

YANG, D. et al. β-Defensins: Linking innate and adaptive immunity through dendritic and T cell CCR6. **Science**, v. 286, n. 5439, p. 525–528, 1999.

YANG, L. et al. Genome-wide analysis reveals differential selection involved with copy number variation in diverse Chinese Cattle. **Scientific Reports**, v. 7, 2017.

YAU, C.; HOLMES, C. C. **CNV discovery using SNP genotyping arraysCytogenetic and Genome Research**, 2008.

ZHAN, B. et al. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. **BMC Genomics**, v. 12:557, 2011.

ZHANG, F. et al. Copy Number Variation in Human Health, Disease, and Evolution. **Annual Review of Genomics and Human Genetics**, v. 10, p. 451–481, 2009.

ZHANG, Q. et al. Identification of copy number variations in Qinchuan cattle using BovineHD Genotyping Beadchip array. **Molecular Genetics and Genomics**, v. 290, p. 319–327, 2015.

ZHOU, W. et al. A genome-wide detection of copy number variation using SNP genotyping arrays in Beijing-You chickens. **Genetica**, v. 142, n. 5, p. 441–450, 2014.

ZHOU, Y. et al. Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations. **Scientific Reports**, v. 6:29219, 2016.

FIGURES

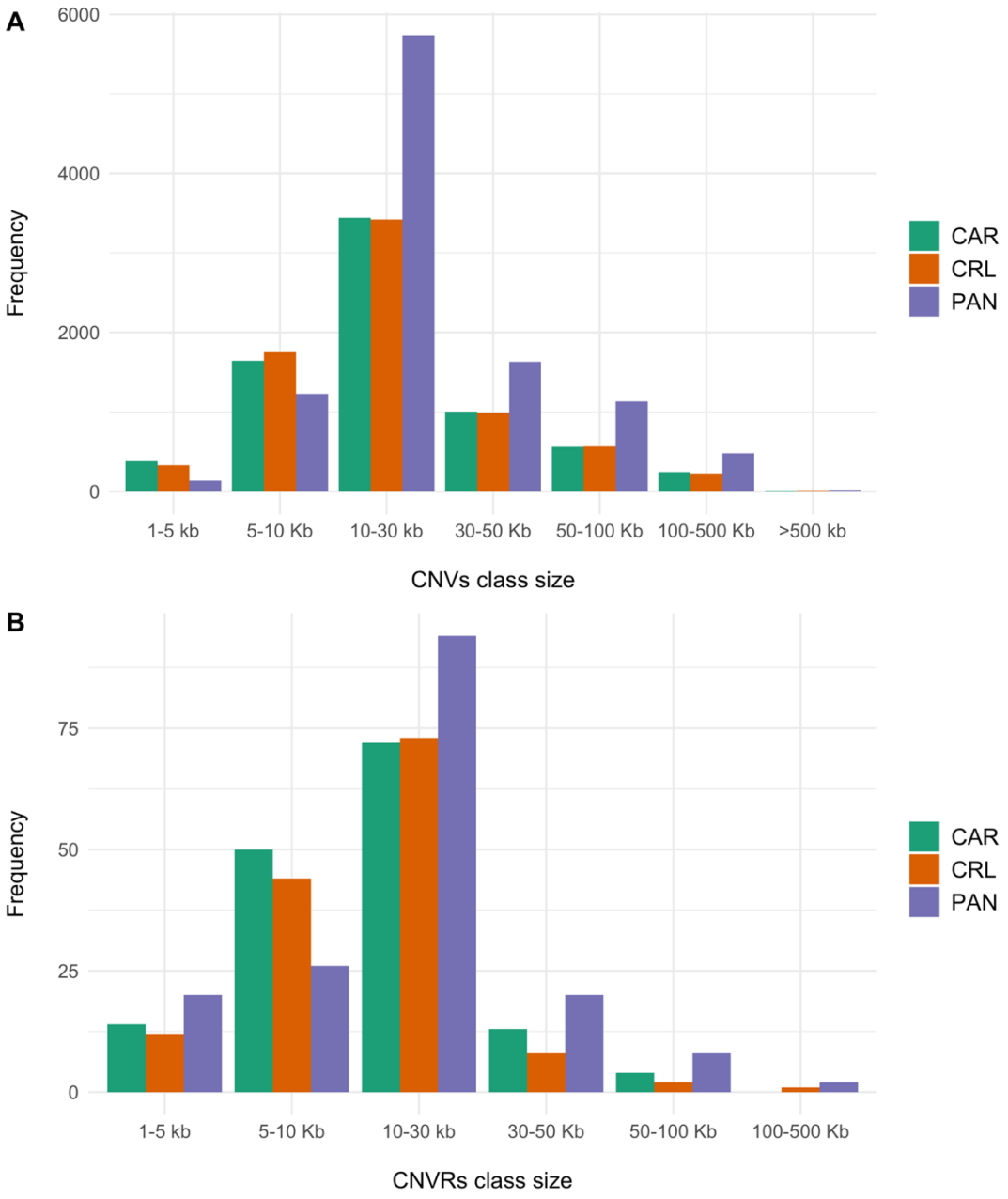


Figure 1. **A.** Copy number variations (CNVs) length class size range distribution for Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle breeds **B.** Copy number variations regions (CNVRs) length class size range distribution for Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle breeds.

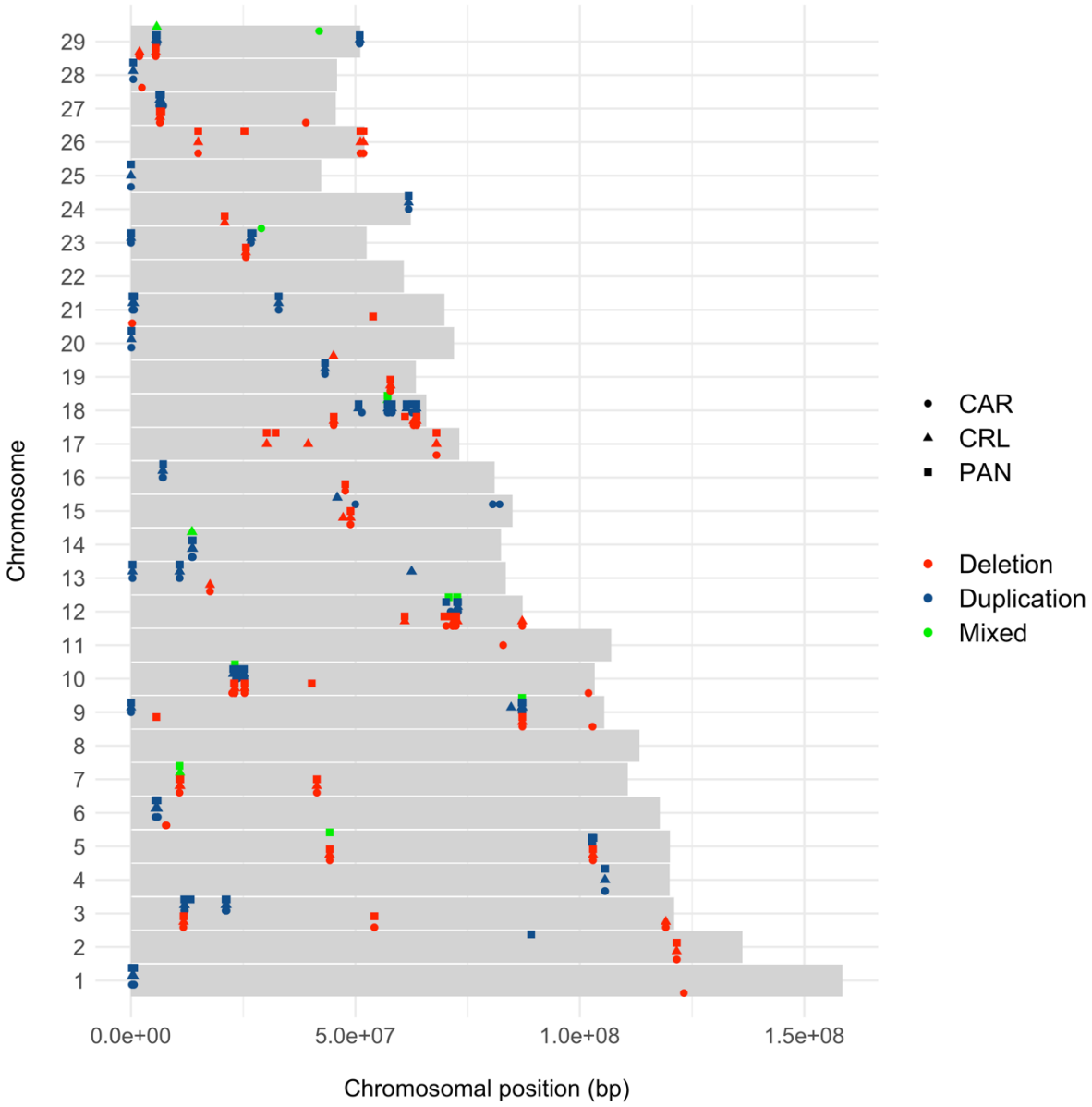


Figure 2. Copy number variation regions (CNVRs) scattering in the Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle genomes according to autosomal length (ARS-UCD1.2). Dots depicting the breeds: circle (CAR), triangle (CRL), and square (PAN). Dots depicting the CNVRs: deletion (red), duplication (blue), and mixed (green) events.

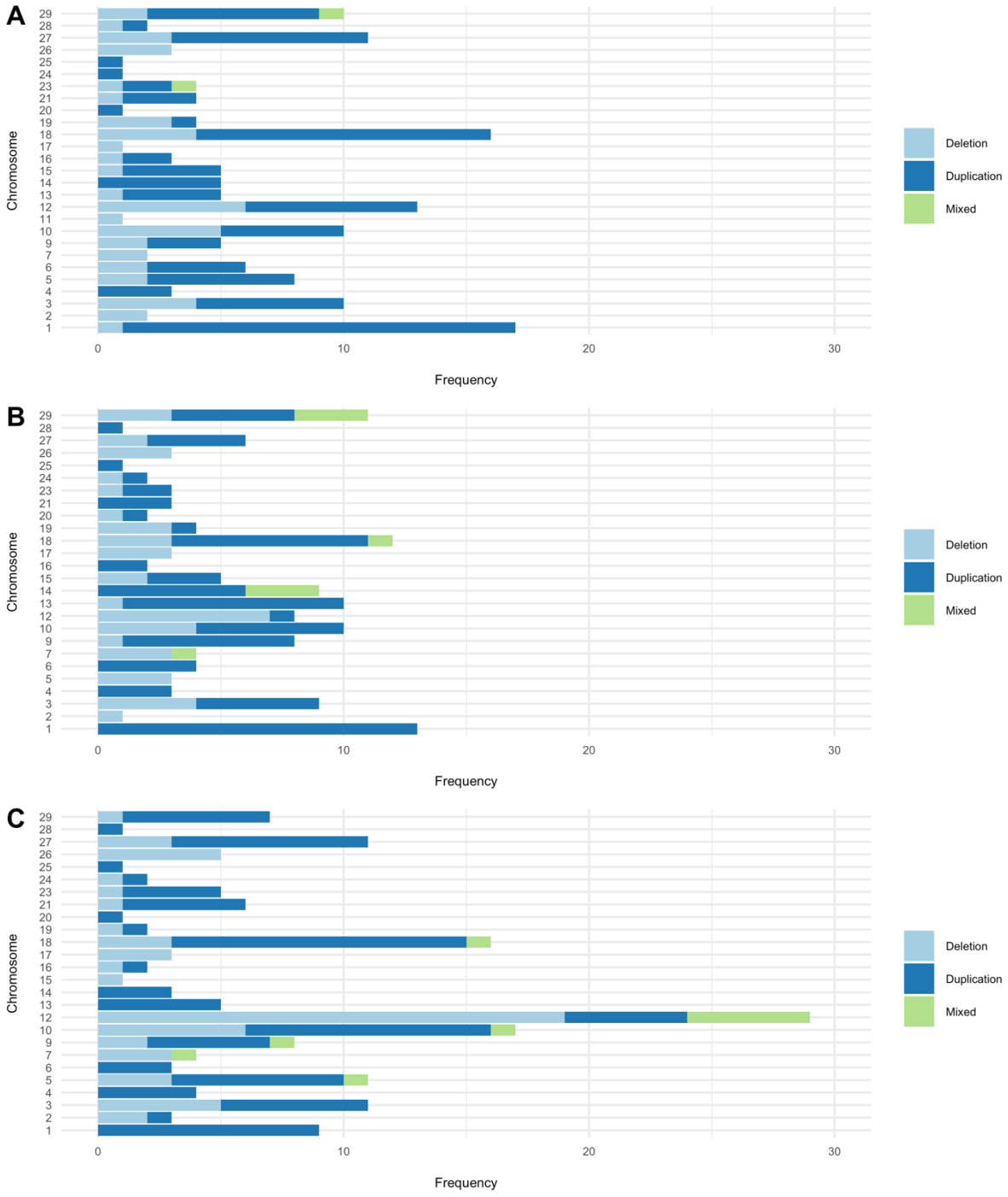


Figure 3. Frequency distribution of copy number variation regions (CNVRs) according to CNVRs event (deletion, duplication, and mixed). **A.** Caracu Caldeano cattle **B.** Crioulo Lageano cattle **C.** Pantaneiro cattle.

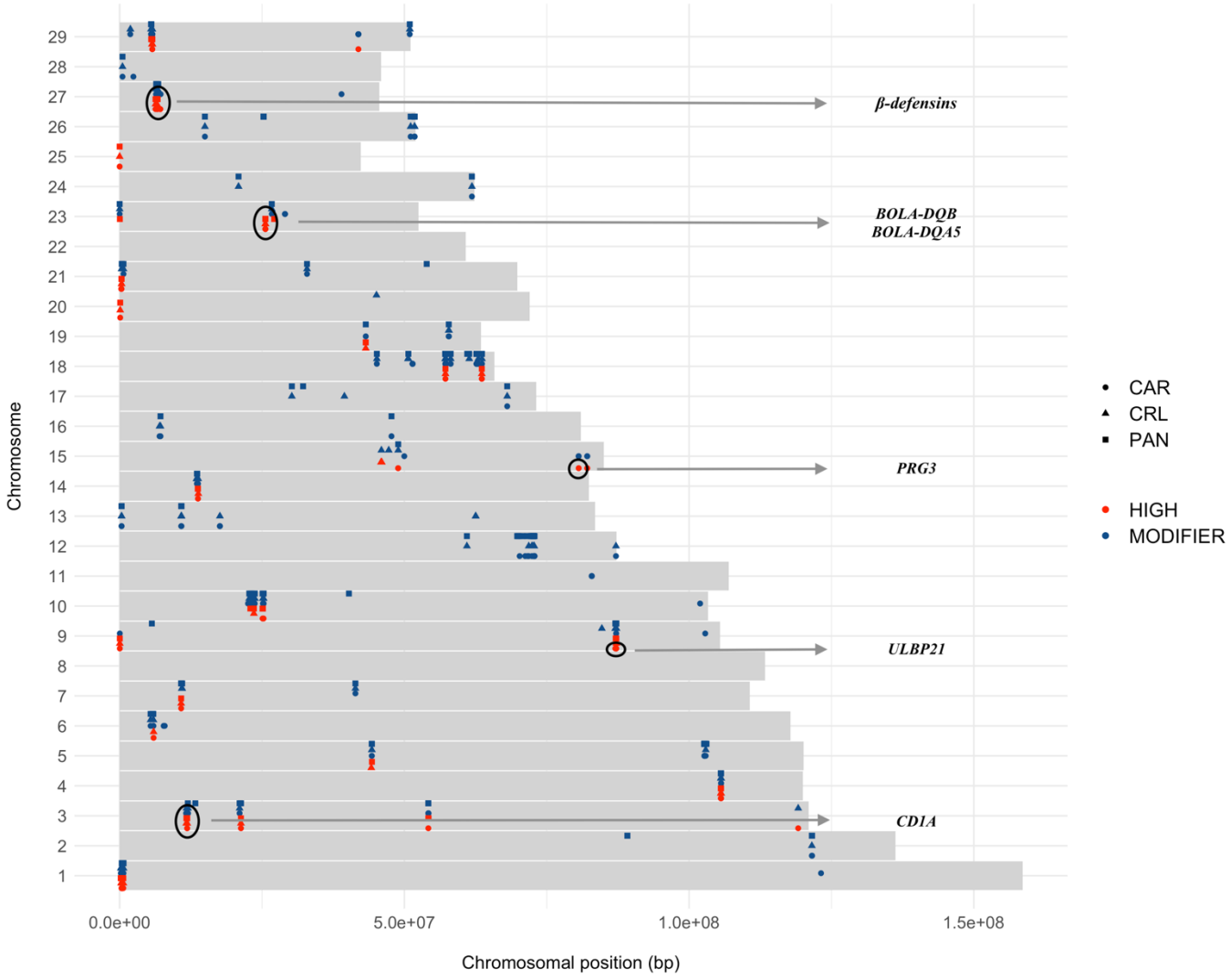


Figure 4. Variants scattering in the Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle genomes according to autosomal length (ARS-UCD1.2). Dots depicting the breeds: circle (CAR), triangle (CRL), and square (PAN). Dots depicting the putative variant impact: high (red) and modifier (blue).

CAPÍTULO 6 – FINAL CONSIDERATIONS

In general, one can conclude that inbreeding estimates based on information derived from pedigree data are not the most appropriate method for capturing former inbreeding events, and it is not uncommon for pedigree records to contain errors (i.e., errors of annotation and absence or loss of data). An incomplete and shallow pedigree cannot account for inbreeding caused by distant ancestors since it does not extend back several generations. Therefore, the use of genomic information can contribute to a more correct estimation of genetic similarity between individuals. Inbreeding coefficients estimated from molecular information, especially those derived from ROH (F_{ROH}), should be used as an accurate estimator of ancient individual inbreeding levels since they can disentangle with a greater accuracy both past and recent relatedness (i.e., age of inbreeding) based on the length of such ROH segment.

The use of molecular information has introduced significant advances into the analyses of inbreeding coefficients, however, a recurrent limitation in studies involving ROH relies on the sensitivity of shallow density panels in detecting such segments. This shortcoming may be responsible for increasing the likelihood of biased and false-positive results in ROH-based estimates of autozygosity. In chapter three, we addressed this concern and deliberated that some results might not reflect the true level of autozygosity since some small ROH remain undetected when using shallow SNP arrays due to the lack of power in accurately determining them. Therefore, results should be interpreted carefully since the SNP array used to generate the data for ROH analysis can strongly influence ROH identification in several livestock species.

Genetic diversity is necessary for populations to evolve in response to environmental changes, and to make sure that the breeding program remains viable in the future, it is essential to monitor and maintain such genetic diversity by controlling heterozygosity levels. It is noteworthy to highpoint that our results have shown low genomic autozygotic levels in breeds in which the development occurred from a narrow genetic base with a limited number of progenitors to disseminate the breed, as well as in those considered endangered. These results might be mainly attributed to: (i) the expansion of the breeding programs and progeny testing; (ii) slight selection pressure

and herd management focused on maintaining genetic diversity, especially for the locally adapted cattle breeds; (iii) formerly closed herds start using semen of proven sires, increasing the overall genetic exchange; (iv) introduction of new genes through genic combinations to explore the complementarity amongst the breeds, especially for the composite breeds; and (v) formation of new herds associated with the diversification in the use of sires.

The genetic characterization of tropically and locally adapted cattle breeds is essential to preserve their genomic diversity, and it is a preliminary step for the development of conservation programs to boost the sustainable use of these genetic resources. Putative signals of selection based on several approaches (ROH, selection signatures, and CNVRs) were detected for regions containing genes largely involved in defense response to bacteria, immune and inflammatory response, homeostasis, and cattle resilience to harsh environments. Our findings improve the knowledge about the genome biology of such cattle breeds and provide candidate genes and genomic regions encompassing relevant traits as well as useful information for future conservation, association, or selection approaches.

APPENDIX A

Appendix 1A. Autozygosity islands across the Nellore cattle genome

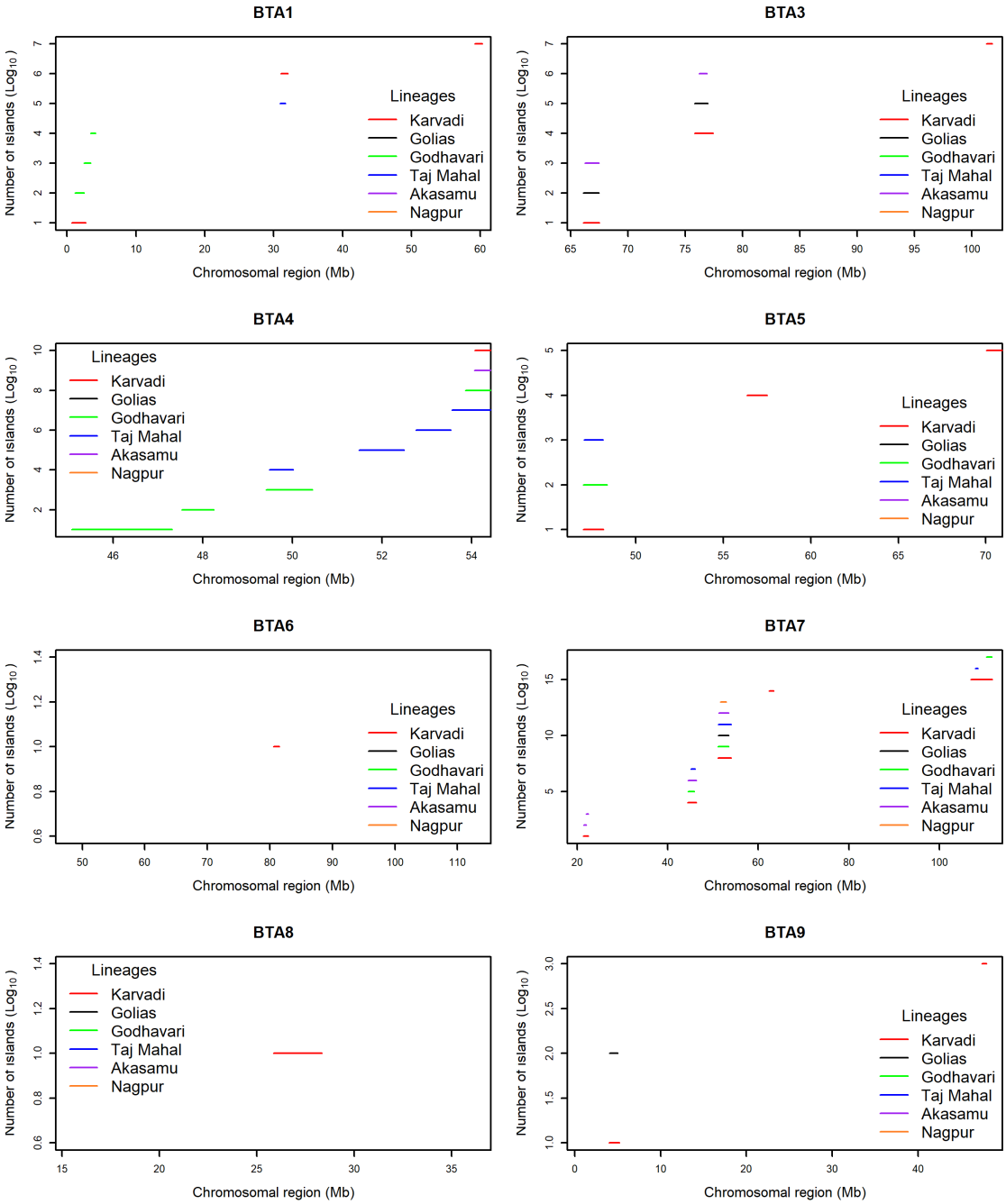
BTA[1]	Start (bp)	End (bp)	Length (bp)
1	850,000	2,464,000	1,614,000
1	30,950,000	32,090,000	1,140,000
1	39,450,000	40,200,000	750,000
1	40,340,000	40,870,000	530,000
1	59,210,000	60,300,000	1,090,000
3	66,110,000	67,510,000	1,400,000
3	75,810,000	77,440,000	1,630,000
3	101,400,000	101,800,000	400,000
4	54,280,000	55,800,000	1,520,000
5	47,000,000	48,130,000	1,130,000
5	56,360,000	57,500,000	1,140,000
5	70,060,000	71,090,000	1,030,000
6	80,560,000	81,390,000	830,000
7	21,390,000	22,480,000	1,090,000
7	39,680,000	40,180,000	500,000
7	43,510,000	44,180,000	670,000
7	44,450,000	46,300,000	1,850,000
7	51,140,000	54,040,000	2,900,000
7	62,370,000	63,450,000	1,080,000
7	84,410,000	85,140,000	730,000
7	107,000,000	111,700,000	4,700,000
8	25,940,000	28,340,000	2,400,000
9	3,971,000	5,182,000	1,211,000
10	45,560,000	46,090,000	530,000
10	52,840,000	55,180,000	2,340,000
12	25,540,000	27,080,000	1,540,000
12	27,350,000	30,040,000	2,690,000
12	34,990,000	37,070,000	2,080,000
12	37,080,000	39,800,000	2,720,000
12	56,790,000	57,850,000	1,060,000
13	50,200,000	50,960,000	760,000
13	62,650,000	66,140,000	3,490,000
14	23,240,000	25,800,000	2,560,000
15	80,230,000	81,420,000	1,190,000
16	66,730,000	70,880,000	4,150,000
17	35,300,000	36,480,000	1,180,000

Appendix 1A. Continuation

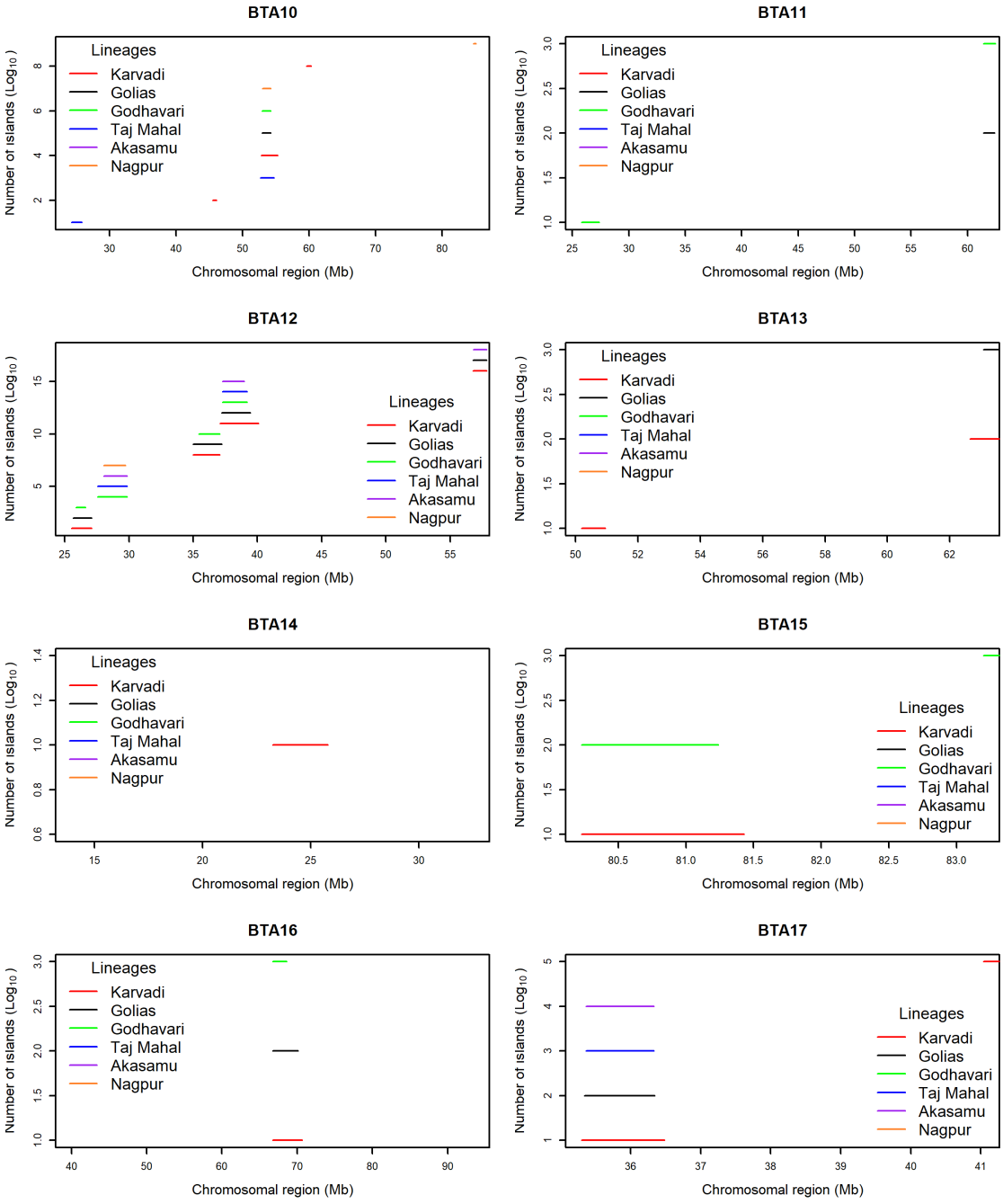
17	38,360,000	39,210,000	850,000
17	41,020,000	42,160,000	1,140,000
19	27,190,000	28,140,000	950,000
19	33,780,000	35,400,000	1,620,000
19	42,680,000	44,010,000	1,330,000
20	13,670,000	14,460,000	790,000
20	30,480,000	31,620,000	1,140,000
20	36,560,000	37,640,000	1,080,000
20	56,890,000	58,010,000	1,120,000
20	70,600,000	71,890,000	1,290,000
21	8,725	1,916,000	1,907,275
21	64,790,000	65,890,000	1,100,000
22	15,320,000	16,220,000	900,000
22	16,600,000	17,850,000	1,250,000
22	34,220,000	34,840,000	620,000
22	43,540,000	43,880,000	340,000
23	14,910	1,253,000	1,238,090
23	36,560,000	37,640,000	1,080,000
24	42,930,000	44,750,000	1,820,000
24	61,530,000	61,880,000	350,000
26	1,984,000	3,214,000	1,230,000
26	15,670,000	16,640,000	970,000
26	21,270,000	23,010,000	1,740,000
26	41,730,000	42,340,000	610,000
27	4,845,000	6,405,000	1,560,000
29	38,730,000	39,820,000	1,090,000

[1] BTA: *Bos taurus* autosome.

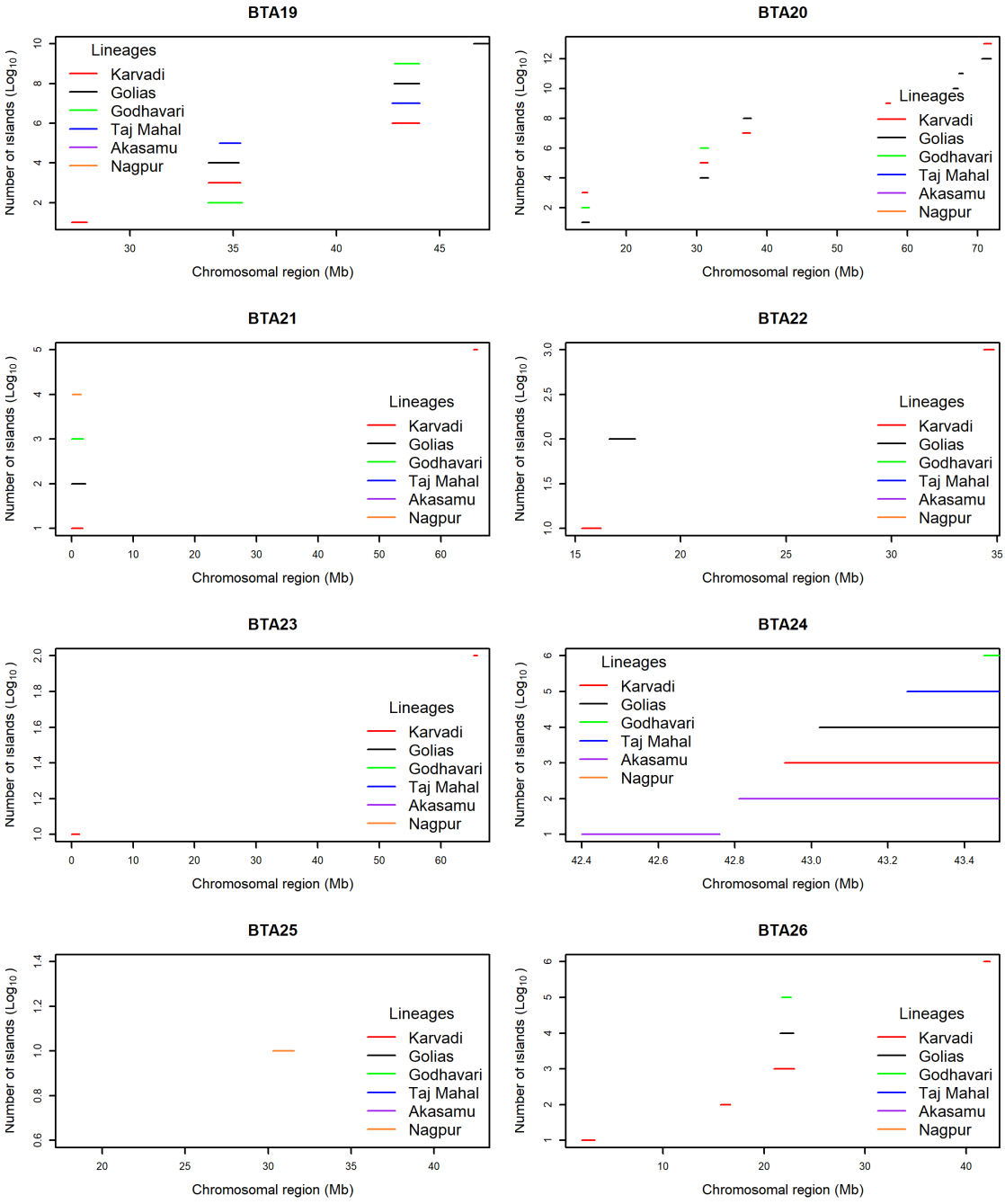
Appendix 2A. Autozygosity islands within the Nellore lineages by chromosome: Karvadi (red), Golias (Black), Godhavari (Green), Taj Mahal (blue), Akasamu (purple), and Nagpur (yellow).



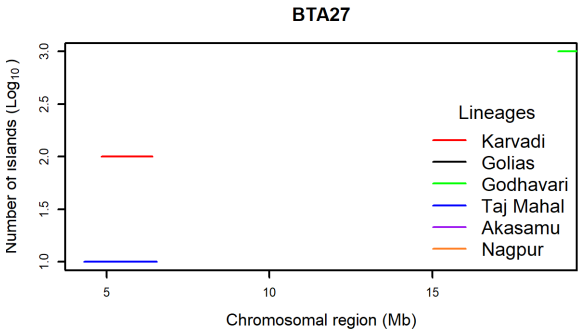
Appendix 2A. Continuation

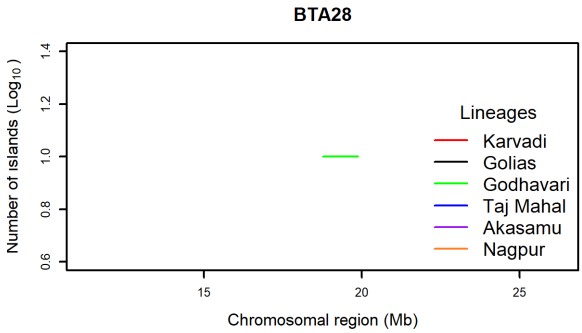


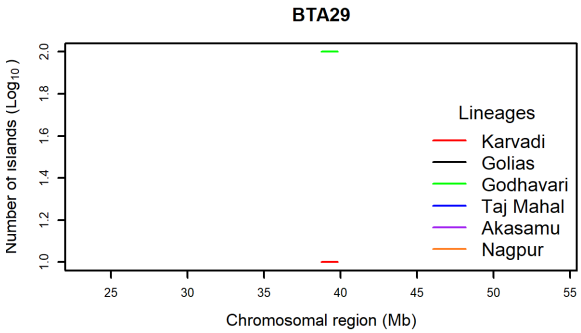
Appendix 2A. Continuation



Appendix 2A. Continuation







Appendix 3A. Overlapping autozygosity islands within the Nellore lineages.

BTA[1]	Start (bp)	End (bp)	Length (bp)	Lineages
1	1,185,000	2,466,000	1,281,001	Karvadi, Godhavari
1	2,472,000	2,725,000	253,001	Karvadi, Godhavari
1	31,050,000	31,680,000	630,001	Taj Mahal, Karvadi
3	66,110,000	66,239,999	130,000	Karvadi, Golias
3	66,240,000	67,450,000	1,210,001	Karvadi, Golias, Akasamu
3	75,830,000	76,209,999	380,000	Karvadi, Golias
3	76,210,000	76,880,000	670,001	Karvadi, Golias, Akasamu
3	76,880,001	76,980,000	100,000	Karvadi, Golias
4	49,490,000	50,020,000	530,001	Godhavari, Taj Mahal
4	53,860,000	54,059,999	200,000	Taj Mahal, Godhavari
4	54,060,000	54,069,999	10,000	Taj Mahal, Godhavari, Akasamu
4	54,070,000	55,690,000	1,620,001	Taj Mahal, Godhavari, Akasamu, Karvadi
4	55,690,001	55,800,000	110,000	Taj Mahal, Godhavari, Karvadi
4	55,800,001	55,820,000	20,000	Taj Mahal, Godhavari
5	47,000,000	47,049,999	50,000	Godhavari, Karvadi
5	47,050,000	48,110,000	1,060,001	Godhavari, Karvadi, Taj Mahal
5	48,110,001	48,130,000	20,000	Godhavari, Karvadi
7	21,410,000	21,990,000	580,001	Karvadi, Akasamu
7	22,020,000	22,430,000	410,001	Karvadi, Akasamu
7	44,470,000	44,489,999	20,000	Karvadi, Godhavari
7	44,490,000	45,119,999	630,000	Karvadi, Godhavari, Akasamu
7	45,120,000	45,830,000	710,001	Karvadi, Godhavari, Akasamu, Taj Mahal

Appendix 3A. Continuation

7	45,830,001	46,050,000	220,000	Karvadi, Akasamu, Taj Mahal
7	46,050,001	46,300,000	250,000	Karvadi, Akasamu
7	51,140,000	51,209,999	70,000	Karvadi, Godhavari
7	51,210,000	51,229,999	20,000	Karvadi, Godhavari, Golias
7	51,230,000	51,249,999	20,000	Karvadi, Godhavari, Golias, Taj Mahal
7	51,250,000	51,609,999	360,000	Karvadi, Godhavari, Golias, Taj Mahal, Akasamu
7	51,610,000	52,930,000	1,320,001	Karvadi, Godhavari, Golias, Taj Mahal, Akasamu, Nagpur
7	52,930,001	53,440,000	510,000	Karvadi, Godhavari, Golias, Taj Mahal, Akasamu
7	53,440,001	53,490,000	50,000	Karvadi, Godhavari, Taj Mahal, Akasamu,
7	53,490,001	54,040,000	550,000	Karvadi, Taj Mahal
7	108,000,000	108,500,000	500,001	Karvadi, Taj Mahal
7	110,400,000	111,600,000	1,200,001	Karvadi, Godhavari
9	4,033,000	5,005,000	972,001	Karvadi, Golias
10	52,840,000	52,919,999	80,000	Taj Mahal, Karvadi
10	52,920,000	52,969,999	50,000	Taj Mahal, Karvadi, Golias
10	52,970,000	53,009,999	40,000	Taj Mahal, Karvadi, Golias, Godhavari
10	53,010,000	54,210,000	1,200,001	Taj Mahal, Karvadi, Golias, Godhavari, Nagpur
10	54,210,001	54,230,000	20,000	Taj Mahal, Karvadi, Golias, Godhavari
10	54,230,001	54,700,000	470,000	Taj Mahal, Karvadi
11	61,420,000	62,390,000	970,001	Godhavari, Golias
12	25,670,000	25,889,999	220,000	Karvadi, Golias
12	25,890,000	26,610,000	720,001	Karvadi, Godhavari
12	26,610,001	27,080,000	470,000	Karvadi, Golias

Appendix 3A. Continuation

12	27,580,000	28,039,999	460,000	Godhavari, Taj Mahal
12	28,040,000	29,740,000	1,700,001	Godhavari, Nagpur
12	29,740,001	29,860,000	120,000	Godhavari, Akasamu
12	34,990,000	35,449,999	460,000	Golias, Karvadi
12	35,450,000	37,060,000	1,610,001	Golias, Karvadi, Godhavari
12	37,060,001	37,070,000	10,000	Golias, Karvadi
12	37,080,000	37,220,000	140,001	Golias, Karvadi
12	37,230,000	37,299,999	70,000	Karvadi, Golias
12	37,300,000	38,960,000	1,660,001	Karvadi, Akasamu
12	38,960,001	39,200,000	240,000	Karvadi, Taj Mahal
12	39,200,001	39,430,000	230,000	Karvadi, Golias
12	56,790,000	56,819,999	30,000	Karvadi, Golias
12	56,820,000	57,830,000	1,010,001	Karvadi, Golias, Akasamu
12	57,830,001	57,840,000	10,000	Karvadi, Golias
13	63,080,000	64,510,000	1,430,001	Karvadi, Golias
15	80,230,000	81,240,000	1,010,001	Karvadi, Godhavari
16	66,730,000	68,600,000	1,870,001	Karvadi, Godhavari
16	68,600,001	70,090,000	1,490,000	Karvadi, Golias
17	35,340,000	35,359,999	20,000	Karvadi, Golias
17	35,360,000	36,330,000	970,001	Karvadi, Golias, Taj Mahal, Akasamu
17	36,330,001	36,340,000	10,000	Karvadi, Golias
19	33,800,000	34,329,999	530,000	Godhavari, Karvadi, Golias
19	34,330,000	35,280,000	950,001	Godhavari, Karvadi, Golias, Taj Mahal

Appendix 3A. Continuation

19	35,280,001	35,350,000	70,000	Godhavari, Karvadi, Taj Mahal
19	42,680,000	42,779,999	100,000	Taj Mahal, Karvadi
19	42,780,000	42,799,999	20,000	Taj Mahal, Karvadi, Golias
19	42,800,000	44,000,000	1,200,001	Taj Mahal, Karvadi, Golias, Godhavari
19	44,000,001	44,010,000	10,000	Taj Mahal, Karvadi, Golias
20	13,670,000	13,679,999	10,000	Golias, Godhavari
20	13,680,000	14,450,000	770,001	Golias, Godhavari, Karvadi
20	14,450,001	14700,000	250,000	Golias, Godhavari
20	30,510,000	31,600,000	1,090,001	Golias, Godhavari, Karvadi
20	31,600,001	31,630,000	30,000	Golias, Godhavari
20	36,660,000	37,620,000	960,001	Karvadi, Golias
20	70,860,000	71,890,000	1,030,001	Golias, Karvadi
21	8,725	112,599	103,875	Golias, Godhavari, Karvadi
21	112,600	1,483,000	1,370,401	Golias, Godhavari, Karvadi, Nagpur
21	1,483,001	1,790,000	307,000	Golias, Godhavari, Karvadi
21	1,790,001	1,916,000	126,000	Golias, Godhavari
24	42,930,000	43,019,999	90,000	Akasamu, Karvadi
24	43,020,000	43,249,999	230,000	Akasamu, Karvadi, Golias
24	43,250,000	43,449,999	200,000	Akasamu, Karvadi, Golias, Taj Mahal
24	43,450,000	43,930,000	480,001	Akasamu, Karvadi, Golias, Taj Mahal, Godhavari
24	43,930,001	44,030,000	100,000	Akasamu, Karvadi, Golias, Taj Mahal
24	44,030,001	44,080,000	50,000	Akasamu, Karvadi
26	21,590,000	21,749,999	160,000	Karvadi, Golias

Appendix 3A. Continuation

26	21,750,000	22,660,000	910,001	Karvadi, Godhavari
26	22,660,001	22,930,000	270,000	Karvadi, Golias
27	4,845,000	6,405,000	1,560,001	Taj Mahal, Karvadi
29	38,730,000	39,810,000	1,080,001	Karvadi, Godhavari

[1] BTA: *Bos taurus* autosome.

Appendix 4A. Non-overlapping autozygosity islands within the Nellore lineages

BTA[1]	Start (bp)	End (bp)	Length (bp)	Lineage	Genes
1	59,210,000	60,300,000	1,090,001	Karvadi	*DRD3, TIGIT, ZBTB20,*
3	101,300,000	101,800,000	500,001	Karvadi	*TESK2, TOE1, MUTYH, HPDL, ZSWIM5, UROD, HECTD3, EIF2B3, TCH2*
4	45,080,000	47,310,000	2,230,001	Godhavari	*RELN, ORC5, LHFPL3, KMT2E, SRPK2, PUS7, RINT1*
4	47,530,000	48,250,000	720,001	Godhavari	*NAMPT*
4	49,410,000	49,489,999	80,000	Godhavari	*LAMB4*
4	51,490,000	52,490,000	1,000,001	Taj Mahal	*ST7, CAPZA2, MET, CAV1, CAV12, TES*
4	52,760,000	53,530,000	770,001	Taj Mahal	*TFEC*
4	55,820,001	55,940,000	120,000	Godhavari	*LSMEM1, IFRD1*
5	56,360,000	57,500,000	1,140,001	Karvadi	*R3HDM2, STAC3, NDUFA4L2, SHMT2, NXPH4, LRP1, STAT6, NAB2, NEMP1, MYO1A, TAC3, ZBTB39, GPR182, RDH16, SDR9C7,*
5	70,040,000	71,090,000	1,050,001	Karvadi	*TCP11L2, POLR3B, RFX4, RIC8B, TMEM263, MTERF2, CRY1, BTBD11*
6	80,560,000	81,500,000	940,001	Karvadi	-
7	62,370,000	63,440,000	1,070,001	Karvadi	*SH3TC2, ABLIM3, AFAP1L1, GRPEL2, PCYOX1L, IL17B, CSNK1A1, ARHGEF37, PPARGC1B, PDE6A, SLC26A2, HMGXB3, CSF1R*
8	25,850,000	28,340,000	2,490,001	Karvadi	*SH3GL2, CNTLN, BNC2*
9	47,470,000	47,980,000	510,001	Karvadi	-

Appendix 4A. Continuation

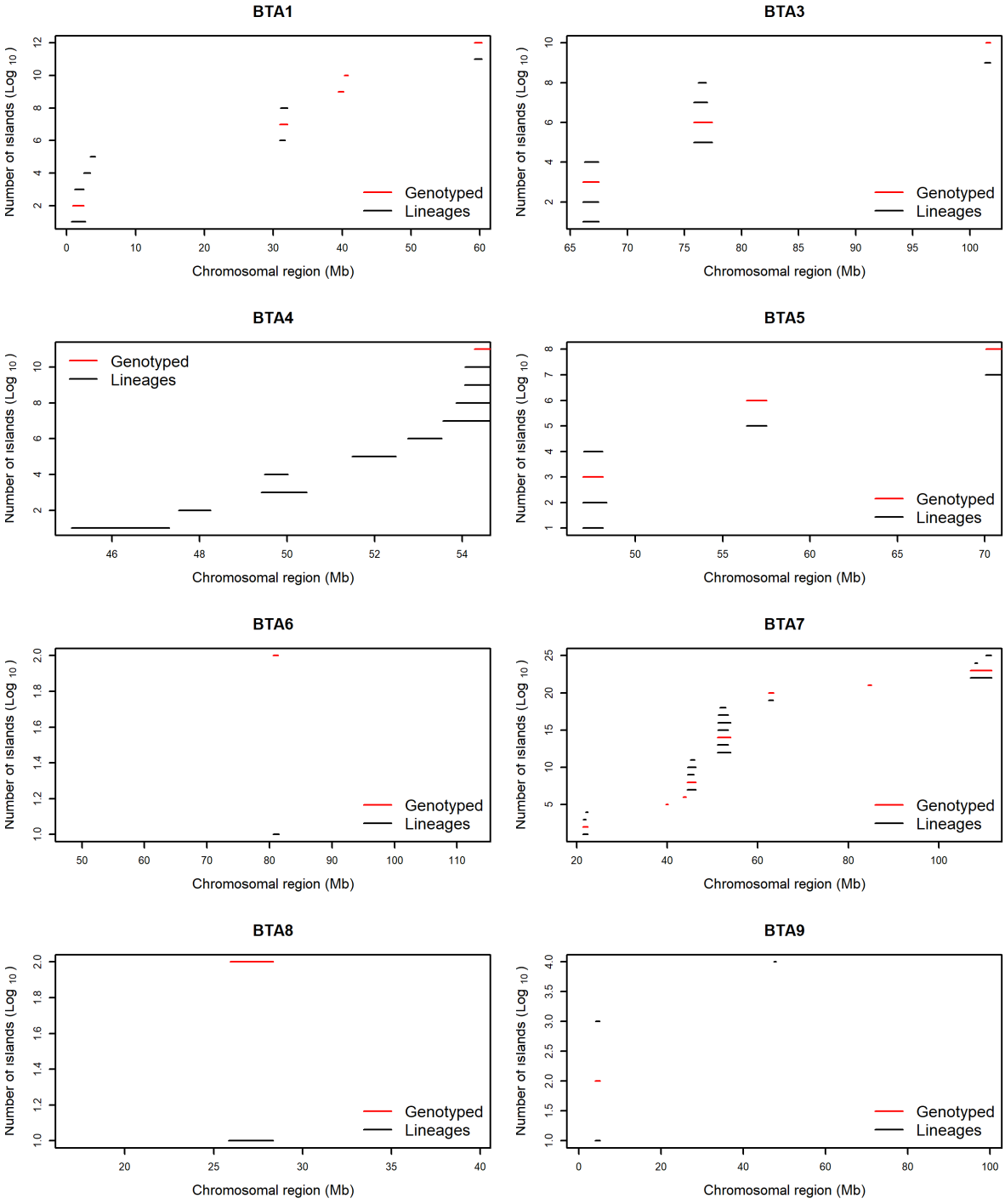
10	24,300,000	25,840,000	1,540,001	Taj Mahal	*TRAV17, TRAV178*
10	45,550,000	46,090,000	540,001	Karvadi	*ZNF609, TRIP4, PCLAF, CSNK1G1, PPIB, SNX22, SNX1, FAM96A, DAPK2*
10	59,670,000	60,340,000	670,001	Karvadi	*SPPL2A, TRPM7, USP50, USP8, GABPB1, HDC, SLC27A2*
10	84,740,000	85,100,000	360,001	Nagpur	*DCAF4, ZFYVE1, RBM25, PSEN1, PAPLN*
11	25,830,000	27,390,000	1,560,001	Godhavari	*THADA, PLEKHH2, DYNC2LI1, ABCG5, ABCG8, LRPPRC, PPM1B, SLC3A1, PREPL, CAMKMT, SIX3, SIX2*
13	50,200,000	50,960,000	760,001	Karvadi	-
14	23,240,000	25,800,000	2,560,001	Karvadi	*NPBWR1, OPRK1, ATP6V1H, RGS20, TCEA1, LYPLA1, MRPL15, POLR2K, SOX17, RP1, XKR4, TMEM68, TGS1, LYN, RPS20,*
15	83,200,000	84,040,000	840,001	Godhavari	*LPXN, CNTF, GLYAT, GAT, GLYATL2, FAM111B, DTX4, MPEG1, OR5A1*
16	70,090,001	70,670,000	580,000	Karvadi	*KCNK2, CENPF, PTPN14*
17	41,040,000	42,020,000	980,001	Karvadi	*C17H4orf45, FNIP2, PPID, ETFDH, C17H4orf46, RXFP1, TMEM144, FAM198B*
19	27,190,000	27,930,000	740,001	Karvadi	*PSMB6, GLTPD2, VMO1, TM4SF5, ZMYND15, CXCL16, MED11, ARRB2, PELP1, ALOX15, ALOX12E, ALOX12, RNASEK, C19H17orf49, BCL6B, SLC16A13*
19	46,630,000	47,700,000	1,070,001	Golias	*MAPT, KANSL1, CDC27, MYL4, ITGB3, EFCAB3, METTL2A, TLK2*
20	56,950,000	57,560,000	610,001	Karvadi	*MARCH11, FBXL7*

Appendix 4A. Continuation

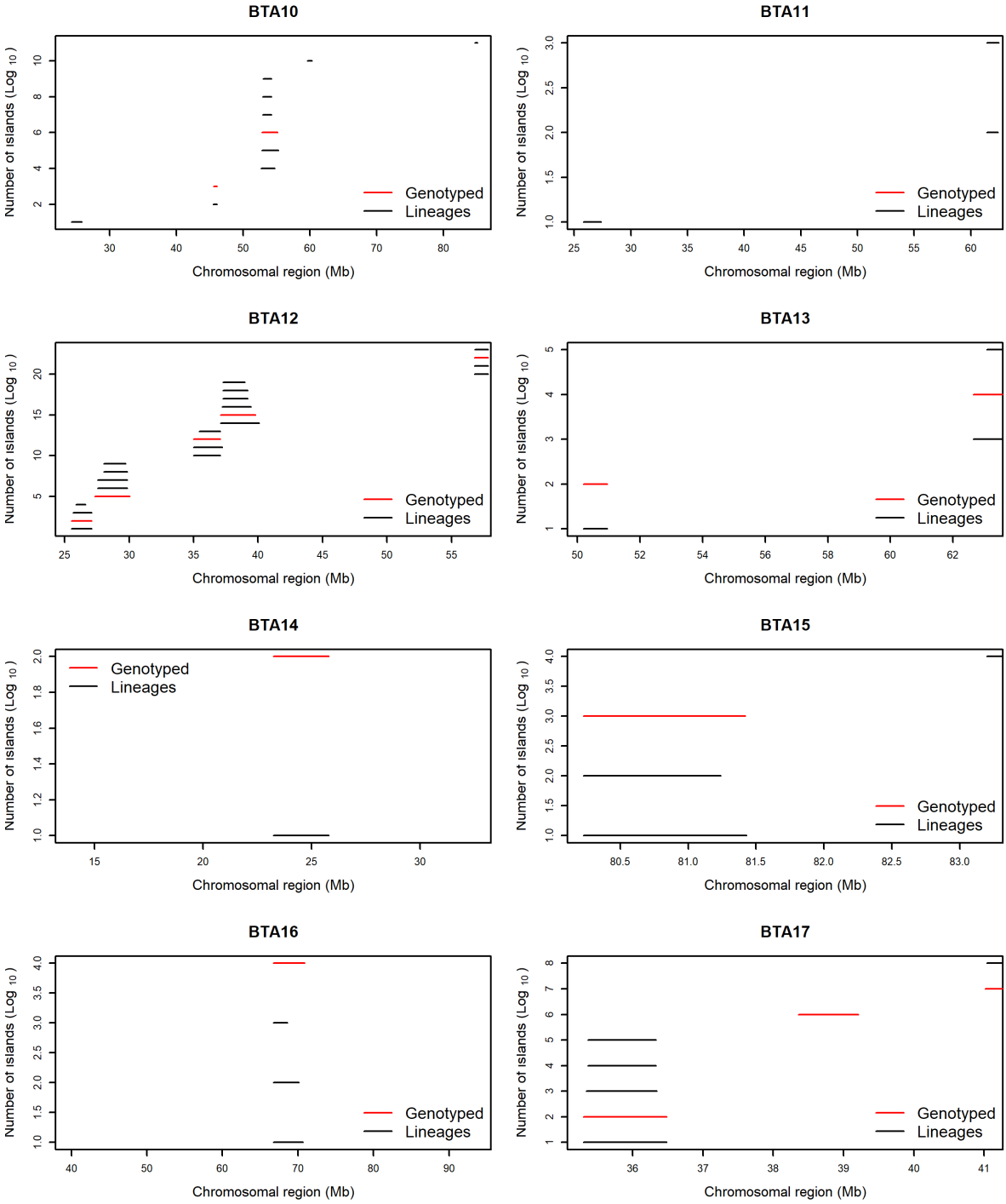
20	66,510,000	67,210,000	700,001	Golias	*PAPD7, SRD5A1, NSUN2, MED10*
20	67,330,000	67,880,000	550,001	Golias	*ICE1*
21	65,280,000	65,890,000	610,001	Karvadi	*BCL11B*
22	34,360,000	34,840,000	480,001	Karvadi	*KBTBD8*
23	43,440	1,253,000	1,209,561	Karvadi	*KHDRBS2*
23	65,280,000	65,890,000	610,001	Karvadi	-
24	42,400,000	42,760,000	360,001	Akasamu	*APCDD1, NAPG, PIEZO2*
25	30,270,000	31,560,000	1,290,001	Nagpur	-
26	1,961,000	3,250,000	1,289,001	Karvadi	*ZWINT*
26	15,670,000	16,650,000	980,001	Karvadi	*PLCE1, NOC3L, TBC1D12, HELLS, CYP2C18, CYP2C87, CYP2C19, PDLIM1*
26	41,730,000	42,340,000	610,001	Karvadi	*FGFR2, ATE1, NSMCE4A, TACC2*
27	18,860,000	19,960,000	1,100,001	Godhavari	*MTMR7, VPS37A, CNOT7, ZDHHC2, MICU3, FGF20*
28	18,760,000	19,880,000	1,120,001	Godhavari	*ADO, EGR2, NRBF2, JMJD1C, REEP3*

[1] BTA: *Bos taurus* autosome.

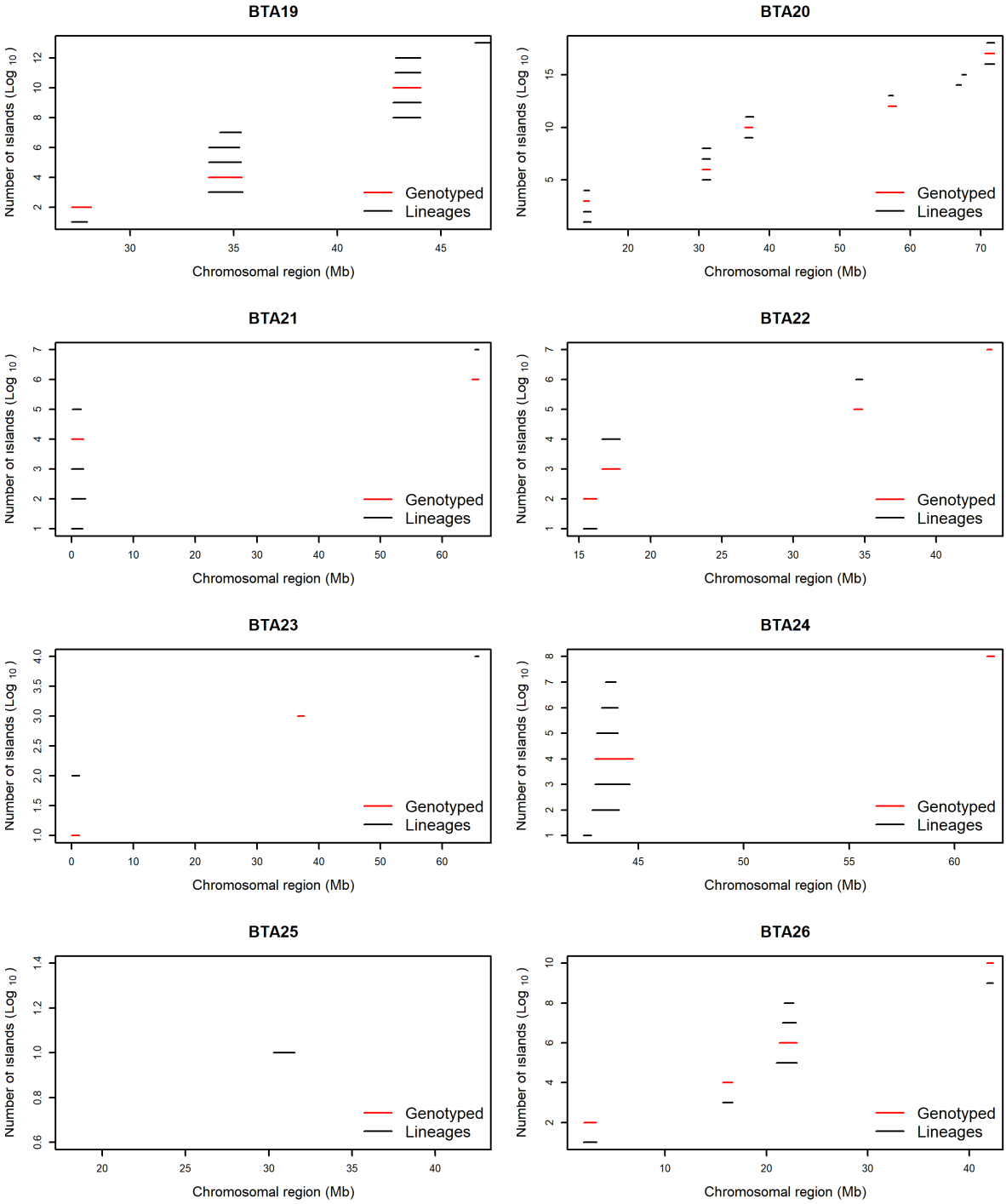
Appendix 5A. Autozygosity islands within the genotyped animals (red) and those with lineages records (black).



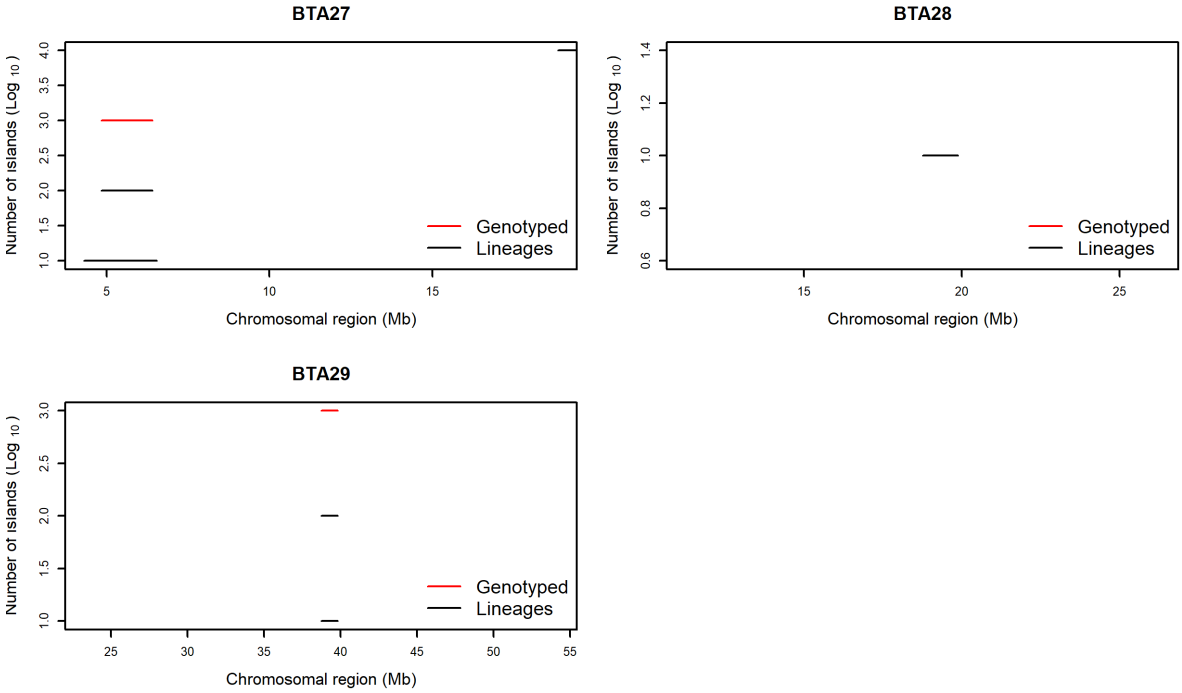
Appendix 5A. Continuation



Appendix 5A. Continuation



Appendix 5A. Continuation



Appendix 6A. Gene Ontology terms and KEGG pathways annotation analysis enriched (P<0.01) based on autozygosity islands set of genes identified for the genotyped animals (*n=9,386*)

Terms	Genes
GO Biological Process	
(GO:0042742) defense response to bacteria	*ELANE, ROMO1, TAP, DEFB4A, LEAP2, DEFB6, DEFB5, DEFB7, EBD, PENK, LAP, DEFB13, DEFB10, DEFB1*
(GO:0030163) protein catabolic process	*PAG17, PAG19, MGC157405, PAG16, MGC157408, PAG21, PAG20, PAG4, PAG1*
(GO:0070200) establishment of protein localization to telomere	*NABP2, WRAP53, BRCA2, TERT*
(GO:0040014) regulation of multicellular organism growth	*FGFR2, DRD3, GDF5, GAMT, AFG3L2, STAT3*
(GO:0045647) negative regulation of erythrocyte differentiation	*STAT5A, LDB1, STAT5B, HSPA9*
(GO:0030901) midbrain development	*FGFR2, KAT2A, RFX4, WLS, PITX3, UQCRQ*
GO Molecular Function	
(GO:0008289) lipid binding	*BPIFB1, BPIFB2, BPIFA3, BPIFB3, BPIFB4, BPIFA1, BPIFB5, BPIFB6, BPIFA2A, BPIFA2B, FER, BPIFA2C, STARD13*
(GO:0004190) aspartic-type endopeptidase activity	*PAG17, PAG19, MGC157405, PAG16, MGC157408, PAG21, PAG20, PAG4, PAG1*
GO Cellular Component	
(GO:0005776) autophagosome	*TBC1D12, MAP1LC3A, BECN1, NBR1, RAB24, USP33, TP53INP2, GABARAP*
(GO:0005634) nucleus	*RALY, RNMT, BTRC, STAT5A, STAT5B, DNASE1L3, TRMT1L, MIER2, RBMS2, ITCH, DND1, PITX3, CRY1, TGS1, SPATA24, PAN2, POLL, MAGEL2, LBX1, CSNK1G1, SUCLG2, CSNK1G2, PTBP1, PPARGC1B, DPCD, HSPB9, RFC3, NABP2, ARRB2, ZWINT, PYGO1, RAD18, MAPK7, FGFR2, STK11, SLF2, ZNF131, NOC3L, MUM1, AFAP1L1, IFI35, IRAK3, HECTD3, MNS1, TCTEX1D4, TCF3, HELLS, REEP6, SREBF1, PLAG1, DVL2, TRIP4, RFX4, PTPN2, BECN1, MICU2, SRA1, RFX7, ARID3A, BRCA2, DONSON, SMYD2, FXR2, SPRYD4,*

	BRCA1, CIDEC, ATE1, R3HDM2, NSMCE4A, PPIB, R3HDM4, NEDD4, MLX, PPID, CPNE1, PPRC1, BCL6B, CIRBP, MAB21L1, NCOR1, TCF12, RAI1, FAM96A, ELF3, PHF23, NFKB2, CBFA2T2, LATS2, AES, TUBB6, FAM83G, IRAK2, EGR1, LYN, ELP5, TP53, TLE2, MBD3, HMGA2, SENP3, ZNF341, ZNF692, FANCD2, TPPP, NAB2, TOP3A, RBM39, UBB, GADD45B, CUEDC2, IRX4, NACA, POLR2E, NDN, POLR2K, ADAD1, PAXBP1, IVNS1ABP, STAT6, RNF126, RAX2, SNRK, BCL11B, TCEA1, KDM3B, NIM1K, TLX1, CSNK1A1, SHMT2, PDS5B, VHL, CS, CDC25C, TRIM23, CENPK, STAT3, STAT2, GPS2, RNF112, RGS20, PSMG2, BNC2, PSPC1, ZBTB4, PDC, GAMT, APBB3, NFIC, OGG1, TJP3, SCAND1, ALKBH5, MPHOSPH8, TP53INP2, ZMYND15
(GO:0005815) microtubule organizing center	*CSNK1A1, SUCLG2, FLII, TCTEX1D4, AK5, RBM39, MAPRE1, PXK, LATS2, FNIP2*
(GO:0005730) nucleolus	*EIF6, ZNF554, MIDN, RNMT, NOC3L, NFS1, TIMM13, ZNF346, YBX2, CRYL1, WDR55, URB1, NPM3, TCEA1, RSL24D1, SDR9C7, TERT, RPS23, HSPA9, IK, VHL, TP53, ARID3A, THUMPD3, FGF22, TACC2, ACADVL, SENP3, PLK3, LRP1, NOLC1, TIMELESS, FANCD2, PPID, LLPH, PSPC1, ZZZ3, NFIC, ARL4D, MPHOSPH8, VPS2*
KEGG pathway	
(bta01100) Metabolic pathways	*PTGES3, IMPAD1, IMPA2, PTGS2, ALOX12E, CYP2C18, SYNJ1, SAT2, PIP5K1C, ACSS2, UQCRQ, PRIM1, NDUFS7, GSS, NDUFS6, CRYL1, PIGL, UQCR11, PIGB, SUCLG2, PIGU, ATP6V1H, ACADVL, MAN2A1, NME5, PLCE1, G6PC, ALOX15, MTMR14, NNT, AOC2, UROD, AOC3, ALOX12, COASY, NAGLU, POLR2E, AHCY, HSD17B1, POLR2K, NDUFB8, NFS1, HMGCS1, CYP2C87, ALDH3A2, POLR2A, PLPP2, ALDH3A1, GLS2, SAO, LPCAT1, PEMT, HSD17B6, DNMT3B, SHMT1, SHMT2, NDUFA2, KL, NDUFA4L2, CS, AK5, ACLY, POLR3B, GART, PLA2G4A, GGT7, MBOAT1, ATP6V0A1, GAMT, CYP8B1, RDH16*

Appendix 7A. Gene Ontology terms annotation analysis enriched (p<0.01) based on copy number variation regions (CNVRs) and autozygosity islands overlapping regions set of genes identified for the genotyped animals (*n=9,386*)

Gene Ontology	*n* (Genes)	P-value	Genes
Biological Process			
GO:0040018	5	0.003	*NIPBL, STAT5A, SLC6A3, STAT5B, HMGA2*
GO:0070200	3	0.006	*NABP2, WRAP53, TERT*
GO:0042742	7	0.006	*DEFB6, DEFB5, DEFB7, EBD, LAP, DEFB1, LEAP2*
GO:0007286	6	0.009	*NME5, RNF17, STK11, ADAD1, AFF4, ZMYND15*
Molecular Function			
GO:0004871	9	0.003	*STAT6, GNAL, GNA11, STAT5A, ACAP1, STAT5B, CXXC5, STAT3, GNG7*
KEGG			
bta05166	12	0.007	*KAT2A, EGR1, DVL2, E2F3, APC2, STAT5A, STAT5B, TP53, NFKB2, TCF3, TERT, GPS2*

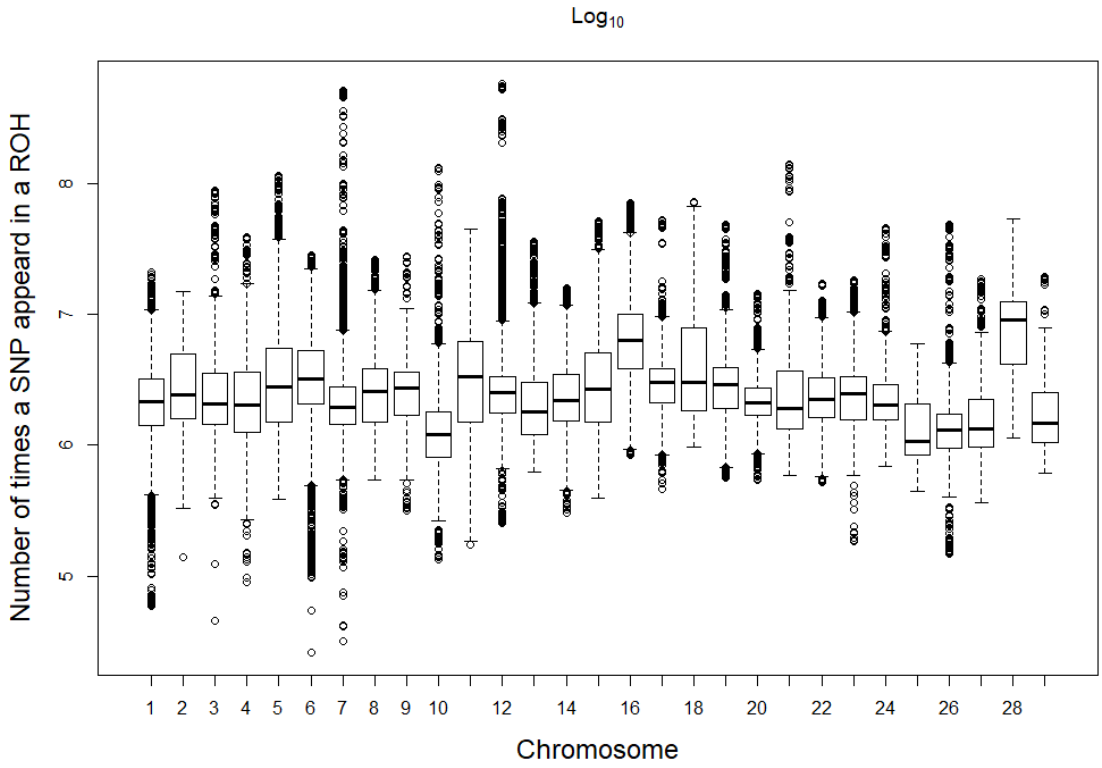
Appendix 8A. Runs of homozygosity islands described in several cattle breeds located within those observed in the present study

Author	Cattle Breed	BTA[2]	Physical Position (bp)
(SÖLKNER et al., 2014)	Brahman, Gyr, and Nellore	7	51,502,500:52,353,000[1]
		12	28,434,000:29:628,100
		21	1,360,390:1,853,150[1]
(GASPA et al., 2014)	Italian Holstein	21	898,385:1,829,761[1]
		26	211,146,794:23,000,155
(SZMATOŁA et al., 2016)	Holstein	7	42,440,064:43,592,173[1]
		7	51,574,295:52,419,683[1]
		14	24,220,070:25,351,733
		20	28,329,720:32,293,167
		22	22,004,775:23,984,012
		29	37,782,301:39,905,644
	Red Polish	1	31,206,393:31,659,179
		7	51,574,295:54,081,460[1]
	Simmental	7	42,645,056:45,383,502[1]
		7	51,157,314:53,101,552[1]
		14	23,853,811:24,326,513
	Limousin	1	31,239,593:32,036,293
		5	47,752,157:49,103,647
		7	42,765,700:43,808,593[1]
		7	53,101,552:53,859,609[1]
		14	23,122,719:28,548,600
(PERIPOLLI et al., 2018)	Gir	6	70,117,799:81,603,050

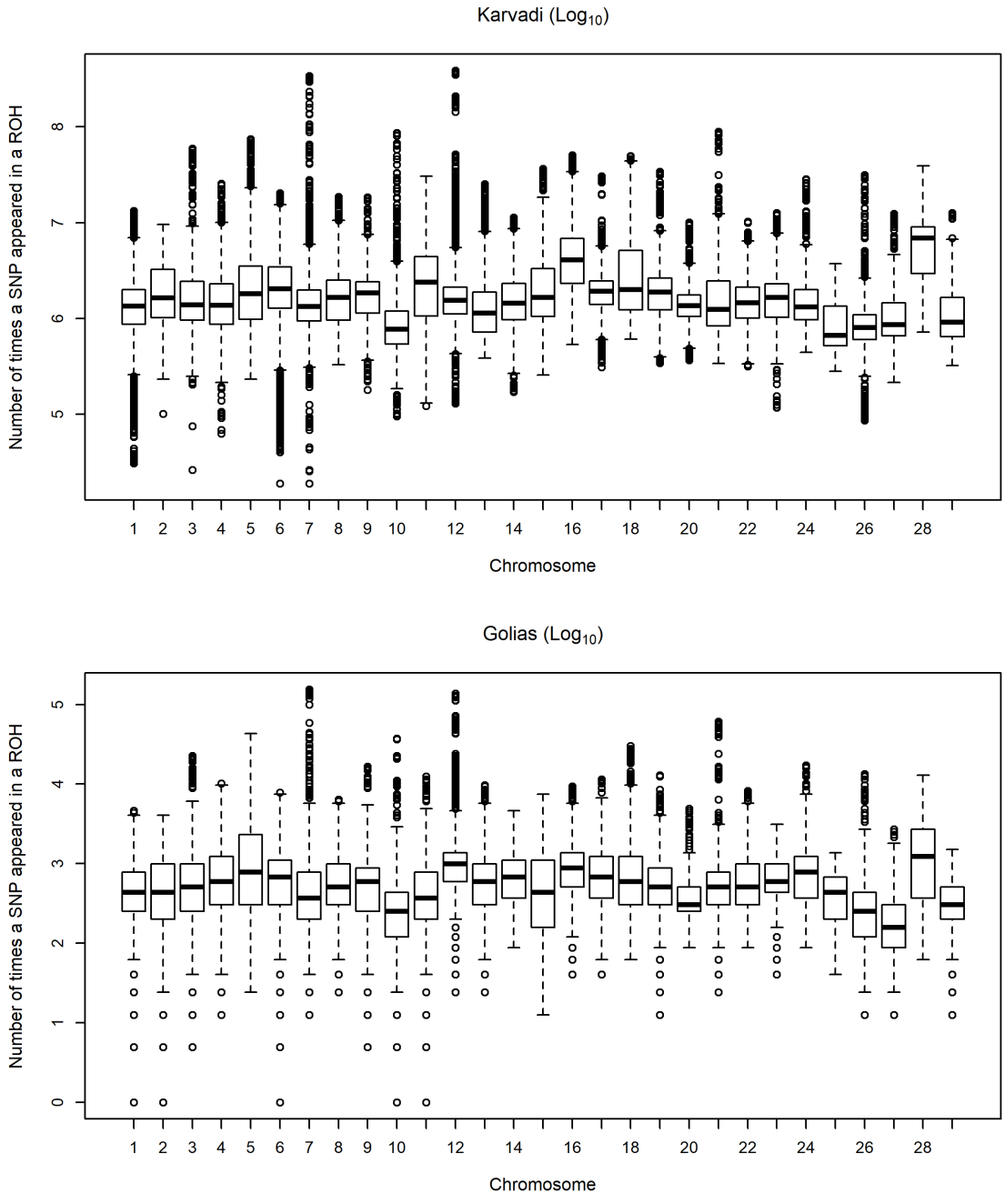
[1] Autozygosity islands overlapping between these studies – current study;

[2] BTA: *Bos taurus* autosome.

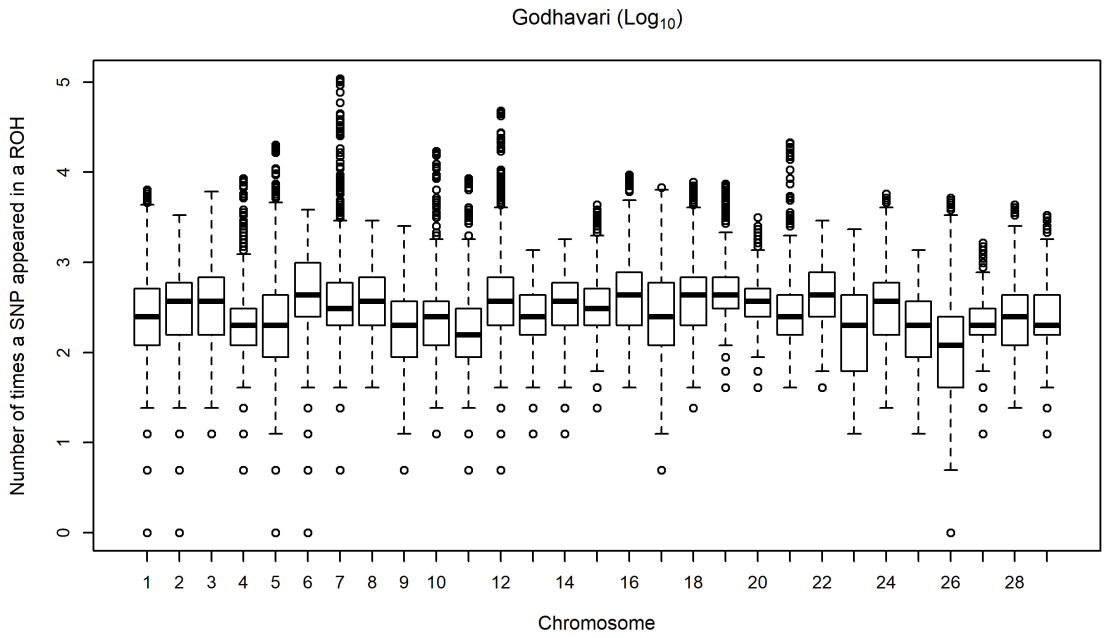
Appendix 9A. Outliers SNPs for the genotyped animals (*n=9,386*) according to Boxplot distribution.

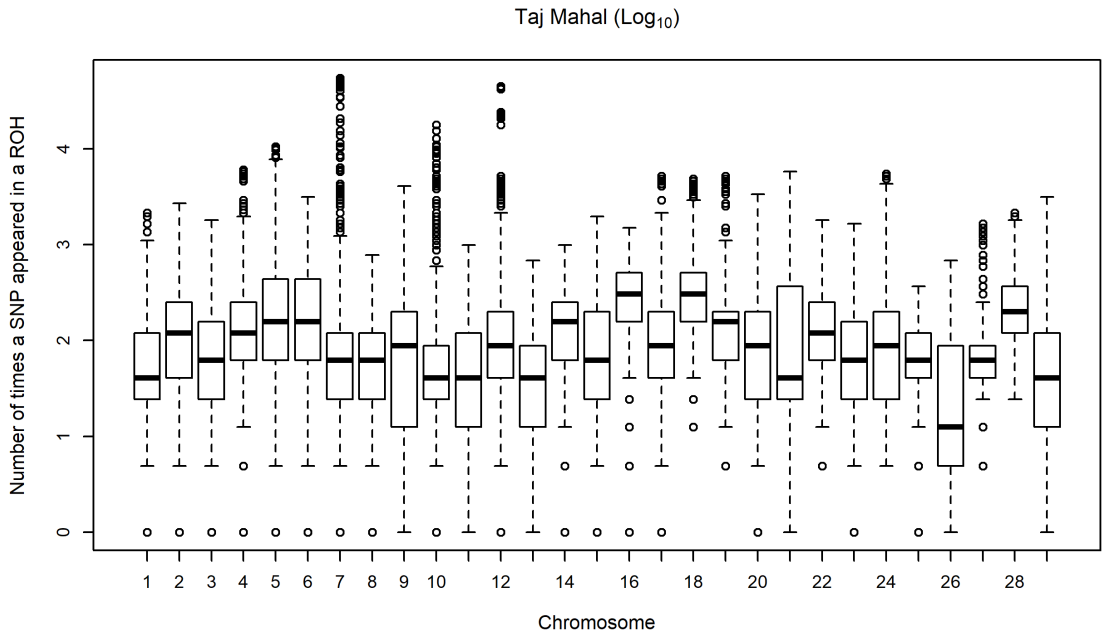


Appendix 10A. Outliers SNPs for each Nellore lineage (*n=8,646*) according to Boxplot distribution.

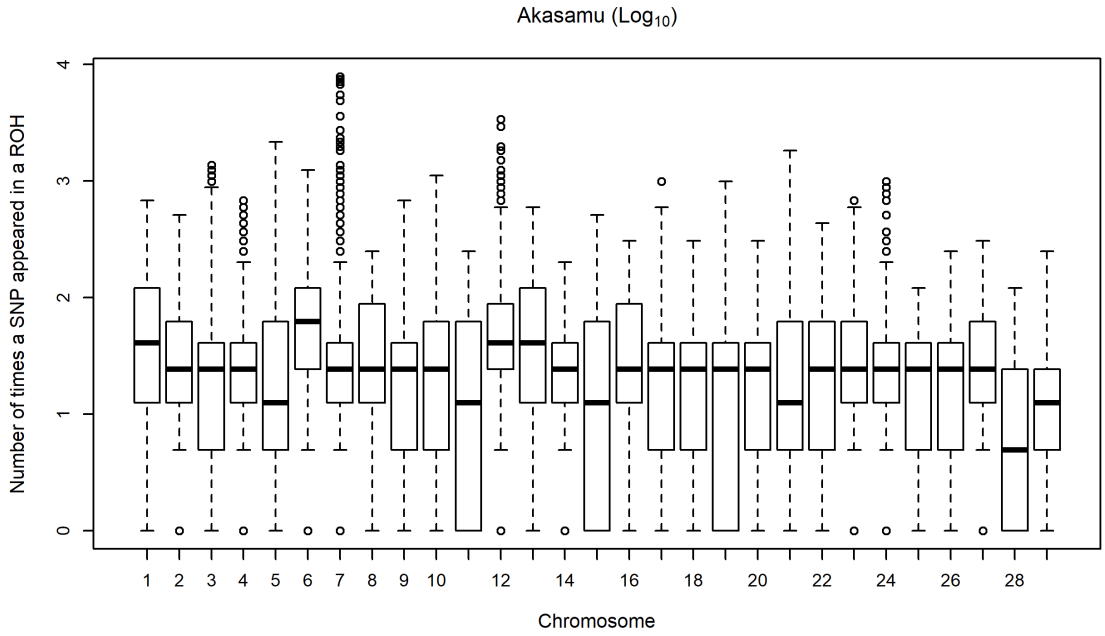


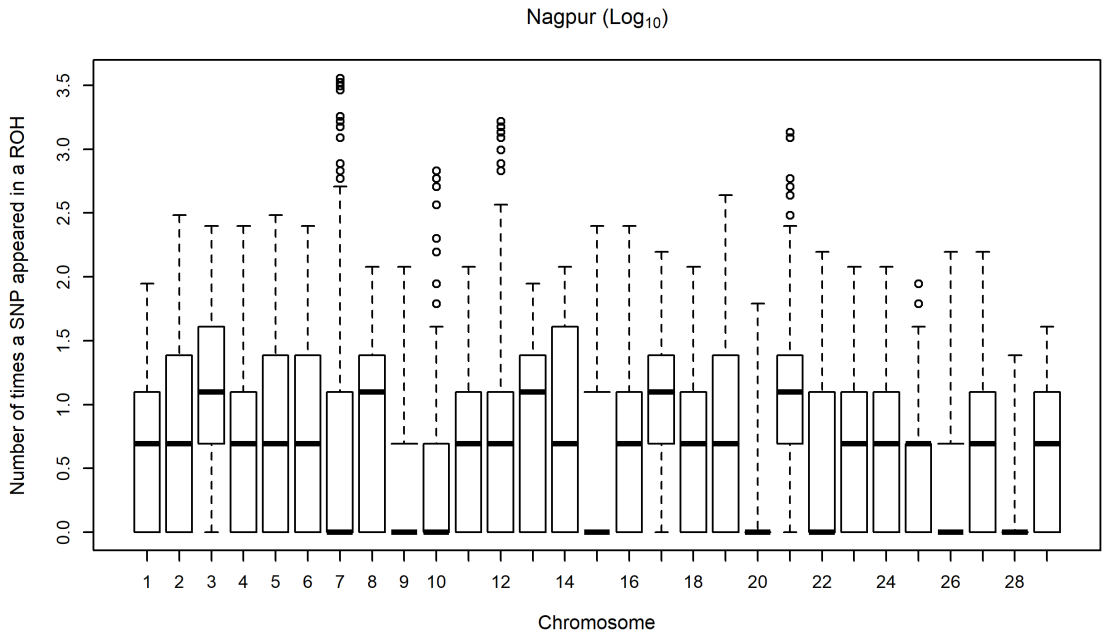
Appendix 10A. Continuation





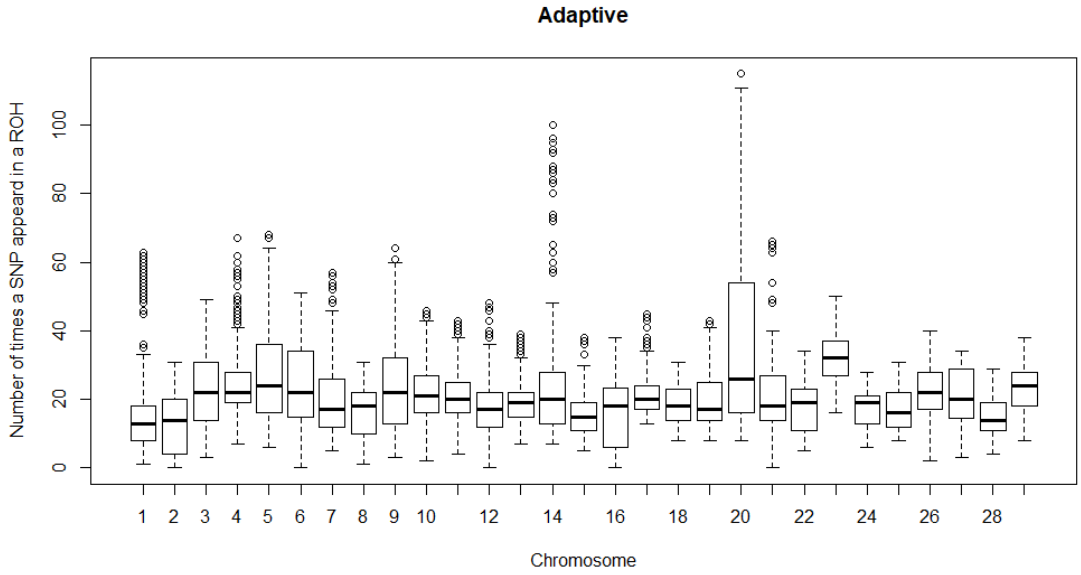
Appendix 10A. Continuation

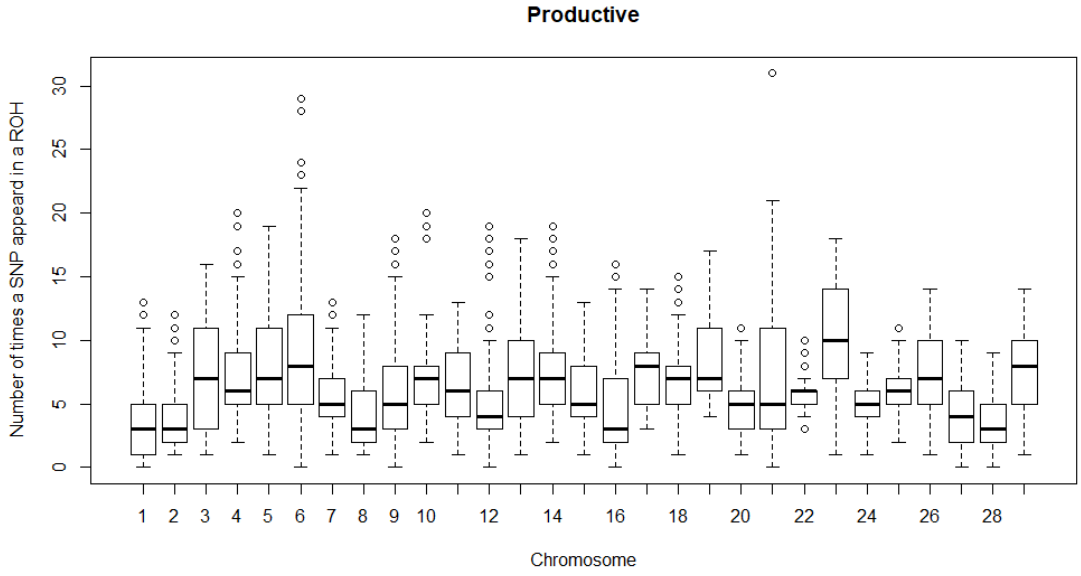




APPENDIX B

Appendix 1B. Outliers SNPs for the composite Montana Tropical® beef cattle according to Boxplot distribution for the adaptive and productive biological type.





Appendix 2B. Autozygosity islands across the genome of the composite Montana Tropical® beef cattle for each biological type (adaptive and productive).

Adaptive biological type[1]		
BTA[2]	**Start (bp)**	**End (bp)**
1	199,195	6,154,638
4	48,280,680	50,212,515
4	69,943,128	71,356,324
4	71,520,063	73,146,717
5	75,299,940	75,309,500
7	44,620,188	46,574,843
11	54,532,980	56,657,354
11	57,356,797	59,902,171
12	17,681,556	20,420,911
13	41,848,011	45,583,950
14	23,817,572	27,751,888
15	727,265	3,709,693
17	1,354,436	3,383,964
19	56,670,683	57,149,037
21	166,024	4,408,359
Productive biological type[3]		
2	5,306,838	7,492,224
4	49,259,497	50,212,515
5	38,239,272	39,615,850
7	108,741,970	112,359,264
9	10,894,290	11,390,953
10	43,562,935	47,842,705
12	17,289,717	20,420,911
12	86,918,646	88,677,992
14	24,475,213	25,887,784
14	51,586,769	54,525,313

Appendix 2B. Continuation

16	25,673,002	25,954,959
18	17,732,724	21,287,904
18	23,319,689	26,209,903
20	32,293,167	33,722,626
21	166,024	3,887,470
22	32,861,744	37,203,531
25	38,961,935	40,226,964

[1] Adaptive biological type comprises animals 4444, 4822, and 4840 according to the NABC system, [2] BTA = *Bos taurus* autosome; [3] Productive biological type comprises animals 4444, 4624, and 4642 according to the NABC system.

Appendix 3B. Gene Ontology (GO) terms and KEGG pathways annotation analysis enriched (P<0.05) based on autozygosity islands set of genes identified in adaptive biological type.

Term	*n*	P-value	Genes
GO Biological Process			
GO:2000117~negative regulation of cysteine-type endopeptidase activity	6	2.41E-07	*CSTL1, CST8, CST11, MGC133636, CST7, CST3*
GO:0007218~neuropeptide signaling pathway	6	0.003195	*NPVF, KISS1R, PENK, NPY, CYSLTR2, NPY2R*
GO:0051258~protein polymerization	3	0.008433	*FGG, FGA, FGB*
GO:0030521~androgen receptor signaling pathway	3	0.023875	*MED4, MED16, UBE3A*
GO:0048240~sperm capacitation	3	0.023875	*SLC26A3, DLD, PCSK4*
GO:0043161~proteasome-mediated ubiquitin-dependent protein catabolic process	6	0.026827	*RNF126, UBXN2B, BTBD2, KCTD2, CDC34, SOD1*
GO:0060337~type I interferon signaling pathway	2	0.028299	*IFNAR2, IFNAR1*
GO:0042391~regulation of membrane potential	4	0.039691	*SLC26A4, HCN2, SLC26A3, DLD*
GO:0072378~blood coagulation, fibrin clot formation	2	0.042148	*FGG, FGB*
GO:0030168~platelet activation	3	0.045415	*FGG, FGA, FGB*
GO Cellular Component			
GO:0005577~fibrinogen complex	3	0.004404	*FGG, FGA, FGB*
GO:0070469~respiratory chain	3	0.020366	*UQCR11, CYCS, UQCRQ*
GO:0001669~acrosomal vesicle	4	0.032617	*RCBTB2, STK31, FNDC3A, ATP8B3*
GO Molecular Function			
GO:0004869~cysteine-type endopeptidase inhibitor activity	6	4.20E-05	*CSTL1, CST8, CST11, MGC133636, CST7, CST3*
GO:0002020~protease binding	6	8.42E-05	*CSTL1, CST8, CST11, MGC133636, CST7, CST3*

GO:0016491~oxidoreductase activity	8	0.001111	*AKR1C3, GDI2, AKR1C4, AKR1E2, 20ALPHA-HSD, SDR16C6, FADS6, CRYZL1*
GO:0030674~protein binding, bridging	3	0.022860	*FGG, FGA, FGB*
GO:0005234~extracellular-glutamate-gated ion channel activity	3	0.022860	*GRIK1, GRIN3B, GRIA4*
GO:0004252~serine-type endopeptidase activity	7	0.025630	*AZU1, GZMM, PRTN3, PRSS57, ELANE, CFD, PCSK4*
GO:0004905~type I interferon receptor activity	2	0.029411	*IFNAR2, IFNAR1*
KEGG pathway			
bta04080:Neuroactive ligand-receptor interaction	11	0.004433	*GPR83, KISS1R, GABRB3, GRIK1, CYSLTR2, LPAR6, NPY2R, MLNR, GABRA5, GRIN3B, GRIA4*
bta04610:Complement and coagulation cascades	5	0.015365	*FGG, THBD, FGA, FGB, CFD*
bta04650:Natural killer cell mediated cytotoxicity	6	0.017777	*PIK3CG, IFNAR2, GRB2, SHC2, IFNGR2, IFNAR1*
bta04910:Insulin signaling pathway	6	0.031561	*PIK3CG, PRKAR2B, GRB2, CALML5, SHC2, PYGB*
bta04024:cAMP signaling pathway	7	0.043454	*PIK3CG, HCN2, NPY, TIAM1, GRIN3B, GRIA4, CALML5*
bta04630:Jak-STAT signaling pathway	6	0.046193	*PIK3CG, IFNAR2, GRB2, IL10RB, IFNGR2, IFNAR1*

Appendix 4B. Gene Ontology (GO) terms and KEGG pathways annotation analysis enriched (P<0.05) based on autozygosity islands set of genes identified in productive biological type.

Term	*n*	P-value	Genes
GO Biological Process			
GO:0045892~negative regulation of transcription, DNA-templated	10	0.002841	*MAGEL2, FOXK1, PRICKLE1, YAF2, NAB1, BRD7, ZNF12, RB1, MT3, ZNF423*
GO:0045893~positive regulation of transcription, DNA-templated	9	0.009351	*MED4, TRIP4, FOXK1, YAF2, BRD7, MSTN, TOX3, MT3, ZNF423*
GO:0007417~central nervous system development	4	0.015791	*NRCAM, NDN, LYN, LIG4*
GO:0010332~response to gamma radiation	3	0.023923	*TRIM13, PRKAA1, LIG4*
GO:0048247~lymphocyte chemotaxis	3	0.026120	*CCL22, CX3CL1, CCL17*
GO:0006511~ubiquitin-dependent protein catabolic process	5	0.027327	*CYLD, HERPUD1, USP3, FBXO4, AMFR*
GO:2001242~regulation of intrinsic apoptotic signaling pathway	2	0.034112	*CYLD, DAPK2*
GO:0035987~endodermal cell differentiation	3	0.035670	*MMP15, LAMB1, MMP2*
GO:0002548~monocyte chemotaxis	3	0.043578	*CCL22, CX3CL1, CCL17*
GO:0045087~innate immune response	7	0.047361	*NLRC5, CYLD, NOD2, LYN, C6, TRIM13, FER*
GO Cellular Component			
GO:0005856~cytoskeleton	7	0.011123	*ACTB, NOD2, FRMD6, TLN2, TRIM9, DRC7, TPM1*
GO Molecular Function			
GO:0008270~zinc ion binding	23	0.002099	*ZDHHC4, TRIP4, SETDB2, USP3, CA12, TRIM13, ZCRB1, PHF11, MMP15, MMP2, PJA2, MAN2A1, CYLD, ADAMTS9, PRICKLE1, MT1A, YAF2, RSPRY1, TRIM9, MT2A, CDADC1, AMFR, MT3*
GO:0004842~ubiquitin-protein transferase activity	7	0.025942	*MAGEL2, UBE3A, TRIM9, TRIM13, RNF216, AMFR, HERC1*

Appendix 4B. Continuation

GO:0016874~ligase activity	5	0.026398	*UBE3A, TRIM9, TRIM13, HERC1, SUCLA2*
GO:0048020~CCR chemokine receptor binding	3	0.028200	*CCL22, CX3CL1, CCL17*
GO:0008504~monoamine transmembrane transporter activity	2	0.035541	*SLC6A2, SLC29A4*
GO:0052689~carboxylic ester hydrolase activity	3	0.041220	*CES1, CES5A, BREH1*
KEGG pathway			
bta04611:Platelet activation	6	0.009873	*ACTB, ADCY7, LYN, TLN2, COL3A1, COL5A2*
bta04978:Mineral absorption	4	0.010366	*MT1A, MT2A, MT1E, SLC40A1*
bta00640:Propanoate metabolism	3	0.029177	*SUCLG2, HIBCH, SUCLA2*
bta04062:Chemokine signaling pathway	6	0.042000	*CCL22, ADCY7, LYN, GNG2, CX3CL1, CCL17*

APPENDIX C

Appendix 1C. Distribution of the functional consequences of the called variants *(n*=33,328,447 SNPs) using the Variant Effect Predictor (VEP) tool.

Consequence	Observations
3_prime_UTR_variant	64,831
5_prime_UTR_variant	13,391
coding_sequence_variant	131
downstream_gene_variant	928,061
intergenic_variant	22,388,630
intron_variant	8,617,335
mature_miRNA_variant	175
missense_variant	88,366
non_coding_transcript_exon_variant	12,574
non_coding_transcript_variant	49
splice_acceptor_variant	471
splice_donor_variant	481
splice_region_variant	20,848
start_lost	208
stop_gained	1,111
stop_lost	58
stop_retained_variant	93
synonymous_variant	126,119
upstream_gene_variant	1,065,515

Appendix 2C. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis enriched (p<0.01) on DAVID tool based on variants with high consequence on protein sequence set of genes.

Term	Benjamini	Count	Genes
GO:0004984~olfactory receptor activity	6.7E-43	161	*LOC508626, LOC784858, LOC100299084, LOC100849008, LOC788573, LOC504623, OR51I1, LOC523060, LOC524702, LOC781264, OR4E2, LOC788554, OR11G2, LOC618173, LOC528807, LOC100299372, LOC519294, OR2A12, LOC785565, OR4D5, LOC524985, LOC100850909, LOC522554, OR13C3, LOC787625, LOC785582, LOC782941, LOC509073, LOC512973, OR2H1, LOC100140748, LOC530990, OR2L13, OR5B3, LOC783885, OR10K2, LOC532075, OR10AD1, LOC617417, LOC100847301, LOC100298850, OR7D2, LOC526276, OR6C2, LOC788476, LOC514057, LOC787659, OR6T1, LOC100295806, OR5D13, LOC786846, LOC100298773, OR2AP1, OR51D1, LOC100847281, LOC616306, LOC619021, LOC785623, OR4D2, LOC782338, LOC100337063, OR2B11, LOC511103, LOC104973083, LOC789031, OR10J3, LOC104969161, LOC786596, LOC618523, LOC785910, LOC508604, LOC523090, LOC107131130, LOC782645, OR4C46, LOC787945, LOC782373, LOC787946, LOC788524, LOC788790, LOC517252, LOC785903, LOC788512, LOC107131158, LOC782792, LOC782555, LOC788693, LOC784455, LOC785277, LOC512627, LOC783518, LOC785683, LOC104969845, LOC615810, LOC616517, LOC788287, OR2A5, LOC516409, LOC518561, LOC532441, LOC788675, LOC101902265, LOC783488, LOC540082, OR4F15, LOC788027, OR10J1, LOC107131149, LOC615150, OR2J3, LOC100140382, LOC528373, LOC516132, LOC783311, LOC785723, OR4C15, LOC787500, LOC785712, LOC100299725, OR10Z1, LOC523139, LOC787991, OR9A4, LOC520162, OR6K3, LOC507971, LOC789936, LOC783313, LOC519616, LOC504567, OR52I2, LOC618675, LOC516940, LOC788633, LOC788874, OR8K1, LOC618660, LOC100298322, LOC509510, LOC782152, LOC782475, OR4C6, LOC532100, LOC506486, LOC540354, LOC616705, OR6Y1, LOC516467, LOC100336916, OR14J1, LOC512296, OR10A3, LOC100301320, LOC783845, LOC614591, LOC524658, LOC784434, LOC617592, LOC784925, LOC789134, LOC510351*
bta04740:Olfactory transduction	1.6E-40	182	*LOC508626, LOC784858, LOC100299084, LOC789358, LOC787867, LOC100849008, LOC788573, LOC504623, OR13F1, OR51I1, LOC523060, LOC100298645, LOC524702, LOC785392, LOC781264, LOC788554, OR4E2, OR11G2, LOC618173, LOC528807, LOC100299372, LOC100139408, OR2A12, LOC785565, LOC519294, OR4D5, LOC784595, LOC789367, LOC615281, LOC524985, LOC100850909, LOC522554, OR13C3, LOC100301297, LOC787625, LOC785582, LOC782941, LOC618816, LOC509073, LOC787150, LOC512973, OR2H1, LOC100140748, LOC530990, OR2L13, OR5B3, LOC783885, OR10K2, LOC532075, OR10AD1, LOC100298850, LOC100847301, LOC617417, OR7D2, LOC526276, OR6C2, LOC788476, LOC787659, LOC514057, LOC100300099, OR6T1, LOC100295806, OR5D13, LOC786846, LOC100298773, OR2AP1, OR51D1, LOC616306, LOC785623, OR4D2, LOC782338, LOC100337063, OR2B11, LOC511103, LOC789031, OR10J3, LOC790121, LOC104969161, LOC786596, LOC619067, LOC618523, LOC785910, LOC508604, LOC523090, LOC782645, OR4C46, LOC787945, LOC782373, LOC787946, CNGB1, LOC788790, LOC517252, LOC785903, LOC782792, LOC782555, LOC788693, LOC784455, OR8G5, LOC785277, LOC512627, LOC617333, LOC783518, LOC785683, LOC104969845, LOC788287, LOC616517, LOC615810, OR2A5, LOC516409, LOC522385, LOC518561, LOC532441,*

			LOC788675, LOC101902265, LOC100140912, LOC783488, OR13F1, LOC540082, OR4F15, LOC788027, OR10J1, LOC615150, OR2J3, LOC100140382, LOC528373, LOC516132, LOC510150, LOC783311, LOC785723, OR4C15, LOC785712, LOC100299725, LOC618717, OR10Z1, LOC523139, LOC787991, OR9A4, LOC520162, OR6K3, LOC507971, LOC789936, LOC783313, LOC519616, LOC504567, OR10AG1, OR52I2, LOC618675, LOC785207, LOC516940, LOC788633, LOC788874, OR8K1, LOC618660, LOC100298322, LOC509510, LOC782152, LOC782475, OR4C6, LOC532100, LOC506486, LOC540354, LOC616705, OR6Y1, LOC516467, LOC788037, LOC100336916, OR14J1, LOC512296, OR10A3, LOC617571, LOC100301320, LOC783845, LOC788898, LOC785779, LOC614591, LOC524658, LOC781403, LOC784434, LOC617592, LOC784925, LOC789134, LOC510351
GO:0004930~G-protein coupled receptor activity	3.4E-34	163	*LOC508626, LOC784858, LOC100299084, TAS2R40, LOC100849008, LOC788573, LOC504623, OR51I1, LOC523060, LOC524702, LOC781264, LOC788554, OR4E2, OR11G2, LOC618173, LOC528807, LOC100299372, LOC519294, OR2A12, LOC785565, OR4D5, LOC524985, LOC100850909, LOC522554, OR13C3, LOC787625, LOC785582, LOC782941, LOC509073, LOC512973, OR2H1, LOC100140748, LOC530990, OR2L13, OR5B3, LOC783885, OR10K2, LOC532075, OR10AD1, LOC617417, LOC100847301, LOC100298850, OR7D2, LOC526276, OR6C2, LOC788476, LOC514057, LOC787659, OR6T1, LOC100295806, OR5D13, LOC786846, LOC100298773, OR2AP1, OR51D1, LOC100847281, LOC785618, LOC616306, LOC619021, LOC785623, OR4D2, LOC782338, LOC100337063, OR2B11, LOC511103, LOC104973083, LOC789031, OR10J3, MGC137098, LOC104969161, LOC786596, ADGRL4, LOC618523, LOC785910, LOC508604, LOC523090, LOC107131130, OR4C46, LOC787945, LOC782373, LOC787946, LOC788524, LOC788790, LOC517252, LOC785903, LOC788512, LOC107131158, LOC782792, LOC782555, LOC788693, LOC784455, LOC785277, LOC512627, LOC783518, LOC785683, LOC788287, LOC616517, LOC615810, OR2A5, LOC516409, LOC518561, LOC788675, LOC101902265, LOC783488, LOC540082, OR4F15, LOC788027, OR10J1, LOC107131149, OR2J3, LOC100140382, LOC528373, LOC516132, LOC783311, GHRHR, LOC785723, OR4C15, LOC787500, LOC785712, LOC100299725, QRFPR, OR10Z1, LOC523139, LOC787991, OR9A4, LOC520162, OR6K3, LOC507971, LOC789936, LOC783313, LOC519616, LOC504567, OR52I2, LOC618675, LOC516940, LOC788633, LOC788874, OR8K1, LOC618660, LOC100298322, LOC509510, LOC782152, LOC782475, OR4C6, LOC532100, LOC506486, LOC616705, OR6Y1, LOC516467, LOC100336916, OR14J1, T2R65A, LOC512296, OR10A3, LOC100301320, LOC783845, LOC614591, LOC524658, LOC784434, LOC617592, LOC784925, LOC789134, LOC510351*

GO:0007186~G-protein coupled receptor signaling pathway	1.2E-29	135	*LOC784858, LOC508626, LOC100299084, LOC100849008, LOC788573, LOC504623, LOC523060, CALCRL, LOC524702, RGS9, LOC781264, LOC788554, OR4E2, OR11G2, LOC100299372, LOC785565, LOC519294, OR2A12, OR4D5, LOC524985, LOC100850909, LOC522554, OR13C3, LOC787625, LOC785582, LOC509073, LOC512973, OR2H1, LOC100140748, LOC783885, OR10K2, OR2L13, OR5B3, LOC532075, OR10AD1, LOC100847301, LOC100298850, OR7D2, LOC526276, LOC788476, LOC514057, OR5D13, LOC100295806, LOC786846, LOC100298773, LOC100847281, LOC616306, LOC619021, LOC785623, OR4D2, LOC100337063, OR2B11, LOC511103, LOC789031, OR10J3, LOC104969161, LOC786596, LOC618523, LOC785910, LOC508604, VAV1, LOC107131130, OR4C46, LOC787946, LOC788524, LOC788790, LOC517252, LOC785903, LOC788512, LOC107131158, LOC782792, LOC782555, LOC788693, LOC784455, LOC785277, LOC512627, LOC783518, LOC785683, LOC616517, LOC788287, LOC615810, OR2A5, LOC516409, LOC788675, LOC101902265, LOC540082, OR4F15, LOC107131149, OR10J1, OR2J3, LOC528373, LOC100140382, LOC516132, LOC783311, LOC785723, OR4C15, LOC787500, LOC785712, LOC100299725, OR10Z1, LOC523139, OR9A4, LOC520162, OR6K3, LOC507971, LOC789936, LOC783313, LOC519616, LOC504567, LOC618675, LOC516940, LOC788874, OR8K1, LOC618660, LOC100298322, LOC509510, LOC782152, LOC782475, OR4C6, LOC506486, LOC532100, LOC616705, OR6Y1, LOC516467, OR14J1, LOC100336916, LOC512296, OR10A3, LOC504773, LOC100301320, LOC783845, LOC614591, LOC784434, LOC617592, LOC784925*

GO:0016021~integral component of membrane	7.2E-13	354	*FCAMR, DERL1, LOC100299084, TAS2R40, CD151, LOC788573, GYPB, LOC781264, FUT6, CISD2, LOC613867, LOC788554, LOC509854, LOC618173, LOC528807, BCL2L1, OR2A12, LOC785565, ABCA10, LOC524985, LOC100850909, LOC787625, LOC785582, LOC782941, HSD17B12, KCNK16, OR2H1, LOC100140748, MARC1, OR5B3, OR2L13, OR10K2, TMPO, FAAH, LOC787659, LOC788205, LOC514057, LOC512150, OR6T1, LOC100295806, LOC527385, SVOP, MFSD14A, LOC100140174, RTN3, ACVRL1, LOC785618, LOC100847738, SLC4A10, LOC785623, FER1L6, LOC782338, OR2B11, LOC104973083, LOC789031, SMIM8, MGC137098, LCLAT1, LOC786596, AIFM2, ADGRL4, MFN1, LRIG3, ROS1, NOX5, OR4C46, LOC107131130, LOC782373, LOC788524, BOSTAUV1R419, CNGB1, SERINC1, FYCO1, TMEM86B, LOC788512, LOC782555, LOC788693, OCA2, LOC512627, ATP8A1, LOC783518, FRMD5, PIK3IP1, CYP4A11, LOC616517, CDHR4, OR2A5, LOC518561, LOC516409, LOC788675, LOC783488, LOC514257, LOC540082, TMC6, OR4F15, LOC788027, SYNE1, MHC class II associated), Hsp40) member C15, CLCN2, TLR5, BOLA-NC1, GHRHR, OR4C15, SLC36A3, LOC787500, FREM2, LOC100299725, LRCH3, LOC787991, IL1RL1, LOC520162, SLC38A9, SLC13A5, MUSK, LOC519616, LOC504567, OR52I2, LOC516940, PIEZO1, PARL, LOC788633, OR8K1, TMIGD2, LOC788634, LOC100335205, LOC100297846, DIRC2, LOC100298322, NPR1, LOC782152, MIC1, LOC782475, LOC506486, MBOAT2, CYP4B1, OR6Y1, CDH22, LOC516467, SLC30A5, OR14J1, T2R65A, FAT3, MGC157082, VRK1, ABCA2, TMEM176A, CLMN, FRRS1L, REV1, LOC784434, LOC100295883, THSD7A, LOC615051, SECTM1A, RMND1, LOC784925, LOC789134, RYR2, LOC100850276, LOC510351, LOC784858, LOC508626, PTPRB, LOC100849008, LOC504623, PLG, LOC523060, OR51I1, CALCRL, UPK3BL, MEGF8, CYB5D2, LOC524702, DSCAML1, PCDH17, KCNH1, OR4E2, LOC100299372, TCEB3, PCDHB11, LOC519294, OR4D5, LRRC4, LOC101905933, LOC785804, ITPR2, PTPRC, CD1A, LOC522554, OR13C3, TGFBR3, PTPRQ, LOC509073, CORIN, PILRA, LOC512973, PRLR, LOC530990, SMCO2, MGC127055, TGFBR2, LOC532075, ISG12, CD207, OR10AD1, B3GNT5, TMEM26, SIDT2, LOC100298850, LOC100847301, LOC617417, OR7D2, LOC526276, OR6C2, SLCO1A2, JKAMP, LOC788476, TRPV3, PQLC2, OR5D13, SMIM11A, LOC786846, LOC100298773, OR2AP1, GSG1L2, OR51D1, LOC100847281, CD46, CLEC7A, LOC522174, ASIC2, LOC616306, LOC619021, MANSC4, OR4D2, TMEM104, LOC100337063, LOC100336589, SDK1, LOC511103, ATG9B, SKINT1, ABC1, OR10J3, UBE2J1, NRCAM, LOC104969161, CCR6, LOC618523, LOC785910, LOC508604, LOC523090, VSTM1, LOC787945, LOC787946, LOC514011, MCOLN3, GRAMD3, SLC6A12, LOC788790, LOC517252, GPAT3, LOC509972, CHIC2, LOC785903, LOC107131158, BOLA-DQA2, LOC782792, EMCN, TMEM116, RYK, LOC785277, TMCO5B, LOC785683, LOC788287, LOC615810, TMEM237, LOC101902265, TIMMDC1, ULBP3, ABCB1, LOC100139826, LOC107131149, PIGN, OR10J1, OR2J3, LOC100140382, LOC528373, IL31RA, LOC516132, NRADD, LOC783311, LOC785236, DPEP3, GALNT5, BOLA-DYA, LOC785723, FKBP8, CDHR3, LOC785712, LMAN1, OR10Z1, UCP1, LOC523139, LOC786796, TMEM63C, USH2A, LOC536660, NKG2C, OR6K3, LOC507971, LOC516101, LOC789936, LOC783313, LOC618675, RYR3, LOC788874, CYP4A22, AGER, LOC618660, IGSF23, TLR3, GP5, LOC509510, OR4C6, LOC532100, GPR89A, LOC616705, CYB561A3, LOC509034, LOC100336916, PTGFRN, SELP, OR10A3, LOC512296, LOC100301320, ALPI, LOC524658, TMEM192, TRPC2, TRPC4, SCARA3, KLRF1, BOSTAUV1R403, SLC2A11, TYRP1, C8H9orf135, GYPA, DPP10, LOC618633*

GO:0005886~plasma membrane	4.6E-12	218	*LOC100299084, LOC788573, STAMBP, LOC781264, TCAF2, LOC788554, LOC618173, LOC528807, OR2A12, LOC785565, CEP89, LOC524985, LOC100850909, NUP35, ALOX15, LOC787625, LOC785582, LOC782941, OR2H1, LOC100140748, OR2L13, OR5B3, OR10K2, LOC787659, LOC514057, OR6T1, LOC100295806, ERAP1, LOC785623, LOC782338, OR2B11, LOC104973083, LOC789031, CPNE2, LOC786596, REPS1, PLEKHH2, LOC107131130, OR4C46, LOC782373, LOC788524, TLN1, LOC788512, LOC782555, XRCC5, RFFL, LOC788693, LOC512627, EPB42, ATP8A1, LOC783518, PIK3IP1, ENPP6, LOC616517, CDHR4, OR2A5, LOC516409, LOC518561, LOC788675, LOC783488, LOC540082, OR4F15, LOC788027, SCN7A, BOLA-NC1, GHRHR, OR4C15, LOC787500, LOC100299725, LOC787991, LOC520162, LOC519616, LOC504567, OR52I2, LOC516940, LOC788633, OR8K1, LOC788634, LOC100298322, NPR1, PPIL2, LOC782152, LOC782475, MIC1, LOC506486, CDH22, OR6Y1, LLGL2, LOC516467, OR14J1, FAT3, LOC784434, LOC784925, LOC789134, LOC510351, LOC100850276, LOC784858, LOC508626, LOC100849008, LOC504623, LOC523060, OR51I1, CALCRL, LOC524702, RGS9, PKP2, PCDH17, OR4E2, LOC100299372, LOC519294, PCDHB11, OR4D5, PDZK1, ITPR2, CD1A, LOC522554, OR13C3, LOC509073, PRLR, LOC512973, LOC530990, TGFBR2, LOC532075, OR10AD1, LOC100298850, LOC100847301, LOC617417, OR7D2, LOC526276, OR6C2, SLCO1A2, LOC788476, MYO10, CSNK1D, OR5D13, LOC786846, LOC100298773, PLA2G3, OR2AP1, OR51D1, LOC100847281, CD46, CLEC7A, SPG11, LOC616306, LOC619021, TJP2, OR4D2, LOC100337063, IL16, SYTL1, LOC511103, OR10J3, BAIAP2L2, NRCAM, PAK1IP1, NMT1, LOC104969161, LOC618523, LOC785910, LOC508604, LOC523090, OPCML, LOC787945, LOC787946, MCOLN3, CATSPER1, MAP3K7, LOC788790, LOC517252, LOC785903, CHIC2, LOC107131158, LOC782792, TICAM2, RYK, LOC785277, XPC, LOC785683, LOC615810, LOC788287, HDAC11, LOC101902265, PIK3C2G, LOC107131149, OR10J1, OR2J3, LOC528373, LOC100140382, LOC516132, LOC783311, ANXA2, LOC785723, LOC785712, CDHR3, OR10Z1, COG3, LOC523139, LOC536660, OR6K3, LOC507971, LOC789936, LOC783313, LOC618675, NMT2, LOC788874, LOC618660, LOC509510, OR4C6, LOC532100, LOC616705, PTGFRN, LOC100336916, LOC512296, OR10A3, LOC100301320, ALPI, LOC524658*
GO:0050907~detection of chemical stimulus involved in sensory perception	4.0E-11	38	*LOC107131158, LOC504567, LOC788693, LOC618675, LOC523060, LOC512627, LOC524702, LOC785683, LOC781264, LOC100298322, OR4E2, LOC616306, LOC785623, OR4D2, LOC100337063, OR4F15, OR4C6, LOC519294, OR4D5, OR10J1, LOC100140382, OR10J3, LOC516467, LOC786596, LOC618523, LOC785910, LOC785723, LOC785582, OR4C15, LOC509073, LOC785712, LOC512973, OR4C46, OR10K2, LOC788790, OR6K3, LOC517252, LOC784925*
GO:0007608~sensory perception of smell	1.6E-10	43	*LOC784858, LOC508626, LOC100849008, LOC785277, LOC783518, LOC615810, LOC616517, LOC788287, OR2A5, LOC516409, LOC788675, LOC101902265, LOC100299372, LOC619021, LOC782475, OR2A12, LOC785565, OR2B11, LOC506486, LOC532100, OR2J3, LOC511103, LOC528373, LOC789031, OR6Y1, LOC616705, LOC100850909, LOC783311, LOC104969161, OR10A3, LOC522554, LOC100301320, OR13C3, LOC787500, OR2H1, OR10Z1, LOC787946, LOC784434, OR2L13, CNGB1, OR10AD1, LOC507971, TTC8*
GO:0004888~transmembrane signaling receptor activity	2.6E-10	41	*FCAMR, LOC107131158, LOC504567, LOC788693, LOC618675, LOC523060, LOC512627, LOC524702, LOC785683, LOC781264, LOC100298322, OR4E2, TLR3, LOC616306, LOC785623, OR4D2, LOC100337063, OR4F15, OR4C6, LOC519294, OR4D5, OR10J1,*

			LOC100140382, OR10J3, LOC516467, LOC786596, TLR5, LOC618523, LOC785910, LOC785723, LOC785582, OR4C15, LOC509073, LOC785712, LOC512973, OR4C46, OR10K2, LOC788790, OR6K3, LOC517252, LOC784925
GO:0005549~odorant binding	3.2E-05	30	*LOC788476, LOC100299084, LOC782555, LOC782792, LOC514057, LOC788573, LOC504623, LOC516940, OR5D13, LOC788874, OR8K1, LOC618660, LOC788554, LOC782152, LOC524985, OR14J1, LOC100336916, LOC787625, LOC100299725, LOC100140748, LOC788524, OR5B3, LOC523139, LOC532075, LOC520162, LOC789936, LOC526276, LOC519616, LOC785903, LOC788512*

Appendix 3C. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis enriched (p<0.01) on DAVID tool based on deleterious variants (SIFT score < 0.05) set of genes.

Term	Benjamini	Count	Genes
GO:0007186~G-protein coupled receptor signaling pathway	3.6E-97	567	*PIK3R6, LOC100299084, LOC788572, LOC539468, LOC788573, OR2T33, LOC781758, LOC527450, OR2W1, LOC616658, LOC788552, LOC518442, LOC788554, LOC789504, LOC785565, LOC511753, ADGRE5, LOC618828, LOC538552, LOC787625, LOC618817, LOC785582, LOC100140748, LOC788587, LOC519492, LOC100300085, OR2L13, LOC787642, LOC107132626, LOC523389, LOC788583, OR2M5, LOC104972581, LOC508806, GPR161, LOC100295806, LOC782366, LOC509895, LOC785623, LOC787665, LOC785624, LOC515704, LOC781804, LOC521350, LOC100139733, PREX1, LOC786596, OR2A2, LOC787694, LOC527414, LOC527415, LOC785639, OR12D2, LOC101905743, LOC785647, LOC504888, OR4C46, LOC788524, ENTPD2, LOC100301231, OR7A10, LOC788512, LOC782555, OR4C16, LOC782554, LOC788693, LOC784455, OR5L1, OR2W3, LOC783518, LOC616517, OR2A5, LOC516409, OR5D14, LOC614592, LOC788675, LOC508766, LOC524304, LOC785431, LOC788723, LOC789690, LOC506981, LOC617507, LOC615605, OR4C15, LOC530825, LOC787500, OR2AT4, OR4C3, OR8K3, LOC788704, OR10V1, OR10C1, LOC508785, LOC788713, LOC526508, LOC519616, OR4D11, LOC788626, OR5M3, OR2D3, LOC100301071, LOC782475, LOC787543, OR6Y1, LOC516467, LOC510901, LOC104968790, OR2AE1, LOC100301104, LOC614591, LOC786467, LOC528722, OR10P1, LOC539574, LOC784434, LOC787574, LOC107132445, OR2M4, LOC617592, LOC509817, LOC516274, LOC508626, LOC784858, LOC789766, LOC516273, LOC785811, LOC614090, RGR, OR2A14, LOC524702, LOC530485, OR13C8, LOC511509, LOC782678, LOC519294, LOC618593, OR5K1, LOC515887, OR1G1, OR2G3, OR4N5, OR1J2, LOC522554, LOC781446, OR9Q2, OR13C3, LOC787898, OR1E1, LOC507378, LOC512973, LOC785848, OR4N2, LOC507383, LOC614143, LOC789812, LOC100299808, LOC789815, OR6K2, OR12D3, LOC100298850, LOC100847301, OR7D2, LOC789817, LOC100298119, LOC541022, LOC618554, LOC787932, OR2AG2, LOC522582, LOC786846, LOC784787, OR1D5, LOC100298773, LOC614021, LOC514864, LOC100847281, LOC101906611, LOC512948, LOC509641, LOC100337063, LOC509633, LOC104970118, LOC104969161, OR10T2, OR2V1, LOC521645, LOC781509, LOC618523, LOC788778, LOC785910, LOC508604, OR7G3, LOC785914, LOC539185, LOC522609, OR5AU1, LOC787946, LOC100847240, LOC539172, LOC508589, LOC100847239, LOC788790, LOC785899, LOC517252, LOC514818, LOC790683, LOC785903, LOC100851523, CASR, LOC529511, LOC782792, OR1L3, LOC782797, LOC529518, LOC101902679, LOC528422, LOC615808, LOC785683, LOC100138976, LOC615810, LOC784706, LOC515045, LOC506121, OR10J1, CCL8, OR2J3, RRH, LOC100139830, LOC516132, OR11H6, LOC104968964, LOC615852, LOC523768, LOC523769, LOC100336980, LOC511657, LOC785723, LOC614895, LOC785712, LOC526765, OR6N1, LOC511678, LOC523753, LOC789943, LOC789936, LOC506202, LOC785755, LOC100298103, OR4A16, LOC509526, LOC618675, LOC509525, LOC788874, OR2T11, OR8B4, LOC789957, LOC618660, LOC618660, LOC616716, LOC784332, LOC613390, LOC513384, LOC100300488, LOC517667, OR7A17, LOC510100, OR11L1, LOC100848076, RGS11, LOC104968568, LOC524160, LOC515540, LOC523258,*

			OR9K2, LOC514434, LOC518869, LOC509025, LOC787247, LOC788323, OR5D18, LOC504773, LOC512296, LOC100301320, LOC509510, LOC615901, OR4C6, LOC513914, LOC508420, LOC784652, OR6P1, LOC616705, LOC100336916, OR10A3, LOC783843, LOC783845, LOC510625, LOC508468, LOC101904538, OR4C13, LOC527779, LOC616755, OR4K5, OR8S1, PIK3R5, OR10A6, LOC529425, LOC523680, LOC787816, ADGRL3, LOC784681, LOC513884, LOC532291, OR5A1, LOC100299628, INPP5K, LOC506533, LOC513175, MTNR1A, LOC508392, LOC781264, OR4D6, LOC515090, LOC785149, LOC509369, OR11G2, LOC506549, OR2A12, LOC615009, LOC524985, OR13A1, LOC100850909, LOC504344, LOC509323, LOC782261, LOC618124, LOC510293, LOC618112, OR1Q1, OR2H1, OR5B3, OR10K2, LOC783885, LOC783884, LOC618140, LOC526047, LOC520835, LOC522775, LGR4, LOC538966, LOC100337392, LOC782255, LOC783998, LOC786149, LOC617122, LOC514057, LOC782288, LOC532238, LOC100299556, LOC782301, LOC789041, LOC510257, OR8D2, LOC785082, OR2B11, LOC520938, INSR, LOC613799, LOC789031, LOC525964, LOC531304, LOC782866, LOC786133, LOC107131130, LOC783951, OR2T4, LOC508315, LOC788998, OR2D2, OR2T12, OR4X1, OR4S1, LOC616125, LOC504501, LOC789246, LOC787041, LOC787071, OR4F15, LOC540082, LGR5, MC1R, MC4R, LOC520181, LOC613726, LOC785944, LOC790152, LOC785946, CXCL10, LOC513062, OR9A4, LOC505546, LOC520162, PREX2, LOC789193, LOC504567, OR8G2, LOC509280, LOC516940, LOC788079, LOC524903, LOC532486, OR8K1, LOC618091, LOC617016, LOC100298322, CXCL3, LOC782152, LOC527077, LOC506486, LOC784108, LOC617011, LOC539064, OR4M1, LOC513101, LOC532501, LOC784957, LOC788089, LOC521749, LOC784897, OR14J1, LOC530231, LOC618052, LOC783002, LOC513151, OR2C3, LOC618070, LOC788055, OR5AR1, LOC618064, LOC509267, OR5P3, LOC514235, LOC784925, OR8D4, LOC783205, AREG, LOC101904987, LOC100849008, LOC538744, LOC783203, LOC526335, LOC504623, OR5V1, LOC523060, LOC104968576, OR9G1, RGS9, LOC783210, LOC781968, LOC787428, PROKR2, RGS18, OR4E2, LOC100299372, LOC507882, OR4D5, LOC790274, AGT, LOC787423, LOC509073, LOC510984, LOC526294, LOC526286, LOC530175, OR2V2, LOC532075, LOC100300302, LOC782009, OR10AD1, LOC526276, LOC788476, LOC789288, LOC533983, LOC515414, LOC100299320, OR5D13, LOC619026, LOC787385, OR5AS1, ADGRD1, OR6B1, OR5L2, LOC101904911, LOC616306, LOC619021, OR4D2, OR1K1, LOC530068, LOC527216, LOC527217, LOC513494, LOC511103, LOC789300, OR10J3, LOC617388, RAPGEF2, OR8H3, LOC100299289, LOC532031, LOC788438, OR7A5, LOC523083, PF4, GPR179, LOC100299275, LOC784214, LOC101904323, CCR5, LOC780976, LOC509124, OR9G9, LOC107131158, LOC515482, LOC524282, OR5I1, LOC508980, LOC100301421, LOC785277, LOC517722, OR7G2, LOC781828, LOC784302, LOC788287, LOC526177, LOC101902265, OR1B1, OR7D4, LOC107131149, OR10G2, LOC100140382, OR1J1, LOC528373, LOC504766, LOC528914, OR2B6, LOC786202, LOC783311, LOC100300446, LOC788246, LOC514546, LOC518816, LOC104968488, LOC532208, OR2T2, LOC617297, MC5R, LOC104968493, OR10Z1, LOC523139, LOC513334, LOC513333, LOC527248, OR6K3, LOC507971, LOC528343, GPR180, LOC783313, LOC510112
GO:0005886~plasma membrane	3.5E-46	1,192	*LOC100299084, RALGAPA2, LOC788572, LOC539468, LOC788573, ALDH3A1, LOC782430, STX1A, OR2W1, LOC788552, LOC518442, LOC788554, AKAP5, LOC528807, RAP1GAP2, AFAP1L2, MPHOSPH8, ADCY7, CEP89, ADGRE5, PEAK1, CLDN23, OR6F1, BLVRB, LRPAP1, LOC788587, LTA4H, STK10, BFSP1, LOC107132626, LOC788583, OR2M5, LOC100336852, PACSIN3,*

TMEM184A, DAPP1, LOC508806, CRB1, AZGP1, PALLD, KATNB1, LOC782366, OR52B4, LOC526412, ERAP1, LOC512672, C2CD5, LOC526411, LOC782338, OR52E4, COLQ, TRHDE, LOC100139733, RXFP2, LOC786596, CPNE2, LOC101905743, OR12D2, PLEKHH2, LOC504888, PLCH2, LOC782373, LOC788524, LOC100301231, ENPP1, LOC788512, LOC518464, LOC510931, LOC782555, OR4C16, LOC782554, NCR1, LOC788693, LOC516396, PAQR9, EPB41, EPB42, CD320, OR2W3, TCHP, ENPP6, CDHR4, LOC516409, LOC518561, LOC788675, LOC516414, LOC508766, LOC786512, LOC524304, TMEM119, MGST2, LOC788723, SYT5, GHRHR, LOC617507, LOC506981, LOC615605, OR4C15, LOC530825, LOC100299247, OR2AT4, LOC788704, CELSR1, OR10V1, SNAP23, LOC788713, LOC508785, LOC526508, FGFBP1, GPA33, OR4D11, LOC788626, ITSN1, LOC788633, LOC788634, LOC100301071, BoLA, LOC782475, OR6Y1, ATP11A, IGSF3, LOC516467, LOC510901, MARCKSL1, CEP41, LOC104968790, GABRD, LOC100301104, OR2AE1, C23H6orf25, PIK3C2B, LOC786467, LOC528722, OR10P1, IGF2R, LOC539574, SUSD3, LOC784434, EMP1, OR2M4, PRX, IGFLR1, RAB3GAP2, MYO1B, LOC508626, LOC784858, LOC516274, LOC516273, SLC39A11, LOC614090, KIAA0922, LOC617878, OR2A14, LOC524702, LOC530485, LOC782678, LOC526621, IFNGR1, KIAA0319, OSBPL3, ITPR3, NSFL1C, OR4N5, OR1J2, LOC528515, CARMIL1, LOC522554, LHCGR, OR9Q2, PKP3, USP6NL, OR1E1, LOC507378, GRM7, LOC512973, CDH7, LOC507383, LOC614143, TSC1, OR52W1, IFNGR2, OR6K2, TGFBR2, CSNK2B, LOC100298850, LOC100847301, OR7D2, RAPGEF3, MYO10, LOC786846, LOC522582, OR52B2, LOC784787, LOC100298773, LOC614021, SMPD1, LOC100847281, LOC514864, MCMBP, LOC512948, ADI1, TJP2, GPRC5B, LOC100337063, BAIAP2L2, FYB, PAK1IP1, LOC104969161, OR10T2, LOC100298808, LOC782624, LOC508595, C2CD4C, SLC2A9, GABRA6, LOC788778, LOC508604, OR7G3, LOC522609, LOC539185, BAG3, SLC26A8, LOC526713, LOC100847240, ARHGEF5, LOC539172, LOC508589, SCN8A, LOC100847239, LOC788790, VNN1, LOC514818, LOC790683, PLCG1, LOC100851523, LOC782792, OR1L3, SLC39A3, LOC782797, LOC528422, LOC615808, COPB1, LOC615810, CDHR1, LOC784706, C1QBP, ZHX2, LOC515045, OR10J1, OR2J3, LOC100139830, LOC516132, PPIP5K1, CDH23, LOC615852, ABRA, LOC100336980, CLDN25, PTAFR, OR52H1, LOC526765, MISP, OR10K1, RAB3B, TTF1, ARHGAP10, OR52B6, SH3D19, MTUS1, HSPB1, ANKRD27, MPL, ATP10B, ATP8B4, DGKQ, PARM1, LOC788864, LOC617817, SCN5A, LOC788874, OR2T11, OR8B4, LOC788872, GCA, LOC615901, LOC613909, LOC784652, LOC508420, OR6P1, ADAM12, LOC100336916, OR10A3, SLC51B, LOC510625, JSP.1, LOC508468, HRAS, ADD1, SULF2, LOC524658, OR8S1, OR10A6, MICALL2, DXO, NFASC, UNC5C, HFE, MTFR1, LOC784681, ZDHHC5, NEURL1, OR6C4, OR5A1, LOC100299628, INPP5K, STAMBP, TCAF2, LOC781264, LOC508392, OR4D6, EQTN, LOC785149, LOC618173, ACIN1, OR2A12, LIG4, TULP1, LOC524985, LOC100850909, EPB41L2, ALOX15, LOC618124, ADORA1, LOC510293, LOC782941, LOC618112, ARFGAP2, OR1Q1, OR2H1, ARVCF, OR5B3, SIGLEC15, LOC618140, LOC520835, LOC615941, TDP1, RASSF1, LOC522775, RELL2, CDC42EP3, LOC538966, LOC100337392, UPK1A, UPK1B, LOC507560, LOC514057, CDH20, LOC100299556, EPB41L3, LOC510257, LOC789041, RASGRP4, FCGR2B, ANXA6, GPR37, LOC785082, JMJD6, LOC520938, OR2B11, PROM2, LOC613799, LOC789031, LOC531304, PCDH20, OR6S1, LOC782866, GPC5, PLCE1, KCNN4, SEMA4A, PLXDC1, SYT11, LAMP3, LOC507550, LOC508315, SLC9A4, SYT15, LOC788998, OR4X1, OR4S1, NFKBIA, LOC616125, LOC519071, LOC789246, NDC1, KIRREL3, PIK3IP1, LOC787041, LOC787071, OR4F15, SCN7A, BOLA-NC1,

IFNAR1, IFNAR2, LOC613726, CLEC12B, GUCY2D, GUCY2C, ABCC6, LOC100337265, LOC505546, LOC618010, PREX2, LOC789193, DUOX1, OR8G2, ZACN, LOC507428, LOC516940, LOC524903, LOC532486, NMUR2, LOC618091, FCAR, LOC527077, LOC539064, PANX2, TMEM8A, LOC784957, LOC532501, LOC784897, OR14J1, LOC618052, LOC783002, CARMIL2, LOC618050, OR2C3, PTPN13, LOC618070, SGIP1, LOC618064, ABTB1, OR5P3, PRSS12, LOC514235, LOC618075, OR8D4, LOC784925, LOC789134, LOC510351, CACNA1H, LOC783205, DSG2, LOC100849008, LOC538744, CACNA1B, LOC783203, CACNA1G, UCHL1, OR6C68, OR5V1, DYSF, LOC523060, LOC104968576, RASSF3, MPHOSPH9, LOC783210, SYT12, CATIP, LOC787428, FEN1, PROKR2, LOC521142, RGS18, HYLS1, LOC100299372, UBR2, CD1A, SMPDL3B, LOC787423, PCDHGC3, TMEM8B, C2CD4D, LOC530994, LOC530990, LOC510046, TRIM69, OR6C2, SSNA1, SLCO1A2, LOC789288, SLC24A1, TNFAIP8L3, MME, LOC100299320, LOC787385, OR5D13, S100A12, OR5AS1, SLC11A1, LOC512399, LOC505646, OR5L2, CLEC7A, SPG11, OR6C75, OCLN, LOC616306, OR4D2, DENND4C, LOC527216, LOC527217, INPP5J, LOC789300, RAPGEF2, LOC100299289, TTYH1, LOC523090, SHANK3, FLNB, LOC523083, SPACA4, ERLIN2, MCOLN3, GPR179, LOC100299275, LOC101904323, IFT122, STXBP5L, IL22RA1, NEDD4, SNF8, MMP25, GGT5, OR5I1, LOC100301421, DNAAF1, LRRC45, LOC785277, NOTCH3, OR7G2, NFE2L2, LY9, LOC101902265, OR1B1, OR7D4, KIF14, OR10G2, RILPL1, OR1J1, LOC528914, LOC783311, LOC514546, SCN10A, LOC518816, LOC104968488, MC5R, OR10Z1, COG3, LOC523139, LOC783323, LOC527248, OR6K3, LOC507971, AIP, LOC783313, LOC510112, TRH, OR51E2, TTYH3, LOC523258, TLR9, IL10RB, OR9K2, CORO1B, ABCA3, LOC514434, LOC518869, RGS3, OPALIN, LOC787247, GLDC, TPCN2, LOC512296, CDH13, LOC100301320, STX19, PCDHA13, LOC613390, GPRC5D, OR52L1, LOC531174, LOC510100, LOC785162, OR11L1, DOC2G, RGS11, LOC104968569, LOC507662, LOC104968568, CLEC4D, STXBP2, LOC515619, KCNN3, OR2T33, LOC781758, CCDC62, LOC787584, MST1R, LOC527450, LOC616658, NALCN, GPR142, DSG3, LOC789504, CCDC70, LOC785565, ITIH4, PKD1L3, ACHE, LOC511753, LOC618828, LOC538552, LOC787625, LOC785582, LOC618817, OR52K1, OR52M1, LOC100140748, PAQR5, LOC519492, OR2L13, LOC100300085, LOC787642, F3, EFR3B, LRRC8B, LOC104972581, ANPEP, LOC787659, SLC9A3R2, OR6T1, LOC100295806, OR52N2, KIAA1524, LOC783446, LYVE1, TAS1R2, LOC509895, CDCA5, LOC785623, SH3TC2, NEDD1, AKAP12, LOC787665, SYTL2, LOC515704, LOC785624, MBD3L1, SLCO1B3, LOC781804, LOC511823, LOC521350, CYTH2, OR2A2, LOC527414, LOC787694, LOC785639, LOC527415, PCDH9, TNFRSF19, PLCD4, REPS1, MSH6, LOC785647, OR4C46, DYX1C1, RGS20, ADCY6, ENTPD2, OR7A10, LANCL1, XRCC5, OR5L1, GPR37L1, ATP8A1, LOC783518, FCHO2, LOC616517, EHBP1, OR2A5, ERAP2, OR5D14, LOC614592, LOC783488, LOC785431, ABCA12, TAS1R1, RELL1, AP2M1, PDPN, SLC9A1, XRN1, TMEM198, LOC789690, FAM126A, VAV2, LOC787500, TNKS1BP1, SPG20, CA9, OR4C3, OR8K3, AOC1, LOC787518, GHR, OR10C1, GH1, LOC519616, POLE, OR52I2, DAPK1, PCDH10, TMEM59, LOC787535, OR5M3, SCN4A, OR2D3, LOC783558, LOC787543, PKD2L1, PRKD3, FAT3, GPRIN1, ILDR1, CAPS, PLSCR5, LOC787574, LOC509817, LOC783598, LOC100850276, LOC783597, FAM155A, BCAM, LOC789766, LOC521676, LOC785811, IGSF8, OR51I1, ESAM, LOC783616, OR13C8, C14H8orf37, LOC511509, SPHK1, CAMK1G, LOC100138866, OR5K1, LOC618593, LOC519294, LOC515887, JAML, PDZK1, TTC7B, OR1G1, OR56B4, OR2G3, LOC781446, OR13C3, RRP8, LOC787898, TMEM8C, LOC785848, OR4N2,

LOC789812, LOC100299808, LOC789815, STAU1, PTPRR, CD101, WBP4, OR12D3, LOC783671, LOC789817, LOC100298119, LOC616942, AGRN, PLA2R1, LOC541022, LOC618554, LOC787932, OR2AG2, OR2AP1, PLA2G3, OR1D5, OR51D1CD46, LOC511570, LOC101906611, LOC511569, CCDC8, LOC509641, UPK3A, MIP, IL16, PMP22, SYTL1, SCN1A, NRCAM, LOC506317, LOC104970118, AMER3, KNTC1, SLCO2B1, OR2V1, LOC781509, LOC521645, ACOX1, LOC618523, SORT1, LOC785910, LNPEP, LOC519176, ITPR1, PTPRO, LOC785914, ARHGAP33, OR5AU1, HN1L, LOC787945, LOC787946, CATSPER1, LOC785899, LOC517252, LOC785903, LOC529511, CASR, ATP10D, CNST, LOC529518, CDC42EP5, PCDHB1, LOC101902679, AXIN2, LOC785683, LOC100138976, MADD, LOC506121, IFNLR1, ANXA2, LOC523768, LOC523769, LOC511657, LOC785723, LY6D, CDH3, LVRN, CDH2, LOC614895, CD44, LOC785712, OR6N1, LOC511678, LOC523753, PSEN1, CARMIL3, LOC789943, LOC789936, UPF3A, LOC506202, OR4A16, LOC100298103, LOC509526, LANCL2, FAM129A, LOC618675, OR6C76, LOC509525, ZP2, SMURF1, XPR1, ZP4, LOC789957, ZP3, LOC618660, LOC618660, LOC787779, LOC616716, CELSR2, LOC509510, OR4C6, LOC513914, MCC, LOC616705, TAS2R38, PLIN2, TBC1D10B, LOC783843, AOC3, LOC787830, NBEA, ALPI, BST1, LOC101904538, OR4C13, LOC527779, LOC616755, OR4K5, HSP90B1, PPP4C, CACNA1A, LOC523680, LOC787816, TMPRSS6, LRP4, LOC513884, LOC532291, OR51Q1, LOC506533, GUCA1B, OSBPL6, WDR19, LOC513175, OR2G2, LOC515090, LOC509369, VAV3, RIF1, TMEM173, LOC506549, APLP1, LOC615014, ADCY4, LOC615009, DAGLB, OR13A1, LOC504344, LOC509323, FSHR, LOC782261, ENPEP, GPC1, CDH9, PLSCR1, FASN, LY6G6F, OR10K2, SMAD3, LOC783884, LOC526047, SURF2, LOC782255, KCNN1, LOC617122, LOC786149, IGSF11, LOC782288, LOC532238, LOC782301, RGS22, LOC782296, OR8D2, DSC1, LOC527918, LOC104973083, ECE1, DSG1, LOC788210, DSC2, LOC525964, STX3, SLC4A1AP, CDH26, PAK1, ANXA3, RXFP1, LOC786133, LOC107131130, OR2T4, SYT17, CELSR3, GCC1, DNAJB4, CTNND1, ELL, OR2D2, WLS, NDRG1, OR2T12, RFFL, LYPD4, LOC504501, OR6C1, LOC540082, LOC788027, PCDH15, LOC532436, MDH2, SLC9A9, MAPT, MC1R, PCDHB14, MC4R, LOC790152, LOC785944, PCDHA6, LOC785946, LOC513062, MAP4, LOC787991, LOC520162, SPATA20, GABRR3, SEMA4D, LOC504567, TMPRSS5, MELK, LOC509280, LOC788079, RHBDF2, RIN1, LRRFIP1, OR8K1, LOC504551, RAMP1, LOC617016, LOC100298322, ALDH3B1, LOC782152, LOC506486, LOC784108, CLDN18, LOC617011, LOC513101, OR4M1, LOC788089, ALOX15B, LOC521749, LOC530231, CLIC4, OR10J3, LOC513151, LYPD3, MLKL, CDH6, LOC107133172, LOC615040, LOC788055, OR5AR1, LOC509267, LOC506452, PCDH12, LOC615276, LOC101904987, LOC526335, LOC504623, RELN, PKN2, RGS9, STAT2, PKP2, LOC781968, OR4E2, LOC784187, PARD6B, OR4D5, LOC615284, CLDN9, LOC790274, PKN1, BAMBI, LOC509073, ANO8, LOC510984, PRLR, LOC526294, SELPLG, LOC788372, FNIP2, DSG4, LOC526286, PCDHGA2, LOC530175, OR2V2, LOC532075, LOC100300302, LOC782009, OR10AD1, LOC617417, LOC526276, SLC17A5, LOC100298658, LOC788357, LOC788476, ASTL, LOC533983, LOC515414, LOC619026, OR51S1, IL10RA, OR51F2, ADGRD1, OR6B1, ACPP, LOC101904911, LOC619021, OR1K1, LOC530068, LOC782046, OR52E8, LOC511103, PKP4, PIGR, OR52K2, OR10J3, NPHS1, LOC617388, CDC42EP1, NMT1, OR8H3, LOC532031, LOC788438, OR7A5, DRD5, LOC509128, CPNE7, LOC517799, FURIN, LOC509124, OR9G9, ORC1, SERINC5, LOC107131158, LOC524282, LOC515482, LOC508980, SYT6, UPK2, LOC517722, XPC, LOC781828, LOC788285, LOC788287, LOC526177,

			LOC788258, PIK3C2G, LOC788263, LOC107131149, LOC528373, LOC100140382, SCN11A, LOC504766, OR2B6, LOC786202, SEMA4F, LOC788242, STXBP5, LOC100300446, LOC788246, LOC532208, LOC617302, OR2T2, TLR4, CDHR3, LOC617297, RPH3AL, ACLY, LOC513334, LOC513333, LOC528343, CD163, BRCA1, PCDH8, GRIN2A, LOC515540, ENTPD8, LOC509025, LOC788323, VSTM4, FANCG, OR5D18, SELE, SYT4, DCHS1, ADGRG3, OR52D1, LOC100300488, OR7A17, SOD1, CACNA1E, LOC100848076, OR51A7, LOC524160
GO:0007608~sensory perception of smell	8.6E-40	171	*LOC508626, LOC789766, LOC784858, LOC532291, LOC516274, LOC101904987, LOC516273, LOC100849008, LOC100299628, LOC614090, LOC783203, OR5V1, OR2T33, OR2A14, OR2W1, LOC787428, OR13C8, LOC616658, LOC785149, LOC100299372, LOC506549, OR2A12, LOC785565, OR2G3, LOC100850909, LOC522554, OR13C3, LOC782261, LOC618828, LOC787898, LOC510293, OR2H1, LOC789812, LOC789815, LOC526286, LOC100300085, OR2L13, OR2V2, LOC618140, LOC107132626, OR10AD1, LOC789817, OR2M5, LOC104972581, LOC515414, LOC618554, LOC782288, LOC787932, LOC100299320, LOC100299556, OR2AG2, LOC782366, MKKS, LOC782301, LOC789041, LOC784787, LOC614021, OR6B1, LOC101904911, LOC619021, LOC527217, LOC515704, OR2B11, LOC520938, LOC613799, LOC511103, LOC789031, LOC781804, LOC104970118, LOC104969161, OR2V1, LOC527414, LOC100299289, LOC527415, LOC523083, LOC787946, OR2T4, NAV2, LOC100299275, CNGB1, LOC101904323, PDE1C, LOC790683, OR2D2, OR2T12, LOC782554, LOC529518, LOC785277, LOC517722, OR2W3, LOC783518, LOC615810, LOC616517, LOC788287, OR2A5, LOC516409, LOC788675, LOC614592, LOC101902265, LOC785431, OR10G2, OR2J3, LOC528373, LOC504766, CNGA4, OR2B6, LOC783311, LOC615852, LOC100336980, LOC518816, LOC506981, LOC532208, LOC787500, OR2T2, LOC614895, LOC617297, LOC526765, OR6N1, OR10Z1, LOC523753, NXNL2, LOC789943, OR10C1, LOC507971, LOC528343, LOC506202, LOC100298103, BBS4, LOC509280, LOC788079, LOC524903, OR2T11, LOC616716, OR2D3, LOC514434, LOC100301071, LOC788323, LOC782475, LOC787543, LOC506486, LOC784108, LOC508420, LOC784652, LOC616705, OR6Y1, LOC530231, LOC618052, OR10A3, LOC783002, LOC100301320, OR2AE1, LOC527779, OR2C3, LOC613390, OR10P1, LOC539574, OR10A6, LOC618070, LOC788055, LOC784434, OR2M4, LOC618064, LOC510100, UBR3, GJB4, LOC523680, LOC787816, LOC514235, LOC784681, TTC8, LOC524160*
GO:0050907~detection of chemical stimulus involved in sensory perception	1.3E-34	134	*LOC785811, LOC788572, LOC526335, LOC523060, LOC781758, LOC524702, LOC781264, LOC530485, LOC527450, OR4D6, LOC781968, OR4E2, LOC511509, LOC519294, LOC618593, LOC515887, OR4D5, OR1G1, LOC504344, OR4N5, LOC781446, LOC538552, OR1E1, LOC618124, LOC785582, LOC509073, LOC618112, LOC512973, LOC526294, OR1Q1, OR4N2, LOC507383, LOC788587, OR10K2, LOC520835, OR6K2, OR12D3, LOC100300302, LOC508806, LOC619026, OR1D5, LOC785082, LOC616306, LOC785623, OR4D2, OR1K1, LOC530068, LOC100337063, OR10J3, LOC617388, LOC531304, LOC786596, OR2A2, OR10T2, LOC521645, LOC618523, LOC785639, LOC788778, LOC785910, OR12D2, LOC522609, LOC785647, OR4C46, LOC100847240, LOC788790, LOC508315, LOC517252, LOC514818, LOC107131158, LOC515482, OR4C16, OR4X1, OR1L3, OR4S1, LOC788693, LOC785683, LOC526177, LOC787071, LOC784706, LOC508766, LOC506121, OR4F15, OR10J1, LOC100140382, LOC788723, LOC523768, LOC789690, LOC785723, LOC104968488, OR4C15, LOC615605, OR2AT4, LOC785712, OR4C3, LOC790152,*

			LOC788704, LOC527248, LOC788713, LOC508785, OR6K3, LOC789193, LOC510112, LOC504567, OR4A16, LOC618675, OR4D11, LOC515540, LOC523258, LOC532486, LOC100298322, LOC509025, LOC518869, LOC615901, LOC787247, OR4C6, OR4M1, LOC516467, LOC788089, LOC784957, LOC532501, LOC521749, LOC510901, LOC104968790, LOC508468, LOC513151, OR4C13, LOC786467, OR4K5, OR8S1, LOC100848076, LOC509817, LOC513884, LOC784925, LOC104968568
GO:0016021~integral component of membrane	3.2E-34	1,958	*NDUFB5, LEPR, LOC788573, LOC104975794, ALDH3A1, STX1A, LRRN4CL, LOC788552, LOC788554, ADAM15, ADAM9, EVA1C, ADCY7, ABCA10, LRFN2, OXA1L, GALNT9, COX7A1, LOC788587, LRRC66, C18H19orf12, SLC25A13, LOC788583, MARC2, TMEM176B, CRB1, LOC508806, TMEM236, SORL1, MFSD14A, OR52B4, RTN2, RTN3, ABCG8, RTN4, PTPRS, SLC22A18, FAM210B, CHST11, MARVELD3, FNDC3A, LOC504888, CLMP, IMMT, LOC788524, FNDC3B, LOC100301231, LOC788512, ADGRF1, LOC788693, NCR1, PRF1, LOC516396, OCA2, OR2W3, CD320, CLCN6, CLEC12A, LOC516409, LOC788675, LOC508766, LOC516414, SYNE3, LOC524304, TMC4, TYW1, MGST2, LOC788723, GCNT1, GHRHR, LOC615605, KIAA0319L, SEMA5B, OR10V1, LOC788713, TM4SF1, OR4D11, PKD1, LOC788626, KREMEN1, LOC788633, GP1BA, LOC788634, TMIGD2, AADACL2, LOC100301071, BoLA, TMEM220, ABCA6, MPDU1, ITFG1, OR6Y1, LOC516467, TSR3, ICAM3, GABRD, LOC100301104, LOC528722, OR10P1, IGF2R, SUSD3, TAAR8, SLC27A5, TMED4, ICAM1, NDUFC2, SLC35F2, TMEM61, LOC784858, LOC508626, LOC516273, MANSC1, SLC39A11, KIAA0922, RIC3, OR2A14, MEGF8, CREB3L3, KCNH1, KCNA4, KLRA1, TCEB3, KCNQ3, KIAA0319, ITPR3, CD276, LOC528515, OR4N5, OR1E1, TGFBR3, PILRA, CDH7, TSC1, SLC44A4, PPP1R3A, MRVI1, OR6K2, LOC100847301, HSD17B7, PQLC2, C10H14orf37, LOC784787, LOC100847281, KRTCAP3, LOC512948, CDCP2, EBPL, GPRC5B, NPY2R, SDK1, LMF1, LOC508595, LOC788778, CNTN6, LOC100847240, LOC524771, LOC788790, LOC100851523, OR1L3, LRRC24, SLC39A3, GALNTL6, LOC528422, LOC615810, LOC784706, ITGA3, ULBP3, OR2J3, TMPRSS15, OSBPL8, LOC516132, TMEM136, CDH23, NRADD, PDE3B, CMTM2, LOC516101, SLITRK1, MPL, SEZ6L, LOC788864, MRGPRF, ERMARD, LOC615901, TEX38, LOC100847151, LOC784652, LOC508420, OR6P1, CYB561A3, TMED6, LOC508468, GOLT1A, LOC524658, OR8S1, TRPC2, TRPC4, UNC5C, SLC2A11, HFE, LOC784681, ZDHHC5, FCAMR, OR6C4, OR5A1, ADGRF5, MGC138914, KCNAB3, LOC508392, OR4D6, EQTN, SLC27A6, LOC785149, BTN1A1, BCL2L1, LOC524985, ADAM10, UNC5B, GDPD5, MARC1, LOC520835, VLDLR, LOC615941, TAP1, VCAM1, RELL2, UPK1A, LOC512150, SLMAP, NFAM1, LOC512149, XKR5, FCGR1A, LOC789041, GPR135, GPR37, LOC785080, P2RY11, SLC4A10, CYP2D14, OR2B11, LOC520938, PROM2, LOC789031, FUT10, PCDH20, ZDHHC23, OR6S1, DISP3, PLXDC1, ADGRD2, ZDHHC12, SYT15, LOC616125, NDUFC1, LOC789246, NDC1, PIK3IP1, CLEC1A, PCNX3, GRAMD1C, FA2H, PTPRF, TMC6, OR4F15, TMEM266, BOSTAUV1R417, TAS2R16, TAS2R46, BTNL2, CLCN2, TAS2R10, BOTA-T2R10B, FAM234B, IFNAR2, CLEC12B, GUCY2D, GUCY2C, TMEM94, GALNT1, LOC505546, SFT2D1, IL1RL1, TREM1, LOC789193, CD300A, OR8G2, ZACN, RNF5, LOC516940, SLC5A1, PIEZO1, LOC524903, LOC532486, SCARB1, NFXL1, LOC784957, LOC532501, LOC784897, TAS2R3, OR14J1, T2R10C, T2R65A, TAS2R42, T2R12, TMEM40, MTX1, GPR160, TAS2R1, NDST4, TMEM82, OR8D4, LOC784925, LOC789134, DSG2, PTPRB, OR5V1, SEC14L3, DYSF, ERBB2, LOC508153, NUP210, SYT12, SDC4, LOC521142, DISP2, LOC785379, LRTM2, CNTNAP4, TMEM89, B3GALNT2, KEL, PCDHGC3, LOC789334, PNPLA3, SLCO1A2, MARCO, LOC789288, MME, OR5D13,*

SLC11A1, LOC512399, LOC505646, TAAR9, OR5L2, CLEC7A, OCLN, LOC616306, BOLA-DQB, BOLA-DQA5, TMEM104, SLC34A1, LOC789300, TMPRSS13, GOLGB1, ERLIN2, ABCB5, TAPBPL, GRAMD3, LOC101904323, SLC43A3, PREB, STXBP5L, IL22RA1, SLC15A1, TRPM2, BOLA-DQA2, EMCN, PODXL, BOLA-DRB3, LOC100301421, SLC25A38, LOC785277, NOTCH3, FREM1, TMEM254, LY9, NAT14, EVI2A, BOSTAUV1R424, PIGN, LOC528914, LOC785236, BOLA-DYB, BOLA-DYA, SLC9B2, FKBP8, MC5R, KLHL31, C25H16orf54, LOC507971, SLC25A11, PCNX2, VTI1B, TLR9, LOC521081, OR9K2, ABCA3, OPALIN, PTPRD, TPCN2, LOC512296, LOC100301320, TMEM206, LRP3, LMTK2, FAM173B, OR52L1, TM2D2, LOC785162, NDFIP2, DERL1, GPR3, KCNN3, LOC781758, MST1R, CEACAM1, MEP1B, RNF112, CCPG1, GPR142, LOC509854, DSG3, LOC789504, TRPM3, LOC785565, IL4R, SLC39A7, LOC618828, LOC538552, LOC618817, RTL1, LOC785582, HSD17B12, OR52K1, OR52M1, GHITM, F3, PTCRA, GOLIM4, LRRC8B, ANPEP, GPR161, TMEM247, MPV17L2, FZD9, LOC509895, TAS1R2, LYVE1, LOC785623, SLCO1B3, MGC137098, LOC521350, SLC16A4, AIFM2, PCDH9, TNFRSF19, KCNC4, LOC785647, SAMM50, PIGW, FAM171B, ADCY6, ENTPD2, BOSTAUV1R421, TMEM79, ERAP2, KCNG4, LOC614592, OR5D14, PIGZ, LYSMD4, LOC785431, RELL1, BSCL2, TP53I13, PDPN, LOC789690, TMEM198, SOAT2, TMEM63A, SLC22A10, KIAA1024L, SDC1, SV2C, SLC36A3, OR8K3, B3GNT6, TIMM21, TMEM131, GHR, SLC38A9, CLCN7, OR52I2, KLRK1, LOC100297846, OR5M3, SLC25A45, OR2D3, LOC785500, TSPAN16, IGSF9B, FAIM2, ILDR1, DBH, FRRS1L, DDR1, CD5, LOC509817, LOC100850276, LOC789766, LOC521676, SLC29A4, LOC785811, ASPHD1, OR13C8, CD19, NIPAL2, CD96, SLC17A8, LOC100138866, LOC618593, JAML, SCD, OR1G1, OR2G3, LOC781446, OR13C3, TMEM8C, GRAMD1B, STYK1, LOC785848, LOC789812, TYR, LOC789815, MGC127055, CD101, SEC22B, MEP1A, GPNMB, LOC789817, LOC100298119, LOC618565, NUP210L, INPP5B, EVC2, PVRIG, MOGAT2, KIT, PTPRK, LOC541022, LOC618554, DSE, OR2AG2, MGC139164, LOC101906604, CD46, LOC101906611, LOC509641, HRH3, PMP22, MOG, NRCAM, CES5A, LOC781509, LOC521645, LOC618523, LOC785910, C2CD2, SLC15A3, PTPRO, LOC785914, VSTM1, KDELR3, LOC514011, MAVS, SLC25A26, TMEM242, PAM, LOC785899, TMEM255B, LOC517252, LOC785903, LOC618737, ADGRG6, LOC529511, CNST, MAN2A1, CPT1B, LOC529518, PCDHB1, LOC101902679, LOC785683, LOC100138976, FNDC9, TMEM161A, CD3E, LOC101902670, DSB, KIAA1024, MADD, IGSF9, OTOP1, LOC785723, SLC1A2, CDH3, CDH2, TAAR2, LOC614895, CD44, LOC785712, OR6N1, TMC8, ERMP1, LOC789943, LOC789936, DUOX2, SIGLEC10, OR4A16, LOC100298103, LOC509526, LOC618675, ZP2, TMEM260, SLC24A5, CLRN3, LOC525599, ZP4, ZP3, LOC618660, AGER, IGSF23, LOC509506, ATP13A1, LOC781439, LOC509510, OR4C6, LOC513914, SOAT1, LOC614821, ALPI, TMC5, OR4C13, SLC38A10, SCARA3, TMPRSS6, ERGIC3, LRP4, LOC513884, FAM174B, LOC618633, THSD1, LOC513175, TMEM130, SLC16A10, LOC509369, TAAR5, SI, GGCX, TMEM173, DNAJC18, APLP1, LOC615014, LOC615009, DAGLB, LOC504344, ALG9, LOC509323, LOC782261, FUT2, TBXAS1, ENPEP, CDH9, CMTM8, PLSCR1, LOC526047, TAPBP, TMEM156, LOC782251, LOC782255, FMO3, LOC100138004, KCNN1, FAM189A2, LOC782288, EXT2, LOC782301, LOC782296, OR8D2, DSC1, MERTK, ECE2, ECE1, DSG1, DSC2, LOC525964, CHL1, SEC14L5, STX3, FMO2, LOC786133, CSF1, TEX2, TMEM25, FAM173A, OR2D2, OR2T12, SLC29A3, LOC504501, TMPRSS3, CHST12, PKHD1, NTN5, LOC509184, WBP1, MSR1, UNC5CL, MAPT, MC1R, PCDHB14, MC4R, SLAMF7, FADS2, LOC790152, LOC101902937, LOC785946, DCST1, DCST2, LOC504567, LDLR, VN1R1,

LOC509280, LOC504551, LOC100298322, UGT2B10, LOC782152, LOC513101, OR4M1, LOC521749, LOC530231, DNAJB12, OR10J3, LOC513151, ECEL1, CDH6, LOC107133172, LOC615040, THSD7A, LOC509267, CTAGE5, LOC615051, SECTM1A, ADGRF4, LOC615276, LOC526335, LOC504623, SLC41A1, IMPG2, JAM3, LOC781968, OR4E2, C3H1orf210, TMEM125, VSIG10L, OR4D5, LOC781977, LOC615284, HEPACAM2, GPR62, LOC790274, YIPF7, LOC615247, CDKAL1, LOC781988, PTPRQ, KCNC1, LOC509073, PRLR, KCNS3, CDH15, DSG4, LOC526286, OR2V2, LOC530175, CD244, LOC782009, DAG1, SIDT2, LOC100298658, SLC17A5, LOC526276, LOC533983, LOC619026, PGAP1, OR51S1, OR51F2, CLEC5A, IL10RA, OR6B1, LOC530077, ACPP, LOC522174, LOC619021, CD302, OR1K1, LOC530068, LOC782046, OR52E8, OR52K2, PIGR, OR10J3, SNX19, LOC509128, FURIN, LOC517799, LOC509124, OR9G9, SERINC5, GJA10, LOC508980, COLEC12, SYT6, SIRPB1, UPK2, LOC517722, LOC781828, GPR63, SMIM6, SMIM5, ABCB1, LOC504766, SEMA4F, LOC786202, STXBP5, GALNT5, MS4A5, TMEM175, OR2T2, CDHR3, MEGF9, KTN1, LOC513334, LOC513333, IL11RA, GPR180, CD163, PCDH8, ABCB9, C29H11orf24, TLR2, TLR3, PMEL, GLT6D1, SELP, SELE, SEC1, SMPD2, LOC513384, ADGRG3, LOC101907000, SLAMF1, LOC100299084, LOC539468, LOC100336869, LOC782430, OR2W1, ADGRE3, LOC518442, DNAH3, TOR1AIP2, CLDN23, TMEM268, OR2M5, LOC100336852, CCDC155, CYP4F2, LOC782366, LETMD1, ARL10, LOC526412, GRIA2, LOC526411, LOC782338, OR52E4, TRHDE, LOC100139733, ITPRIPL1, RXFP2, ADGRL4, SLC30A9, ABHD1, OR12D2, LOC782373, PI16, TMEM86B, LOC518464, SFT2D2, LOC510931, LOC782554, OR4C16, PAQR9, LOC782545, ILVBL, LOC518561, QPCTL, TOR1AIP1, LOC786512, MXRA8, SYNE1, CARD19, SYT5, LCAT, LOC506981, LOC617507, KLRD1, OR4C15, LOC530825, LOC100299247, OR2AT4, THSD7B, CELSR1, LOC526508, TMPRSS9, GPA33, SPTLC1, DCAF17, RHOBTB3, LOC782475, CNTFR, NPC1L1, DHCR7, IGSF3, KLRJ1, TRIM13, LOC510901, LOC104968790, OR2AE1, C23H6orf25, LOC100299180, LOC539574, LGR6, EMP1, LOC614090, LOC617878, SOGA3, LOC530485, LOC782678, TMED8, LOC526621, LOC522554, OR9Q2, OTOP3, SGCB, LOC782699, LOC507378, GRM7, FCER2, LOC614143, OR52W1, IFNGR2, BLA-DQB, PLD6, CANX, CD80, FER1L5, TGFBR2, LOC518161, CR2, LOC100298850, OR7D2, DEGS1, LOC786846, LOC522582, LOC100298773, OR52B2, LOC514864, DCBLD1, CD4, KCNS2, ASIC2, CYP3A5, LOC100337063, MYADML2, MEGF10, LOC104969161, LOC100298808, OR10T2, LOC782624, GABRA6, SLC2A9, OR7G3, FLVCR2, CLEC10A, LOC539185, ACSBG2, SLC15A4, GPR156, IL21R, LOC526713, FCGR2A, LOC539172, GSG1, LOC514818, WBP1L, LOC790683, TMEM263, LOC782792, TNFSF12, LOC782797, AADACL3, PIEZO2, TIMMDC1, IL1RAP, LOC515045, OR10J1, C15H11orf87, IL31RA, LY75, TLR6, LOC100336980, LRIG2, CLDN25, PTAFR, OR52H1, OR10K1, ULBP11, LOC786796, SIRPA, TSPAN33, ATP10B, GALNT3, PARM1, LOC617817, CLCN3, OR2T11, EGFR, CLEC6A, LOC613909, ADGRA1, SLC9A8, MOGAT1, PDZK1IP1, ADAM12, GGT6, LOC100336916, ABCB4, ATL1, SLC51B, JSP.1, LOC510625, GCNT2, MFSD10, OR10A6, LMF2, GYPA, SLAMF9, LOC100299628, TM4SF4, LOC613867, CLCA1, LOC618173, LOC782957, DSEL, CSMD2, ACSL5, TRPM5, LOC618124, LOC782941, LOC510293, LOC618112, OR1Q1, METTL7A, OR2H1, TMEM144, TMEM19, OR5B3, SIGLEC15, RNFT2, LOC618140, MFSD9, LOC538966, LOC100337392, F2RL3, HIGD1B, LOC613822, LOC514057, LOC507560, CDH20, LOC100299556, NDUFB4, ANXA6, LOC613799, COX7A2, LOC531304, LOC782866, SLC17A1, SEMA4A, KCNN4, NFE2L1, TMC3, LAMP3, ZDHHC18, MFN2, FYCO1, SIGMAR1, OR4X1, OR4S1, SRD5A2, LOC519071, KIRREL3, CHRNA6,

LOC787041, LOC514257, CLCA4, TMED3, ADGRF2, DNAJC15, ADTRP, TMEM138, AMIGO2, AGPAT4, BOLA-NC1, FZD3, MS4A18, LOC613726, ABCC6, LOC100337265, LOC100139049, FAM134A, LOC618010, FAM159A, DUOX1, LAYN, LOC100848815, LOC507428, SLC25A36, PARL, NMUR2, LOC527077, TMEM201, LOC539064, SMPD3, PANX2, TMEM14A, PROM1, B4GAT1, LOC783002, ILDR2, LOC618050, DPY19L4, OR2C3, AGMO, TMEM184B, LOC618070, LOC618064, GALNT15, OR5P3, LOC514235, LOC618076, LOC618075, LOC510351, LOC783205, AREG, LINGO3, GJC2, LOC100849008, LOC538744, B3GNT8, LOC783203, SLC45A2, LOC523060, LOC104968576, BCL2L10, DSCAML1, FAM171A1, LOC783210, LOC787428, PDGFRB, SYNGR3, CD7, LOC100299372, DRAM2, KCNA10, CLSTN2, LOC787423, CORIN, LOC530994, LOC530990, DUOXA2, LOC510046, GALNT11, NDUFAF6, RNF122, OR6C2, LOC100299320, LOC787385, AMN, IL2RB, OR5AS1, STX8, OR6C75, OR4D2, LOC527216, SPAG4, LOC100299289, MGC133647, ROBO2, ACE3, LOC523083, SLC23A3, GPR179, LOC100299275, FAM205C, POM121C, CTNS, CLSTN3, ADGRG5, MMP25, TMEM116, IL9R, TM7SF3, OR7G2, TMEM71, SLC16A13, OR1B1, SYNGR1, OR1J1, LOC783311, LOC514546, PLLP, CMTM5, LOC104968488, CYP4V2, KIRREL2, OR10Z1, TMEM207, LOC523139, LOC783323, LOC527248, KLRG1, LOC783313, LOC531152, LOC510112, SUSD1, OR51E2, VSIG10, LOC523258, TMEM92, CNR1, TMEM134, NOTCH1, LOC514434, LOC518869, TMEM215, LOC787247, GPR89A, LRRN2, CLSTN1, B3GNT7, ABCB11, LOC100337457, MS4A14, STX19, PNPLA6, PCDHA13, RHBDD3, LOC613390, GPRC5D, FOXRED1, LOC510100, OR11L1, GALNT10, LOC507662, LOC104968568, TECRL, CD151, CLEC4D, SLC5A4, ABCA7, LOC515619, OR2T33, SPNS1, LOC783380, DLL3, LOC787584, GYPB, MAN2A2, CCDC149, LOC527450, LOC616658, STOML1, RHBDD2, LOC511753, TARP, SLC15A5, SFXN3, LOC787625, LOC100140748, PAQR5, LOC519492, LOC100300085, LOC787642, CCDC90B, NCLN, SIGIRR, LOC104972581, LOC787659, OR6T1, LOC100295806, LOC527385, ACSL6, OR52N2, LOC783446, CALCR, ASTN1, SLITRK3, FER1L6, GBP4, LOC787665, SYNE2, LOC511823, HSD17B2, LOC527414, LOC787694, FIBCD1, KLRG2, EXTL2, LOC100295750, NOTCH4, POPDC2, LOC101904668, OR5L1, GPR37L1, CHST13, TMEM143, ATP8A1, LOC783518, LOC540544, LOC616517, CYP4A11, PLEKHB1, PTPRG, ADGRG7, LOC100140640, TAS1R1, SLC9A1, GPR61, TUSC5, LOC787500, ADAMDEC1, FREM2, BOLA, RNF121, LOC787518, OR10C1, LOC519616, LRRC3B, FFAR3, PCDH10, LOC787535, ZDHHC2, SLC38A11, LOC783558, LOC787543, FAT3, GALNT8, IKBIP, LOC787574, LOC100295883, AADAC, LOC783597, FAM155A, BCAM, TM6SF2, FUT4, OR51I1, LOC511498, ESAM, ADAM3A, LOC783616, CSF3R, LOC511509, CERS6, EVC, LOC519294, LOC515887, OR56B4, GPR151, LOC787898, LRP6, OR4N2, LOC100299808, HSD17B3, TMEM221, PTPRR, ABCB8, OR12D3, TMEM211, LOC616942, CMTM4, SLC37A1, LOC787932, OR2AP1, OR51D1, GSG1L2, PIGV, FCMR, LOC511570, TRPM1, NAALAD2, MFSD8, LOC104970118, OR2V1, SLCO2B1, SORT1, ITPR1, OR5AU1, LOC787945, LOC787946, HGSNAT, PTPRZ1, CH25H, SLC34A3, LOC101904614, TSPAN18, OTOP2, LOC506121, KMO, IFNLR1, LOC523768, LOC511657, LOC523769, LMAN1, LOC511678, LOC523753, PSEN1, SGCG, CXCL16, LOC506202, HEG1, DAGLA, XPR1, CYP4A22, TMEM132B, LOC616716, LOC787779, RPRML, TPBG, LOC616705, TAS2R38, PRRG4, GPR162, LOC783843, LOC787830, LRRC52, LOC527779, LOC616755, UGT2A1, OMA1, LOC787816, LOC523680, GLTPD2, ADGRL3, RNF180, BOLA-DRA, C8H9orf135, LOC532291, TMEM62, OR51Q1, LOC515090, OR2G2, LOC506549, UNC5D, OR13A1, KIAA1549L, NCAM2, PXMP4, LY6G6F, LOC783884, WBSCR28, LOC617122,

IGSF11, CORO7, LOC532238, PIGO, TMEM5, LOC100140174, XYLT2, LOC100847738, LOC527918, LOC104973083, FAM205A, LOC788210, EXTL3, TMTC4, LOC100140223, CMTM7, CDH26, PTPRU, RXFP1, BVES, LOC107131130, OR2T4, NIPAL4, CIST1, SERINC1, TMEM74, WLS, LHFPL3, OR6C1, FAM189B, LOC540082, LOC788027, LOC532436, GPR84, APMAP, SLC9A9, PCDHA6, LOC520162, ABCC12, GABRR3, GK2, TMEM132A, LOC788079, TREM2, RAMP1, LOC617016, LOC506486, CLDN18, LOC784108, LOC788089, COX10, REV1, SLC28A1, NDST2, SLC12A2, LOC506452, MCHR1, ADAM1B, LOC100296421, PCDH12, LOC101904987, P2RY6, CXCR6, UPK3BL, EVI2B, LOC515333, TMEM108, TCTN3, LDLRAD1, CLDN9, ISLR2, ANO8, LOC510984, LOC788372, PCDHGA2, ISG12B, LOC532075, AKAP1, CD207, LOC100300302, CLEC4F, LOC617417, TMEM139, LOC788357, LOC788476, LOC515414, KCNV2, ADGRD1, KIR2DS1, RNF170, LOC101904911, LOC100336589, NIPAL1, LOC511103, ABCA13, SKINT1, SLC46A1, NPHS1, LOC617388, CCR6, LOC532031, LOC788438, PSME4, MGST1, OR7A5, LRRC19, GPAT3, LOC524282, LOC788285, LOC788287, RTP2, AMIGO3, TMEM237, LOC788258, LOC539818, NRM, LOC788263, LOC107131149, LOC528373, LOC100140382, OR2B6, LOC788242, LOC100300446, SLC47A2, LOC788246, LOC100300442, MFSD7, LOC532208, LOC617297, OTOF, SLC25A48, TLR10, CLRN2, SEMA5A, NKG2C, LOC528343, LRP10, TREML2, RYR3, GRIN2A, LOC515540, GJA3, BDKRB1, ENTPD8, LOC788323, CACHD1, VSTM4, OR5D18, TMEM132E, UMODL1, LOC100336507, SYT4, MCTP2, OR52D1, LOC100300488, LOC100848076, OR51A7, TAS2R40, NPBWR1, LOC788572, MOGAT2, LOC506533, ABCC5, LOC781264, SORCS1, NALCN, FUT6, SLC30A6, LRTM1, LOC528807, OR2A12, MAN1C1, TMIGD3, ADCY4, ADGRE5, TM9SF1, LOC100850909, ATP6V0A4, FSHR, SLC17A4, OR6F1, ADORA1, POMT2, SUSD6, OR2L13, OR10K2, CD180, LOC107132626, LOC522775, UPK1B, LOC100140261, TMEM184A, PRRT3, TMPO, NXPE2, LOC786149, TMBIM1B, MBOAT1, LOC510257, FCGR2B, RNF182, GAL3ST2, LOC785082, LOC785618, LOC512672, GJA9, SEL1L, LOC507581, LOC785624, LOC515704, SLC39A13, LOC781804, GRAMD1A, LOC786596, OR2A2, BOSTAUV1R420, LOC785639, LOC527415, TMEM225, CASC4, DERL3, LOC101905743, TMC7, ROS1, NOX5, CYB5R2, OR4C46, CELSR3, TIMD4, BOSTAUV1R419, LOC507550, CNGB1, TMCO4, LOC508315, ELFN1, LOC101905757, SLC9A4, ENPP1, LOC788998, OR7A10, LOC782555, IL12RB1, PCNX1, FAM234A, FOLH1B, CSPG5, MYOF, ZAN, SLC16A6, CDHR4, OR2A5, LOC787071, MSMO1, MGAT4A, LOC783488, CEACAM20, ABCA12, RTN1, CRB2, PCDH15, TMEM119, LSMEM2, MPZL3, BCAR4, TLR5, IFNAR1, FMO1, CA9, OR4C3, VKORC1, LOC529196, LOC785944, ASPH, LOC788704, LOC513062, STX16, LOC787991, TMEM50A, LOC508785, SLC13A5, FAM187B, MUSK, HCRTR1, TMPRSS5, GPR183, SDR42E1, C2CD2L, RHBDF2, OR8K1, TMEM59, LOC618091, TMEM158, SLC47A1, TMEM161B, LOC617011, CYP4B1, TMEM8A, ATP11A, LOC618052, CKAP4, MGC157082, DCT, ABCA2, HEPHL1, TMC2, CLMN, LOC786467, ADAM19, LOC788055, LOC784434, OR2M4, CNNM3, OR5AR1, GJB4, IGFLR1, LOC783598, SLC35F4, LOC516274, IGSF8, OR6C68, PLG, LOC524702, LOC781494, LOC784187, LOC512464, OR5K1, OCSTAMP, IFNGR1, LOC101905933, PTPRC, CD1A, SDCBP, OR1J2, LHCGR, ACBD5, BAMBI, SLC13A3, TMEM8B, LOC100298868, LOC512973, LOC526294, SELPLG, LOC507383, CSF2RB, LPCAT3, OR10AD1, LOC783671, TMEM52B, B3GNT5, DGCR2, SLC11A2, LRRC15, KCNG2, MS4A7, SLC46A2, LRRC55, OR1D5, SLC8A2, LOC614021, LOC523126, LOC515418, LOC511569, UPK3A, NRP2, LOC527217, SIGLEC1, ATG9B, SLC34A2, NSG1, LOC506317, COX11, OR8H3, PIGQ, CD86, HTR3B, LNPEP, LOC508604,

			LOC519176, LOC523090, RMDN2, DLL1, SLC36A1, LOC522609, SLC25A39, IL1RL2, SLC26A8, EXTL1, DRD5, MCOLN3, MIA3, ANO3, CYP7B1, LOC508589, CCR5, LOC100847239, TMTC3, GJA8, CD2, LOC107131158, CASR, ATP10D, LOC515482, OR5I1, TMCO5B, POMGNT2, LOC615808, CDHR1, LOC526177, GALNT12, TREML1, LOC101902265, OR7D4, CD200R1L, OR10G2, LOC618733, LOC100139830, LRIT1, LOC615852, LOC518816, LOC100138951, UGT3A2, LOC526765, TAAR1, OR52B6, C13H20orf24, C28H10orf35, SVOPL, ACE, TNFRSF1A, SNX14, OR6K3, LOC510716, MGAM, SPNS2, LRIT2, ARID1B, ATP8B4, LRRC25, OR6C76, LOC509525, TM9SF2, LOC788874, OR8B4, LOC788872, LOC789957, IL10RB, HILPDA, LOC618660, CELSR2, HHATL, RNF186, ABCA9, LOC509025, ZDHHC3, THBD, COX4I2, TMEM213, LOC509034, CERS4, LOC511180, OR10A3, AOC3, GINM1, DCHS1, LOC615303, LOC101904538, KLRB1, OR4K5, ATRN, NFASC, FAM69C, LOC531174, OR7A17, CCDC136, LOC104968569, LRIG1, LOC524160, MUC13
GO:0004888~transmembrane signaling receptor activity	3.3E-33	155	*FCAMR, LOC785811, LOC788572, LOC526335, LOC523060, LOC781758, LOC524702, LOC530485, LOC527450, LOC781264, OR4D6, LOC781968, OR4E2, LOC511509, LOC519294, LOC618593, LOC515887, OR4D5, OR1G1, LOC504344, OR4N5, LOC781446, LOC538552, OR1E1, LOC618124, LOC785582, LOC509073, LOC618112, LOC512973, LOC526294, OR1Q1, OR4N2, LOC507383, LOC788587, OR10K2, LOC520835, OR6K2, OR12D3, LOC100300302, LGR4, PLA2R1, LOC508806, NFAM1, LOC619026, OR1D5, LYVE1, LOC785082, LOC616306, LOC785623, OR4D2, OR1K1, LOC530068, LOC100337063, OR10J3, LOC617388, LOC531304, LOC786596, OR2A2, OR10T2, LOC521645, LOC618523, LOC785639, LOC785910, LOC788778, OR12D2, LOC522609, LOC785647, OR4C46, LOC100847240, LOC788790, LOC508315, LOC517252, LOC514818, LOC107131158, LOC515482, OR4X1, OR4C16, OR4S1, OR1L3, LOC788693, LOC785683, SEMA6A, LOC526177, CD3E, LOC784706, LOC787071, LOC508766, LOC506121, OR4F15, OR10J1, LOC100140382, LOC788723, LY75, TLR6, TLR5, LGR5, LOC523768, LOC789690, LOC785723, LOC104968488, LOC615605, OR4C15, LOC100848575, TLR4, OR2AT4, LOC785712, OR4C3, LOC790152, LOC788704, TLR10, LOC527248, LOC788713, LOC508785, OR6K3, LOC789193, SEMA4D, LOC510112, LOC504567, OR4A16, OR4D11, LOC618675, LOC515540, LOC523258, TLR9, LOC532486, TLR2, LOC100298322, TLR3, LOC509025, LOC518869, THBD, LOC615901, LOC787247, OR4C6, OR4M1, LOC532501, LOC516467, LOC784957, LOC788089, LOC521749, LOC510901, SELE, LOC104968790, LOC508468, CD79B, LOC513151, OR4C13, LOC786467, OR4K5, OR8S1, LOC100848076, LOC509817, LOC104968568, LOC784925, LOC513884*
GO:0005549~odorant binding	3.0E-29	138	*LOC783205, OR5A1, LOC100299084, LOC538744, LOC788573, LOC504623, LOC104968576, LOC508392, LOC783210, LOC788552, LOC515090, LOC509369, LOC788554, LOC518442, LOC782678, OR5K1, LOC511753, LOC615009, LOC524985, OR13A1, LOC790274, LOC509323, OR9Q2, LOC787625, LOC618817, LOC510984, LOC785848, LOC100140748, LOC614143, LOC104969118, LOC100299808, LOC519492, OR5B3, LOC783884, LOC787642, LOC530175, LOC526047, LOC532075, LOC522775, LOC782009, LOC100337392, LOC788583, LOC526276, LOC782255, LOC788476, LOC789288, LOC514057, LOC787385, OR5D13, OR5AS1, LOC514864, LOC509895, OR5L2, OR8D2, LOC527216, LOC789300, LOC100139733, LOC525964, LOC782866, OR8H3, LOC781509, LOC787694, LOC532031, LOC788438, LOC101905743, LOC785914, LOC539185, OR5AU1, LOC786133, LOC788524, LOC508589, LOC509124, LOC785899, OR9G9, LOC100301231, LOC788998, LOC785903, LOC788512, LOC529511, LOC524282, LOC782792,*

			LOC782555, OR5I1, LOC782797, OR5L1, LOC528422, LOC781828, LOC100138976, LOC615808, OR5D14, LOC515045, LOC524304, LOC100139830, LOC100300446, LOC511657, LOC617507, LOC613726, OR8K3, LOC785944, LOC523139, OR10V1, LOC513334, LOC513062, LOC513333, LOC505546, LOC520162, LOC789936, LOC519616, OR8G2, LOC516940, LOC788874, OR8B4, OR8K1, LOC618091, LOC789957, LOC618660, LOC618660, OR9K2, OR5M3, LOC617016, LOC782152, LOC527077, LOC513914, LOC539064, LOC617011, OR6P1, OR5D18, LOC784897, OR14J1, LOC100336916, LOC783843, LOC616755, LOC528722, LOC513384, LOC100300488, OR5AR1, OR11L1, OR5P3, OR8D4
GO:0050911~detection of chemical stimulus involved in sensory perception of smell	6.9E-10	45	*LOC533983, LOC786149, LOC541022, LOC788626, LOC100301421, LOC100295806, OR7G2, LOC100298773, LOC787041, LOC100847281, LOC512948, LOC509641, LOC789504, LOC509510, OR7D4, LOC787665, LOC785624, LOC107131149, LOC513101, LOC516132, LOC521350, LOC786202, LOC512296, LOC788246, LOC510625, LOC787423, LOC100301104, LOC530825, LOC507378, OR7G3, LOC504888, OR7A5, LOC785946, LOC511678, LOC787574, OR7A17, LOC509267, LOC100847239, LOC100298850, LOC100847301, OR7D2, LOC100298119, LOC526508, LOC783313, OR7A10*
GO:0007165~signal transduction	2.4E-05	209	*DARADD, CD48, OR51Q1, LOC515619, TGFBRAP1, LOC787584, RASSF8, AKAP5, LOC618173, LOC528807, UNC5D, GRK6, RASAL2, PDE11A, SNX27, LOC782941, RPS6KC1, UNC5B, OR52K1, APBB1IP, OR52M1, RASSF1, SIGIRR, ARHGAP21, CHRNG, ARHGAP18, OR52N2, PDE10A, OR52B4, LOC783446, LYVE1, LOC526412, NCOA5, SOX8, SH2B1, LOC526411, AKAP12, OR52E4, LOC527918, LOC511823, TRHDE, RREB1, ZNF516, PRKAB2, LOC782373, SRGAP1, CAP2, DTHD1, SPARCL1, CHRNA5, CTGF, SYDE2, LOC518464, NDRG1, LANCL1, LOC516396, LOC519071, PARK2, CHRNA6, LOC518561, TRAF6, IQGAP3, LOC783488, LOC516414, LOC532436, CHRNA10, PDPN, SH3BP1, UNC5CL, RIN2, LOC100299247, TICAM1, GUCY2C, LOC100337265, RASSF9, LOC618010, GAPVD1, GABRR3, ARHGAP39, OR52I2, LOC507428, RIN1, LOC788633, LOC504551, ARAP2, ZNF536, LOC783558, TRIM13, LOC618050, RASSF7, LOC107133172, DLC1, LOC506452, RASSF6, SARM1, GAS6, LOC783598, LOC618075, LOC789134, LOC783597, LOC510351, LOC615276, LOC521676, INPP1, OR51I1, LOC617878, RASSF3, MAGI3, LOC783616, GRK1, LOC784187, LOC526621, RTKN, SAG, OR56B4, RASSF4, LOC528515, PKN1, RHPN1, CHRNB3, LOC530994, GRK4, LOC530990, OR52W1, STARD13, CSNK2B, LOC783671, LOC617417, PLA2R1, INPP5B, INPP4B, MYO10, SLC39A12, WISP1, ARHGAP28, OR51S1, OR51F2, OR52B2, ANGPTL3, LOC512399, LOC505646, OR51D1, STOML3, LOC511570, LOC511569, TRIM38, ARHGAP27, LOC782046, OR52E8, OR52K2, FYB, LOC508595, LNPEP, LOC523090, LOC519176, PDE6A, LOC517799, PDE5A, PDE6B, PDE1C, TENM3, SYDE1, NDRG2, TOM1L1, LOC788263, SRGAP3, NRADD, NDRG4, LVRN, OR52H1, ARHGAP10, PDE3B, LOC783323, OR52B6, ARHGAP20, PLEKHH3, CAMK1, PDE9A, RPS6KL1, IMPA2, TRH, ARHGAP24, ARHGAP11A, OR51E2, LANCL2, LOC788864, LOC617817, PDE4C, LOC788872, ANK3, LOC787779, ZNF831, THBD, LOC613909, VOPP1, SALL4, UCN, LOC787830, ARHGAP22, PLPPR4, RADIL, SRGAP2, LOC524658, ARHGAP12, LOC531174, HIVEP2, PPP2R5C, LOC507662, OR51A7*

Appendix 4C. Analysis of Molecular Variance.

a) Analysis of Molecular Variance between Pantaneiro (PAN) and Crioulo Lageano (CRL) cattle breeds

Call: pegas::amova(formula = gen_dist ~ info_factor, is.squared = TRUE)

	SSD	MSD	df
info_factor	34.34943	34.34943	1
Error	642.59911	29.20905	22
Total	676.94853	29.43254	23

Variance components:

	sigma2	P.value
info_factor	0.42836	0
Error	29.20905	

Phi-statistics:

info_factor.in.GLOBAL
0.01445351

Variance coefficients:

a
12

b) Analysis of Molecular Variance of the four breeds

Call: pegas::amova(formula = gen_dist ~ Breeds, is.squared = TRUE)

	SSD	MSD	df
Breeds	194.6287	64.87624	3
Error	1272.8329	28.28517	45
Total	1467.4616	30.57212	48

Variance components:

	sigma2	P.value
Breeds	2.9883	0
Error	28.2852	

Phi-statistics:

Breeds.in.GLOBAL
0.09555297

Variance coefficients:

a
122.449

Appendix 5C. Annotated candidate sweep regions retrieved from the top 1% of the empirical distribution generated by the within-population DCMS statistic.

BTA[1]	Start (bp)	End (bp)	DCMS score	Genes
1	4,200,000	4,250,000	19.66	
1	6,150,000	6,200,000	21.97	*BACH1*
1	6,250,000	6,300,000	10.02	
1	8,300,000	8,350,000	18.80	
1	18,150,000	18,200,000	18.36	
1	36,350,000	36,400,000	22.55	
1	41,600,000	41,650,000	14.83	*ARL6, EPHA6*
1	42,000,000	42,050,000	8.89	
1	50,050,000	50,100,000	12.82	
1	51,650,000	51,700,000	9.89	
1	59,350,000	59,400,000	10.46	*TIGIT*
1	59,400,000	59,450,000	29.94	
1	60,500,000	60,550,000	10.03	
1	69,250,000	69,300,000	36.23	*KALRN*
1	81,800,000	81,850,000	16.80	
1	82,900,000	82,950,000	16.77	*VPS8*
1	82,950,000	83,000,000	11.97	*VPS8*
1	83,000,000	83,050,000	9.25	
1	102,000,000	102,050,000	32.04	
1	111,450,000	111,500,000	15.03	
1	111,600,000	111,650,000	20.37	
1	112,250,000	112,300,000	14.08	*KCNAB1*
1	120,300,000	120,350,000	15.84	*CPB1*
1	120,650,000	120,700,000	14.25	
1	128,150,000	128,200,000	10.93	*GRK7*
1	129,200,000	129,250,000	28.05	
1	149,150,000	149,200,000	20.75	
1	149,250,000	149,300,000	13.25	
1	153,050,000	153,100,000	12.73	
2	4,400,000	4,450,000	20.06	*SAP130*
2	9,500,000	9,550,000	17.89	
2	10,200,000	10,250,000	65.36	
2	14,600,000	14,650,000	10.81	*PPP1R1C*
2	14,700,000	14,750,000	21.95	*ITPRID2*
2	16,850,000	16,900,000	8.71	*CWC22*
2	29,500,000	29,550,000	10.14	
2	40,600,000	40,650,000	11.10	
2	73,200,000	73,250,000	16.60	
2	95,050,000	95,100,000	10.27	

Appendix 5C. Continuation

2	98,300,000	98,350,000	8.34	*KANSL1L*
2	101,950,000	102,000,000	26.76	
2	102,000,000	102,050,000	14.30	
2	103,100,000	103,150,000	27.79	*VWC2L*
2	103,300,000	103,350,000	22.62	*BARD1*
2	103,600,000	103,650,000	11.03	*ABCA12*
2	122,950,000	123,000,000	12.42	*NKAIN1*
3	2,150,000	2,200,000	15.02	
3	4,450,000	4,500,000	20.41	*PBX1*
3	13,700,000	13,750,000	10.94	*ETV3*
3	13,750,000	13,800,000	13.70	*ETV3, ETV3L*
3	16,250,000	16,300,000	15.80	*AQP10, ATP8B2*
3	21,500,000	21,550,000	21.75	*ANKRD34A, ANKRD34A, LIX1L*
3	21,550,000	21,600,000	12.07	*RBM8A, LIX1L, PEX11B, ITGA10, ANKRD35*
3	23,250,000	23,300,000	12.39	
3	23,550,000	23,600,000	13.28	*REG4*
3	37,050,000	37,100,000	12.84	
3	37,100,000	37,150,000	15.75	
3	37,200,000	37,250,000	29.24	
3	41,650,000	41,700,000	8.84	
3	42,250,000	42,300,000	9.04	
3	44,350,000	44,400,000	34.87	*PLPPR5*
3	44,400,000	44,450,000	11.78	*PLPPR5*
3	44,450,000	44,500,000	13.28	*PLPPR5*
3	44,500,000	44,550,000	26.64	
3	44,550,000	44,600,000	25.40	
3	46,600,000	46,650,000	9.85	
3	46,650,000	46,700,000	14.00	
3	46,700,000	46,750,000	8.25	
3	47,150,000	47,200,000	25.29	
3	47,200,000	47,250,000	16.91	
3	47,700,000	47,750,000	11.49	
3	47,950,000	48,000,000	21.74	
3	52,150,000	52,200,000	9.15	*HFM1*
3	55,400,000	55,450,000	9.19	
3	87,150,000	87,200,000	9.63	*FGGY*
3	87,200,000	87,250,000	23.82	*FGGY*
3	87,250,000	87,300,000	14.70	*FGGY*
3	103,450,000	103,500,000	22.84	
3	105,500,000	105,550,000	19.62	
4	900,000	950,000	10.37	

Appendix 5C. Continuation

4	16,00,000	1,650,000	22.08	
4	1,950,000	2,000,000	22.92	
4	2,950,000	3,000,000	14.41	
4	12,200,000	12,250,000	10.93	*PPP1R9A*
4	34,150,000	34,200,000	13.99	
4	61,600,000	61,650,000	12.58	
4	61,650,000	61,700,000	10.97	
4	69,500,000	69,550,000	9.35	
4	74,050,000	74,100,000	19.45	
4	101,550,000	101,600,000	16.77	*CHRM2*
4	101,600,000	101,650,000	13.18	
4	101,750,000	101,800,000	14.53	*PTN*
4	114,700,000	114,750,000	14.18	*NUB1, WDR86*
4	115,150,000	115,200,000	15.36	
4	115,200,000	115,250,000	10.91	*GALNTL5*
4	115,550,000	115,600,000	15.05	
4	116,150,000	116,200,000	11.73	
4	116,450,000	116,500,000	19.99	
4	116,650,000	116,700,000	8.54	
4	116,750,000	116,800,000	9.06	
5	800,000	850,000	9.19	*TSPAN8*
5	3,700,000	3,750,000	22.56	
5	6,650,000	6,700,000	21.51	
5	6,700,000	6,750,000	10.04	
5	7,150,000	7,200,000	13.52	
5	9,450,000	9,500,000	8.88	*PPP1R12A*
5	14,550,000	14,600,000	13.42	*SLC6A15*
5	14,650,000	14,700,000	13.35	
5	36,000,000	36,050,000	19.94	*NELL2*
5	36,150,000	36,200,000	10.50	*TMEM117*
5	36,700,000	36,750,000	15.85	*TMEM117*
5	42,900,000	42,950,000	8.58	*PTPRR*
5	46,400,000	46,450,000	13.20	
5	49,950,000	50,000,000	13.29	*SRGAP1*
5	55,650,000	55,700,000	16.93	
5	61,500,000	61,550,000	15.34	
5	61,550,000	61,600,000	16.18	
5	65,300,000	65,350,000	19.76	*ANO4*
5	66,400,000	66,450,000	17.84	
5	66,450,000	66,500,000	15.14	
5	66,500,000	66,550,000	16.31	*IGF1*
5	66,700,000	66,750,000	12.94	

Appendix 5C. Continuation

5	66,850,000	66,900,000	10.62	
5	66,950,000	67,000,000	11.90	*PAH*
5	67,000,000	67,050,000	11.99	*PAH*
5	67,050,000	67,100,000	17.30	*ASCL1*
5	67,100,000	67,150,000	37.08	
5	104,300,000	104,350,000	11.87	*TAPBPL, CD27, LTBR,*
5	104,350,000	104,400,000	10.17	*LTBR, SCNN1A*
6	550,000	600,000	8.17	
6	2,200,000	2,250,000	8.26	
6	2,550,000	2,600,000	15.81	
6	13,600,000	13,650,000	15.30	
6	15,100,000	15,150,000	8.39	
6	15,450,000	15,500,000	11.87	
6	15,600,000	15,650,000	12.04	
6	15,750,000	15,800,000	14.20	
6	21,650,000	21,700,000	10.06	
6	25,500,000	25,550,000	20.31	
6	25,750,000	25,800,000	10.30	
6	25,850,000	25,900,000	13.86	*DDIT4L*
6	37,650,000	37,700,000	9.74	*PYURF, HERC5*
6	46,000,000	46,050,000	24.80	
6	55,050,000	55,100,000	10.59	
6	55,100,000	55,150,000	15.09	
6	64,950,000	65,000,000	16.50	*GUF1, GNPDA2*
6	70,700,000	70,750,000	9.27	*LNX1*
6	78,050,000	78,100,000	9.80	
6	78,100,000	78,150,000	14.48	
6	78,150,000	78,200,000	10.51	
6	78,200,000	78,250,000	9.90	
6	100,250,000	100,300,000	9.53	
7	400,000	450,000	23.45	*FLT4*
7	8,350,000	8,400,000	16.53	
7	8,400,000	8,450,000	9.84	*CYP4F2*
7	29,000,000	29,050,000	20.69	
7	69,450,000	69,500,000	16.81	
7	69,500,000	69,550,000	15.64	
7	79,250,000	79,300,000	9.40	
7	84,250,000	84,300,000	8.90	*ATG10*
7	84,300,000	84,350,000	9.08	*ATG10*
7	86,800,000	86,850,000	9.07	
7	86,850,000	86,900,000	9.70	
7	96,600,000	96,650,000	11.18	

Appendix 5C. Continuation

8	14,050,000	14,100,000	18.88	
8	15,800,000	15,850,000	17.76	
8	16,250,000	16,300,000	12.18	*LINGO2*
8	16,700,000	16,750,000	12.45	*MOB3B*
8	21,450,000	21,500,000	11.65	
8	54,750,000	54,800,000	10.11	
8	54,800,000	54,850,000	25.98	
8	59,250,000	59,300,000	19.26	
8	74,100,000	74,150,000	12.12	
8	77,350,000	77,400,000	19.15	*GALT, CCL27, CCL19, IL11RA*
8	85,700,000	85,750,000	8.79	*BICD2*
8	104,000,000	104,050,000	26.07	
9	2,450,000	2,500,000	16.12	
9	5,250,000	5,300,000	12.47	
9	10,550,000	10,600,000	15.68	
9	24,350,000	24,400,000	13.41	*RSPO3*
9	24,400,000	24,450,000	9.47	*RSPO3*
9	24,700,000	24,750,000	14.27	
9	43,800,000	43,850,000	12.93	*QRSL1, RTN4IP1*
9	45,600,000	45,650,000	16.65	
9	45,750,000	45,800,000	31.48	*HACE1*
9	52,000,000	52,050,000	13.85	
9	52,050,000	52,100,000	11.56	
9	73,000,000	73,050,000	11.44	
9	77,150,000	77,200,000	8.79	*ARFGEF3*
9	98,650,000	98,700,000	12.88	*PRKN*
9	98,700,000	98,750,000	9.35	*PRKN*
10	6,200,000	6,250,000	20.42	
10	13,150,000	13,200,000	30.46	*RF00582, DIS3L, TIPIN*
10	13,200,000	13,250,000	15.16	*RF00582, TIPIN, MAP2K1*
10	13,500,000	13,550,000	9.01	*SMAD6*
10	19,250,000	19,300,000	9.99	*ARIH1*
10	33,900,000	33,950,000	15.77	
10	35,250,000	35,300,000	19.15	
10	38,500,000	38,550,000	16.88	*CCNDBP1*
10	41,100,000	41,150,000	33.63	*EPB42*
10	41,150,000	41,200,000	31.27	
10	61,900,000	61,950,000	15.34	*FBN1*
10	61,950,000	62,000,000	8.32	*FBN1*
10	62,000,000	62,050,000	15.01	*FBN1*
10	69,900,000	69,950,000	15.60	*EXOC5, AP5M1*

Appendix 5C. Continuation

10	69,950,000	70,000,000	13.26	
10	74,300,000	74,350,000	16.10	
10	76,650,000	76,700,000	8.66	*SYNE2*
10	81,700,000	81,750,000	12.76	
10	82,100,000	82,150,000	11.09	*SMOC1, SLC8A3*
10	82,200,000	82,250,000	11.89	*SLC8A3*
10	82,250,000	82,300,000	9.04	*SLC8A3*
10	89,500,000	89,550,000	14.20	*NGB, POMT2*
10	99,600,000	99,650,000	10.77	
10	99,650,000	99,700,000	19.87	
10	103,150,000	103,200,000	16.41	*TTC7B*
10	103,900,000	103,950,000	10.13	*CTDSPL2*
11	1,650,000	1,700,000	9.17	*NPHP1*
11	22,300,000	22,350,000	9.57	
11	26,450,000	26,500,000	13.46	
11	40,200,000	40,250,000	9.96	
11	44,250,000	44,300,000	39.70	*SH3RF3*
11	48,350,000	48,400,000	14.71	*REEP1*
11	53,900,000	53,950,000	11.66	
11	66,000,000	66,050,000	8.52	
11	69,150,000	69,200,000	12.71	
11	72,150,000	72,200,000	9.75	*FNDC4, GCKR, IFT172*
11	72,400,000	72,450,000	8.79	*ATRAID, SLC5A6, CAD*
11	79,400,000	79,450,000	18.89	
11	80,850,000	80,900,000	20.16	
11	85,750,000	85,800,000	22.74	
12	11,100,000	11,150,000	12.87	*CNMD, SUGT1*
12	11,250,000	11,300,000	29.11	*ELF1*
12	33,050,000	33,100,000	11.44	
12	33,100,000	33,150,000	21.18	*GPR12*
12	33,150,000	33,200,000	12.44	*WASF3*
12	36,900,000	36,950,000	17.65	*ATP12A*
12	49,850,000	49,900,000	8.53	
12	54,150,000	54,200,000	19.37	
12	82,500,000	82,550,000	8.63	
12	82,550,000	82,600,000	10.28	
12	82,650,000	82,700,000	12.78	
12	85,450,000	85,500,000	16.86	
13	4,400,000	4,450,000	17.75	
13	6,300,000	6,350,000	11.46	
13	7,600,000	7,650,000	13.53	*SEL1L2*
13	11,600,000	11,650,000	9.23	

Appendix 5C. Continuation

13	23,100,000	23,150,000	26.97	*MLLT10*
13	29,750,000	29,800,000	27.98	*CDNF, HSPA14*
13	43,900,000	43,950,000	11.75	*AKR1C3*
13	43,950,000	44,000,000	13.67	*AKR1C4*
13	48,250,000	48,300,000	8.82	*SHLD1*
13	63,000,000	63,050,000	16.68	*BPIFA2A*
13	63,100,000	63,150,000	13.05	
13	73,350,000	73,400,000	15.45	*JPH2, OSER1*
13	76,200,000	76,250,000	14.61	
13	76,300,000	76,350,000	8.93	*EYA2*
13	77,750,000	77,800,000	16.05	
13	78,000,000	78,050,000	12.48	*STAU1*
14	100,000	150,000	12.58	
14	300,000	350,000	9.35	
14	550,000	600,000	10.62	
14	2,600,000	2,650,000	19.09	*LY6E*
14	2,650,000	2,700,000	11.91	
14	4,550,000	4,600,000	13.25	*TRAPPC9*
14	15,100,000	15,150,000	13.32	
14	22,250,000	22,300,000	21.25	*SNTG1*
14	25,950,000	26,000,000	12.10	
14	26,700,000	26,750,000	8.50	*TOX*
14	28,200,000	28,250,000	11.34	
14	28,250,000	28,300,000	14.68	
14	29,950,000	30,000,000	12.64	
14	31,050,000	31,100,000	8.33	*CYP7B1*
14	35,650,000	35,700,000	12.53	*SLCO5A1*
14	40,500,000	40,550,000	18.42	
14	42,350,000	42,400,000	21.19	
14	46,050,000	46,100,000	9.72	
14	55,000,000	55,050,000	10.02	
14	55,300,000	55,350,000	8.81	
14	55,350,000	55,400,000	15.86	
14	56,150,000	56,200,000	15.02	
14	83,400,000	83,450,000	10.28	*ENPP2*
15	4,200,000	4,250,000	9.94	
15	5,300,000	5,350,000	18.39	
15	5,450,000	5,500,000	11.37	
15	35,250,000	35,300,000	27.38	*KCNC1, SERGEF*
15	35,300,000	35,350,000	14.33	*MYOD1, KCNC1*
15	35,350,000	35,400,000	31.66	
15	35,400,000	35,450,000	12.18	*OTOG*

Appendix 5C. Continuation

15	36,750,000	36,800,000	20.28	*SOX6*
15	36,800,000	36,850,000	23.46	*SOX6*
15	42,300,000	42,350,000	20.54	
15	43,150,000	43,200,000	9.17	*SBF2*
15	43,200,000	43,250,000	38.89	*SBF2*
15	49,750,000	49,800,000	19.21	
15	49,850,000	49,900,000	9.64	
15	54,450,000	54,500,000	28.78	*P4HA3, PGM2L1*
15	54,500,000	54,550,000	11.44	*PGM2L1*
15	61,250,000	61,300,000	8.40	
15	62,500,000	62,550,000	15.64	*DCDC1*
15	63,450,000	63,500,000	9.90	
15	64,150,000	64,200,000	29.23	
15	64,200,000	64,250,000	13.30	
15	64,350,000	64,400,000	10.87	*QSER1*
15	64,400,000	64,450,000	20.72	*QSER1*
15	66,000,000	66,050,000	9.98	*EHF*
15	66,100,000	66,150,000	21.32	
15	81,150,000	81,200,000	19.05	
15	81,500,000	81,550,000	8.72	
15	81,650,000	81,700,000	16.79	*LRRC55*
15	81,750,000	81,800,000	21.37	*TNKS1BP1*
15	84,700,000	84,750,000	15.73	
16	7,000,000	7,050,000	16.24	
16	27,300,000	27,350,000	20.03	*TLR5*
16	27,400,000	27,450,000	9.75	*SUSD4*
16	38,450,000	38,500,000	9.13	*KIFAP3*
16	41,050,000	41,100,000	9.56	
16	41,350,000	41,400,000	17.86	
16	64,350,000	64,400,000	20.87	*CACNA1E*
17	44,50,000	4,500,000	18.14	
17	4,500,000	4,550,000	12.11	
17	7,050,000	7,100,000	12.86	*LRBA*
17	10,550,000	10,600,000	21.18	*ARHGAP10*
17	10,650,000	10,700,000	8.78	
17	27,150,000	27,200,000	10.75	
17	36,150,000	36,200,000	18.84	
17	36,200,000	36,250,000	9.58	
17	36,300,000	36,350,000	27.23	
17	36,500,000	36,550,000	8.50	
17	37,550,000	37,600,000	19.49	
17	37,600,000	37,650,000	15.26	

Appendix 5C. Continuation

17	37,650,000	37,700,000	9.09	
17	39,250,000	39,300,000	24.11	
17	41,450,000	41,500,000	10.69	
17	48,100,000	48,150,000	16.36	
17	48,150,000	48,200,000	8.90	
17	49,250,000	49,300,000	8.71	*GLT1D1*
17	66,450,000	66,500,000	16.84	*CORO1C*
17	68,450,000	68,500,000	18.81	*TPST2*
17	68,550,000	68,600,000	17.41	
17	70,500,000	70,550,000	24.63	
17	70,750,000	70,800,000	13.05	*AP1B1*
17	71,150,000	71,200,000	9.92	*MTMR3*
18	250,000	300,000	10.89	
18	2,600,000	2,650,000	23.60	*BCAR1*
18	3,350,000	3,400,000	24.41	*CNTNAP4*
18	3,500,000	3,550,000	10.04	*CNTNAP4*
18	3,650,000	3,700,000	20.43	
18	3,800,000	3,850,000	9.74	
18	25,350,000	25,400,000	10.78	
18	26,050,000	26,100,000	8.81	*CSNK2A2, CFAP20*
18	32,950,000	33,000,000	10.17	
18	33,100,000	33,150,000	8.82	
18	33,150,000	33,200,000	14.69	
18	33,450,000	33,500,000	13.73	
18	34,700,000	34,750,000	25.73	*PDP2, RRAD, CDH16*
18	37,800,000	37,850,000	15.07	
18	38,000,000	38,050,000	18.36	
18	44,850,000	44,900,000	11.89	*LSM14A*
18	60,000,000	60,050,000	15.67	
18	61,950,000	62,000,000	9.52	
19	2,550,000	2600,000	9.13	
19	12,700,000	12,750,000	15.21	
19	12,750,000	12,800,000	15.19	*APPBP2*
19	18,350,000	18,400,000	9.26	*ADAP2*
19	18,400,000	18,450,000	26.39	*TEFM, ATAD5*
19	18,500,000	18,550,000	9.76	*SUZ12*
19	25,050,000	25,100,000	12.02	*NCBP3*
19	25,500,000	25,550,000	14.78	*SPNS3*
19	27,550,000	27,600,000	18.19	*ACADVL, PHF23, GABARAP, ELP5, DLG4, DVL2, CTDNEP1*
19	27,600,000	27,650,000	16.03	*ELP5, CLDN7, SLC2A4, EIF5A, YBX2*

Appendix 5C. Continuation

19	50,450,000	50,500,000	23.66	*TBCD*
19	56,100,000	56,150,000	15.44	*RNF157*
19	62,750,000	62,800,000	9.47	
20	5,050,000	5,100,000	11.29	
20	9,000,000	9,050,000	16.54	
20	13,900,000	13,950,000	14.67	*PPWD1, TRIM23*
20	13,950,000	14,000,000	8.64	*PPWD1, CENPK*
20	20,500,000	20,550,000	8.32	*RAB3C*
20	28,700,000	28,750,000	9.27	*EMB*
20	30,000,000	30,050,000	10.22	
20	30,050,000	30,100,000	12.68	
20	30,100,000	30,150,000	42.64	
20	30,200,000	30,250,000	11.04	
20	30,250,000	30,300,000	10.67	
20	30,300,000	30,350,000	13.31	
20	30,350,000	30,400,000	16.20	
20	30,400,000	30,450,000	12.36	
20	30,450,000	30,500,000	28.29	
20	31,400,000	31,450,000	15.54	*CCL28*
20	32,300,000	32,350,000	14.33	
20	33,850,000	33,900,000	10.92	
20	34,000,000	34,050,000	10.85	
20	34,050,000	34,100,000	25.35	
20	34,100,000	34,150,000	27.26	
20	34,150,000	34,200,000	21.60	
20	47,450,000	47,500,000	14.71	
20	60,100,000	60,150,000	10.26	
20	66,350,000	66,400,000	11.39	
20	67,150,000	67,200,000	27.20	
20	67,200,000	67,250,000	25.83	
20	67,250,000	67,300,000	18.81	
20	71,100,000	71,150,000	8.66	*LPCAT1*
21	1,650,000	1,700,000	9.28	
21	2,150,000	2,200,000	34.67	
21	6,550,000	6,600,000	10.54	*ADAMTS17*
21	12,350,000	12,400,000	13.06	
21	12,450,000	12,500,000	12.68	
21	33,300,000	33,350,000	11.28	
21	33,350,000	33,400,000	9.11	
21	33,450,000	33,500,000	10.44	
21	36,550,000	36,600,000	18.12	
21	40,750,000	40,800,000	24.71	

Appendix 5C. Continuation

21	55,250,000	55,300,000	15.88	*TOGARAM1, PRPF39*
21	56,700,000	56,750,000	14.84	
21	61,050,000	61,100,000	12.57	
21	61,450,000	61,500,000	8.90	
21	61,750,000	61,800,000	14.16	
21	63,000,000	63,050,000	15.35	*PAPOLA*
21	63,250,000	63,300,000	13.97	*VRK1*
21	65,600,000	65,650,000	9.67	
21	65,750,000	65,800,000	14.43	
21	65,800,000	65,850,000	12.52	*BCL11B*
22	500,000	550,000	15.58	*VOPP1*
22	14,950,000	15,000,000	23.73	
22	47,750,000	47,800,000	10.07	*CACNA1D*
22	48,650,000	48,700,000	8.27	*ITIH1, NEK4*
23	4,500,000	4,550,000	8.56	*BMP5*
23	13,100,000	13,150,000	23.89	
23	13,200,000	13,250,000	17.34	
23	13,250,000	13,300,000	11.65	
23	13,300,000	13,350,000	13.39	
23	13,350,000	13,400,000	9.44	
23	13,400,000	13,450,000	12.61	
23	13,450,000	13,500,000	11.26	
23	19,550,000	19,600,000	17.90	*RCAN2*
23	19,600,000	19,650,000	18.87	*RCAN2*
23	19,650,000	19,700,000	22.48	*RCAN2*
23	20,000,000	20,050,000	15.90	*MEP1A*
23	20,100,000	20,150,000	14.75	*ADGRF5*
23	20,300,000	20,350,000	17.75	
23	20,350,000	20,400,000	9.69	
23	20,450,000	20,500,000	8.30	*TNFRSF21*
23	38,750,000	38,800,000	12.40	
23	38,800,000	38,850,000	10.81	
23	38,900,000	38,950,000	13.31	
23	38,950,000	39,000,000	11.07	*RNF144B*
23	48,750,000	48,800,000	9.42	*F13A1*
24	19,750,000	19,800,000	31.62	
24	34,700,000	34,750,000	8.83	
24	50,650,000	50,700,000	11.83	*MAPK4*
25	23,00,000	2,350,000	8.81	*SRRM2, FLYWCH2*
25	6,250,000	6,300,000	12.42	
25	20,500,000	20,550,000	14.97	
25	20,550,000	20,600,000	8.50	

Appendix 5C. Continuation

25	22,950,000	23,000,000	10.50	*LCMT1*
25	23,000,000	23,050,000	12.55	*LCMT1*
26	500,000	550,000	8.99	*PCDH15*
26	5,150,000	5,200,000	10.04	
26	7,850,000	7,900,000	13.22	*PRKG1*
26	17,900,000	17,950,000	17.27	
26	22,400,000	22,450,000	12.23	*KCNIP2, OGA, ARMH3*
26	29,400,000	29,450,000	9.89	
26	29,750,000	29,800,000	22.70	
26	32,500,000	32,550,000	27.19	
26	32,600,000	32,650,000	8.35	
26	48,900,000	48,950,000	17.67	
26	49,300,000	49,350,000	30.79	
26	50,100,000	50,150,000	34.59	
26	50,150,000	50,200,000	8.28	
26	50,200,000	50,250,000	13.02	
27	50,000	100,000	9.77	*CLN8*
27	2,400,000	2,450,000	27.13	
27	15,300,000	15,350,000	18.89	*FAM149A, CYP4V2, KLKB1*
27	15,350,000	15,400,000	12.52	*F11*
27	15,400,000	15,450,000	14.04	*MTNR1A*
27	15,450,000	15,500,000	20.48	*MTNR1A, FAT1*
27	19,100,000	19,150,000	13.56	
27	19,350,000	19,400,000	10.32	
27	37,250,000	37,300,000	9.12	*RNF170, HOOK3*
27	37,550,000	37,600,000	9.61	
27	40,050,000	40,100,000	13.88	*TOP2B*
27	43,450,000	43,500,000	17.19	
28	4,600,000	4,650,000	8.85	
28	7,850,000	7,900,000	15.23	
28	11,400,000	11,450,000	10.37	
28	20,950,000	21,000,000	8.28	
28	21,000,000	21,050,000	8.44	
28	21,200,000	21,250,000	9.03	
28	27,300,000	27,350,000	13.55	
28	27,350,000	27,400,000	16.08	
28	34,350,000	34,400,000	14.41	
28	43,900,000	43,950,000	12.76	
29	550,000	600,000	19.41	*PANX1*
29	4,400,000	4,450,000	23.07	
29	5,450,000	5,500,000	11.53	
29	5,550,000	5,600,000	11.40	

Appendix 5C. Continuation

29	6,400,000	6,450,000	34.08	*TYR*
29	39,850,000	39,900,000	17.82	*PAG9*

[1] BTA: *Bos taurus* autosome

Appendix 6C. Annotated candidate sweep regions retrieved from the top 1% of the empirical distribution generated by the cross-population DCMS statistic.

BTA[1]	Start (bp)	End (bp)	DCMS score	Genes
1	5,650,000	5,700,000	12.97	
1	12,250,000	12,300,000	20.97	
1	13,650,000	13,700,000	15.18	
1	46,350,000	46,400,000	13.90	*PCNP*
1	46,600,000	46,650,000	14.79	
1	49,350,000	49,400,000	16.13	
1	54,200,000	54,250,000	13.56	*MORC1*
1	55,700,000	55,750,000	13.85	
1	66,950,000	67,000,000	16.27	*EAF2*
1	70,050,000	70,100,000	14.66	*HEG1*
1	104,600,000	104,650,000	13.50	
1	116,750,000	116,800,000	12.75	
1	122,400,000	122,450,000	25.27	
1	122,450,000	122,500,000	23.62	
1	134,350,000	134,400,000	18.75	*PPP2R3A*
1	134,550,000	134,600,000	13.58	
1	134,900,000	134,950,000	12.95	
1	134,950,000	135,000,000	15.15	
1	136,300,000	136,350,000	18.13	
1	136,350,000	136,400,000	24.10	
1	136,400,000	136,450,000	19.24	*SLCO2A1*
1	136,450,000	136,500,000	17.94	
1	136,750,000	136,800,000	17.42	
1	138,250,000	138,300,000	14.00	*DNAJC13*
1	150,300,000	150,350,000	13.45	*MORC3*
2	450,000	500,000	12.79	*OCA2*
2	5,250,000	5,300,000	13.71	*CYP27C1*
2	36,050,000	36,100,000	16.20	*RBMS1*
2	36,100,000	36,150,000	21.50	
2	125,300,000	125,350,000	13.44	*GMEB1, YTHDF2*
3	6,450,000	6,500,000	19.44	
3	6,800,000	6,850,000	18.27	*DDR2*
3	8,200,000	8,250,000	23.46	*MPZ, SDHC, PCP4L1*
3	8,250,000	8,300,000	19.64	*NR1I3, TOMM40L*
3	40,250,000	40,300,000	15.38	
3	52,550,000	52,600,000	16.20	
3	52,600,000	52,650,000	17.44	
3	53,200,000	53,250,000	12.86	
3	58,650,000	58,700,000	14.42	*CYR61*

Appendix 6C. Continuation

3	64,100,000	64,150,000	14.66	
3	64,150,000	64,200,000	12.79	
3	64,450,000	64,500,000	16.66	
3	64,500,000	64,550,000	13.31	
3	64,800,000	64,850,000	14.79	
3	64,850,000	64,900,000	15.86	
3	68,200,000	68,250,000	13.72	
3	68,850,000	68,900,000	13.19	
3	69,400,000	69,450,000	17.62	
3	69,450,000	69,500,000	14.42	
3	73,350,000	73,400,000	15.76	
3	77,250,000	77,300,000	18.16	
3	81,350,000	81,400,000	13.09	
3	81,400,000	81,450,000	17.50	
3	81,450,000	81,500,000	14.35	
3	81,600,000	81,650,000	24.92	*UBE2U*
3	81,650,000	81,700,000	20.80	*ROR1*
3	81,700,000	81,750,000	21.86	
3	81,750,000	81,800,000	21.71	
3	81,800,000	81,850,000	25.79	
3	81,850,000	81,900,000	20.19	
3	81,900,000	81,950,000	21.09	
3	81,950,000	82,000,000	26.43	
3	82,050,000	82,100,000	19.70	
3	82,350,000	82,400,000	19.51	*EFCAB7, ITGB3BP*
3	82,400,000	82,450,000	19.78	
3	82,450,000	82,500,000	27.87	*ALG6*
3	87,900,000	87,950,000	12.64	*MYSM1*
3	87,950,000	88,000,000	13.52	
3	96,150,000	96,200,000	13.74	*FAF1*
4	13,750,000	13,800,000	15.23	
4	17,050,000	17,100,000	13.80	*GLCCI1*
4	17,100,000	17,150,000	17.62	
4	17,150,000	17,200,000	19.12	*ICA1*
4	17,200,000	17,250,000	17.36	
4	17,250,000	17,300,000	16.30	
4	17,300,000	17,350,000	23.93	
4	17,350,000	17,400,000	16.27	
4	18,000,000	18,050,000	14.89	
4	30,100,000	30,150,000	17.37	
4	30,150,000	30,200,000	15.23	
4	41,850,000	41,900,000	14.28	

Appendix 6C. Continuation

4	42,500,000	42,550,000	14.69	
4	62,450,000	62,500,000	14.06	*DPY19L1*
4	62,600,000	62,650,000	12.76	*NPSR1*
4	101,050,000	101,100,000	14.61	
4	101,600,000	101,650,000	15.02	
4	102,800,000	102,850,000	14.22	
4	102,850,000	102,900,000	14.08	
4	111,000,000	111,050,000	15.51	
4	113,750,000	113,800,000	15.32	*GIMAP7*
4	117,050,000	117,100,000	13.30	
4	117,250,000	117,300,000	18.02	
4	117,300,000	117,350,000	15.84	
4	117,350,000	117400,000	26.19	
4	117,400,000	117,450,000	26.08	
4	117,450,000	117,500,000	16.18	
5	3,700,000	3,750,000	13.11	
5	4,100,000	4,150,000	21.18	
5	10,100,000	10,150,000	15.49	*PTPRQ*
5	11,350,000	11,400,000	13.35	
5	11,950,000	12,000,000	15.37	*METTL25*
5	12,100,000	12,150,000	22.30	
5	12,200,000	12,250,000	13.13	
5	12,250,000	12,300,000	17.32	
5	12,700,000	12,750,000	12.67	
5	12,900,000	12,950,000	13.55	
5	17,150,000	17,200,000	13.87	
5	31,800,000	31,850,000	17.75	
5	36,850,000	36,900,000	12.83	*TWF1, IRAK4*
5	38,750,000	38,800,000	13.04	*YAF2, GXYLT1*
5	40,800,000	40,850,000	17.40	*LRRK2*
5	40,850,000	40,900,000	16.39	
5	40,900,000	40,950,000	20.63	
5	40,950,000	41,000,000	13.00	
5	68,400,000	68,450,000	14.03	*CHST11*
5	72,750,000	72,800,000	13.41	*LARGE1*
5	72,800,000	72,850000	21.95	
5	72,900,000	72,950,000	16.34	
5	74,750,000	74,800,000	13.29	
5	74,800,000	74,850,000	12.84	
5	86,200,000	86,250,000	14.98	
5	106,950,000	107,000,000	14.75	*TSPAN11*
5	113,150,000	113,200,000	14.93	*POLR3H, CSDC2, PMM1*

Appendix 6C. Continuation

5	113,200,000	113,250,000	15.95	*DESI1, XRCC6*
5	118,650,000	118,700,000	13.48	
5	119,050,000	119,100,000	17.08	
5	119,100,000	119,150,000	13.82	
5	120,850,000	120,900,000	16.93	*BRD1*
5	120,900,000	120,950,000	13.73	*CRELD2, ALG12*
5	120,950,000	121,000,000	13.10	*PIM3, IL17REL*
6	1,650,000	1,700,000	13.17	
6	1,700,000	1,750,000	12.66	
6	6,200,000	6,250,000	14.21	
6	9,550,000	9,600,000	16.07	
6	16,550,000	16,600,000	15.36	
6	16,700,000	16,750000	14.06	*GAR1, LRIT3, RRH*
6	17,150,000	17,200,000	13.68	
6	17,550,000	17,600,000	18.02	*COL25A1*
6	17,600,000	17,650,000	12.80	
6	22,750,000	22,800,000	12.84	
6	23,050,000	23,100,000	13.22	*SLC9B2, BDH2*
6	24,100,000	24,150,000	13.85	
6	24,150,000	24,200,000	13.33	
6	38,150,000	38,200,000	15.33	
6	60,350,000	60,400,000	13.98	
6	70,150,000	70,200,000	13.63	*RASL11B, SCFD2*
6	70,350,000	70,400,000	14.92	
6	70,550,000	70,600,000	12.91	*FIP1L1*
6	70,600,000	70,650,000	12.99	
6	76,000,000	76,050,000	12.97	
6	78,000,000	78,050,000	15.58	
6	91,850,000	91,900,000	16.09	
6	95,000,000	95,050,000	12.81	*FRAS1*
6	102,500,000	102,550,000	12.73	*ARHGAP24*
6	102,650,000	102,700,000	17.25	*MAPK10*
7	22,400,000	22,450,000	13.30	*TIMM13, LMNB2, GADD45B*
7	25,650,000	25,700,000	16.90	*ADAMTS19*
7	25,700,000	25,750,000	20.64	
7	26,050,000	26,100,000	14.64	
7	26,200,000	26,250,000	18.95	*SLC27A6*
7	26,300,000	26,350,000	14.51	
7	26,350,000	26,400,000	12.90	
7	29,050,000	29,100,000	16.01	
7	57,050,000	57,100,000	12.97	
7	58,200,000	58,250,000	13.36	

Appendix 6C. Continuation

7	84,500,000	84,550,000	16.71	*RPS23, ATG10*
7	84,550,000	84,600,000	19.10	
7	84,700,000	84,750,000	16.02	
7	85,500,000	85,550,000	13.23	*XRCC4*
7	85,550,000	85,600,000	15.76	
7	86,500,000	86,550,000	13.48	*EDIL3*
7	86,550,000	86,600,000	12.93	
7	86,650,000	86,700,000	14.12	
7	86,700,000	86,750,000	14.76	
7	87,550,000	87,600,000	12.73	
7	87,600,000	87,650,000	13.78	
8	600,000	650,000	15.53	*PALLD*
8	15,400,000	15,450,000	13.91	
8	32,600,000	32,650,000	17.24	
8	55,400,000	55,450,000	14.74	
8	57,350,000	57,400,000	12.80	
8	58,750,000	58,800,000	12.93	
8	75,050,000	75,100,000	18.53	*DPYSL2*
9	23,800,000	23,850,000	12.84	*SNAP91*
9	23,900,000	23,950,000	12.96	
9	28,850,000	28,900,000	14.64	*PKIB*
9	43,150,000	43,200,000	13.66	*PDSS2*
9	43,200,000	43,250,000	20.89	
9	43,250,000	43,300,000	14.23	
9	51,150,000	51,200,000	13.19	*FAXC*
9	68,050,000	68,100,000	20.29	
9	98,600,000	98,650,000	16.00	*PRKN*
9	98,650,000	98,700,000	13.80	
10	6,250,000	6,300,000	14.36	
10	19,550,000	19,600,000	16.75	
10	25,000,000	25,050,000	12.65	*TRAV17*
10	27,150,000	27,200,000	15.64	
10	27,800,000	27,850,000	17.43	
10	30,750,000	30,800,000	14.85	
10	30,800,000	30,850,000	13.56	
10	34,550,000	34,600,000	13.22	
10	38,600,000	38,650,000	15.26	
10	38,650,000	38,700,000	13.31	
10	82,700,000	82,750,000	16.55	*MAP3K9*
10	84,350,000	84,400,000	13.05	*RGS6*
10	85,750,000	85,800,000	15.02	*ENTPD5, BBOF1*
10	87,600,000	87,650,000	16.45	*WDR36*

Appendix 6C. Continuation

10	87,650,000	87,700,000	13.53	
10	87,700,000	87,750,000	18.21	
10	87,750,000	87,800,000	21.74	
10	87,800,000	87,850,000	14.63	
10	102,900,000	102,950,000	14.92	*PSMC1, NRDE2*
11	50,000	100,000	14.06	
11	100,000	150,000	20.90	
11	650,000	700,000	13.61	
11	1,000,000	1,050,000	31.24	
11	1,750,000	1,800,000	13.78	*MALL*
11	2,500,000	2,550,000	13.42	*ARID5A*
11	4,400,000	4,450,000	21.25	*TXNDC9*
11	8,100,000	8,150,000	12.69	
11	10,000,000	10,050,000	28.86	*M1AP*
11	17,100,000	17,150,000	13.77	
11	17,300,000	17,350,000	17.78	
11	17,350,000	17,400,000	17.04	
11	17,450,000	17,500,000	14.33	
11	17,950,000	18,000,000	14.64	
11	19,600,000	19,650,000	19.31	*EIF2AK2, SULT6B1*
11	19,650,000	19,700,000	20.71	*NDUFAF7, CEBPZ*
11	19,700,000	19,750,000	23.65	*PRKD3*
11	19,750,000	19,800,000	22.09	
11	19,800,000	19,850,000	16.22	*QPCT*
11	19,850,000	19,900,000	17.79	
11	19,900,000	19,950,000	29.66	
11	19,950,000	20,000,000	22.92	
11	21,750,000	21,800,000	12.94	*MAP4K3*
11	22,100,000	22,150,000	15.59	*THUMPD2*
11	22,150,000	22,200,000	14.82	
11	22,200,000	22,250,000	21.54	
11	22,250,000	22,300,000	23.53	
11	22,300,000	22,350,000	15.41	
11	22,350,000	22,400,000	18.56	
11	22,400,000	22,450,000	16.58	*SLC8A1*
11	22,450,000	22,500,000	21.04	
11	22,600,000	22,650,000	18.15	
11	22,650,000	22,700,000	13.27	
11	22,700,000	22,750,000	13.39	
11	22,750,000	22,800,000	13.91	
11	23,050,000	23,100,000	16.46	
11	23,100,000	23,150,000	18.68	

Appendix 6C. Continuation

11	26,950,000	27,000,000	18.37	*CAMKMT*
11	40,500,000	40,550,000	14.41	*VRK2*
11	44,050,000	44,100,000	16.06	
11	44,700,000	44,750,000	13.23	*SULT1C2, SULT1C4, GCC2*
11	44,750,000	44,800,000	16.82	*SULT1C3*
11	44,800,000	44,850,000	19.54	
11	44,850,000	44,900,000	20.25	*SLC5A7*
11	44,900,000	44,950,000	14.63	
11	47,300,000	47,350,000	13.81	*EIF2AK3*
11	47,350,000	47,400,000	20.41	
11	53,900,000	53,950,000	13.97	
11	64,950,000	65,000,000	12.92	
11	66,500,000	66,550,000	12.87	*WDR92*
11	67,250,000	67,300,000	13.24	*GKN3P, GKN2, GKN1*
11	67,450,000	67,500,000	13.36	*ANTXR1*
11	67,700,000	67,750,000	16.02	*GFPT1*
11	67,750,000	67,800,000	13.58	*NFU1*
11	68,550,000	68,600,000	15.07	*PCYOX1*
11	69,450,000	69,500,000	14.31	
11	69,650,000	69,700,000	13.72	*YPEL5*
11	78,850,000	78,900,000	12.79	*LAPTM4A, MATN3*
11	81,050,000	81,100,000	15.52	*SMC6, VSNL1*
11	81,100,000	81,150,000	15.71	
11	85,650,000	85,700,000	12.90	
11	93,150,000	93,200,000	14.98	
11	93,250,000	93,300,000	13.61	
11	93,350,000	93,400,000	15.18	
11	98,050,000	98,100,000	16.99	*GARNL3*
11	98,600,000	98,650,000	16.52	*PIP5KL1, DPM2, FAM102A*
11	98,650,000	98,700,000	14.77	
12	3,500,000	3,550,000	14.15	
12	10,950,000	11,000,000	19.84	
12	11,050,000	11,100,000	13.05	*CNMD*
12	37,550,000	37,600,000	13.50	
12	52,200,000	52,250,000	16.17	
13	12,100,000	12,150,000	14.36	
13	14,650,000	14,700,000	16.14	
13	20,100,000	20,150,000	12.79	
13	63,750,000	63,800,000	15.63	
13	78,450,000	78,500,000	13.09	*B4GALT5*
14	2,700,000	2,750,000	13.46	*GML*

Appendix 6C. Continuation

14	2,800,000	2,850,000	17.56	*LYNX1, LYPD2, SLURP1, THEM6, LY6D*
14	3,000,000	3,050,000	13.02	*ADGRB1*
14	3,150,000	3,200,000	27.46	*TSNARE1*
14	4,600,000	4,650,000	14.52	*TRAPPC9, KCNK9*
14	4,700,000	4,750,000	12.64	
14	20,400,000	20,450,000	14.17	
14	23,250,000	23,300,000	14.83	*NPBWR1*
14	27,750,000	27,800,000	13.73	
14	27,900,000	27,950,000	13.58	*RAB2A*
14	28,100,000	28,150,000	16.80	*CHD7*
14	30,050,000	30,100,000	15.81	
14	30,100,000	30,150,000	13.35	
14	30,150,000	30,200,000	12.87	
14	30,200,000	30,250,000	15.39	
14	30,300,000	30,350,000	13.57	
14	30,350,000	30,400,000	14.08	
14	30,450,000	30,500,000	15.13	
14	31,450,000	31,500,000	14.21	
14	31,500,000	31,550,000	14.41	
14	35,850,000	35,900,000	13.03	*PRDM14*
14	37,600,000	37,650,000	12.89	
14	37,650,000	37,700,000	18.82	
14	37,700,000	37,750,000	23.71	*TRPA1*
14	37,750,000	37,800,000	27.19	
14	37,800,000	37,850,000	17.77	
14	37,850,000	37,900,000	20.91	
14	38,000,000	38,050,000	15.05	*KCNB2*
14	38,400,000	38,450,000	19.19	
14	38,450,000	38,500,000	15.95	
14	38,500,000	38,550,000	17.29	*TERF1*
14	38,550,000	38,600,000	18.82	*SBSPON*
14	38,600,000	38,650,000	21.94	
14	38,650,000	38,700,000	13.11	
14	38,700,000	38,750,000	18.39	*RPL7, RDH10*
14	39,000,000	39,050,000	15.76	
14	39,150,000	39,200,000	13.35	
14	39,200,000	39,250,000	17.18	*UBE2W*
14	44,250,000	44,300,000	14.53	
14	46,250,000	46,300,000	14.78	
14	46,300,000	46,350,000	22.48	*PAG1*
14	46,350,000	46,400,000	13.71	

Appendix 6C. Continuation

14	52,900,000	52,950,000	13.99	*CSMD3*
14	52,950,000	53,000,000	13.22	
14	55,550,000	55,600,000	13.90	
14	56,500,000	56,550,000	15.81	
14	75,000,000	75,050,000	13.27	
14	84,250,000	84,300,000	13.94	*SNTB1*
15	5,650,000	5,700,000	15.86	*THAP12*
15	8,400,000	8,450,000	14.80	
15	10,150,000	10,200,000	13.32	
15	10,900,000	10,950,000	12.85	
15	27,100,000	27,150,000	16.52	
15	27,150,000	27,200,000	12.81	
15	27,200,000	27,250,000	15.44	
15	27,250,000	27,300,000	13.23	
15	27,300,000	27,350,000	13.38	
15	32,150,000	32,200,000	13.54	
15	56,500,000	56,550,000	14.12	
15	66,300,000	66,350,000	13.04	*PDHX*
15	66,400,000	66,450,000	13.21	
15	83,750,000	83,800,000	13.62	
15	83,900,000	83,950,000	23.53	
15	83,950,000	84,000,000	28.49	*OR5A1, OR4D6*
16	950,000	1,000,000	12.91	
16	9,250,000	9,300,000	16.05	
16	9,300,000	9,350,000	19.05	
16	9,700,000	9,750,000	14.74	
16	9,750,000	9,800,000	14.56	
16	12,450,000	12,500,000	16.83	
16	25,600,000	25,650,000	12.75	
16	43,000,000	43,050,000	14.70	*DISP3*
16	44,000,000	44,050,000	15.04	*DFFA, CORT, CENPS, PGD*
16	45,000,000	45,050,000	13.59	*SPSB1*
16	81,400,000	81,450,000	16.16	*KIF21B*
17	10,50,000	1,100,000	12.83	
17	11,00,000	1,150,000	16.03	
17	13,00,000	1,350,000	15.56	
17	49,00,000	4,950,000	16.69	
17	101,50,000	10,200,000	15.54	*NR3C2, ARHGAP10*
17	105,00,000	10,550,000	13.19	
17	111,00,000	11,150,000	13.77	
17	111,50,000	11,200,000	13.76	
17	116,50,000	11,700,000	13.42	*TTC29*

Appendix 6C. Continuation

17	12,550,000	12,600,000	14.62	
17	12,950,000	13,000,000	22.42	*SMAD1*
17	17,750,000	17,800,000	14.70	
17	31,950,000	32,000,000	19.77	
17	39,950,000	40,000,000	14.50	
17	48,950,000	49,000,000	15.04	
18	5,250,000	5,300,000	16.42	*CLEC3A, WWOX*
18	34,350,000	34,400,000	15.05	*BEAN1, TK2, CKLF*
18	35,550,000	35,600,000	13.35	*DPEP3, SLC12A4, DPEP2*
18	44,950,000	45,000,000	13.21	*GPI, KIAA0355*
18	49,000,000	49,050,000	14.10	
18	52,450,000	52,500,000	14.68	
18	53,500,000	53,550,000	13.28	*FOSB, RTN2, PPM1N, VASP, OPA3*
19	500,000	550,000	15.95	
19	14,700,000	14,750,000	13.76	
19	59,100,000	59,150,000	20.48	
19	59,150,000	59,200,000	15.40	
20	15,000,000	15,050,000	14.49	
20	33,400,000	33,450,000	14.35	*C6*
20	38,000,000	38,050,000	15.55	*NADK2, RANBP3L*
20	69,150,000	69,200,000	15.66	
21	200,000	250,000	12.97	
21	7,650,000	7,700,000	14.42	*TTC23*
21	8,450,000	8,500,000	13.61	
21	18,350,000	18,400,000	15.68	
21	24,700,000	24,750,000	14.10	*ADAMTSL3*
21	24,750,000	24,800,000	16.44	
21	24,800,000	24,850,000	15.10	
21	24,850,000	24,900,000	15.93	
21	29,950,000	30,000,000	12.97	
21	30,750,000	30,800,000	15.61	*OTUD7A*
21	34,300,000	34,350,000	17.10	*CYP1A2, CYP1A1*
21	34,950,000	35,000,000	14.24	*ISLR2, PML*
21	35,200,000	35,250,000	15.02	
21	35,250,000	35,300,000	14.77	*GZMB*
21	35,550,000	35,600,000	14.89	*STXBP6*
21	36,850,000	36,900,000	17.34	
21	37,350,000	37,400,000	13.91	
21	37,400,000	37,450,000	16.53	
21	38,850,000	38,900,000	12.96	
21	46,050,000	46,100,000	17.56	*NFKBIA*

Appendix 6C. Continuation

21	61,200,000	61,250,000	15.16	
21	62,150,000	62,200,000	27.42	
21	62,300,000	62,350,000	21.83	
21	62,800,000	62,850,000	21.80	*ATG2B*
21	62,850,000	62,900,000	27.27	*GSKIP, AK7*
21	68,950,000	69,000,000	13.71	*RCOR1*
23	2,250,000	2,300,000	14.25	
23	8,750,000	8,800,000	21.30	*UHRF1BP1*
23	8,800,000	8,850,000	24.96	*TAF11, ANKS1A*
23	14,500,000	14,550,000	17.03	
23	14,550,000	14,600,000	16.47	
23	14,600,000	14,650,000	17.46	
23	14,750,000	14,800,000	19.68	
23	14,800,000	14,850,000	17.17	
23	14,850,000	14,900,000	16.03	
23	19,100,000	19,150,000	13.20	
23	22,500,000	22,550,000	14.57	
23	22,600,000	22,650,000	19.98	
23	22,650,000	22,700,000	18.54	
23	22,700,000	22,750,000	14.02	
23	22,750,000	22,800,000	12.94	
23	30,450,000	30,500,000	13.22	
23	32,100,000	32,150,000	13.11	*CARMIL1*
23	50,150,000	50,200,000	14.83	
23	50,200,000	50,250,000	17.09	*SLC22A23*
23	50,750,000	50,800,000	13.21	*MYLK4*
24	3,600,000	3,650,000	14.02	
24	28,300,000	28,350,000	23.39	
24	28,400,000	28,450,000	16.81	
24	31,650,000	31,700,000	14.47	
24	31,700,000	31,750,000	13.13	
24	34,800,000	34,850,000	16.64	*MIB1*
24	34,850,000	34,900,000	16.39	
24	35,950,000	36,000,000	13.03	*ENOSF1, YES1*
24	36,050,000	36,100,000	13.18	
24	36,100,000	36,150,000	12.86	*ADCYAP1*
24	54,950,000	55,000,000	12.99	
25	1,250,000	1,300,000	14.06	*CRAMP1, JPT2*
25	1,300,000	1,350,000	12.70	*NME3, MAPK8IP3*
25	12,900,000	12,950,000	19.30	*ERCC4*
25	12,950,000	13,000,000	16.97	
25	13,200,000	13,250,000	13.88	*MRTFB*

Appendix 6C. Continuation

25	13,250,000	13,300,000	13.13	
25	19,050,000	19,100,000	13.32	*DNAH3*
25	24,600,000	24,650,000	19.85	
25	24,650,000	24,700,000	19.72	
25	24,700,000	24,750,000	14.79	
25	24,750,000	24,800,000	15.61	
25	24,800,000	24,850,000	12.91	
25	33,700,000	33,750,000	12.87	*LAT2, EIF4H*
25	33,750,000	33,800,000	17.38	*LIMK1, ELN*
25	37,200,000	37,250,000	14.43	*CYP3A5*
26	3,350,000	3,400,000	18.47	
26	3,400,000	3,450,000	38.40	
26	28,400,000	28,450,000	15.12	
26	33,450,000	33,500,000	12.98	
26	33,700,000	33,750,000	15.62	
26	33,950,000	34,000,000	28.33	
26	34,050,000	34,100,000	23.14	
26	34,100,000	34,150,000	19.76	
26	34,600,000	34,650,000	13.60	*NHLRC2*
26	35,600,000	35,650,000	12.84	*FAM160B1*
26	39,450,000	39,500,000	14.63	
26	41,150,000	41,200,000	14.20	
27	6,300,000	6,350,000	14.98	*GPM6A*
27	8,450,000	8,500,000	13.08	
27	16,800,000	16,850,000	15.04	
27	21,150,000	21,200,000	15.50	
27	22,450,000	22,500,000	13.83	
27	41,250,000	41,300,000	13.10	
27	42,650,000	42,700,000	14.15	
27	42,700,000	42,750,000	13.98	
27	42,750,000	42,800,000	13.82	
27	43,050,000	43,100,000	13.85	
27	44,350,000	44,400,000	13.20	
27	44,600,000	44,650,000	12.65	
27	44,850,000	44,900,000	17.51	
27	44,900,000	44,950,000	12.90	
27	44,950,000	45,000,000	13.05	
28	23,300,000	23,350,000	13.08	*CTNNA3*
28	29,450,000	29,500,000	14.74	*DNAJC9, FAM149B1*
28	29,550,000	29,600,000	17.17	*MSS51, ANXA7*
28	29,600,000	29,650,000	22.27	*USP54, PPP3CB*
28	29,650,000	29,700,000	15.06	

Appendix 6C. Continuation

28	29,800,000	29,850,000	16.15	*SYNPO2L, SEC24C, FUT11*
28	29,850,000	29,900,000	15.40	*NDST2, CAMK2G*
28	35,500,000	35,550,000	17.71	
29	7,550,000	7,600,000	14.11	*RAB38*
29	27,100,000	27,150,000	15.75	

[1] BTA: *Bos taurus* autosome

Appendix 7C. Runs of homozygosity (ROH) hotspots for Gir (GIR), Caracu Caldeano (CAR), Crioulo Lageano (CRL), and Pantaneiro (PAN) cattle breeds.

Breed	BTA[1]	Start (bp)	End (bp)	Length (bp)	*n*[2]
GIR	1	8,009,125	8,161,172	152,047	7
GIR	1	8,161,173	8,161,316	143	6
GIR	1	8,161,317	8,351,676	190,359	7
GIR	1	8,351,677	8,351,753	76	6
GIR	1	8,351,754	8,523,590	171,836	7
GIR	5	47,981,719	48,094,216	112,497	6
GIR	7	51,862,717	52,407,969	545,252	6
GIR	7	52,407,970	52,610,293	202,323	7
GIR	11	11,810,949	11,810,958	9	6
GIR	11	11,810,959	11,818,746	7,787	7
GIR	11	11,818,747	12,386,367	567,620	8
GIR	11	12,386,368	12,386,770	402	6
GIR	11	61,381,295	61,540,581	159,286	6
GIR	12	28,859,813	29,230,533	370,720	8
GIR	12	29,230,534	29,438,853	208,319	7
GIR	13	50,334,994	50,712,729	377,735	7
GIR	13	50,712,730	50,999,757	287,027	8
GIR	13	50,999,758	50,999,908	150	7
GIR	13	50,999,909	51,008,138	8,229	8
GIR	13	51,008,139	51,233,528	225,389	7
GIR	13	51,233,529	51,233,569	40	6
GIR	13	51,233,570	51,542,914	309,344	7
GIR	13	51,550,903	51,579,041	28,138	7
GIR	13	51,579,042	51,811,696	232,654	6
GIR	15	40,143,056	40,339,106	196,050	6
GIR	18	14,030,008	14,030,892	884	6
GIR	18	14,030,893	14,042,615	11,722	7
GIR	18	14,042,616	14,547,319	504,703	8
GIR	18	14,547,320	14,562,105	14,785	7
GIR	18	14,562,106	14,669,281	107,175	6
GIR	21	39,765,789	40,280,058	514,269	7
GIR	22	23,981,119	24,153,382	172,263	6
GIR	25	35,817,478	36,090,255	272,777	6
GIR	25	36,090,547	36,264,728	174,181	6
CAR	1	31,196,696	31,766,947	570,251	6
CAR	1	31,767,052	31,819,565	52,513	6
CAR	1	31,819,591	31,890,922	71,331	6
CAR	1	40,145,004	40,354,391	209,387	6

Appendix 7C. Continuation

CAR	1	40,397,295	40,554,725	157,430	6
CAR	1	40,842,012	40,859,667	17,655	6
CAR	1	40,859,668	41,041,955	182,287	7
CAR	1	41,041,956	41,214,884	172,928	6
CAR	1	41,372,977	41,438,665	65,688	6
CAR	1	41,438,791	41,926,758	487,967	6
CAR	1	42,531,526	43,080,442	548,916	6
CAR	1	65,445,527	65,775,494	329,967	6
CAR	1	107,715,631	107,762,628	46,997	6
CAR	1	112,007,588	112,894,640	887,052	6
CAR	1	112,896,656	113,258,608	361,952	7
CAR	1	113,258,609	113,258,810	201	6
CAR	1	113,258,811	113,411,032	152,221	7
CAR	1	114,378,068	114,398,377	20,309	7
CAR	1	114,398,378	114,573,918	175,540	6
CAR	1	114,574,168	114,624,249	50,081	6
CAR	1	127,003,536	127,048,180	44,644	6
CAR	1	131,272,996	131,779,541	506,545	6
CAR	1	139,370,703	139,989,186	618,483	6
CAR	2	2,005,842	2,184,901	179,059	6
CAR	2	122,146,022	122,155,885	9,863	6
CAR	2	122,155,886	122,179,878	23,992	8
CAR	2	122,179,879	122,181,649	1,770	9
CAR	2	122,181,650	122,181,755	105	10
CAR	2	122,181,756	122,261,385	79,629	11
CAR	2	122,261,386	122,261,588	202	10
CAR	2	122,261,589	122,679,462	417,873	11
CAR	2	122,679,463	122,684,312	4,849	10
CAR	2	122,684,313	122,700,477	16,164	9
CAR	2	122,700,478	122,717,442	16,964	8
CAR	2	122,717,443	122,794,571	77,128	7
CAR	2	122,794,572	122,822,417	27,845	6
CAR	3	37,395,143	37,451,695	56,552	6
CAR	3	37,451,754	37,553,730	101,976	6
CAR	3	38,872,099	39,545,921	673,822	6
CAR	3	39,908,304	40,053,907	145,603	6
CAR	3	72,570,558	73,011,972	441,414	6
CAR	3	75,342,655	75,477,285	134,630	6
CAR	3	75,637,207	75,973,113	335,906	6
CAR	3	76,987,264	77,188,457	201,193	6
CAR	3	77,188,792	77,488,414	299,622	6
CAR	3	96,794,337	96,798,358	4,021	6

Appendix 7C. Continuation

CAR	4	82,720,483	83,209,388	488,905	6
CAR	4	83,684,877	83,699,651	14,774	6
CAR	4	83,699,652	84,205,537	505,885	7
CAR	4	84,205,538	84,206,754	1,216	6
CAR	5	19,599,748	19,777,874	178,126	6
CAR	5	19,813,150	20,040,300	227,150	6
CAR	5	24,533,038	24,533,237	199	6
CAR	5	24,533,238	24,797,751	264,513	7
CAR	5	24,797,752	25,311,116	513,364	8
CAR	5	25,311,117	25,372,627	61,510	6
CAR	5	25,547,304	25,559,218	11,914	6
CAR	5	25,559,219	25,559,281	62	7
CAR	5	25,559,282	26,568,749	1,009,467	6
CAR	5	26,570,526	26,964,250	393,724	6
CAR	5	30,155,125	30,768,544	613,419	7
CAR	5	30,768,545	31,010,198	241,653	8
CAR	5	31,764,550	32,531,884	767,334	8
CAR	5	32,531,885	32,532,083	198	6
CAR	5	34,093,535	34,156,848	63,313	6
CAR	5	34,156,849	34,347,247	190,398	7
CAR	5	38,226,162	38,761,636	535,474	9
CAR	5	38,761,637	38,761,745	108	8
CAR	6	33,367,500	33,814,384	446,884	7
CAR	6	33,814,385	33,825,533	11,148	6
CAR	7	20,828,563	21,299,326	470,763	6
CAR	7	50,817,400	50,976,273	158,873	6
CAR	7	50,976,304	50,992,075	15,771	6
CAR	7	50,992,076	51,121,550	129,474	7
CAR	7	51,121,551	51,121,584	33	6
CAR	7	51,121,585	51,245,385	123,800	7
CAR	7	51,245,386	51,519,790	274,404	6
CAR	7	51,601,408	51,858,394	256,986	6
CAR	7	51,859,274	51,862,716	3,442	6
CAR	7	51,862,717	52,421,216	558,499	7
CAR	7	56,530,662	56,796,839	266,177	6
CAR	7	56,797,041	56,870,746	73,705	6
CAR	7	56,870,747	56,924,413	53,666	7
CAR	7	56,924,414	57,031,791	107,377	8
CAR	7	57,034,767	57,035,036	269	6
CAR	7	57,035,037	57,587,281	552,244	7
CAR	7	58,992,195	58,995,658	3,463	6
CAR	7	58,995,659	59,811,492	815,833	7

Appendix 7C. Continuation

CAR	7	59,811,493	59,811,574	81	6
CAR	7	59,811575	59,955,402	143,827	7
CAR	7	59,955,403	60,881,686	926,283	6
CAR	7	60,882,025	61,028,980	146,955	6
CAR	7	61,029,088	61,350,702	321,614	6
CAR	7	62,798,166	62,999,316	201,150	6
CAR	7	62,999,317	63,260,902	261,585	7
CAR	7	63,260,903	63,850,608	589,705	6
CAR	7	63,859,698	63,979,201	119,503	6
CAR	7	64,554,630	64,859,219	304,589	6
CAR	8	6,075,754	6,284,776	209,022	6
CAR	8	6,284,834	6,968,436	683,602	6
CAR	8	7,165,430	8,176,301	1,010,871	6
CAR	8	8,424,812	8,516,297	91,485	6
CAR	8	10,487,314	10,924,528	437,214	6
CAR	8	10,924,529	11,220,806	296,277	7
CAR	8	15,720,454	15,720,503	49	6
CAR	8	15,720,504	15,756,354	35,850	7
CAR	8	15,756,355	15,756,400	45	6
CAR	8	15,756,401	15,758,462	2,061	7
CAR	8	15,758,463	16,058,715	300,252	6
CAR	8	16,058,940	16,168,528	109,588	6
CAR	8	80,266,390	80,286,169	19,779	6
CAR	8	80,286,170	80,658,566	372,396	7
CAR	8	80,658,567	80,940,908	282,341	6
CAR	8	102,878,448	102,963,154	84,706	6
CAR	9	3,973,619	4,202,200	228,581	6
CAR	10	46,063,098	46,670,321	607,223	8
CAR	10	46,670,322	46,671,018	696	7
CAR	10	46,671,019	46,671,850	831	6
CAR	10	52,022,344	52,663,085	640,741	6
CAR	10	52,663,086	53,332,592	669,506	7
CAR	10	53,332,593	53,332,622	29	6
CAR	10	53,332,623	53,356,914	24,291	7
CAR	10	53,356,915	53,359,631	2,716	6
CAR	10	53,359,632	53,823,584	463,952	7
CAR	10	53,823,585	53,825,571	1,986	6
CAR	10	54,553,448	54,553,637	189	6
CAR	10	54,553,638	54,808,123	254,485	7
CAR	10	54,808,124	55,354,998	546,874	9
CAR	10	55,354,999	55,357,584	2,585	7
CAR	10	55,357,585	55,642,643	285,058	6

Appendix 7C. Continuation

CAR	10	68,841,576	69,422,504	580,928	6
CAR	11	10,682,899	11,022,535	339,636	6
CAR	11	11,548,113	12,131,519	583,406	6
CAR	11	12,397,970	12,465,465	67,495	6
CAR	11	56,854,996	56,866,347	11,351	6
CAR	11	56,866,348	56,867,733	1,385	7
CAR	11	56,867,734	56,988,771	121,037	8
CAR	11	56,988,772	57,327,185	338,413	7
CAR	11	57,327,186	57,430,866	103,680	8
CAR	11	57,430,867	57,485,957	55,090	6
CAR	11	67,347,892	67,348,449	557	6
CAR	11	67,348,450	67,568,020	219,570	7
CAR	11	67,568,021	67,836,384	268,363	6
CAR	11	67,836,500	68,709,496	872,996	6
CAR	11	78,438,393	78,666,057	227,664	6
CAR	12	813,538	813,775	237	6
CAR	12	25,849,855	26,092,231	242,376	6
CAR	12	26,092,232	26,794,634	702,402	7
CAR	12	28,077,251	28,628,347	551,096	6
CAR	12	28,692,283	29,232,632	540,349	6
CAR	12	29,721,403	29,885,850	164,447	6
CAR	12	29,885,851	30,468,100	582,249	7
CAR	12	30,468,101	30,907,679	439,578	6
CAR	12	37,776,618	38,484,490	707,872	8
CAR	12	38,484,491	38,796,087	311,596	7
CAR	12	38,796,088	38,812,167	16,079	6
CAR	13	64,202,501	64,695,379	492,878	6
CAR	14	1,424,815	2,016,241	591,426	6
CAR	14	23,956,515	23,965,224	8,709	6
CAR	14	27,842,552	27,844,672	2,120	6
CAR	14	33,672,659	33,926,829	254,170	6
CAR	14	33,926,830	34,537,194	610,364	7
CAR	14	34,537,195	34,667,602	130,407	6
CAR	14	36,762,280	36,837,671	75,391	6
CAR	14	36,852,214	36,854,841	2,627	6
CAR	14	36,854,842	36,879,323	24,481	7
CAR	14	36,879,324	36,941,740	62,416	6
CAR	14	36,941,741	37,358,113	416,372	7
CAR	14	37,358,114	37,416,117	58,003	6
CAR	14	52,487,754	52,597,635	109,881	6
CAR	14	52,597,763	52,914,848	317,085	6
CAR	15	6,324,230	7,195,975	871,745	6

Appendix 7C. Continuation

CAR	15	9,545,866	9,573,960	28,094	7
CAR	15	9,573,961	9,597,755	23,794	8
CAR	15	9,597,756	9,600,452	2,696	9
CAR	15	9,600,453	9,726,133	125,680	8
CAR	15	9,726,134	9,801,691	75,557	7
CAR	15	9,801,692	10,068,052	266,360	8
CAR	15	10,068,053	10,135,065	67,012	7
CAR	15	10,135,066	10,135,103	37	6
CAR	15	10,135,104	10,214,897	79,793	7
CAR	15	10,214,898	10,215,039	141	6
CAR	15	10,215,040	10,263,084	48,044	7
CAR	15	10,263,085	10,781,621	518,536	8
CAR	15	10,781,622	11,019,849	238,227	7
CAR	15	11,019,850	11,020,294	444	6
CAR	15	11,826,702	11,842,542	158,40	7
CAR	15	11,842,543	12,074,345	231,802	8
CAR	15	12,074,346	12,114,984	40,638	9
CAR	15	12,114,985	12,419,869	304,884	8
CAR	15	12,419,870	12,545,569	125,699	7
CAR	15	12,545,570	12,545,657	87	6
CAR	15	12,545,658	12,654,057	108,399	7
CAR	15	12,654,058	12,775,732	121,674	8
CAR	15	12,775,733	13,196,628	420,895	7
CAR	15	13,196,629	13,196,676	47	6
CAR	15	13,196,677	13,435,825	239,148	7
CAR	15	14,319,954	14,452,802	132,848	6
CAR	15	17,891,119	18,349,908	458,789	6
CAR	15	24,440,233	24,507,340	67,107	6
CAR	15	24,507,462	24,533,762	26,300	6
CAR	15	30,642,603	30,884,582	241,979	6
CAR	15	30,884,646	31,026,610	141,964	6
CAR	15	32,378,196	32,412,580	34,384	6
CAR	15	32,412,581	33,043,764	631,183	7
CAR	15	33,043,765	33,188,677	144,912	6
CAR	15	33,256,299	33,811,495	555,196	6
CAR	15	33,814,024	33,889,227	75,203	6
CAR	15	35,365,655	35,645,332	279,677	6
CAR	15	40,280,681	40,621,174	340,493	7
CAR	15	40,621,175	40,990,058	368,883	8
CAR	16	45,362,028	45,375,191	13,163	6
CAR	16	45,375,192	45,375,207	15	7
CAR	16	45,375,208	45,604,259	229,051	8

Appendix 7C. Continuation

CAR	16	45,604,260	45,686,021	81,761	9
CAR	16	45,686,022	45,937,552	251,530	10
CAR	16	45,937,553	45,943,769	6,216	9
CAR	16	45,943,770	45,960,024	16,254	8
CAR	16	45,960,025	45,986,332	26,307	7
CAR	16	66,908,951	67,272,163	363,212	6
CAR	16	67,272,164	67,449,284	177,120	7
CAR	16	67,449,285	67,778,714	329,429	6
CAR	17	35,257,618	35,362,484	104,866	6
CAR	17	35,554,519	36,099,448	544,929	6
CAR	17	57,750,909	57,775,729	24,820	7
CAR	17	57,775,730	58,353,512	577,782	9
CAR	17	58,353,513	58,378,739	25,226	8
CAR	17	58,378,740	58,395,672	16,932	7
CAR	18	13,399,574	13,974,237	574,663	7
CAR	18	13,974,238	14,710,261	736,023	8
CAR	18	14,710,262	14,710,298	36	7
CAR	18	14,712,548	14,712,620	72	8
CAR	18	14,712,621	15,237,691	525,070	8
CAR	18	15,237,692	15,242,733	5,041	6
CAR	18	15,242,734	15,246,177	3,443	7
CAR	18	15,246,178	15,420,066	173,888	8
CAR	18	15,420,067	15,483,091	63,024	9
CAR	18	15,483,092	15,950,073	466,981	10
CAR	18	15,950,074	15,950,183	109	9
CAR	18	15,950,184	16,005,371	55,187	10
CAR	18	16,005,372	16,193,358	187,986	9
CAR	18	16,193,359	16,193,363	4	8
CAR	18	16,193,364	16,213,223	19,859	7
CAR	18	16,213,224	16,238,851	25,627	6
CAR	18	17,726,503	17,790,887	64,384	6
CAR	18	17,790,954	17,831,279	40,325	6
CAR	18	30,932,697	31,246,807	314,110	6
CAR	18	31,496,321	31,556,726	60,405	6
CAR	18	34,718,675	35,481,561	762,886	6
CAR	19	8,243,042	8,485,703	242,661	6
CAR	19	13,619,635	13,619,703	68	6
CAR	19	13,619,704	14,045,588	425,884	7
CAR	19	14,045,589	14,170,579	124,990	6
CAR	20	37,181,866	38,196,797	1,014,931	6
CAR	20	38,196,798	38,245,558	48,760	7
CAR	20	38,245,559	38,245,680	121	6

Appendix 7C. Continuation

CAR	20	38,245,681	38,245,850	169	7
CAR	20	38,245,851	38,824,729	578,878	8
CAR	20	38,824,730	38,925,477	100,747	6
CAR	20	38,925,478	38,989,472	63,994	8
CAR	20	38,989,473	38,989,524	51	7
CAR	20	38,989,525	39,074,324	84,799	8
CAR	20	39,074,325	39,074,448	123	7
CAR	20	39,074,449	39,267,427	192,978	8
CAR	20	39,267,428	39,267,756	328	7
CAR	20	39,267,757	39,277,399	9,642	8
CAR	20	39,277,400	39,654,239	376,839	7
CAR	20	39,654,240	39,654,385	145	6
CAR	20	39,654,386	40,448,675	794,289	7
CAR	20	40,448,676	40,448,698	22	6
CAR	20	40,448,699	40,774,734	326,035	7
CAR	20	40,774,735	40,832,180	57,445	6
CAR	20	40,832,388	40,929,100	96,712	6
CAR	20	40,929,101	41,127,598	198,497	7
CAR	20	41,127,599	41,329,260	201,661	6
CAR	20	41,330,024	41,592,059	262,035	6
CAR	20	41,592,101	42,963,673	1,371,572	7
CAR	20	42,963,674	43,013,603	49,929	6
CAR	21	128	405,288	405,160	7
CAR	21	405,289	527,579	122,290	8
CAR	21	527,580	678,853	151,273	7
CAR	21	678,854	701,115	22,261	6
CAR	21	2,494,700	2,888,773	394,073	6
CAR	21	2,889,071	3,417,065	527,994	6
CAR	21	5,934,125	5,934,255	130	7
CAR	21	5,934,256	6,109,014	174,758	8
CAR	21	6,109,015	6,109,027	12	9
CAR	21	6,109,028	6,143,313	34,285	10
CAR	21	6,143,314	6,625,173	481,859	11
CAR	21	32,085,986	32,489,360	403,374	6
CAR	21	63,180,480	63,672,252	491,772	7
CAR	21	63,672,253	63,672,302	49	6
CAR	21	69,497,866	70,227,464	729,598	7
CAR	21	70,227,465	70,387,387	159,922	6
CAR	25	268,557	923,374	654,817	6
CAR	27	18,102,351	18,409,057	306,706	6
CAR	28	44,002,081	44,139,843	137,762	6
CAR	28	44,139,844	44,140,364	520	7

Appendix 7C. Continuation

CAR	28	44,140,365	44,412,320	271,955	8
CAR	28	44,412,321	44,412,344	23	6
CRL	7	51,169,994	51,232,930	62,937	6
CRL	7	51,232,931	51,457,547	224,617	7
CRL	7	51,457,548	51,518,446	60,899	8
CRL	7	51,518,447	51,616,914	98,468	7
CRL	7	51,616,915	51,854,215	237,301	8
CRL	7	51,854,216	51,855,512	1,297	7
CRL	7	51,855,513	51,855,687	175	6
CRL	7	51,858,972	51,860,157	1,186	6
CRL	7	51,861,003	51,862,691	1,689	6
CRL	7	51,862,692	51,862,716	25	7
CRL	7	51,862,717	52,421,216	558,500	8
CRL	7	52,421,217	52,594,780	173,564	7
CRL	7	52,594,781	52,596,240	1,460	6
CRL	16	45,266,368	45,363,085	96,718	6
CRL	16	45,363,086	45,942,100	579,015	7
CRL	16	45,942,101	45,968,573	26,473	6
PAN	2	42,661,906	42,920,413	258,508	6
PAN	9	39,491,705	39,508,902	17,198	6
PAN	9	39,508,903	39,679,487	170,585	7
PAN	9	39,679,488	39,816,061	136,574	6
PAN	9	39,816,062	40,019,264	203,203	6
PAN	10	54,791,922	54,792,371	450	6
PAN	10	54,792,456	55,077,956	285,501	6
PAN	16	45,393,866	45,677,279	283,414	6
PAN	20	2,961,920	3,538,471	576,552	6
PAN	20	3,538,601	3,642,020	103,420	6
PAN	20	14,189,696	14,515,193	325,498	6
PAN	25	1,341,386	1,345,563	4,178	6
PAN	25	1,345,564	1,384,740	39,177	6
PAN	25	1,384,741	1,415,096	30,356	7
PAN	25	1,415,097	1,872,379	457,283	6
PAN	25	1,872,380	1,890,276	17,897	6

[1] BTA: *Bos taurus* autosome

[2] *n*= Number of animals sharing the same runs of homozygosity (ROH) region

Appendix 8C. Overlapping of the putative sweep regions identified from the top 1% of the within-population DCMS statistic with candidate regions under positive selection previously reported in other cattle populations.

BTA[1]	Start (bp)	End (bp)	Reference[2]
1	6,187,555	6,194,716	Somavilla et al. 2014
1	8,300,000	8,350,000	Iso-Touru et al. 2016 Xu et al. 2015
1	18,150,000	18,200,000	Boitard et al. 2016
1	82,900,000	82,949,999	Boitard et al. 2016
1	82,950,001	82,960,514	Boitard et al. 2016
1	111,450,000	111,500,000	Stella et al. 2010
1	111,600,000	111,650,000	Stella et al. 2010
1	112,250,000	112,300,000	Boitard et al. 2016
2	4,400,000	4,450,000	González-Rodríguez et al. 2016
2	9,500,000	9,550,000	González-Rodríguez et al. 2016
2	10,200,000	10,250,000	González-Rodríguez et al. 2016
2	73,200,000	73,250,000	González-Rodríguez et al. 2016
2	98,300,000	98,350,000	Stella et al. 2010
2	101,950,000	101,991,288	Xu et al. 2015
2	103,100,035	103,112,630	Mei et al. 2018
2	103,112,631	103,142,370	Boitard et al. 2016 Mei et al. 2018
2	103,142,371	103,149,975	Mei et al. 2018
3	44,433,212	44,449,999	Stella et al. 2010
3	44,450,001	44,499,999	Stella et al. 2010
3	44,500,001	44,549,999	Stella et al. 2010
3	44,550,001	44,600,000	Stella et al. 2010
3	55,417,575	55,450,000	Xu et al. 2015
3	103,450,047	103,499,908	Mei et al. 2018
3	105,500,001	105,512,029	Wang et al. 2019 Stella et al. 2010
3	105,512,030	105,550,000	Stella et al. 2010 Wang et al. 2019 Xu et al. 2015
4	12,200,000	12,221,673	Xu et al. 2015
4	12,221,674	12,250,000	Xu et al. 2015 Boitard et al. 2016
4	61,600,000	61,649,999	Boitard et al. 2016
4	61,650,001	61,700,000	Boitard et al. 2016
4	69,500,000	69,550,000	Rothammer et al. 2013
4	114,700,000	114,725,049	Iso-Touru et al. 2016

Appendix 8C. Continuation

4	114,725,050	114,750,000	Iso-Touru et al. 2016 Kim et al. 2017
4	115,150,000	115,199,999	Iso-Touru et al. 2016
4	115,200,001	115,250,000	Iso-Touru et al. 2016
4	115,550,000	115,565,142	Iso-Touru et al. 2016
4	116,150,000	116,151,004	Boitard et al. 2016 O`Brien et al. 2014 Stella et al. 2010
4	116,151,005	116,200,000	Stella et al. 2010
4	116,450,000	116,450,012	Stella et al. 2010
4	116,450,013	116,499,998	Stella et al. 2010 Mei et al. 2018
4	116,499,999	116,500,000	Stella et al. 2010 Mei et al. 2018
4	116,650,000	116,699,972	Stella et al. 2010 Iso-Touru et al. 2016
4	116,699,973	116,700,000	Stella et al. 2010 Iso-Touru et al. 2016
4	116,750,000	116,800,000	Stella et al. 2010 Iso-Touru et al. 2016
5	6,650,000	6,699,999	Wang et al. 2019
5	36,700,109	36,749,697	Mei et al. 2018
5	46,400,000	46,450,000	Xu et al. 2015
5	55,650,000	55,700,000	Xu et al. 2015
5	61,500,000	61,549,999	Xu et al. 2015
5	61,550,001	61,577,603	Xu et al. 2015
5	66,400,000	66,400,298	Iso-Touru et al. 2016
5	66,400,299	66,449,929	Iso-Touru et al. 2016 Mei et al. 2018
5	66,449,930	66,449,999	Iso-Touru et al. 2016
5	66,450,001	66,499,999	Iso-Touru et al. 2016
5	66,500,001	66,550,000	Iso-Touru et al. 2016
6	13,600,017	13,649,974	Mei et al. 2018
6	37,650,000	37,700,000	Rothammer et al. 2013
6	46,000,000	46,050,000	Iso-Touru et al. 2016
6	64,950,000	65,000,000	Stella et al. 2010
6	70,700,000	70,725,190	Rothammer et al. 2013
6	70,725,191	70,729,074	Kim et al. 2017 Rothammer et al. 2013
6	70,729,075	70,750,000	Rothammer et al. 2013 Xu et al. 2015 Kim et al. 2017

Appendix 8C. Continuation

6	78,227,807	78,250,000	Xu et al. 2015
7	69,500,056	69,549,984	Mei et al. 2018
7	79,250,000	79,300,000	Iso-Touru et al. 2016
8	21,450,000	21,500,000	Xu et al. 2015
8	54,750,000	54,799,999	Iso-Touru et al. 2016
8	54,800,001	54,850,000	Iso-Touru et al. 2016
8	59,250,000	59,250,007	Xu et al. 2015
8	59,250,008	59,299,967	Xu et al. 2015 Mei et al. 2018
8	59,299,968	59,300,000	Xu et al. 2015
8	104,000,000	104,050,000	Stella et al. 2010
9	24,700,523	24,750,000	Stella et al. 2010
9	43,800,000	43,850,000	Rothammer et al. 2013
9	45,600,000	45,630,507	Xu et al. 2015 Rothammer et al. 2013
9	45,630,509	45,650,000	Xu et al. 2015 Rothammer et al. 2013
9	45,750,000	45,800,000	Xu et al. 2015 Rothammer et al. 2013
9	52,000,000	52,007,228	Iso-Touru et al. 2016 Rothammer et al. 2013
9	52,007,229	52,049,999	Rothammer et al. 2013
9	52,050,001	52,100,000	Rothammer et al. 2013
9	98,650,000	98,699,999	Stella et al. 2010
9	98,700,001	98,750,000	Stella et al. 2010
10	13,200,056	13,250,000	Mei et al. 2018
10	33,900,000	33,908,723	Zhao et al. 2015
10	33,908,724	33,950,000	Xu et al. 2015 Zhao et al. 2015
10	35,250,000	35,274,738	Iso-Touru et al. 2016 Zhao et al. 2015
10	35,274,739	35,300,000	Iso-Touru et al. 2016
10	38,500,000	38,526,183	Xu et al. 2015
10	38,526,184	38,550,000	Xu et al. 2015 Kim et al. 2017
10	69,900,000	69,949,999	Boitard et al. 2016
10	69,950,001	69,950,459	Boitard et al. 2016
10	76,650,000	76,676,183	Kim et al. 2017
10	76,676,185	76,700,000	Kim et al. 2017
10	103,150,000	103,176,183	Iso-Touru et al. 2016
10	103,176,184	103,200,000	Kim et al. 2017 Iso-Touru et al. 2016

Appendix 8C. Continuation

11	44,250,000	44,250,014	Xu et al. 2015
11	44,250,015	44,300,000	Xu et al. 2015 Mei et al. 2018
11	48,350,000	48,400,000	Kim et al. 2017
11	66,000,000	66,050,000	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	69,150,000	69,200,000	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	72,150,000	72,200,000	González-Rodríguez et al. 2016
11	80,850,000	80,900,000	Xu et al. 2015
11	85,750,000	85,800,000	Iso-Touru et al. 2016
13	4,400,000	4,450,000	Boitard et al. 2016
13	6,313,308	6,350,000	Iso-Touru et al. 2016
13	11,600,000	11,650,000	Xu et al. 2015
13	48,250,000	48,300,000	Stella et al. 2010
13	63,000,000	63,050,000	Stella et al. 2010
13	63,100,000	63,146,419	Stella et al. 2010
13	63,146,420	63,150,000	Stella et al. 2010 Xu et al. 2015
13	78,000,001	78,050,000	Liao et al. 2013
14	550,000	600,000	Kim et al. 2017
14	25,950,000	25,958,784	Pitt et al. 2019 O`Brien et al. 2014
14	25,958,785	25,959,446	Pitt et al. 2019 Boitard et al. 2016
14	25,959,447	26,000,000	Pitt et al. 2019
14	26,700,000	26,750,000	Pitt et al. 2019
14	29,950,000	30,000,000	Wang et al. 2019
15	5,300,000	5,350,000	Stella et al. 2010
15	5,450,000	5,500,000	Stella et al. 2010
15	36,750,000	36,797,454	Stella et al. 2010 Zhao et al. 2015
15	36,797,455	36,799,999	Stella et al. 2010
15	36,800,001	36,850,000	Stella et al. 2010
15	63,450,000	63,497,946	Boitard et al. 2016
15	64,400,061	64,449,981	Mei et al. 2018
16	41,384,834	41,400,000	Zhao et al. 2015
16	64,350,000	64,400,000	Iso-Touru et al. 2016
17	7,050,000	7,100,000	Stella et al. 2010
17	37,646,557	37,649,999	Stella et al. 2010
17	37,650,001	37,700,000	Stella et al. 2010
17	68,450,000	68,500,000	Iso-Touru et al. 2016

Appendix 8C. Continuation

17	68,550,000	68,581,654	Iso-Touru et al. 2016
17	68,581,656	68,600,000	Iso-Touru et al. 2016
18	25,350,000	25,386,867	Rothammer et al. 2013
18	32,950,000	33,000,000	Wang et al. 2019
18	33,472,642	33,500,000	Boitard et al. 2016
19	2,571,047	2,600,000	Xu et al. 2015
19	25,050,000	25,100,000	Stella et al. 2010 Liao et al. 2013
19	27,550,000	27,599,999	Bahbahani et al. 2015 Mei et al. 2018
19	27,600,001	27,600,086	Liao et al. 2013 Bahbahani et al. 2015 Mei et al. 2018
19	27,600,087	27,649,853	Liao et al. 2013 Bahbahani et al. 2015 Mei et al. 2018
19	27,649,854	27,650,000	Liao et al. 2013 Bahbahani et al. 2015 Mei et al. 2018
20	13,900,000	13,949,999	Stella et al. 2010 Xu et al. 2015
20	13,950,001	14,000,000	Stella et al. 2010 Xu et al. 2015
20	20,500,000	20,550,000	Xu et al. 2015
20	32,342,891	32,350,000	O`Brien et al. 2014 Boitard et al. 2016
20	33,850,000	33,900,000	Stella et al. 2010
20	34,000,000	34,049,999	Stella et al. 2010
20	34,050,001	34,099,999	Stella et al. 2010
20	34,100,001	34,149,999	Stella et al. 2010 Wang et al. 2019
20	34,150,001	34,200,000	Stella et al. 2010 Wang et al. 2019
20	47,450,000	47,500,000	Xu et al. 2015
21	1,650,000	1,658,789	Xu et al. 2015 Mei et al. 2018
21	1,658,790	1,700,000	Mei et al. 2018
21	2,150,000	2,200,000	Xu et al. 2015
21	6,550,000	6,591,118	Boitard et al. 2016
21	12,350,000	12,400,000	Stella et al. 2010
21	12,450,000	12,500,000	Stella et al. 2010

Appendix 8C. Continuation

21	33,300,000	33,302,672	Stella et al. 2010 Xu et al. 2015 Iso-Touru et al. 2016
21	33,302,674	33,349,999	Stella et al. 2010 Xu et al. 2015 Iso-Touru et al. 2016
21	33,350,001	33,400,000	Stella et al. 2010 Xu et al. 2015 Iso-Touru et al. 2016
21	33,450,000	33,478,768	Stella et al. 2010 Xu et al. 2015 Iso-Touru et al. 2016
21	33,478,769	33,500,000	Stella et al. 2010 Iso-Touru et al. 2016
21	36,564,029	36,600,000	Stella et al. 2010
21	61,450,000	61,500,000	Stella et al. 2010
21	61,750,000	61,800,000	Stella et al. 2010
21	63,250,000	63,300,000	Wang et al. 2019
23	13,213,563	13,249,999	Boitard et al. 2016
23	13,250,001	13,299,999	Boitard et al. 2016
23	13,300,001	13,349,999	Boitard et al. 2016
23	13,350,001	13,399,999	Boitard et al. 2016
23	13,400,001	13,407,144	Boitard et al. 2016
23	13,423,402	13,449,999	Boitard et al. 2016
23	13,450,001	13,453,356	Boitard et al. 2016
23	13,453,358	13,483,560	Boitard et al. 2016
23	13,483,562	13,500,000	Boitard et al. 2016
26	500,000	550,000	Boitard et al. 2016
27	2,400,000	2,450,000	Boitard et al. 2016
28	11,400,000	11,450,000	Iso-Touru et al. 2016
29	4,400,004	4,449,967	Mei et al. 2018

[1] BTA: *Bos taurus* autosome.

[2] Reference from the common signals found between our analysis and previous signatures of selection regions reported in the literature.

Appendix 9C. Overlapping of the putative sweep regions identified from the top 1% of the cross-population DCMS statistic with candidate regions under positive selection previously reported in other cattle populations.

BTA[1]	Start (bp)	End (bp)	Reference[2]
1	12,250,000	12,297,891	Xu et al. 2015 Boitard et al. 2016
1	12,297,892	12,298,927	Boitard et al. 2016
1	12,298,928	12,300,000	Xu et al. 2015 Boitard et al. 2016
1	70,050,000	70,100,000	Stella et al. 2010
2	450,000	500,000	Iso-Touru et al. 2016
2	5,250,000	5,300,000	González-Rodríguez et al. 2016
2	36,050,000	36,099,999	Iso-Touru et al. 2016
2	36,100,001	36,150,000	Iso-Touru et al. 2016
2	125,300,000	125,325,208	Kim et al. 2017
3	6,800,000	6,800,001	Wang et al. 2019
3	8,200,000	8,249,999	Makina et al. 2015
3	8,250,001	8,300,000	Makina et al. 2015
3	40,250,000	40,267,094	Boitard et al. 2016
3	64,119,260	64,147,990	Boitard et al. 2016
3	73,350,000	73,357,040	Boitard et al. 2016
3	81,800,108	81,849,861	Mei et al. 2018
3	96,150,000	96,195,939	Stella et al. 2010
4	17,050,000	17,099,999	Xu et al. 2015
4	17,100,001	17,149,999	Xu et al. 2015
4	17,150,001	17,199,999	Xu et al. 2015
4	17,200,001	17,200,594	Xu et al. 2015
4	102,800,000	102,821,193	Boitard et al. 2016
4	111,000,000	111,050,000	Iso-Touru et al. 2016
4	113,750,000	113,780,179	Boitard et al. 2016
4	117,050,000	117,065,141	Stella et al. 2010 Iso-Touru et al. 2016
4	117,065,143	117,100,000	Stella et al. 2010 Iso-Touru et al. 2016
4	117,250,000	117,299,999	Iso-Touru et al. 2016
4	117,300,001	117,349,999	Iso-Touru et al. 2016
4	117,350,001	117,399,999	Iso-Touru et al. 2016
4	117,400,001	117,449,999	Iso-Touru et al. 2016
4	117,450,001	117,500,000	Iso-Touru et al. 2016
5	17,150,000	17,200,000	González-Rodríguez et al. 2016
5	31,800,000	31,811,934	Boitard et al. 2016
5	40,850,001	40,899,868	Mei et al. 2018

Appendix 9C. Continuation

5	72,750,000	72,799,999	Stella et al. 2010
5	72,800,001	72,850,000	Stella et al. 2010
5	72,900,000	72,922,419	Stella et al. 2010 Xu et al. 2015 Boitard et al. 2016
5	72,922,420	72,923,279	Stella et al. 2010 Xu et al. 2015
5	72,923,280	72,950,000	Stella et al. 2010 Xu et al. 2015 Boitard et al. 2016
5	106,950,000	107,000,000	Stella et al. 2010
5	113,150,001	113,199,999	Kim et al. 2017
5	120,850,000	120,875,605	Kim et al. 2017
5	120,875,607	120,899,999	Kim et al. 2017
5	120,900,001	120,900,187	Kim et al. 2017
5	120,900,188	120,925,605	Kim et al. 2017 Mei et al. 2018
5	120,925,607	120,949,999	Kim et al. 2017 Mei et al. 2018
5	120,950,001	120,975,606	Stella et al. 2010 Kim et al. 2017 Mei et al. 2018
5	120,975,607	120,999,866	Stella et al. 2010 Mei et al. 2018
5	120,999,867	121,000,000	Liao et al. 2013
6	16,700,000	16,725,191	Kim et al. 2017
6	24,100,000	24,149,999	Iso-Touru et al. 2016
6	24,150,001	24,200,000	Iso-Touru et al. 2016
6	38,150,000	38,200,000	Rothammer et al. 2013 Zhao et al. 2015 González-Rodríguez et al. 2016
6	60,350,000	60,400,000	Iso-Touru et al. 2016
6	70,150,000	70,200,000	Rothammer et al. 2013
6	70,350,000	70,400,000	Rothammer et al. 2013
6	70,550,000	70,599,999	Rothammer et al. 2013
6	70,600,001	70,600,058	Rothammer et al. 2013
6	70,600,059	70,649,948	Rothammer et al. 2013 Mei et al. 2018
6	70,649,949	70,650,000	Rothammer et al. 2013
6	91,850,000	91,900,000	Rothammer et al. 2013
6	95,000,000	95,050,000	Rothammer et al. 2013
7	22,400,000	22,450,000	Iso-Touru et al. 2016

Appendix 9C. Continuation

7	25,650,000	25,670,296	Boitard et al. 2016
7	25,670,297	25,699,999	Somavilla et al. 2014 Boitard et al. 2016
7	25,700,001	25,701,722	Somavilla et al. 2014 Boitard et al. 2016
7	25,701,723	25,750,000	Boitard et al. 2016
7	26,050,000	26,100,000	Zhao et al. 2015
7	26,300,000	26,349,999	Boitard et al. 2016
7	26,350,001	26,400,000	Boitard et al. 2016
7	87,600,024	87,650,000	Mei et al. 2018
8	600,000	650,000	Iso-Touru et al. 2016
8	57,350,000	57,357,077	Iso-Touru et al. 2016
8	57,357,079	57,400,000	Iso-Touru et al. 2016
8	58,750,000	58,770,849	Xu et al. 2015 Wang et al. 2019
8	58,770,850	58,800,000	Wang et al. 2019
9	23,825,194	23,849,999	Kim et al. 2017
9	43,150,000	43,199,999	Rothammer et al. 2013 Boitard et al. 2016
9	43,200,001	43,249,999	Rothammer et al. 2013 Boitard et al. 2016
9	43,250,001	43,265,691	Rothammer et al. 2013 Boitard et al. 2016
9	43,265,692	43,300,000	Rothammer et al. 2013
9	51,150,000	51,200,000	Rothammer et al. 2013 Iso-Touru et al. 2016
9	98,619,640	98,649,999	Stella et al. 2010
9	98,650,001	98,700,000	Stella et al. 2010
10	38,600,000	38,616,677	Xu et al. 2015
10	102,900,000	102,950,000	Xu et al. 2015
11	650,000	700,000	Iso-Touru et al. 2016
11	1,000,000	1,050,000	Iso-Touru et al. 2016 Boitard et al. 2016
11	4,400,000	4,449,999	Kim et al. 2017
11	17,379,718	17,400,000	Kim et al. 2017
11	17,950,000	17,998,249	Boitard et al. 2016
11	17,998,250	18,000,000	Stella et al. 2010 Boitard et al. 2016
11	19,600,000	19,649,999	Iso-Touru et al. 2016
11	19,650,001	19,699,999	Iso-Touru et al. 2016
11	19,700,001	19,749,999	Iso-Touru et al. 2016
11	19,750,001	19,799,999	Iso-Touru et al. 2016

Appendix 9C. Continuation

11	19,800,001	19,849,999	Iso-Touru et al. 2016
11	19,850,001	19,899,999	Iso-Touru et al. 2016
11	19,900,001	19,949,999	Iso-Touru et al. 2016
11	19,950,001	20,000,000	Iso-Touru et al. 2016
11	22,774,051	22,800,000	Zhao et al. 2015
11	26,982,784	27,000,000	Boitard et al. 2016
11	40,500,000	40,550,000	Boitard et al. 2016
11	44,050,000	44,099,840	Mei et al. 2018
11	44,700,000	44,749,999	Xu et al. 2015
11	44,750,001	44,786,826	Xu et al. 2015
11	44,800,010	44,849,952	Mei et al. 2018
11	64,950,000	65,000,000	Rothammer et al. 2013
11	66,500,000	66,519,462	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	66,519,463	66,550,000	Rothammer et al. 2013 Xu et al. 2015 González-Rodríguez et al. 2016
11	67,250,000	67,250,081	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	67,250,082	67,299,983	Rothammer et al. 2013 González-Rodríguez et al. 2016 Mei et al. 2018
11	67,299,984	67,300,000	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	67,450,000	67,479,717	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	67,479,718	67,500,000	Rothammer et al. 2013 González-Rodríguez et al. 2016 Kim et al. 2017
11	67,700,000	67,749,999	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	67,750,001	67,800,000	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	68,550,000	68,600,000	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	69,450,000	69,500,000	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	69,650,000	69,700,000	Rothammer et al. 2013 González-Rodríguez et al. 2016 Wang et al. 2019
11	78,884,570	78,894,156	Somavilla et al. 2014
11	81,100,001	81,149,575	Wang et al. 2019

Appendix 9C. Continuation

11	81,149,576	81,150,000	Boitard et al. 2016 Wang et al. 2019
11	85,650,000	85,700,000	Iso-Touru et al. 2016
11	98,614,186	98,618,476	Xu et al. 2015
11	98,618,478	98,649,999	Xu et al. 2015
11	98,650,001	98,700,000	Xu et al. 2015
12	52,200,000	52,216,886	Xu et al. 2015
13	12,100,000	12,150,000	Xu et al. 2015
13	63,750,000	63,800,000	Xu et al. 2015
14	2,700,000	2,725,151	Kim et al. 2017
14	2,725,152	2,750,000	Kim et al. 2017
14	2,825,152	2,850,000	Kim et al. 2017
14	3,000,000	3,000,001	Wang et al. 2019
14	3,160,330	3,162,081	Somavilla et al. 2014
14	4,700,000	4,703,288	Boitard et al. 2016
14	23,294,853	23,300,000	Xu et al. 2015
14	28,100,000	28,137,734	Boitard et al. 2016
14	35,850,079	35,900,000	Mei et al. 2018
14	37,881,210	37,900,000	Boitard et al. 2016
14	38,000,000	38,027,882	Boitard et al. 2016
14	38,700,064	38,749,871	Mei et al. 2018
14	39,200,134	39,249,997	Mei et al. 2018
14	52,950,001	53,000,000	Kim et al. 2017
14	55,550,000	55,600,000	Zhao et al. 2015
14	75,000,000	75,050,000	Iso-Touru et al. 2016
15	5,650,000	5,664,075	Boitard et al. 2016
15	8,400,000	8,450,000	Stella et al. 2010
15	56,500,000	56,500,114	Xu et al. 2015
15	56,500,115	56,521,929	Xu et al. 2015 Mei et al. 2018
15	56,521,930	56,550,000	Xu et al. 2015 Boitard et al. 2016 Mei et al. 2018
16	43,000,000	43,050,000	Boitard et al. 2016 Mei et al. 2018
16	44,000,000	44,040,550	Boitard et al. 2016
16	45,000,001	45,050,000	Boitard et al. 2016 González-Rodríguez et al. 2016
17	1,346,129	1,350,000	Xu et al. 2015
17	39,950,000	40,000,000	Iso-Touru et al. 2016
18	35,550,000	35,588,704	Iso-Touru et al. 2016

Appendix 9C. Continuation

18	35,588,705	35,600,000	Iso-Touru et al. 2016 Boitard et al. 2016
18	53,500,001	53,550,000	Xu et al. 2015
19	500,000	550,000	Boitard et al. 2016
20	15,000,000	15,050,000	Xu et al. 2015
21	200,000	250,000	Xu et al. 2015
21	24,700,000	24,749,999	Iso-Touru et al. 2016
21	24,750,001	24,799,999	Iso-Touru et al. 2016
21	24,800,001	24,802,672	Iso-Touru et al. 2016
21	24,802,674	24,827,200	Iso-Touru et al. 2016
21	24,827,201	24,849,999	Stella et al. 2010 Iso-Touru et al. 2016
21	24,850,001	24,900,000	Stella et al. 2010 Iso-Touru et al. 2016
21	29,950,000	30,000,000	Iso-Touru et al. 2016 Pitt et al. 2019
21	30,750,000	30,775,020	Iso-Touru et al. 2016 Kim et al. 2017
21	30,775,021	30,800,000	Iso-Touru et al. 2016 Kim et al. 2017
21	34,300,000	34,302,672	Stella et al. 2010 Zhao et al. 2015 Iso-Touru et al. 2016
21	34,302,674	34,350,000	Stella et al. 2010 Zhao et al. 2015 Iso-Touru et al. 2016
21	34,950,000	35,000,000	Stella et al. 2010 Iso-Touru et al. 2016
21	35,200,000	35,249,999	Stella et al. 2010 Iso-Touru et al. 2016
21	35,250,001	35,259,414	Stella et al. 2010 Iso-Touru et al. 2016
21	35,259,415	35,300,000	Iso-Touru et al. 2016
21	36,850,000	36,900,000	Stella et al. 2010
21	37,350,001	37,394,355	Stella et al. 2010 Xu et al. 2015
21	37,394,356	37,399,999	Stella et al. 2010
21	37,400,001	37,450,000	Stella et al. 2010
21	46,050,000	46,060,514	Boitard et al. 2016
21	62,882,706	62,900,000	Boitard et al. 2016
23	14,779,113	14,799,999	Stella et al. 2010
23	14,800,001	14,849,999	Stella et al. 2010

Appendix 9C. Continuation

23	14,850,001	14,900,000	Stella et al. 2010
23	19,100,030	19,149,985	Mei et al. 2018
23	22,500,000	22,524,766	Stella et al. 2010 Boitard et al. 2016
23	22,524,767	22,550,000	Stella et al. 2010
23	22,600,000	22,649,999	Stella et al. 2010
23	22,650,001	22,699,999	Stella et al. 2010
23	22,700,001	22,749,999	Stella et al. 2010
23	22,750,001	22,800,000	Stella et al. 2010
23	30,450,000	30,500,000	Wang et al. 2019
23	50,750,000	50,800,000	Stella et al. 2010
24	28,300,000	28,303,581	Somavilla et al. 2014
24	54,950,000	55,000,000	Stella et al. 2010
26	3,439,753	3,450,000	Somavilla et al. 2014
26	28,400,001	28,425,563	Kim et al. 2017
26	33,459,409	33,500,000	Xu et al. 2015
26	35,600,000	35,650,000	Stella et al. 2010
26	39,467,903	39,470,848	Somavilla et al. 2014
27	6,300,000	6,319,353	Somavilla et al. 2014
27	44,350,000	44,400,000	Stella et al. 2010
27	44,600,000	44,650,000	Stella et al. 2010
27	44,850,000	44,899,999	Stella et al. 2010
27	44,900,001	44,949,999	Stella et al. 2010
27	44,950,001	45,000,000	Stella et al. 2010
28	23,300,000	23,344,569	Boitard et al. 2016
28	29,850,001	29,900,000	Kim et al. 2017
28	35,500,239	35,549,977	Mei et al. 2018
29	7,550,000	7,600,000	Stella et al. 2010 Iso-Touru et al. 2016
29	27,100,022	27,149,999	Mei et al. 2018

[1] BTA: *Bos taurus* autosome

[2] Reference from the common signals found between our analysis and previous signatures of selection regions reported in the literature

Appendix 10C. Overlapping between runs of homozygosity (ROH) hotspots and the top 1% of the within-population DCMS statistic with the candidate regions under positive selection previously reported in other cattle populations.

BTA[1]	Start (bp)	End (bp)	Reference[2]
1	8,300,000	8,350,000	Xu et al. 2015 Iso-Touru et al. 2016
1	112,250,000	112,300,000	Boitard et al. 2016
21	6,550,000	6,591,118	Boitard et al. 2016
21	63,250,000	63,300,000	Wang et al. 2019

[1] BTA: *Bos taurus* autosome.

[2] Reference from the common signals found between our analysis and previous signatures of selection regions reported in the literature.

Appendix 11C. Overlapping between runs of homozygosity (ROH) hotspots and the top 1% of the cross-population DCMS statistic with the candidate regions under positive selection previously reported in other cattle populations.

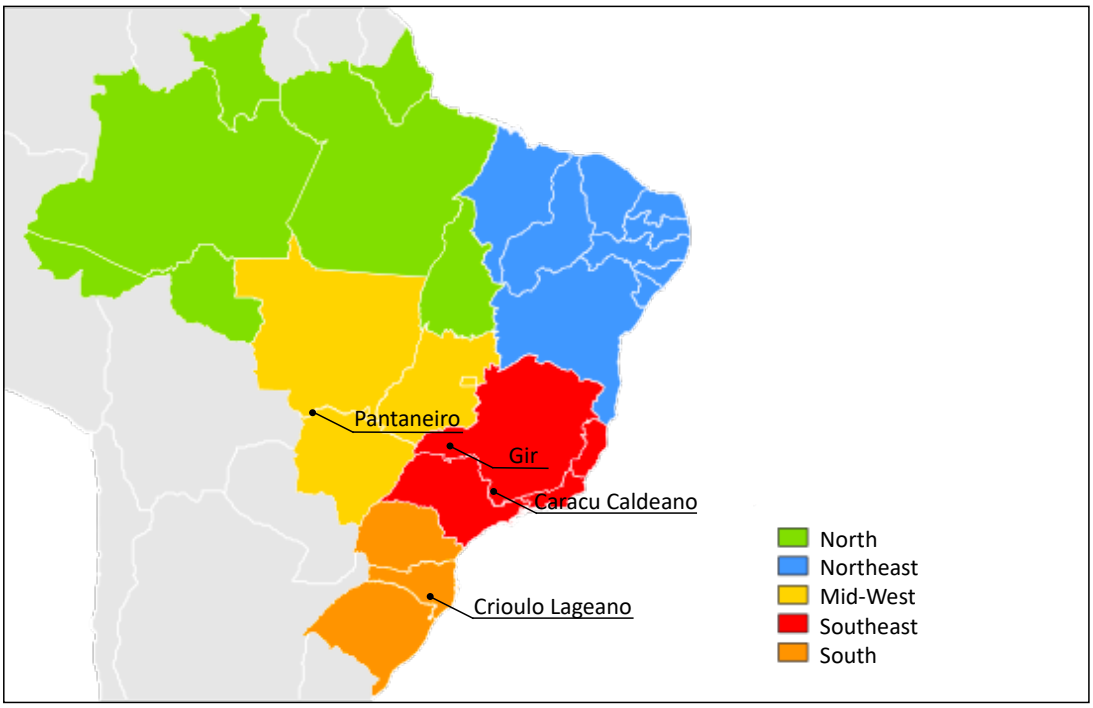
BTA[1]	Start (bp)	End (bp)	Reference[2]
5	31,800,000	31,811,934	Boitard et al. 2016
11	67,450,000	67,479,717	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	67,479,718	67,500,000	Rothammer et al. 2013 González-Rodríguez et al. 2016 Kim et al. 2017
11	67,700,000	67,749,999	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	67,750,001	67,800,000	Rothammer et al. 2013 González-Rodríguez et al. 2016
11	68,550,000	68,600,000	Rothammer et al. 2013 González-Rodríguez et al. 2016
21	200,000	250,000	Xu et al. 2015

[1] BTA: *Bos taurus* autosome

[2] Reference from the common signals found between our analysis and previous signatures of selection regions reported in the literature.

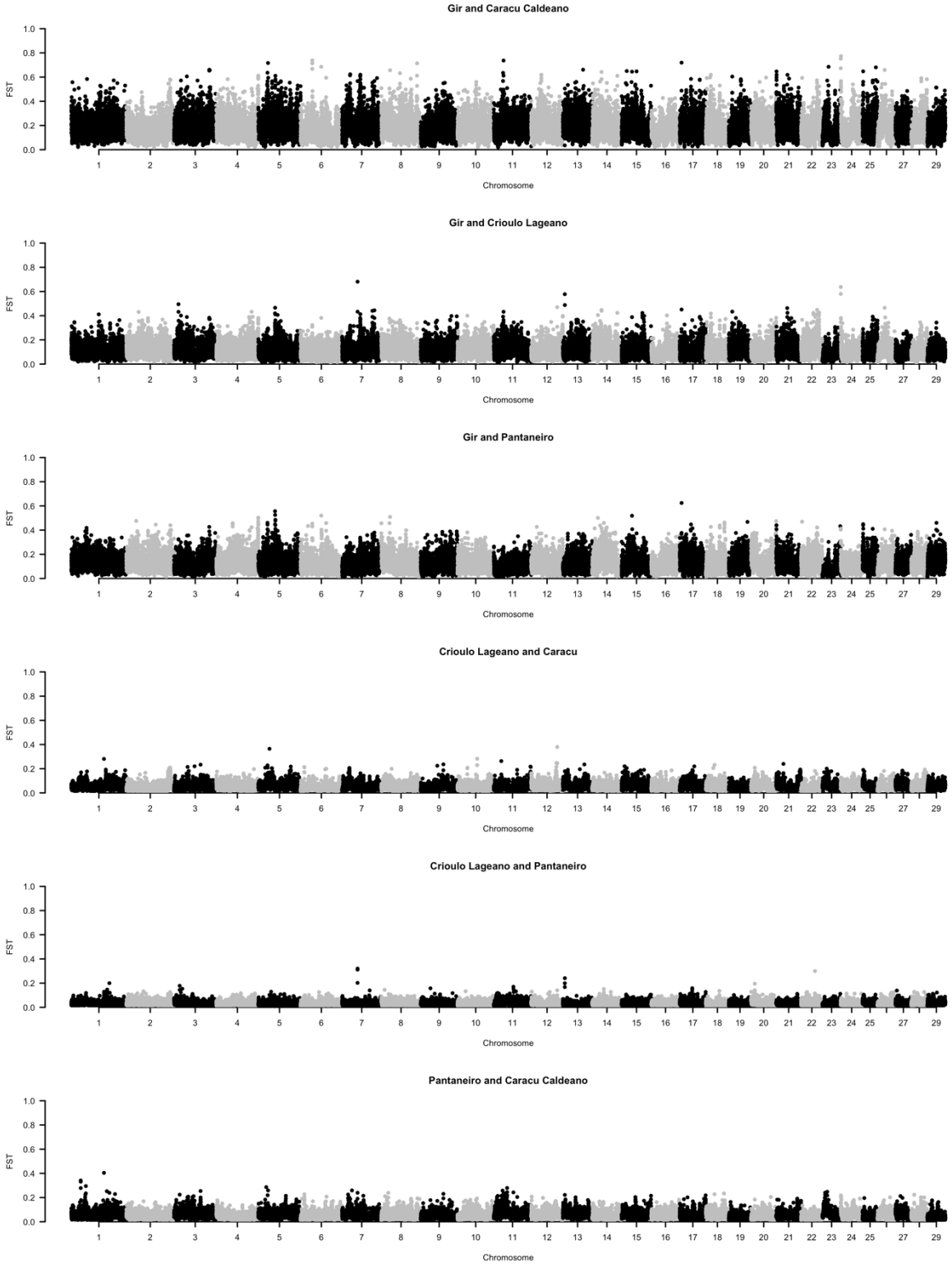
Appendix 12C. Brazilian geographical regions of the four cattle breeds sampled in the study.

(Adapted from https://pt.wikipedia.org/wiki/Ficheiro:Brazil_Labelled_Map.svg)

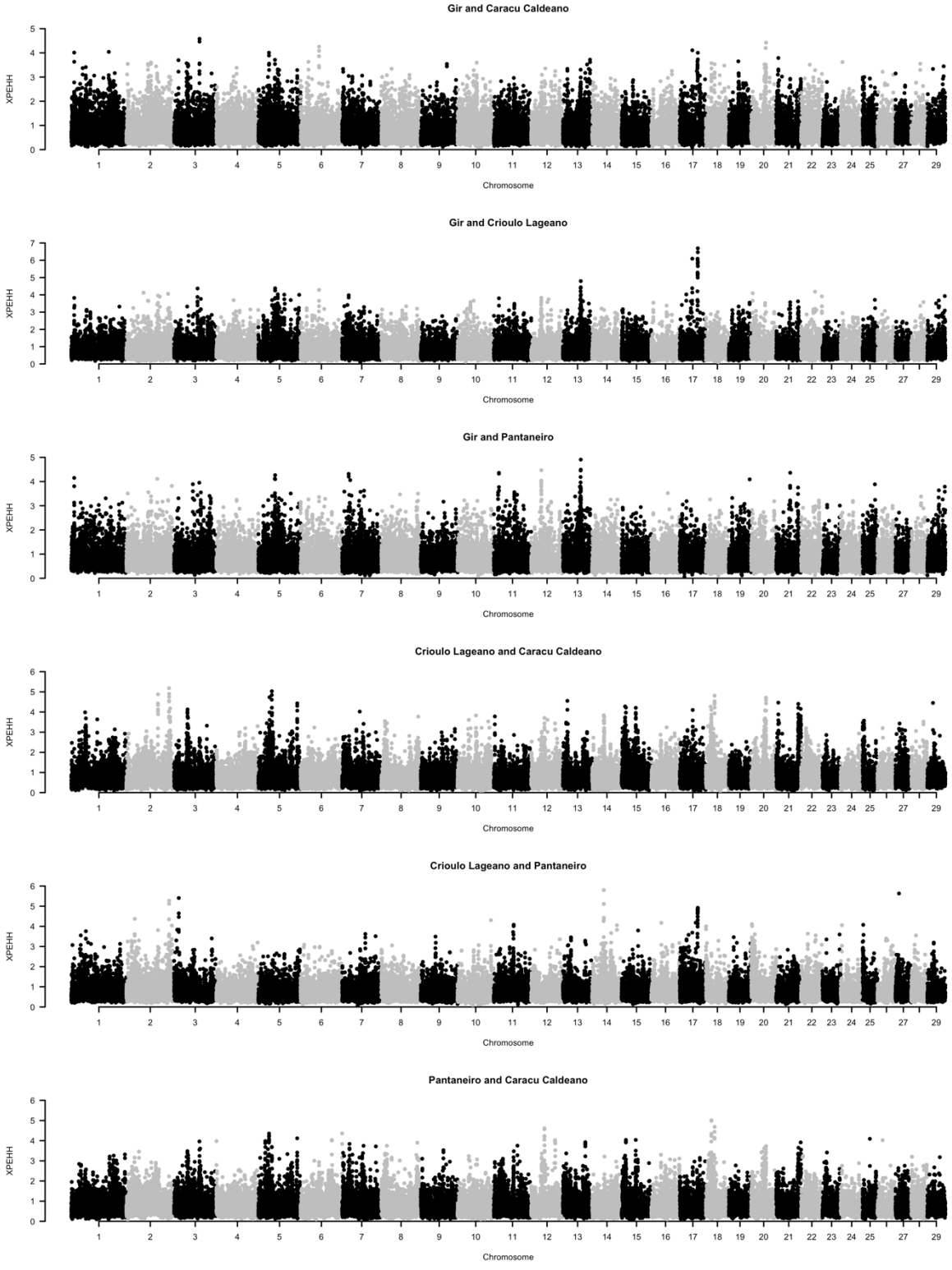


Appendix 13C. Manhattan plot of the independent results for each selective sweep statistical method and population.

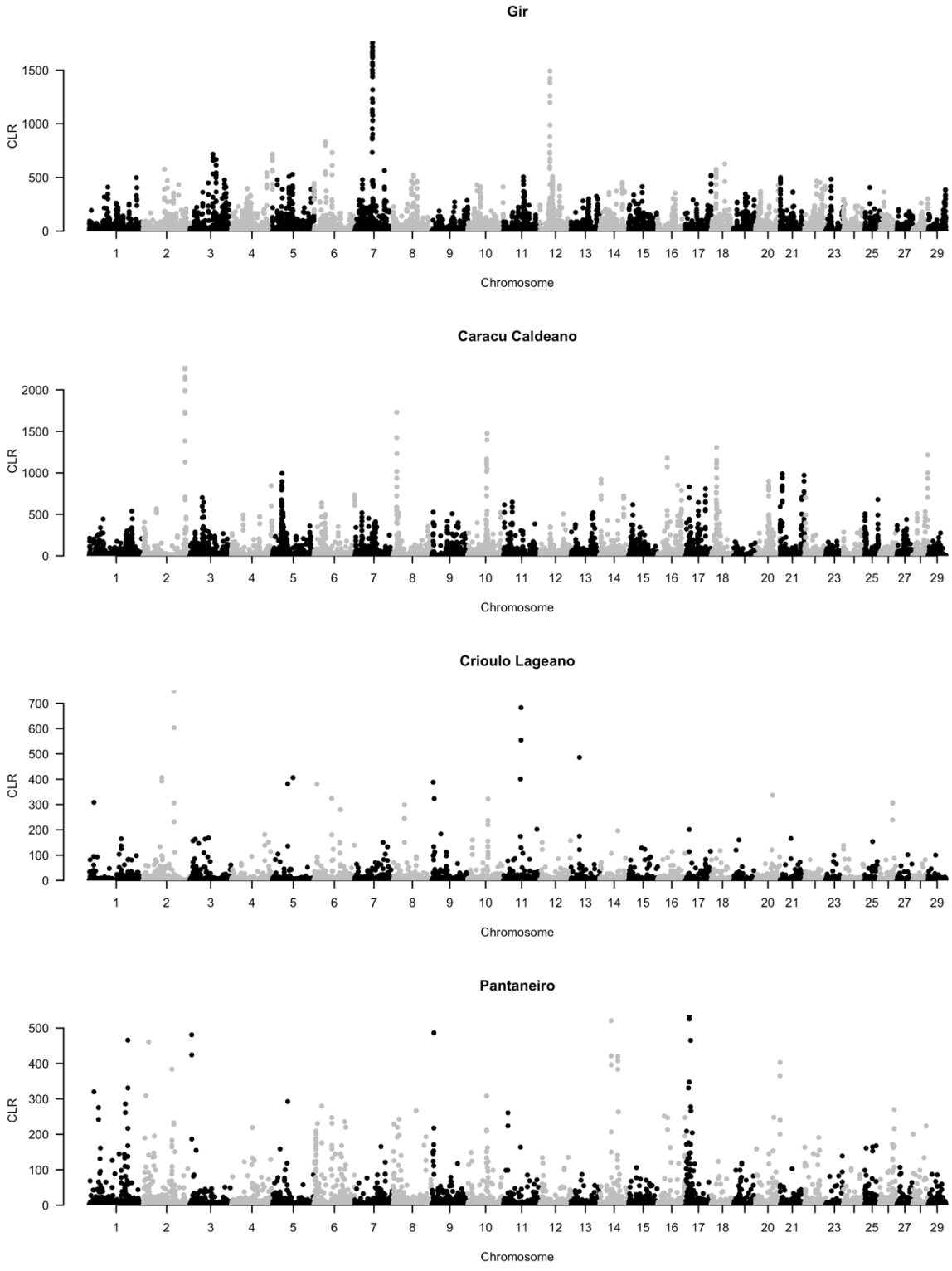
a) FST



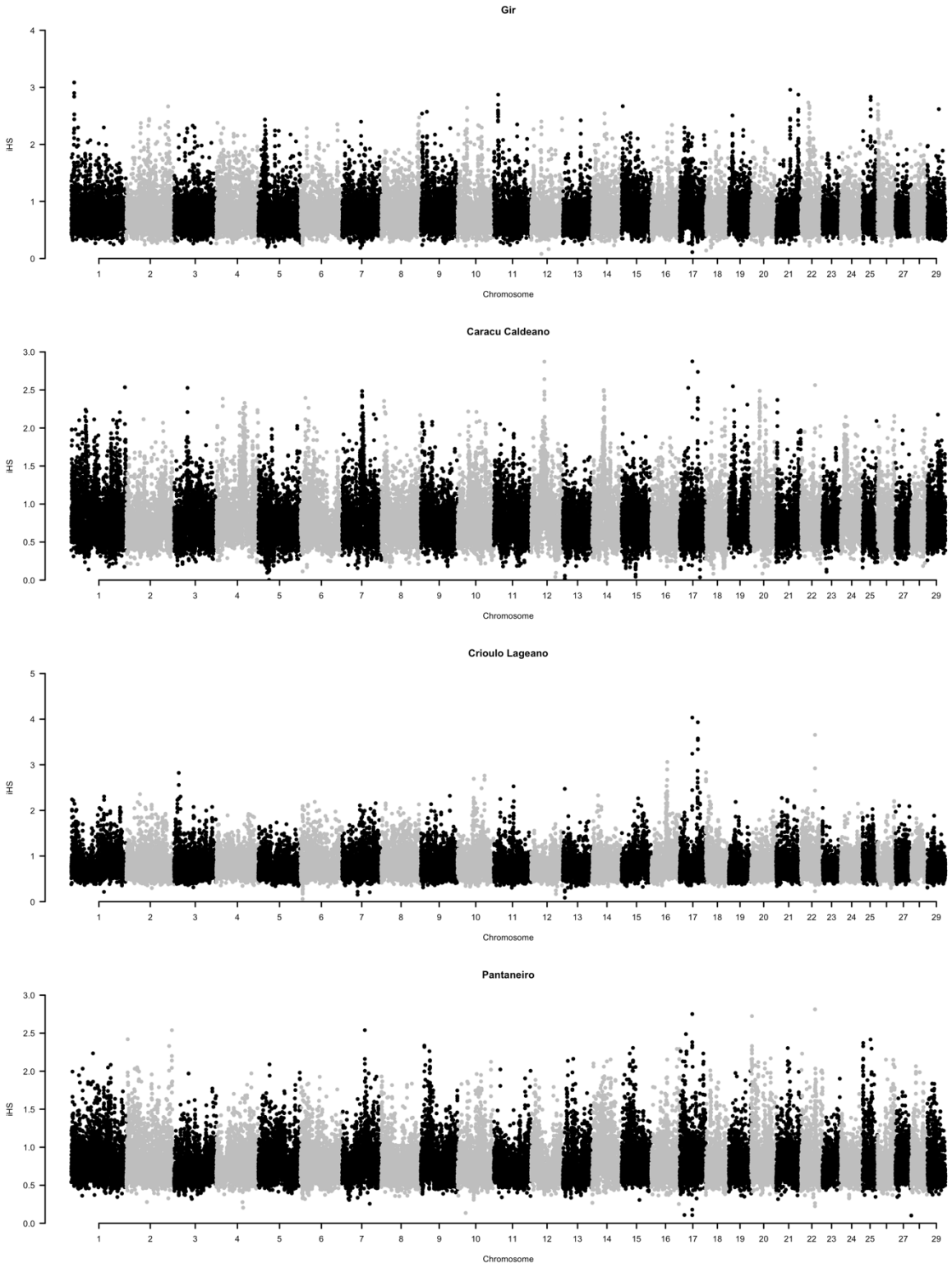
b) XPEHH


Gir and Caracu Caldeano
Gir and Crioulo Lageano
Gir and Pantaneiro
Crioulo Lageano and Caracu Caldeano
Crioulo Lageano and Pantaneiro
Pantaneiro and Caracu Caldeano
XPEHH
Chromosome

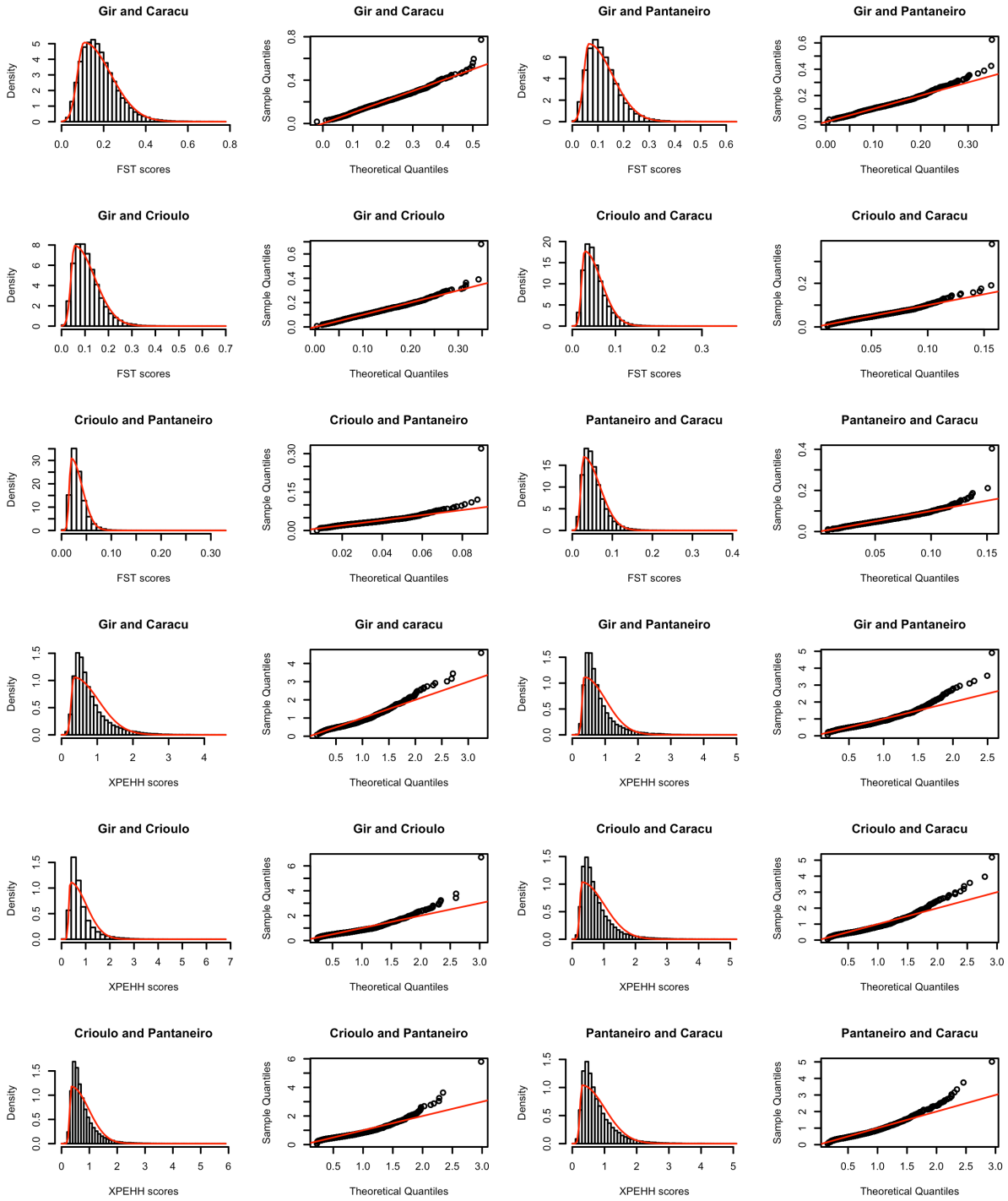
c) CLR

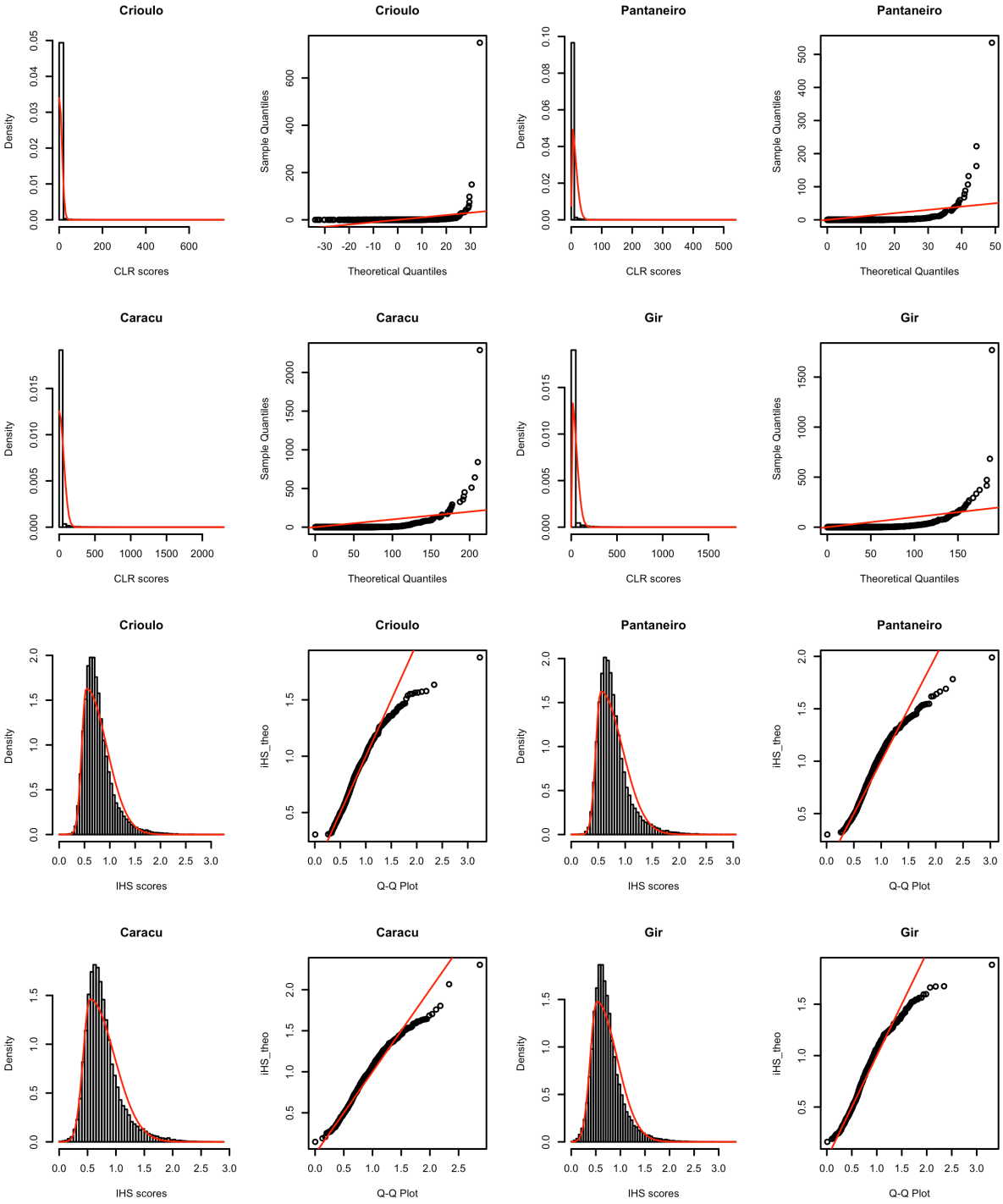

Gir
CLR
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 23 25 27 29
Chromosome
Caracu Caldeano
CLR
2000
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 23 25 27 29
Chromosome
Crioulo Lageano
CLR
700
600
500
400
300
200
100
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 23 25 27 29
Chromosome
Pantaneiro
CLR
500
400
300
200
100
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 23 25 27 29
Chromosome

d) iHS


Gir
iHS
Chromosome
Caracu Caldeano
iHS
Chromosome
Crioulo Lageano
iHS
Chromosome
Pantaneiro
iHS
Chromosome

Appendix 14C. Histogram and quantile-quantile (Q-Q) plots of statistical scores calculated for all four methods derived from a skewness normal distribution.




Crioulo
Density
CLR scores
Crioulo
Sample Quantiles
Theoretical Quantiles
Pantaneiro
Density
CLR scores
Pantaneiro
Sample Quantiles
Theoretical Quantiles
Caracu
Density
CLR scores
Caracu
Sample Quantiles
Theoretical Quantiles
Gir
Density
CLR scores
Gir
Sample Quantiles
Theoretical Quantiles
Crioulo
Density
IHS scores
Crioulo
IHS_theo
Q-Q Plot
Pantaneiro
Density
IHS scores
Pantaneiro
IHS_theo
Q-Q Plot
Caracu
Density
IHS scores
Caracu
IHS_theo
Q-Q Plot
Gir
Density
IHS scores
Gir
IHS_theo
Q-Q Plot

APPENDIX D

Appendix 1D. The distribution and size characteristics of copy number variations (CNVs) in Caracu Caldeano cattle mapped to the ARS-UCD1.2 genome assembly

BTA[1]	BTA Length (Bp)	Sum CNV Length (kb)	*n* CNV	*n* Deletion	*n* Duplication	Mean Length (Kb)	Median Length (kb)	Minimum Length (Kb)	Maximum Lenght (kb)
1	158534110	9286.69	305	191	114	30.44	14.49	2.49	270.49
2	136231102	2425.84	155	110	45	15.65	12.49	2.99	66.49
3	121005158	8008.17	324	197	127	24.71	15.49	2.49	331.99
4	120000601	9795.62	379	223	156	25.84	18.49	2.49	128.49
5	120089316	10053.15	355	241	114	28.31	17.49	2.49	265.99
6	117806340	4880.71	285	225	60	17.12	13.99	2.49	62.49
7	110682743	13095.61	385	280	105	34.01	17.49	2.99	654.49
8	113319770	4655.79	207	137	70	22.49	14.99	2.99	138.49
9	105454467	7652.26	237	156	81	32.28	15.49	1.49	806.99
10	103308737	19998.05	446	258	188	44.83	17.49	2.99	1006.99
11	106982474	2559.84	156	129	27	16.40	9.49	2.49	106.99
12	87216183	22517.54	456	335	121	49.38	20.99	2.99	467.49
13	83472345	4074.32	174	61	113	23.41	19.74	2.49	143.49
14	82403003	4973.83	161	87	74	30.89	17.99	2.49	123.49
15	85007780	15397	496	302	194	31.04	18.49	2.99	336.99
16	81013979	4506.78	213	137	76	21.15	14.99	3.49	253.99
17	73167244	3663.79	208	161	47	17.61	14.74	3.99	57.99
18	65820629	11213.56	436	213	223	25.71	20.99	3.49	141.49
19	63449741	3067.34	157	107	50	19.53	14.99	3.99	84.49
20	71974595	3094.32	171	125	46	18.09	13.49	3.49	80.49
21	69862954	4193.29	209	119	90	20.06	16.99	2.49	176.99
22	60773035	884.93	62	40	22	14.27	11.99	3.49	48.99

Appendix 1D. Continuation

23	52498615	11181.58	420	216	204	26.62	17.49	2.49	374.49
24	62317253	1673.40	94	70	24	17.80	10.11	2.49	82.49
25	42350435	827.45	42	12	30	19.70	20.99	6.49	40.99
26	51992305	3297.3	200	190	10	16.48	12.49	2.49	97.49
27	45612108	8391.77	228	112	116	36.80	22.24	4.49	171.49
28	45940150	3563.9	100	79	21	35.63	19.24	4.49	348.49
29	51098607	7233.77	224	127	97	32.29	18.24	2.99	273.99

[1]BTA = *Bos taurus* autosome

Appendix 2D. The distribution and size characteristics of copy number variations (CNVs) in Crioulo Lageano cattle mapped to the ARS-UCD1.2 genome assembly

BTA[1]	BTA Length (Bp)	Sum CNV Length (kb)	*n* CNV	*n* Deletion	*n* Duplication	Mean Length (Kb)	Median Length (kb)	Minimum Length (Kb)	Maximum Lenght (kb)
1	158534110	9956.67	323	213	110	30.82	14.99	2.49	245.99
2	136231102	3721.81	186	141	45	20.01	11.49	2.49	370.49
3	121005158	8937.63	364	238	126	24.55	15.49	2.99	229.49
4	120000601	8810.14	351	204	147	25.10	15.99	2.49	217.49
5	120089316	12913.61	385	256	129	33.28	20.99	2.49	280.49
6	117806340	8995.68	319	233	86	28.20	14.49	2.49	344.49
7	110682743	10610.07	432	264	168	24.56	16.99	3.99	465.99
8	113319770	4623.26	232	168	64	19.92	13.49	1.99	137.99
9	105454467	8571.27	221	154	67	38.78	14.49	3.49	864.99
10	103308737	17666.56	439	225	214	40.24	17.49	3.49	1004.99
11	106982474	2780.85	148	131	17	18.79	11.49	4.49	106.49
12	87216183	18814.58	420	319	101	44.79	19.74	2.99	807.49
13	83472345	5104.29	207	81	126	24.65	18.49	2.49	219.99
14	82403003	5486.82	174	102	72	31.53	18.74	2.99	106.99
15	85007780	12956.48	518	331	187	25.01	17.49	2.49	334.99
16	81013979	4006.31	185	114	71	21.65	15.49	3.49	262.49
17	73167244	3738.81	186	159	27	20.10	15.99	2.49	98.99
18	65820629	11141.08	416	209	207	26.78	20.49	3.49	148.99
19	63449741	2638.86	138	87	51	19.12	12.74	3.99	85.99
20	71974595	3235.32	174	129	45	18.59	13.24	3.49	82.49
21	69862954	4261.29	209	132	77	20.38	14.49	3.49	91.99
22	60773035	781.44	56	42	14	13.95	12.99	3.49	49.49
23	52498615	8718.67	324	166	158	26.90	19.99	3.49	156.99
24	62317253	1329.40	91	74	17	14.60	8.99	2.99	63.49

Appendix 2D. Continuation

25	42350435	11101.95	49	14	35	22.48	21.99	5.49	55.99
26	51992305	4209.27	222	208	14	18.96	12.49	2.49	204.99
27	45612108	7346.29	210	116	94	34.98	22.74	2.49	285.49
28	45940150	2855.91	84	69	15	33.99	11.49	4.49	342.49
29	51098607	6149.76	231	144	87	26.62	17.49	2.99	197.49

[1]BTA = *Bos taurus* autosome

Appendix 3D. The distribution and size characteristics of copy number variations (CNVs) in Pantaneiro cattle mapped to the ARS-UCD1.2 genome assembly

BTA[1]	BTA Length (Bp)	Sum CNV Length (kb)	*n* CNV	*n* Deletion	*n* Duplication	Mean Length (Kb)	Median Length (kb)	Minimum Length (Kb)	Maximum Lenght (kb)
1	158534110	19589.94	662	523	139	29.59	19.79	2.99	248.99
2	136231102	12876.13	471	340	131	27.33	19.19	4.79	272.39
3	121005158	17296.84	558	378	180	30.99	20.39	4.79	568.79
4	120000601	15396.68	515	360	155	29.89	20.39	2.99	289.79
5	120089316	17043.71	495	330	165	34.43	22.19	4.19	266.39
6	117806340	13002.66	543	420	123	23.94	17.39	2.99	194.39
7	110682743	19909.84	563	389	174	35.36	20.39	2.99	585.59
8	113319770	11727.18	418	300	118	28.05	18.89	1.19	213.59
9	105454467	16477.96	445	331	114	37.02	18.59	4.79	1007.39
10	103308737	22737.09	513	322	191	44.32	20.00	2.99	1007.39
11	106982474	11202.85	347	246	101	32.28	18.59	2.99	500.39
12	87216183	24498.08	517	386	131	47.38	22.19	2.99	773.39
13	83472345	8103.36	231	112	119	35.08	20.39	3.59	496.19
14	82403003	9575.72	275	179	96	34.82	20.39	2.99	420.59
15	85007780	17116.24	562	366	196	30.45	23.39	3.59	202.79
16	81013979	9560.12	272	177	95	35.14	20.39	2.99	387.59
17	73167244	11493.88	321	243	78	35.80	19.79	2.99	362.99
18	65820629	15612.77	429	218	211	36.39	23.39	2.99	373.19
19	63449741	6493.03	162	100	62	40.08	23.69	3.59	778.79
20	71974595	6567.94	251	187	64	26.16	18.59	2.99	260.99
21	69862954	8811.90	293	172	121	30.07	20.99	3.59	313.79
22	60773035	6059.25	150	93	57	40.39	19.79	4.19	386.39
23	52498615	10188.90	298	157	141	34.19	25.79	3.59	172.19
24	62317253	4808.24	158	120	38	30.43	18.89	3.59	263.39

Appendix 3D. Continuation

25	42350435	7072.09	107	22	85	66.09	34.19	6.59	613.19
26	51992305	6482.13	269	219	50	24.09	15.59	2.99	225.59
27	45612108	8583.39	203	107	96	42.28	22.19	4.79	215.39
28	45940150	6068.26	134	87	47	45.28	25.79	4.79	426.59
29	51098607	8611.59	201	114	87	42.84	27.59	2.99	247.19

[1]BTA = *Bos taurus* autosome

Appendix 4D. Copy number variation regions (CNVRs) scattering in the Caracu Caldeano cattle genome

BTA[1]	Start (bp)	End (bp)	Length (bp)	Event
1	206501	246500	40000	Duplication
1	268501	283500	15000	Duplication
1	284501	336500	52000	Duplication
1	338001	370500	32500	Duplication
1	371501	383000	11500	Duplication
1	406501	426500	20000	Duplication
1	428001	430000	2000	Duplication
1	430501	450000	19500	Duplication
1	451001	481500	30500	Duplication
1	498501	501500	3000	Duplication
1	502001	512000	10000	Duplication
1	582001	587500	5500	Duplication
1	588501	595000	6500	Duplication
1	595501	636500	41000	Duplication
1	638501	661500	23000	Duplication
1	665501	672000	6500	Duplication
1	123165501	123174500	9000	Deletion
2	121555001	121566500	11500	Deletion
2	121598501	121613500	15000	Deletion
3	11666001	11671500	5500	Deletion
3	11854001	11874500	20500	Duplication
3	11981501	11988000	6500	Duplication
3	11995001	11996000	1000	Duplication
3	11996501	12008500	12000	Duplication
3	21065001	21072500	7500	Duplication
3	21302001	21316000	14000	Duplication
3	54209001	54225000	16000	Deletion
3	54230001	54238000	8000	Deletion
3	119159501	119168500	9000	Deletion
4	105571001	105591500	20500	Duplication
4	105598001	105611500	13500	Duplication
4	105620001	105633000	13000	Duplication
5	44283001	44298500	15500	Deletion
5	102684501	102691000	6500	Duplication
5	102691501	102702000	10500	Duplication
5	102703501	102710500	7000	Duplication
5	102724001	102753000	29000	Duplication
5	102753501	102783500	30000	Duplication

Appendix 4D. Continuation

5	102785501	102808500	23000	Duplication
5	102919501	102925500	6000	Deletion
6	5433001	5461500	28500	Duplication
6	5924501	5928000	3500	Duplication
6	5928501	5941500	13000	Duplication
6	5960001	5983000	23000	Duplication
6	7728001	7735000	7000	Deletion
6	7912501	7924500	12000	Deletion
7	10801001	10819500	18500	Deletion
7	41402001	41407000	5000	Deletion
9	27501	48500	21000	Duplication
9	87047501	87082500	35000	Duplication
9	87187501	87194500	7000	Deletion
9	87199501	87241000	41500	Duplication
9	102832501	102839500	7000	Deletion
10	22549001	22551000	2000	Deletion
10	23080001	23092500	12500	Deletion
10	23107001	23175000	68000	Deletion
10	23373001	23380000	7000	Duplication
10	23729001	23747000	18000	Duplication
10	25097501	25106000	8500	Duplication
10	25156001	25165000	9000	Duplication
10	25260001	25268000	8000	Duplication
10	25293001	25310500	17500	Deletion
10	101931501	101936500	5000	Deletion
11	82908001	82916500	8500	Deletion
12	70245501	70247000	1500	Deletion
12	71218501	71224000	5500	Duplication
12	71669501	71679500	10000	Deletion
12	71684501	71711500	27000	Deletion
12	71844001	71864000	20000	Deletion
12	72392001	72400000	8000	Deletion
12	72649001	72665000	16000	Duplication
12	72683001	72703500	20500	Duplication
12	72798501	72821500	23000	Duplication
12	72822501	72837000	14500	Duplication
12	72838001	72859500	21500	Duplication
12	72861501	72881500	20000	Duplication
12	87158001	87167000	9000	Deletion
13	325001	347000	22000	Duplication
13	355501	375500	20000	Duplication
13	10800001	10812500	12500	Duplication

Appendix 4D. Continuation

13	10860501	10875000	14500	Duplication
13	17602501	17611000	8500	Deletion
14	13577501	13586500	9000	Duplication
14	13589001	13610000	21000	Duplication
14	13626001	13631000	5000	Duplication
14	13632001	13642500	10500	Duplication
14	13756001	13835500	79500	Duplication
15	48903001	48931000	28000	Deletion
15	49984501	49991500	7000	Duplication
15	80586001	80594000	8000	Duplication
15	80611501	80622500	11000	Duplication
15	82106501	82151500	45000	Duplication
16	7010501	7018500	8000	Duplication
16	7176501	7189500	13000	Duplication
16	47738001	47745500	7500	Deletion
17	68057501	68079500	22000	Deletion
18	45161501	45176000	14500	Deletion
18	51436501	51447000	10500	Duplication
18	57179501	57199500	20000	Duplication
18	57211501	57241000	29500	Duplication
18	57264001	57271500	7500	Duplication
18	57272001	57294000	22000	Duplication
18	57294501	57310500	16000	Duplication
18	57332501	57349500	17000	Duplication
18	58134001	58147500	13500	Duplication
18	62656501	62665500	9000	Duplication
18	62666001	62668000	2000	Duplication
18	62668501	62672500	4000	Duplication
18	62981001	62986500	5500	Deletion
18	62994001	62996000	2000	Deletion
18	63598501	63608000	9500	Duplication
18	63626001	63643000	17000	Deletion
19	43222001	43228500	6500	Duplication
19	57794501	57803000	8500	Deletion
19	57804001	57805500	1500	Deletion
19	57807501	57820000	12500	Deletion
20	96501	119500	23000	Duplication
21	274501	280500	6000	Deletion
21	339001	358000	19000	Duplication
21	673001	697000	24000	Duplication
21	32903501	32937000	33500	Duplication
23	1	28500	28500	Duplication

Appendix 4D. Continuation

23	25595001	25648000	53000	Deletion
23	26718501	26728500	10000	Duplication
23	29037501	29050000	12500	Mixed
24	61859501	61872000	12500	Duplication
25	1	22000	22000	Duplication
26	14973001	14980500	7500	Deletion
26	51098001	51110000	12000	Deletion
26	51793501	51804500	11000	Deletion
27	6300501	6344000	43500	Duplication
27	6383001	6389000	6000	Duplication
27	6445501	6455500	10000	Deletion
27	6456501	6462500	6000	Deletion
27	6552501	6572500	20000	Duplication
27	6650001	6666500	16500	Duplication
27	6667501	6687500	20000	Duplication
27	7138001	7146000	8000	Duplication
27	7147501	7196500	49000	Duplication
27	7202001	7210000	8000	Duplication
27	38962001	38970500	8500	Deletion
28	504001	524000	20000	Duplication
28	2416001	2427500	11500	Deletion
29	1869501	1894500	25000	Deletion
29	5504001	5513000	9000	Deletion
29	5540001	5549000	9000	Duplication
29	5681501	5689000	7500	Duplication
29	5689501	5702000	12500	Duplication
29	5716501	5733500	17000	Duplication
29	5734001	5768500	34500	Duplication
29	5769001	5773000	4000	Duplication
29	41924001	41930500	6500	Mixed
29	50941501	50979500	38000	Duplication

[1]BTA = *Bos taurus* autosome

Appendix 5D. Copy number variation regions (CNVRs) scattering in the Crioulo Lageano cattle genome

BTA[1]	Start (bp)	End (bp)	Length (bp)	Event
1	205501	255000	49500	Duplication
1	256001	267000	11000	Duplication
1	268501	383000	114500	Duplication
1	389501	402500	13000	Duplication
1	403501	439000	35500	Duplication
1	439501	450000	10500	Duplication
1	451001	481500	30500	Duplication
1	498501	512000	13500	Duplication
1	513501	526500	13000	Duplication
1	569001	633000	64000	Duplication
1	637501	653500	16000	Duplication
1	654001	664000	10000	Duplication
1	665501	675000	9500	Duplication
2	121555001	121566500	11500	Deletion
3	11721001	11731500	10500	Deletion
3	11732501	11733500	1000	Deletion
3	11734001	11752500	18500	Deletion
3	11854001	11874000	20000	Duplication
3	11981501	11988000	6500	Duplication
3	11988501	12008500	20000	Duplication
3	21065501	21076500	11000	Duplication
3	21302001	21316000	14000	Duplication
3	119159501	119166000	6500	Deletion
4	105569001	105592500	23500	Duplication
4	105598001	105611500	13500	Duplication
4	105620001	105633000	13000	Duplication
5	44193001	44197500	4500	Deletion
5	44283501	44299000	15500	Deletion
5	102920001	102925500	5500	Deletion
6	5433001	5461000	28000	Duplication
6	5915001	5928000	13000	Duplication
6	5928501	5942500	14000	Duplication
6	5961001	5982000	21000	Duplication
7	10800501	10819500	19000	Deletion
7	10919501	10928500	9000	Mixed
7	10986001	10997000	11000	Deletion
7	41402001	41407000	5000	Deletion
9	27501	48000	20500	Duplication
9	84681001	84690000	9000	Duplication
9	87050001	87056000	6000	Duplication

Appendix 5D. Continuation

9	87057501	87069000	11500	Duplication
9	87071001	87080500	9500	Duplication
9	87187501	87194500	7000	Deletion
9	87202501	87216000	13500	Duplication
9	87218001	87227500	9500	Duplication
10	22767501	22781000	13500	Duplication
10	23080001	23084500	4500	Deletion
10	23111501	23139500	28000	Deletion
10	23156001	23165500	9500	Deletion
10	23373001	23380000	7000	Duplication
10	23571001	23591500	20500	Duplication
10	23725501	23731000	5500	Duplication
10	23731501	23747000	15500	Duplication
10	25156501	25167500	11000	Duplication
10	25293001	25310500	17500	Deletion
12	60979501	60987000	7500	Deletion
12	71844001	71864000	20000	Deletion
12	72392001	72400500	8500	Deletion
12	72401501	72411000	9500	Deletion
12	72460001	72504500	44500	Deletion
12	72719501	72728500	9000	Deletion
12	72798501	72818000	19500	Duplication
12	87157501	87167000	9500	Deletion
13	327501	347000	19500	Duplication
13	356001	375500	19500	Duplication
13	10800001	10813000	13000	Duplication
13	10840501	10846500	6000	Duplication
13	10847501	10853500	6000	Duplication
13	10860501	10874000	13500	Duplication
13	17604501	17610500	6000	Deletion
13	62513501	62521000	7500	Duplication
13	62521501	62526500	5000	Duplication
13	62528001	62539000	11000	Duplication
14	13533001	13541500	8500	Mixed
14	13543001	13545500	2500	Mixed
14	13577501	13585000	7500	Duplication
14	13589001	13617500	28500	Duplication
14	13632001	13642500	10500	Duplication
14	13642501	13643000	500	Mixed
14	13757001	13763000	6000	Duplication
14	13764001	13806500	42500	Duplication
14	13811001	13835500	24500	Duplication

Appendix 5D. Continuation

15	45901501	45909000	7500	Duplication
15	45967001	45976500	9500	Duplication
15	46016001	46032500	16500	Duplication
15	47252001	47269500	17500	Deletion
15	48917001	48931000	14000	Deletion
16	7010501	7018500	8000	Duplication
16	7171501	7188000	16500	Duplication
17	30236501	30258500	22000	Deletion
17	39465501	39472000	6500	Deletion
17	68058001	68079500	21500	Deletion
18	45167501	45176000	8500	Deletion
18	50629501	50641000	11500	Duplication
18	50704501	50747000	42500	Duplication
18	57179501	57193000	13500	Mixed
18	57212001	57234500	22500	Duplication
18	57272501	57294000	21500	Duplication
18	57300001	57310500	10500	Duplication
18	58134001	58147500	13500	Duplication
18	61390501	61399000	8500	Duplication
18	62981001	62986500	5500	Deletion
18	63594501	63604000	9500	Duplication
18	63626001	63643000	17000	Deletion
19	43218501	43228500	10000	Duplication
19	57794001	57803000	9000	Deletion
19	57804001	57826500	22500	Deletion
19	57828501	57829500	1000	Deletion
20	96501	119500	23000	Duplication
20	45099501	45116000	16500	Deletion
21	339001	367500	28500	Duplication
21	673001	697000	24000	Duplication
21	32903501	32937000	33500	Duplication
23	1	28500	28500	Duplication
23	25593501	25648000	54500	Deletion
23	26719501	26728500	9000	Duplication
24	20877001	20882000	5000	Deletion
24	61861501	61870500	9000	Duplication
25	1	22000	22000	Duplication
26	14973001	14980500	7500	Deletion
26	51098001	51110000	12000	Deletion
26	51793501	51804500	11000	Deletion
27	6300501	6342000	41500	Duplication
27	6445501	6455500	10000	Deletion

Appendix 5D. Continuation

27	6456501	6462500	6000	Deletion
27	6552501	6573500	21000	Duplication
27	6651001	6656500	5500	Duplication
27	6657001	6661500	4500	Duplication
28	504001	518000	14000	Duplication
29	1880001	1892000	12000	Deletion
29	1893001	1901500	8500	Deletion
29	5504001	5513000	9000	Deletion
29	5539501	5551000	11500	Duplication
29	5681501	5702000	20500	Duplication
29	5709501	5711500	2000	Mixed
29	5735001	5738500	3500	Mixed
29	5740501	5755500	15000	Mixed
29	5755501	5773000	17500	Duplication
29	50947001	50961500	14500	Duplication
29	50962501	50979500	17000	Duplication

[1]BTA = *Bos taurus* autosome

Appendix 6D. Copy number variation regions (CNVRs) scattering in the Pantaneiro cattle genome

BTA[1]	Start (bp)	End (bp)	Length (bp)	Event
1	209001	371000	162000	Duplication
1	372001	383000	11000	Duplication
1	398501	450000	51500	Duplication
1	451001	483000	32000	Duplication
1	513501	524000	10500	Duplication
1	569001	637000	68000	Duplication
1	637501	676000	38500	Duplication
1	677001	678500	1500	Duplication
1	688001	691000	3000	Duplication
2	89160001	89180000	20000	Duplication
2	121555001	121566500	11500	Deletion
2	121598001	121614000	16000	Deletion
3	11721001	11730500	9500	Deletion
3	11732501	11752500	20000	Deletion
3	11757501	11769000	11500	Deletion
3	11854501	11874500	20000	Duplication
3	11974501	12008500	34000	Duplication
3	13289501	13323500	34000	Duplication
3	21065501	21076500	11000	Duplication
3	21302001	21316000	14000	Duplication
3	21317001	21324500	7500	Duplication
3	54206001	54225000	19000	Deletion
3	54229001	54325000	96000	Deletion
4	105569001	105581500	12500	Duplication
4	105583001	105592500	9500	Duplication
4	105598501	105614500	16000	Duplication
4	105616001	105638000	22000	Duplication
5	44283001	44283500	500	Mixed
5	44283501	44299000	15500	Deletion
5	44302501	44317000	14500	Deletion
5	102645001	102656500	11500	Duplication
5	102658001	102682500	24500	Duplication
5	102684501	102711500	27000	Duplication
5	102726501	102735000	8500	Duplication
5	102738501	102786500	48000	Duplication
5	102789001	102807500	18500	Duplication
5	102919501	102925500	6000	Deletion
5	103025001	103039500	14500	Duplication
6	5433001	5461500	28500	Duplication
6	5784501	5800500	16000	Duplication

Appendix 6D. Continuation

6	5918501	5943000	24500	Duplication
7	10800501	10801000	500	Mixed
7	10801001	10819500	18500	Deletion
7	10986001	10997000	11000	Deletion
7	41402001	41407000	5000	Deletion
9	27501	46000	18500	Duplication
9	5654001	5662000	8000	Deletion
9	87047501	87056000	8500	Duplication
9	87057501	87082500	25000	Duplication
9	87118501	87125500	7000	Mixed
9	87187501	87194500	7000	Deletion
9	87198501	87216500	18000	Duplication
9	87217501	87235000	17500	Duplication
10	22766501	22781000	14500	Duplication
10	22951001	22975500	24500	Deletion
10	23008001	23021000	13000	Duplication
10	23080501	23084500	4000	Deletion
10	23085501	23093000	7500	Deletion
10	23107001	23165500	58500	Deletion
10	23165501	23167500	2000	Mixed
10	23373001	23380000	7000	Duplication
10	23528001	23547500	19500	Duplication
10	23578001	23591500	13500	Duplication
10	23659501	23668000	8500	Duplication
10	23740501	23753000	12500	Duplication
10	25091501	25106000	14500	Duplication
10	25144501	25155500	11000	Duplication
10	25156501	25167000	10500	Duplication
10	25293001	25310500	17500	Deletion
10	40264501	40269500	5000	Deletion
12	60979501	60986500	7000	Deletion
12	69822001	69825000	3000	Deletion
12	69825501	69837500	12000	Deletion
12	69838001	69873500	35500	Deletion
12	69911501	69921500	10000	Deletion
12	69922501	69944000	21500	Deletion
12	70075501	70087500	12000	Deletion
12	70205001	70225000	20000	Duplication
12	70778501	70780000	1500	Mixed
12	70788001	70988500	200500	Deletion
12	71491001	71514000	23000	Deletion
12	71528501	71540500	12000	Deletion

Appendix 6D. Continuation

12	71844001	71864000	20000	Deletion
12	72194001	72265500	71500	Deletion
12	72270001	72297500	27500	Deletion
12	72317501	72330000	12500	Deletion
12	72332001	72362500	30500	Deletion
12	72390501	72408000	17500	Deletion
12	72409001	72423000	14000	Deletion
12	72424501	72457000	32500	Deletion
12	72457501	72504500	47000	Deletion
12	72626501	72629500	3000	Mixed
12	72632001	72634000	2000	Mixed
12	72645001	72647500	2500	Mixed
12	72647501	72662000	14500	Duplication
12	72665501	72707500	42000	Duplication
12	72707501	72709500	2000	Mixed
12	72798501	72847000	48500	Duplication
12	72848001	72878500	30500	Duplication
13	325001	347000	22000	Duplication
13	356001	376000	20000	Duplication
13	10800001	10813000	13000	Duplication
13	10856501	10860000	3500	Duplication
13	10860501	10871000	10500	Duplication
14	13575001	13618500	43500	Duplication
14	13623501	13643000	19500	Duplication
14	13743001	13840000	97000	Duplication
15	48917001	48931000	14000	Deletion
16	7170501	7188000	17500	Duplication
16	47738001	47745500	7500	Deletion
17	30242001	30257500	15500	Deletion
17	32215001	32218500	3500	Deletion
17	68058001	68079500	21500	Deletion
18	45167501	45175000	7500	Deletion
18	50703501	50746500	43000	Duplication
18	57179501	57180500	1000	Mixed
18	57182001	57193500	11500	Duplication
18	57212501	57234500	22000	Duplication
18	57272501	57293500	21000	Duplication
18	57300001	57310500	10500	Duplication
18	57333001	57350000	17000	Duplication
18	58130501	58147000	16500	Duplication
18	61041501	61052000	10500	Deletion
18	61384501	61400500	16000	Duplication

Appendix 6D. Continuation

18	62657001	62679500	22500	Duplication
18	62730001	62752500	22500	Duplication
18	63590501	63597000	6500	Duplication
18	63597501	63604000	6500	Duplication
18	63626001	63643000	17000	Deletion
19	43222001	43234500	12500	Duplication
19	57795001	57824500	29500	Deletion
20	96501	119500	23000	Duplication
21	339001	378000	39000	Duplication
21	380001	381000	1000	Duplication
21	382501	389000	6500	Duplication
21	673001	697000	24000	Duplication
21	32903501	32937000	33500	Duplication
21	53920501	53927000	6500	Deletion
23	1	28500	28500	Duplication
23	37001	53000	16000	Duplication
23	25594501	25645500	51000	Deletion
23	26718501	26728500	10000	Duplication
23	27129501	27149500	20000	Duplication
24	20877501	20882000	4500	Deletion
24	61854501	61881000	26500	Duplication
25	1	22000	22000	Duplication
26	14973001	14980000	7000	Deletion
26	25244501	25254500	10000	Deletion
26	25287001	25290000	3000	Deletion
26	51098001	51110500	12500	Deletion
26	51793501	51804500	11000	Deletion
27	6300501	6344000	43500	Duplication
27	6379501	6414500	35000	Duplication
27	6417001	6427500	10500	Duplication
27	6428001	6445000	17000	Duplication
27	6445501	6455500	10000	Deletion
27	6456501	6462500	6000	Deletion
27	6552501	6573500	21000	Duplication
27	6589001	6622500	33500	Duplication
27	6657501	6670000	12500	Duplication
27	6670501	6688500	18000	Duplication
27	6782001	6806000	24000	Deletion
28	504001	524000	20000	Duplication
29	5504001	5513000	9000	Deletion
29	5534501	5551000	16500	Duplication
29	5583001	5611500	28500	Duplication

Appendix 6D. Continuation

29	5638501	5708000	69500	Duplication
29	5711501	5733500	22000	Duplication
29	5734501	5775500	41000	Duplication
29	50942501	50971000	28500	Duplication

[1]BTA = *Bos taurus* autosome

Appendix 7D. Shared copy number variation regions (CNVRs) among Caracu Caldeano, Crioulo Lageano, and Pantaneiro cattle breeds

BTA[1]	Start (bp)	End (bp)	Length (bp)	Event
1	209001	246500	37500	Duplication
1	268501	283500	15000	Duplication
1	284501	336500	52000	Duplication
1	338001	370500	32500	Duplication
1	372001	383000	11000	Duplication
1	406501	426500	20000	Duplication
1	428001	430000	2000	Duplication
1	430501	439000	8500	Duplication
1	439501	450000	10500	Duplication
1	451001	481500	30500	Duplication
1	582001	587500	5500	Duplication
1	588501	595000	6500	Duplication
1	595501	633000	37500	Duplication
1	638501	653500	15000	Duplication
1	654001	661500	7500	Duplication
1	665501	672000	6500	Duplication
2	121555001	121566500	11500	Deletion
3	11854501	11874000	19500	Duplication
3	11981501	11988000	6500	Duplication
3	11995001	11996000	1000	Duplication
3	11996501	12008500	12000	Duplication
3	21065501	21072500	7000	Duplication
3	21302001	21316000	14000	Duplication
4	105571001	105581500	10500	Duplication
4	105583001	105591500	8500	Duplication
4	105598501	105611500	13000	Duplication
4	105620001	105633000	13000	Duplication
5	44283501	44298500	15000	Deletion
5	102920001	102925500	5500	Deletion
6	5433001	5461000	28000	Duplication
6	5924501	5928000	3500	Duplication
6	5928501	5941500	13000	Duplication
7	10801001	10819500	18500	Deletion
7	41402001	41407000	5000	Deletion
9	27501	46000	18500	Duplication
9	87050001	87056000	6000	Duplication
9	87057501	87069000	11500	Duplication
9	87071001	87080500	9500	Duplication
9	87187501	87194500	7000	Deletion
9	87202501	87216000	13500	Duplication

Appendix 7D. Continuation

9	87218001	87227500	9500	Duplication
10	23080501	23084500	4000	Deletion
10	23111501	23139500	28000	Deletion
10	23156001	23165500	9500	Deletion
10	23373001	23380000	7000	Duplication
10	23740501	23747000	6500	Duplication
10	25156501	25165000	8500	Duplication
10	25293001	25310500	17500	Deletion
12	71844001	71864000	20000	Deletion
12	72392001	72400000	8000	Deletion
12	72798501	72818000	19500	Duplication
13	327501	347000	19500	Duplication
13	356001	375500	19500	Duplication
13	10800001	10812500	12500	Duplication
13	10860501	10871000	10500	Duplication
14	13577501	13585000	7500	Duplication
14	13589001	13610000	21000	Duplication
14	13632001	13642500	10500	Duplication
14	13757001	13763000	6000	Duplication
14	13764001	13806500	42500	Duplication
14	13811001	13835500	24500	Duplication
15	48917001	48931000	14000	Deletion
16	7176501	7188000	11500	Duplication
17	68058001	68079500	21500	Deletion
18	45167501	45175000	7500	Deletion
18	57179501	57180500	1000	Duplication
18	57182001	57193000	11000	Duplication
18	57212501	57234500	22000	Duplication
18	57272501	57293500	21000	Duplication
18	57300001	57310500	10500	Duplication
18	58134001	58147000	13000	Duplication
18	63598501	63604000	5500	Duplication
18	63626001	63643000	17000	Deletion
19	43222001	43228500	6500	Duplication
19	57795001	57803000	8000	Deletion
19	57804001	57805500	1500	Deletion
19	57807501	57820000	12500	Deletion
20	96501	119500	23000	Duplication
21	339001	358000	19000	Duplication
21	673001	697000	24000	Duplication
21	32903501	32937000	33500	Duplication
23	1	28500	28500	Duplication

Appendix 7D. Continuation

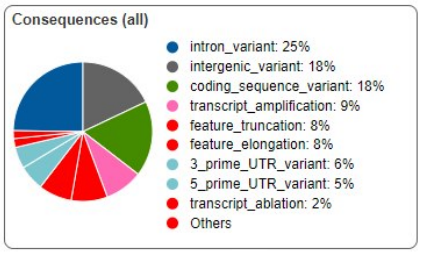
23	25595001	25645500	50500	Deletion
23	26719501	26728500	9000	Duplication
24	61861501	61870500	9000	Duplication
25	1	22000	22000	Duplication
26	14973001	14980000	7000	Deletion
26	51098001	51110000	12000	Deletion
26	51793501	51804500	11000	Deletion
27	6300501	6342000	41500	Duplication
27	6445501	6455500	10000	Deletion
27	6456501	6462500	6000	Deletion
27	6552501	6572500	20000	Duplication
27	6657501	6661500	4000	Duplication
28	504001	518000	14000	Duplication
29	5504001	5513000	9000	Deletion
29	5540001	5549000	9000	Duplication
29	5681501	5689000	7500	Duplication
29	5689501	5702000	12500	Duplication
29	5735001	5738500	3500	Duplication
29	5740501	5755500	15000	Duplication
29	5755501	5768500	13000	Duplication
29	5769001	5773000	4000	Duplication
29	50947001	50961500	14500	Duplication
29	50962501	50971000	8500	Duplication

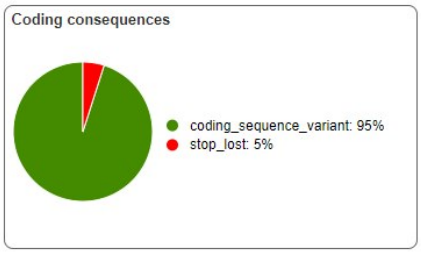
[1]BTA= *Bos taurus* autosome

Appendix 8D.

A

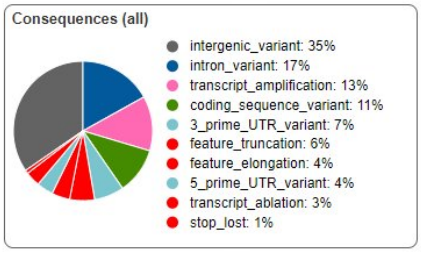
Category	Count
Variants processed	155
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	80
Overlapped transcripts	125
Overlapped regulatory features	-

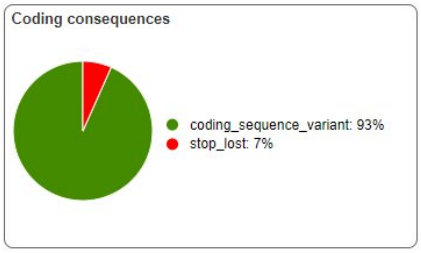




B

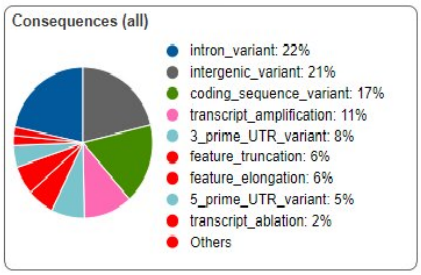
Category	Count
Variants processed	148
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	60
Overlapped transcripts	74
Overlapped regulatory features	-

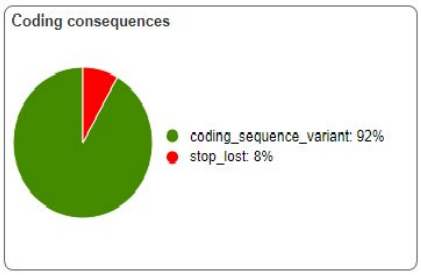




C

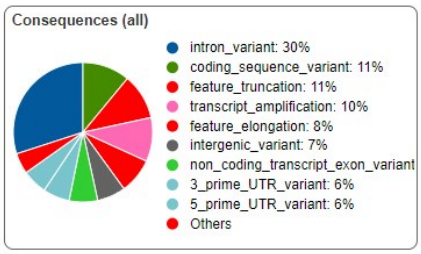
Category	Count
Variants processed	180
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	89
Overlapped transcripts	124
Overlapped regulatory features	-

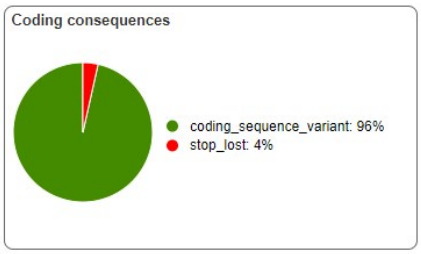




D

Category	Count
Variants processed	105
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	98
Overlapped transcripts	174
Overlapped regulatory features	-





Appendix 9D. Distribution of the variants with high consequence on protein sequence based on copy number variants regions (CNVRs) for the Caracu Caldeano cattle

BTA[1]	Start	End	Event	Consequence	Gene stable ID	Gene Symbol	Gene type
1	338001	370500	duplication	transcript_amplification	ENSBTAG00000006648	-	protein_coding
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000047028	*5S_rRNA*	rRNA
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000053686	*5S_rRNA*	rRNA
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000049697	*5S_rRNA*	rRNA
1	638501	661500	duplication	transcript_amplification	ENSBTAG00000046619	*5S_rRNA*	rRNA
3	11854001	11874500	duplication	transcript_amplification	ENSBTAG00000054063	*CD1A*	protein_coding
3	21302001	21316000	duplication	transcript_amplification	ENSBTAG00000048757	*U1*	snRNA
3	54209001	54225000	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000037634	-	protein_coding
3	119159501	119168500	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,feature_truncation	ENSBTAG00000054589	-	protein_coding
4	105571001	105591500	duplication	transcript_amplification	ENSBTAG00000031234	-	protein_coding
4	105598001	105611500	duplication	transcript_amplification	ENSBTAG00000051786	-	protein_coding
4	105620001	105633000	duplication	transcript_amplification	ENSBTAG00000048597	-	pseudogene
6	5960001	5983000	duplication	transcript_amplification	ENSBTAG00000017045	*FABP2*	protein_coding
7	10801001	10819500	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000026148	-	protein_coding
9	27501	48500	duplication	transcript_amplification	ENSBTAG00000051646	*5S_rRNA*	rRNA
9	87047501	87082500	duplication	transcript_amplification	ENSBTAG00000047902	*ULBP21*	protein_coding
9	87199501	87241000	duplication	transcript_amplification	ENSBTAG00000036061	-	protein_coding
9	87199501	87241000	duplication	transcript_amplification	ENSBTAG00000050143	-	protein_coding
10	25097501	25106000	duplication	transcript_amplification	ENSBTAG00000050668	-	protein_coding
10	25260001	25268000	duplication	transcript_amplification	ENSBTAG00000035530	-	protein_coding
10	25260001	25268000	duplication	transcript_amplification	ENSBTAG00000049741	-	protein_coding

Appendix 9D. Continuation

14	13756001	13835500	duplication	transcript_amplification	ENSBTAG00000049356	-	lncRNA
15	48903001	48931000	deletion	transcript_ablation	ENSBTAG00000048467	*U6*	snRNA
15	48903001	48931000	deletion	transcript_ablation	ENSBTAG00000047714	*OR51A7*	protein_coding
15	80611501	80622500	duplication	transcript_amplification	ENSBTAG00000002007	*PRG3*	protein_coding
15	82106501	82151500	duplication	transcript_amplification	ENSBTAG00000038323	*GLYAT*	protein_coding
18	57211501	57241000	duplication	transcript_amplification	ENSBTAG00000037710	-	protein_coding
18	63598501	63608000	duplication	transcript_amplification	ENSBTAG00000050868	-	protein_coding
18	63598501	63608000	duplication	transcript_amplification	ENSBTAG00000048593	-	protein_coding
20	96501	119500	duplication	transcript_amplification	ENSBTAG00000048439	*5S_rRNA*	rRNA
21	274501	280500	deletion	transcript_ablation	ENSBTAG00000054702	-	protein_coding
21	339001	358000	duplication	transcript_amplification	ENSBTAG00000051425	-	protein_coding
23	25595001	25648000	deletion	transcript_ablation	ENSBTAG00000021077	*BOLA-DQB*	protein_coding
23	25595001	25648000	deletion	transcript_ablation	ENSBTAG00000038128	*BOLA-DQA5*	protein_coding
25	1	22000	duplication	transcript_amplification	ENSBTAG00000048872	*5S_rRNA*	rRNA
27	6300501	6344000	duplication	transcript_amplification	ENSBTAG00000050630	*DEFB13*	protein_coding
27	6552501	6572500	duplication	transcript_amplification	ENSBTAG00000053555	-	protein_coding
27	6667501	6687500	duplication	transcript_amplification	ENSBTAG00000051383	*DEFB7*	protein_coding
27	7138001	7146000	duplication	transcript_amplification	ENSBTAG00000053557	*DEFB4A*	protein_coding
27	7147501	7196500	duplication	transcript_amplification	ENSBTAG00000033545	*EBD*	protein_coding
29	5734001	5768500	duplication	transcript_amplification	ENSBTAG00000054266	*U6*	snRNA
29	5734001	5768500	duplication	transcript_amplification	ENSBTAG00000051538	-	pseudogene
29	41924001	41930500	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000052238	-	protein_coding

[1]BTA= *Bos taurus* autosome

Appendix 10D. Distribution of the variants with high consequence on protein sequence based on copy number variants regions (CNVRs) for the Criulo Lageano cattle

BTA[1]	Start	End	Event	Consequence	Gene stable ID	Gene Symbol	Gene type
1	268501	383000	duplication	transcript_amplification	ENSBTAG00000006648	-	protein_coding
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000047028	*5S_rRNA*	rRNA
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000053686	*5S_rRNA*	rRNA
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000049697	*5S_rRNA*	rRNA
1	654001	664000	duplication	transcript_amplification	ENSBTAG00000046619	*5S_rRNA*	rRNA
3	11721001	11731500	deletion	transcript_ablation	ENSBTAG00000039189	-	protein_coding
3	11734001	11752500	deletion	transcript_ablation	ENSBTAG00000022893	-	protein_coding
3	11854001	11874000	duplication	transcript_amplification	ENSBTAG00000054063	*CD1A*	protein_coding
3	21302001	21316000	duplication	transcript_amplification	ENSBTAG00000048757	*U1*	snRNA
4	105569001	105592500	duplication	transcript_amplification	ENSBTAG00000031234	-	protein_coding
4	105598001	105611500	duplication	transcript_amplification	ENSBTAG00000051786	-	protein_coding
4	105620001	105633000	duplication	transcript_amplification	ENSBTAG00000048597	-	pseudogene
5	44193001	44197500	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000022971	-	protein_coding
6	5961001	5982000	duplication	transcript_amplification	ENSBTAG00000017045	*FABP2*	protein_coding
7	10800501	10819500	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000026148	-	protein_coding
9	27501	48000	duplication	transcript_amplification	ENSBTAG00000051646	*5S_rRNA*	rRNA
9	87218001	87227500	duplication	transcript_amplification	ENSBTAG00000050143	-	protein_coding
10	23571001	23591500	duplication	transcript_amplification	ENSBTAG00000052580	-	protein_coding
14	13811001	13835500	duplication	transcript_amplification	ENSBTAG00000049356	-	lncRNA
15	45901501	45909000	duplication	transcript_amplification	ENSBTAG00000019144	*OR6A2*	protein_coding
15	46016001	46032500	duplication	transcript_amplification	ENSBTAG00000035675	-	protein_coding
15	46016001	46032500	duplication	transcript_amplification	ENSBTAG00000042988	*U6*	snRNA

Appendix 10D. Continuation

18	57212001	57234500	duplication	transcript_amplification	ENSBTAG00000037710	-	protein_coding
18	63594501	63604000	duplication	transcript_amplification	ENSBTAG00000050868	-	protein_coding
19	43218501	43228500	duplication	transcript_amplification	ENSBTAG00000049920	*U2*	snRNA
20	96501	119500	duplication	transcript_amplification	ENSBTAG00000048439	*5S_rRNA*	rRNA
21	339001	367500	duplication	transcript_amplification	ENSBTAG00000048268	-	protein_coding
21	339001	367500	duplication	transcript_amplification	ENSBTAG00000051425	-	protein_coding
23	25593501	25648000	deletion	transcript_ablation	ENSBTAG00000021077	*BOLA-DQB*	protein_coding
23	25593501	25648000	deletion	transcript_ablation	ENSBTAG00000038128	*BOLA-DQA5*	protein_coding
25	1	22000	duplication	transcript_amplification	ENSBTAG00000048872	*5S_rRNA*	rRNA
27	6300501	6342000	duplication	transcript_amplification	ENSBTAG00000050630	*DEFB13*	protein_coding
27	6552501	6573500	duplication	transcript_amplification	ENSBTAG00000053555	-	protein_coding
29	5681501	5702000	duplication	transcript_amplification	ENSBTAG00000009197	-	protein_coding
29	5740501	5755500	deletion	transcript_ablation	ENSBTAG00000051538	-	pseudogene
29	5740501	5755500	duplication	transcript_amplification	ENSBTAG00000051538	-	pseudogene
29	5755501	5773000	duplication	transcript_amplification	ENSBTAG00000054266	*U6*	snRNA

[1]BTA= *Bos taurus* autosome

Appendix 11D. Distribution of the variants with high consequence on protein sequence based on copy number variants regions (CNVRs) for the Pantaneiro cattle.

BTA[1]	Start	End	Event	Consequence	Gene stable ID	Gene Symbol	Gene type
1	209001	371000	duplication	transcript_amplification	ENSBTAG00000006648	-	protein_coding
1	451001	483000	duplication	transcript_amplification	ENSBTAG00000047028	*5S_rRNA*	rRNA
1	451001	483000	duplication	transcript_amplification	ENSBTAG00000053686	*5S_rRNA*	rRNA
1	451001	483000	duplication	transcript_amplification	ENSBTAG00000049697	*5S_rRNA*	rRNA
1	637501	676000	duplication	transcript_amplification	ENSBTAG00000046619	*5S_rRNA*	rRNA
3	11721001	11730500	deletion	transcript_ablation	ENSBTAG00000039189	-	protein_coding
3	11732501	11752500	deletion	transcript_ablation	ENSBTAG00000022893	-	protein_coding
3	11757501	11769000	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000038502	-	protein_coding
3	11854501	11874500	duplication	transcript_amplification	ENSBTAG00000054063	*CD1A*	protein_coding
3	21302001	21316000	duplication	transcript_amplification	ENSBTAG00000048757	*U1*	snRNA
3	54206001	54225000	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000037634	-	protein_coding
3	54229001	54325000	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000037634	-	protein_coding
4	105569001	105581500	duplication	transcript_amplification	ENSBTAG00000031234	-	protein_coding
4	105598501	105614500	duplication	transcript_amplification	ENSBTAG00000051786	-	protein_coding
4	105616001	105638000	duplication	transcript_amplification	ENSBTAG00000048597	-	pseudogene
5	44302501	44317000	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000000198	-	protein_coding
7	10801001	10819500	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000026148	-	protein_coding
9	27501	46000	duplication	transcript_amplification	ENSBTAG00000051646	*5S_rRNA*	rRNA
9	87118501	87125500	deletion	transcript_ablation	ENSBTAG00000038891	-	protein_coding
9	87118501	87125500	duplication	transcript_amplification	ENSBTAG00000038891	-	protein_coding
9	87217501	87235000	duplication	transcript_amplification	ENSBTAG00000050143	-	protein_coding
10	22951001	22975500	deletion	transcript_ablation	ENSBTAG00000046819	-	protein_coding

Appendix 11D. Continuation

10	23528001	23547500	duplication	transcript_amplification	ENSBTAG00000048374	-	protein_coding
10	23528001	23547500	duplication	transcript_amplification	ENSBTAG00000051554	-	protein_coding
10	23578001	23591500	duplication	transcript_amplification	ENSBTAG00000052580	-	protein_coding
10	25091501	25106000	duplication	transcript_amplification	ENSBTAG00000050668	-	protein_coding
10	25144501	25155500	duplication	transcript_amplification	ENSBTAG00000054698	-	protein_coding
14	13743001	13840000	duplication	transcript_amplification	ENSBTAG00000052622	-	protein_coding
14	13743001	13840000	duplication	transcript_amplification	ENSBTAG00000050469	-	protein_coding
14	13743001	13840000	duplication	transcript_amplification	ENSBTAG00000049356	-	lncRNA
18	57212501	57234500	duplication	transcript_amplification	ENSBTAG00000037710	-	protein_coding
18	63597501	63604000	duplication	transcript_amplification	ENSBTAG00000050868	-	protein_coding
19	43222001	43234500	duplication	transcript_amplification	ENSBTAG00000030308	*U2*	snRNA
19	43222001	43234500	duplication	transcript_amplification	ENSBTAG00000050956	*U2*	snRNA
20	96501	119500	duplication	transcript_amplification	ENSBTAG00000048439	*5S_rRNA*	rRNA
21	339001	378000	duplication	transcript_amplification	ENSBTAG00000048268	-	protein_coding
21	339001	378000	duplication	transcript_amplification	ENSBTAG00000051425	-	protein_coding
21	339001	378000	duplication	transcript_amplification	ENSBTAG00000054133	-	protein_coding
23	37001	53000	duplication	transcript_amplification	ENSBTAG00000053364	*5S_rRNA*	rRNA
23	25594501	25645500	deletion	transcript_ablation	ENSBTAG00000021077	*BOLA-DQB*	protein_coding
23	25594501	25645500	deletion	transcript_ablation	ENSBTAG00000038128	*BOLA-DQA5*	protein_coding
23	27129501	27149500	duplication	transcript_amplification	ENSBTAG00000023563	-	protein_coding
25	1	22000	duplication	transcript_amplification	ENSBTAG00000048872	*5S_rRNA*	rRNA
27	6300501	6344000	duplication	transcript_amplification	ENSBTAG00000050630	*DEFB13*	protein_coding
27	6379501	6414500	duplication	transcript_amplification	ENSBTAG00000051796	-	protein_coding
27	6428001	6445000	duplication	transcript_amplification	ENSBTAG00000050419	-	protein_coding
27	6552501	6573500	duplication	transcript_amplification	ENSBTAG00000053555	-	protein_coding
27	6589001	6622500	duplication	transcript_amplification	ENSBTAG00000048737	*DEFB10*	protein_coding
27	6589001	6622500	duplication	transcript_amplification	ENSBTAG00000053326	-	protein_coding

Appendix 11D. Continuation

27	6670501	6688500	duplication	transcript_amplification	ENSBTAG00000051383	*DEFB7*	protein_coding
29	5534501	5551000	duplication	transcript_amplification	ENSBTAG00000048894	-	protein_coding
29	5638501	5708000	duplication	transcript_amplification	ENSBTAG00000054977	-	protein_coding
29	5638501	5708000	duplication	transcript_amplification	ENSBTAG00000009197	-	protein_coding
29	5711501	5733500	duplication	transcript_amplification	ENSBTAG00000049646	-	protein_coding
29	5734501	5775500	duplication	transcript_amplification	ENSBTAG00000054266	*U6*	snRNA
29	5734501	5775500	duplication	transcript_amplification	ENSBTAG00000048802	-	protein_coding
29	5734501	5775500	duplication	transcript_amplification	ENSBTAG00000051538	-	pseudogene

[1]BTA= *Bos taurus* autosome

Appendix 12D. Distribution of the variants with high consequence on protein sequence based on shared copy number variants regions (CNVRs) among the three studied breeds.

BTA[1]	Start	End	Event	Consequence	Gene stable ID	Gene Symbol	Gene type
1	209001	246500	duplication	transcript_amplification	112447072	*LOC112447072*	lncRNA
1	284501	336500	duplication	transcript_amplification	112447074	*LOC112447074*	lncRNA
1	338001	370500	duplication	transcript_amplification	ENSBTAG00000006648	-	protein_coding
1	406501	426500	duplication	transcript_amplification	100138661	*LOC100138661*	lncRNA
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000047028	*5S_rRNA*	rRNA
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000053686	*5S_rRNA*	rRNA
1	451001	481500	duplication	transcript_amplification	ENSBTAG00000049697	*5S_rRNA*	rRNA
1	654001	661500	duplication	transcript_amplification	ENSBTAG00000046619	*5S_rRNA*	rRNA
3	11854501	11874000	duplication	transcript_amplification	ENSBTAG00000054063	*CD1A*	protein_coding
3	11854501	11874000	duplication	transcript_amplification	-	-	cdna
3	21302001	21316000	duplication	transcript_amplification	ENSBTAG00000048757	*U1*	snRNA
3	21302001	21316000	duplication	transcript_amplification	107132278	*LOC107132278*	snRNA
4	10557100 1	10558150 0	duplication	transcript_amplification	ENSBTAG00000031234	-	protein_coding
4	10558300 1	10559150 0	duplication	transcript_amplification	615948	*LOC615948*	protein_coding
4	10559850 1	10561150 0	duplication	transcript_amplification	ENSBTAG00000051786	-	protein_coding
4	10559850 1	10561150 0	duplication	transcript_amplification	100297263	*LOC100297263*	protein_coding
4	10559850 1	10561150 0	duplication	transcript_amplification	789121	*LOC789121*	protein_coding
4	10562000 1	10563300 0	duplication	transcript_amplification	ENSBTAG00000048597	-	pseudogene
7	10801001	10819500	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	ENSBTAG00000026148	-	protein_coding
7	10801001	10819500	deletion	stop_lost,coding_sequence_variant,3_prime_UTR_variant,intron_variant,feature_truncation	100299045	*LOC100299045*	protein_coding
9	27501	46000	duplication	transcript_amplification	ENSBTAG00000051646	*5S_rRNA*	rRNA

Appendix 12D. Continuation

9	27501	46000	duplication	transcript_amplification	112448010	*LOC112448010*	lncRNA
9	87202501	87216000	duplication	transcript_amplification	100336795	*LOC100336795*	protein_coding
9	87218001	87227500	duplication	transcript_amplification	ENSBTAG00000050143	-	protein_coding
14	13764001	13806500	duplication	transcript_amplification	112449613	*LOC112449613*	protein_coding
14	13811001	13835500	duplication	transcript_amplification	ENSBTAG00000049356	-	lncRNA
18	57212501	57234500	duplication	transcript_amplification	ENSBTAG00000037710	-	protein_coding
18	57272501	57293500	duplication	transcript_amplification	104974923	*LOC104974923*	protein_coding
18	63598501	63604000	duplication	transcript_amplification	ENSBTAG00000050868	-	protein_coding
20	96501	119500	duplication	transcript_amplification	ENSBTAG00000048439	*5S_rRNA*	rRNA
21	339001	358000	duplication	transcript_amplification	ENSBTAG00000051425	-	protein_coding
21	673001	697000	duplication	transcript_amplification	112443211	*LOC112443211*	lncRNA
23	1	28500	duplication	transcript_amplification	101906171	*LOC101906171*	protein_coding
23	1	28500	duplication	transcript_amplification	112443722	*LOC112443722*	lncRNA
23	25595001	25645500	deletion	transcript_ablation	ENSBTAG00000021077	*BOLA-DQB*	protein_coding
23	25595001	25645500	deletion	transcript_ablation	ENSBTAG00000038128	*BOLA-DQA5*	protein_coding
23	25595001	25645500	deletion	transcript_ablation	-	-	cdna
24	61861501	61870500	duplication	transcript_amplification	-	-	cdna
24	61861501	61870500	duplication	transcript_amplification	786348	*LOC786348*	protein_coding
25	1	22000	duplication	transcript_amplification	ENSBTAG00000048872	*5S_rRNA*	rRNA
26	51793501	51804500	deletion	transcript_ablation	112444460	*LOC112444460*	lncRNA
27	6300501	6342000	duplication	transcript_amplification	ENSBTAG00000050630	*DEFB13*	protein_coding
27	6300501	6342000	duplication	transcript_amplification	112444638	*LOC112444638*	lncRNA
27	6300501	6342000	duplication	transcript_amplification	112444639	*LOC112444639*	lncRNA
27	6552501	6572500	duplication	transcript_amplification	ENSBTAG00000053555	-	protein_coding
27	6657501	6661500	duplication	transcript_amplification	112444642	*LOC112444642*	lncRNA
29	5681501	5689000	duplication	transcript_amplification	101902282	*TRIM48*	protein_coding
29	5740501	5755500	duplication	transcript_amplification	ENSBTAG00000051538	-	pseudogene

Appendix 12D. Continuation

29	5740501	5755500	duplication	transcript_amplification	523762	*LOC523762*	protein_coding
29	5755501	5768500	duplication	transcript_amplification	ENSBTAG00000054266	*U6*	snRNA
29	5755501	5768500	duplication	transcript_amplification	112444941	*LOC112444941*	snRNA

[1]BTA= *Bos taurus* autosome

Appendix 13D. Annotated genes within the copy number variants regions (CNVRs).

Breed	Gene stable ID	BTA[1]	Gene start (bp)	Gene end (bp)	Gene name	Gene type
CAR	ENSBTAG00000054063	3	11858601	11863244	*CD1A*	protein_coding
CAR	ENSBTAG00000050603	5	102627214	102694488	*WC1.3*	protein_coding
CAR	ENSBTAG00000049861	5	102768087	102819373	*WC1*	protein_coding
CAR	ENSBTAG00000017045	6	5970114	5973346	*FABP2*	protein_coding
CAR	ENSBTAG00000047902	9	87049719	87056668	*ULBP21*	protein_coding
CAR	ENSBTAG00000039329	9	87236522	87247606	*RAET1G*	protein_coding
CAR	ENSBTAG00000014081	11	82736891	83207199	*NBAS*	protein_coding
CAR	ENSBTAG00000002045	11	82902339	82922670	*PSMD13*	protein_coding
CAR	ENSBTAG00000004147	13	17565238	17604498	*FBH1*	protein_coding
CAR	ENSBTAG00000047714	15	48915399	48916337	*OR51A7*	protein_coding
CAR	ENSBTAG00000002007	15	80616545	80651796	*PRG3*	protein_coding
CAR	ENSBTAG00000052095	15	82088599	82108601	*GAT*	protein_coding
CAR	ENSBTAG00000030847	15	82122690	82192681	*GLYATL2*	protein_coding
CAR	ENSBTAG00000038323	15	82127678	82142196	*GLYAT*	protein_coding
CAR	ENSBTAG00000054818	18	62655625	62662605	*KIR3DL2*	protein_coding
CAR	ENSBTAG00000021077	23	25607502	25622150	*BOLA-DQB*	protein_coding
CAR	ENSBTAG00000038128	23	25636255	25643878	*BOLA-DQA5*	protein_coding
CAR	ENSBTAG00000000951	26	51737604	51822703	*JAKMIP3*	protein_coding
CAR	ENSBTAG00000000447	26	14968609	15042514	*FRA10AC1*	protein_coding
CAR	ENSBTAG00000050630	27	6326521	6328629	*DEFB13*	protein_coding
CAR	ENSBTAG00000051383	27	6676076	6678281	*DEFB7*	protein_coding
CAR	ENSBTAG00000053557	27	7138873	7140876	*DEFB4A*	protein_coding
CAR	ENSBTAG00000033545	27	7165176	7180420	*EBD*	protein_coding
CRL	ENSBTAG00000054063	3	11858601	11863244	*CD1A*	protein_coding
CRL	ENSBTAG00000048852	4	105567721	105569006	*TRBV3-1*	protein_coding
CRL	ENSBTAG00000017045	6	5970114	5973346	*FABP2*	protein_coding

Appendix 13D. Continuation

CRL	ENSBTAG00000047902	9	87049719	87056668	*ULBP21*	protein_coding
CRL	ENSBTAG00000009144	13	62503682	62514220	*BPIFA2A*	protein_coding
CRL	ENSBTAG00000019144	15	45906051	45907010	*OR6A2*	protein_coding
CRL	ENSBTAG00000021077	23	25607502	25622150	*BOLA-DQB*	protein_coding
CRL	ENSBTAG00000038128	23	25636255	25643878	*BOLA-DQA5*	protein_coding
CRL	ENSBTAG00000000951	26	51737604	51822703	*JAKMIP3*	protein_coding
CRL	ENSBTAG00000000447	26	14968609	15042514	*FRA10AC1*	protein_coding
CRL	ENSBTAG00000050630	27	6326521	6328629	*DEFB13*	protein_coding
PAN	ENSBTAG00000009725	2	89102213	89173911	*AOX1*	protein_coding
PAN	ENSBTAG00000054063	3	11858601	11863244	*CD1A*	protein_coding
PAN	ENSBTAG00000048852	4	105567721	105569006	*TRBV3-1*	protein_coding
PAN	ENSBTAG00000050603	5	102627214	102694488	*WC1.3*	protein_coding
PAN	ENSBTAG00000049861	5	102768087	102819373	*WC1*	protein_coding
PAN	ENSBTAG00000054245	5	103022590	103086215	*CD163L1*	protein_coding
PAN	ENSBTAG00000024888	6	5745779	5894011	*PDE5A*	protein_coding
PAN	ENSBTAG00000047902	9	87049719	87056668	*ULBP21*	protein_coding
PAN	ENSBTAG00000054818	18	62655625	62662605	*KIR3DL2*	protein_coding
PAN	ENSBTAG00000021077	23	25607502	25622150	*BOLA-DQB*	protein_coding
PAN	ENSBTAG00000038128	23	25636255	25643878	*BOLA-DQA5*	protein_coding
PAN	ENSBTAG00000000951	26	51737604	51822703	*JAKMIP3*	protein_coding
PAN	ENSBTAG00000000447	26	14968609	15042514	*FRA10AC1*	protein_coding
PAN	ENSBTAG00000004612	26	25203123	26181296	*SORCS3*	protein_coding
PAN	ENSBTAG00000050630	27	6326521	6328629	*DEFB13*	protein_coding
PAN	ENSBTAG00000048737	27	6596422	6598413	*DEFB10*	protein_coding
PAN	ENSBTAG00000051383	27	6676076	6678281	*DEFB7*	protein_coding
Shared CNVRs	ENSBTAG00000054063	3	11858601	11863244	*CD1A*	protein_coding
Shared CNVRs	ENSBTAG00000047902	9	87049719	87056668	*ULBP21*	protein_coding

Appendix 13D. Continuation

Shared CNVRs	ENSBTAG00000021077	23	25607502	25622150	*BOLA-DQB*	protein_coding
Shared CNVRs	ENSBTAG00000038128	23	25636255	25643878	*BOLA-DQA5*	protein_coding
Shared CNVRs	ENSBTAG00000000951	26	51737604	51822703	*JAKMIP3*	protein_coding
Shared CNVRs	ENSBTAG00000000447	26	14968609	15042514	*FRA10AC1*	protein_coding
Shared CNVRs	ENSBTAG00000050630	27	6326521	6328629	*DEFB13*	protein_coding

[1]BTA= *Bos taurus* autosome

Appendix 14D. Reported quantitative trait locus **(**QTLs) based on copy number variation regions (CNVRs) identified within each breed (Caracu Caldeano, Crioulo Lageano, and Pantaneiro) and based on shared CNVRs observed in between the three studied breeds.

BTA[1]	Start (bp)	End (bp)	Event	QTL
Caracu Caldeano				
6	5924501	5928000	Duplication	Foot angle QTL [1], Milk fat yield QTL [1], Milk yield QTL [1], Net merit QTL [1], Milk protein percentage QTL [1], Milk protein yield QTL [1], and Rear leg placement - rear and side view QTL [1]
14	13756001	13835500	Duplication	Milk tetracosanoic acid content QTL [2]
17	68057501	68079500	Deletion	Non-return rate QTL [3]
21	673001	697000	Duplication	Calving ease QTL [3]
Crioulo Lageano				
6	5915001	5928000	Duplication	Foot angle QTL [1], Milk fat yield QTL [1], Milk yield QTL [1], Net merit QTL [1], Milk protein percentage QTL [1], Milk protein yield QTL [1], and Rear leg placement - rear and side view QTL [1]
14	13764001	13806500	Duplication	Milk tetracosanoic acid content QTL [2]
17	68058001	68079500	Deletion	Non-return rate QTL [3]
21	673001	697000	Duplication	Calving ease QTL [3]
Pantaneiro				
3	13289501	13323500	Duplication	Milk protein percentage QTL [4]
6	5918501	5943000	Duplication	Foot angle QTL [1], Milk fat yield QTL [1], Milk yield QTL [1], Net merit QTL [1], Milk protein percentage QTL [1], Milk protein yield QTL [1], and Rear leg placement - rear and side view QTL [1]
14	13743001	13840000	Duplication	Milk tetracosanoic acid content QTL [2]
17	68058001	68079500	Deletion	Non-return rate QTL [3]
21	673001	697000	Duplication	Calving ease QTL [3]
29	5583001	5611500	Duplication	Milk palmitic acid content [5]
Shared CNVRs				
6	5924501	5928000	Duplication	Foot angle QTL [1], Milk fat yield QTL [1], Milk yield QTL [1], Net merit QTL [1], Milk protein percentage QTL [1], Milk protein yield QTL [1], and Rear leg placement - rear and side view QTL [1]
14	13764001	13806500	Duplication	Milk tetracosanoic acid content QTL [2]
17	68058001	68079500	Deletion	Non-return rate QTL [3]
21	673001	697000	Duplication	Calving ease QTL [3]

[1]BTA: *Bos taurus* autosome.

1. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics. 2011;12:408.

2. Ibeagha-Awemu EM, Peters SO, Akwanji KA, Imumorin IG, Zhao X. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. Sci Rep. 2016;6:31109.

3. Frischknecht M, Bapst B, Seefried FR, Signer-Hasler H, Garrick D, Stricker C, et al. Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. BMC Genomics. 2017;18.

4. Nayeri S, Sargolzaei M, Abo-Ismail MK, May N, Miller SP, Schenkel F, et al. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 2016;17:75.

5. Buitenhuis B, Janss LLG, Poulsen NA, Larsen LB, Larsen MK, Sørensen P. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics. 2014;15:1112.