RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta tese será disponibilizado somente a partir de 01/05/2024.

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

DETECTION AND FUNCTIONAL ASSESSMENT OF STRUCTURAL VARIANTS USING WHOLE GENOME RE-SEQUENCING DATA IN NELLORE CATTLE

Natalia Andrea Marin Garzon

Animal Scientist

M.Sc. Genetics and Animal Breeding

Jaboticabal-SP

2023

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

DETECTION AND FUNCTIONAL ASSESSMENT OF STRUCTURAL VARIANTS USING WHOLE GENOME RE-SEQUENCING DATA IN NELLORE CATTLE

Student: Natalia Andrea Marín Garzón

Advisor: Profª. Dra. Lucia Galvão de Albuquerque

Co-advisor: Dra. Giovana Vargas

Thesis presented to the School of Agricultural and Veterinarian Sciences– SãoPaulo State University, campus of Jaboticabal in partial fulfillment of requirements for the degree of Doctor in Genetics and Animal Breeding.

Jaboticabal-SP

2023

M337d

Marin-Garzon, Natalia Andrea
Detection and functional assessment of structural variants using whole genome re-sequencing data in Nellore cattle / Natalia Andrea Marin-Garzon. -- Jaboticabal, 2024
110 p. : il., tabs.

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

1. Animal breeding. 2. CNV. 3. DNA. 4. Genome sequencing. 5. Zebu cattle. 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.

Impacto potencial desta pesquisa

O presente estudo permitiu identificar variantes genéticas no genoma de 151 touros representativos da raça Nelore, conhecidas como variações estruturais. Essas variações são compostas por diferentes tipos, dos quais conseguiram-se identificar deleções, duplicações, e inversões de fragmentos de DNA. Também foram identificadas as regiões do genoma onde há maior ocorrência dessas variações na população avaliada, bem com os genes localizados dentre dessas regiões, que estão envolvidos em funções biológicas e moleculares, que são importantes para processos adaptativos, e que também estão associados a características de importância económica em gado de corte, como a eficiência alimentar. Esses resultados contribuem a desvendar os mecanismos genéticos que promoveram a adaptação ambiental do gado Nelore, causado potencialmente por mudanças estruturais e rearranjos de segmentos de DNA funcionalmente importantes.

Potential impact of this research

The present study allowed the identification of genetic variants in the genome of 151 representative Nellore bulls, known as structural variations. These variations correspond to different types, from which deletions, duplications, and inversions of DNA fragments were identified. Genomic regions with the highest occurrence of these variations in the evaluated population were also identified, as well as the genes located within these regions, which are involved in biological and molecular functions that are important for adaptive processes, and which are also associated with economically important traits in cattle, such as feed efficiency. These results contribute to unveiling the genetic mechanisms driving the environmental adaptation of Nellore cattle, potentially caused by structural changes and rearrangements of functionally important DNA segments.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: DETECTION AND FUNCTIONAL ASSESSMENT OF STRUCTURAL VARIANTS USING WHOLE GENOME RE-SEQUENCING DATA IN NELLORE CATTLE

AUTORA: NATALIA ANDREA MARIN GARZON
ORIENTADORA: LUCIA GALVÃO DE ALBUQUERQUE
COORIENTADORA: GIOVANA VARGAS

Aprovada como parte das exigências para obtenção do Título de Doutora em Genética e Melhoramento Animal, pela Comissão Examinadora:

gov.br Documento assinado digitalmente
LUCIA GALVAO DE ALBUQUERQUE
Data: 14/11/2023 10:34:52-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. LUCIA GALVÃO DE ALBUQUERQUE (Participaçao Virtual)
Departamento de Zootecnia / FCAV UNESP Jaboticabal

gov.br Documento assinado digitalmente
DIERCLES FRANCISCO CARDOSO
Data: 23/11/2023 12:13:15-0300
Verifique em https://validar.iti.gov.br

Dr. DIERCLES FRANCISCO CARDOSO (Participaçao Virtual)
Gensys Tecnologia e Sistemas LTDA / Jaboticabal/SP

gov.br Documento assinado digitalmente
RICARDO VIEIRA VENTURA
Data: 07/12/2023 13:22:38-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. RICARDO VIEIRA VENTURA (Participaçao Virtual)
Departamento de Nutrição e Produção Animal / FMVZ USP Pirassununga/SP

Prof. Dr. LUIS GABRIEL GONZALEZ HERRERA (Participaçao Virtual)
Departamento de Produção Animal / Universidad Nacional de Colombia - Medellin/Colômbia

Prof. Dr. RAFAEL ESPIGOLAN (Participaçao Virtual)
Departamento de Zootecnia e Ciências Biológicas / Universidade Federal de Santa Maria - Santa Maria/RS

gov.br Documento assinado digitalmente
RAFAEL ESPIGOLAN
Data: 14/11/2023 11:07:19-0300
Verifique em https://validar.iti.gov.br

Jaboticabal, 31 de outubro de 2023

ABOUT THE AUTHOR

NATALIA ANDREA MARÍN GARZÓN- born in April 12th, 1994 in Bello-Antioquia, Colombia. Daughter of Hilda Natalia Garzón and Carlos Marín. Bachelor of Animal Sciences from the National University of Colombia, campus of Medellin-2017. Her studies were granted with the scholarship “Fondo Para La Educación Superior” from Government of Antioquia (2013-2016). During the last year of undergraduate course, she undertook an internship in Genetics and Animal Breeding at São Paulo State University under the supervision of Prof. Dr. Lucia Galvão de Albuquerque and Prof. Dr. Luis Gabriel González Herrera, granted by the Department of Foreign Affairs of National University of Colombia. She started her M.Sc. studies in 2017 in Genetics and Animal Breeding at FCAV-UNESP with a grant from “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior”, under the supervision of Prof. Dr. Lucia Galvão de Albuquerque and Prof. Dr. Ana Fabricia Braga Magalhães, obtaining her title in 2019. During the same year she started her Ph.D. formation in the current graduate program of Animal Sciences at the same university under the supervision of Prof. Dr. Lucia Galvão de Albuquerque and Dr. Giovana Vargas. Her Ph.D. studies were granted by “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” from 2019 to 2022. In January 2022 she started a position as Research and Development Coordinator at Colombian Association of Zebu Cattle Breeders (ASOCEBU).

I entirely dedicate this work to my mother Hilda, for being my safe harbor every second of my existence, for letting my wings grow and supporting me in the search and reach of my dreams. For being the best example of hope and strength even during the heaviest battles. To her, who with empowerment inspired me to finish my Ph.D.

ACKNOWLEDGMENTS

Firstly, I thank the wise universe that drew the paths I had to take to reach my goals. For making me face difficult situations that surely transform me into a more ethical professional.

I am so grateful to Prof. Dr Lucia for having accepted me in her research group since the beginning of my undergraduate internship. For her teachings and support that were essential for the development of my Ph.D.

I thank …

All my family, but especially to my parents and my brother for their unconditional support. For motivating me every day to continue reaching my dreams, and for demonstrating every single day of my life, the most pure and sincere feeling of love I have ever received.

My life partner and love, Oscar, who has undoubtedly been my greatest support during this journey. For the sincere love and the loyal friendship, for the advises and for motivate me to be a better person and professional.

My co advisor Dr. Giovana Vargas, who even in the distance was always available to help me through all the difficulties, especially for being a friend during the most challenging moments along this journey.

Professor Luis Gabriel González, who motivated me to embark on the area, who continues teaching me every day and who has played an essential role in the development of my professional career.

Larissa Braga and our virtual meetings, that were essential to guide me in the execution of this study. For being always willingness to help me and discuss results.

To my “Albuquerque” brothers and sisters, especially Lucio, Patricia, Marta, Larissa, Gabby, Bruna, and everyone who helped and taught me from the first day I came to the research group. I always keep in my memories the mornings of extremely dark coffee and good conversations, the “churrasco” and pizza days. I am especially grateful to Lucio, for all the help during the realization of this study.

To my dear friend Valentina, who was my greatest support during the dark days that came with the pandemics of Covid-19, and the lovely friends Brazil gave me: Marisa, Karen, Maria Camila, and Jaque, for the support and loyal friendship.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001"

CONTENT

List of Tables

List of figures.

Detection and functional assessment of structural variants using whole genome re-sequencing data in Nellore cattle.

Abstract- Ongoing advances in genome sequencing technologies have enabled the unravelling of many structural variants (SVs) in livestock genomes. Association of SVs with complex traits are promising targets for animal breeding because of their effects on gene expression. The aims of this study were: *i).* to detect structural variants using whole genome resequencing data of 151 representative Nellore bulls by combining calling algorithms *ii)* to discover non redundant and highly frequent regions of structural variants (SVR) in the analyzed bulls. *iii).* to search for positional candidate genes, and quantitative trait loci (QTL) overlapping the most frequent SVR in the population. iv) to assess the functional impact of positional candidate genes overlapping SVR trough enriched gene ontology terms (GO terms) related to biological process (BP), molecular function (MF), cellular component (CC), and biochemical pathways. The whole genome re-sequencing (WGS) data from 151 representative Nellore bulls was used to conduct genome-wide structural variation calling, as well as detection of common SVRs for the analyzed bulls. The gene content and nearby QTLs of SVRs were retrieved from publicly available genomic databases, and functional enrichment analysis of positional candidate genes overlapping the most frequent SVRs were conducted using PANTHER. A total of 215,031 high-confidence SV was obtained, most of them corresponding to copy number variants (CNV) (183,032 deletions-DEL, and 14,013 duplications-DUP), and 17,986 inversions (INV). Total structural variation encompasses, on average, 4.81% of the individual autosomal genome extension. A total of 3,752 non-redundant SVR frequent in more than 5% of bulls were obtained, corresponding in more than 97% to regions of copy number variants (CNVR), and 3% to regions of inversions (INVR). All SVR comprises 13.13% of total autosomal genome extension, which were attributed in 11.4% to CNVR and 1.7% to INVR. Among all SVRs, 532 were shared by more than 50% of bulls and overlapped a total of 130 QTL distributed into 6 QTL types: exterior, health, meat and carcass, milk, production, and reproduction, which are related to a total of 50 economically important traits. Most SVR overlapped QTLs related to residual feed intake, structural soundness, multiple birth, clinical mastitis, and milk energy yield. Regarding gene content, 204 SVRs, overlapped a total of 1,164 positional candidate genes, which were significatively overrepresented into GO terms related to BP, MF, CC and one biochemical pathway. Among the significantly enriched genes, we highlight members of the olfactory receptor (OR) gene family, which play essential roles in mechanisms for adaptation to the environment. These genes were mainly found into regions of inversion and mixed events. Similarly, genes from the defensin family (*DEFB*), that play important roles in the innate immune system of multicellular organisms, and which are known to be caused by duplication events that mammalian genomes have undergone. Other important genes were found in this study, such as the members of the secretory phospholipase A2 family, adhesion G protein-coupled receptors, and zinc finger binding proteins. Most of the genes found in this study have been described as potential candidates for feed efficiency indicator traits, which reflects the biochemical mechanisms in which they are involved that have led to the improved fitness. The results of this study provide important knowledge about the mechanisms driving

changes in the genome in Nellore cattle, the contributed to adaptation to environment.
Keywords: DNA; copy number variation; mobile elements; next generation sequencing; sequence gains and losses; zebu cattle

Detecção e avaliação funcional de variantes estruturais utilizando dados de re-sequenciamento do genoma completo em bovinos Nelore.

Resumo-Os avanços contínuos nas tecnologias de sequenciamento do genoma permitiram o desvendamento de diferentes variantes estruturais (SVs) nos genomas dos animais de interesse zootécnico. A associação de SVs com características complexas são alvos promissores para o melhoramento animal devido aos seus efeitos na expressão gênica. Os objetivos deste estudo foram: i). detectar variantes estruturais usando dados de res-sequenciamento do genoma completo de 151 touros Nelore representativos combinando diferentes algoritmos; ii) descobrir regiões não redundantes e altamente frequentes de variantes estruturais (SVR) nos touros analisados. iii). procurar genes candidatos posicionais e locos de características quantitativas ("QTL") em sobreposição às SVRs mais frequentes na população. iv) avaliar o impacto funcional de genes candidatos posicionais abrigados pelas SVRs através de enriquecimento funcional de termos de ontologia genética (GO terms) relacionados a processo biológico (BP), função molecular (MF), componente celular (CC) e vias bioquímicas. Dados de re-sequenciamento do genoma completo de 151 touros Nelore representativos foram usados para conduzir a detecção de variantes estruturais, bem como para identificar SVRs comuns aos touros analisados. O conteúdo genético e os QTLs próximos de SVRs foram recuperados de bancos de dados genômicos disponíveis publicamente, e a análise de enriquecimento funcional de genes candidatos posicionais que se sobrepõem aos SVRs mais frequentes foi conduzida usando PANTHER. Foram obtidas 215.031 SVs de alta confiança, a maioria correspondendo a variantes de número de cópias (CNV) (183.032 deleções-DEL, e 14.013 duplicações-DUP, e 17.986 inversões (INV). A variação estrutural total abrange, em média, 4,81% da extensão do genoma autossômico individual. Foram obtidos um total de 3.752 SVRs frequentes em mais de 5% dos touros, correspondendo em mais de 97% a regiões de variantes de número de cópias (CNVR) e 3% a regiões de inversões (INVR). Todos os SVRs compreendem 13,13% da extensão total do genoma autossômico, que foram atribuídos em 11,4% a CNVRs e 1,7% a INVRs. Entre todos as SVRs, 532 foram compartilhados por mais de 50% dos touros e sobrepuseram um total de 130 QTL distribuídos em 6 tipos: exterior, saúde, carne e carcaça, leite, produção e reprodução, que estão relacionados a um total de 50 características de importância económica. A maioria dos QTLs sobrepostos às SVRs estão relacionados a consumo residual, solidez estrutural, nascimentos múltiplos, mastite clínica, e produção de energia do leite. Em relação ao conteúdo gênico, 204 SVRs das mais frequentes se sobrepuseram a um total de 1.164 genes candidatos posicionais, que foram significativamente super-representados em termos GO relacionados a BP, MF, CC e uma via bioquímica. Dentre os genes significativamente enriquecidos, destacamos os membros da família de genes dos receptores olfativos (OR), com funções importantes nos mecanismos de adaptação ao ambiente. Esses genes foram encontrados principalmente dentre de regiões com ocorrência de eventos de inversões e eventos mistos, que podem explicar a prevalência nos genomas ao longo da evolução dos mamíferos. De forma similar, os genes da família das defensinas (DEFB) que desempenham papéis importantes no sistema imunológico inato de organismos multicelulares, e que são conhecidos por serem causados por

eventos de duplicação que sofreram os genomas dos mamíferos. Destacam-se também membros da família da fosfolipase A2 secretora, os receptores acoplados à proteína G de adesão e as proteínas de ligação aos dedos de zinco. A maioria dos genes encontrados neste estudo são candidatos potenciais para características indicadoras de eficiência alimentar, o que reflete os mecanismos bioquímicos nos quais estão envolvidos e que levaram à melhoria do valor adaptativo. Os resultados deste estudo fornecem conhecimento importante sobre os mecanismos que impulsionam as mudanças no genoma de bovinos Nelore, as causas dos processos de adaptação e a caracterização das consequências da variação estrutural em relação à diversidade genética.
Palavras-chave: DNA; elementos móveis; gado zebu, ganhos e perdas de sequência, sequenciamento de próxima geração; variação do número de cópias

1. INTRODUCTION

The widespread developments of molecular techniques and bioinformatic algorithms have enabled the unravelling of genetic mechanisms driving differences between individuals, which are mainly attributed to single-nucleotide variants (SNVs), small insertions and deletions (indels;<50 bp), and structural variations (SVs) (Bickhart, 2014; Fernandes Junior et al., 2020; Ho, Urban & Mill., 2020).

Structural variants are generally defined as inherited variations in DNA sequences larger than 50 Bp that exhibit changes in the copy number, orientation, and/or chromosomal location, in relation to the reference genome (Alkan et al., 2011, Escamarís, Docampo & Rabionet., 2015). The balanced SVs types correspond to the mobile elements (inversions and translocations), which cause rearrangements of DNA sequences (Escamarís, Docampo & Rabionet., 2015; Bickhart., 2014). By other side, small insertions, and deletions (indels), and the copy number variants (CNVs), comprise the unbalanced SVs in which gains or losses of DNA segments are observed. CNVs are the most studied SVs types in livestock, and encompass duplication and deletion events ranging in sizes from 1 Kb to 5 Mb (Ho, Urban & Mills., 2020; Zhang et al., 2009; Freeman et al., 2006).

Compared to SNPs, SVs affect larger fractions of the genome and account for the greatest amount of total polymorphic content among individual genomes (Couldrey, Keehan, Johnson, Tiplady, Winkelman, Littlejohn, & Scott, 2017). SVs may have potentially greater effects than SNPs: from modification of genes to disruptions in regulation of gene expression, leading to recessive disorders (Conrad et al., 2010; Zhang et al., 2009).

A variety of based platforms such as comparative genomic hybridization and SNP arrays have been commonly used for SVs wide detection in livestock (Bickhart et al., 2012). However, array-based approaches lack genome coverage, especially in segmental duplication regions, which are known for being hotspots for CNVs formation (Zhao et al., 2013). The advantages of detecting SV using next generation sequencing data (NGS) include, higher coverage and resolution, more accurate prediction of copy numbers, more precise detection of breakpoints, and higher power

to identify novel SVs (Zhao et al., 2013). In cattle, the study of structural variations has reached an increasingly attention because of their association with economically important phenotypes such as fatty acid profile (Lemos et al., 2018a), meet tenderness (Silva et al., 2016), feed conversion ratio (Santana et al., 2016); growth traits (Liu et al., 2020; Zhou et al., 2016), milk production traits (Zhou et al., 2016), and milk somatic cell score (Durán Aguilar et al., 2017).

Although NGS has become more feasible in livestock, most of studies regarding structural variant wide detection are still being conducted from SNP arrays-based approaches. There are few published studies that explored the structural diversity using whole genome sequencing data all with less than 100 animals (Peripolli et al., 2023; Sun et al., 2023; Braga et al., 2022; Butty et al., 2020; Silva et al., 2016, Santana et al., 2016; Bickhart et al., 2016; Boussaha et al., 2015). Furthermore, most of them have been focused on the assessment of copy number variants (CNVs) rather than the other SVs types.

The detection of balanced and unbalanced structural variants using next generation sequencing data may contribute to the unravelling of the molecular mechanisms underlying genetic diversity, breed adaptation and evolution, especially in Nellore cattle in which they have not been well elucidated so far. Therefore, the aims were: *i).* to detect structural variants using whole genome resequencing data of 151 representative Nellore bulls by combining calling algorithms *ii)* to discover non redundant and highly frequent regions of structural variants (SVR) in the analyzed bulls. *iii).* to search for positional candidate genes, and quantitative trait loci (QTL) overlapping the most frequent SVR in the population. iv) to assess the functional impact of positional candidate genes overlapping SVR trough enriched gene ontology terms (GO terms) related to biological process (BP), molecular function (MF), cellular component (CC), and biochemical pathways.

5. FINAL CONSIDERATIONS.

The ongoing developments on NGS-based algorithms have allowed highly confident predictions of structural variants. Here, we detected and assessed deletion, duplication, and inversion events at individual and population level, providing valuable information regarding the structural diversity of the Nellore cattle genome. This contributes to unravel highly prevalent structural variants from the whole genome sequences of a large set of representative ancestors of Brazilian Nellore population. Particularly, this can be the first ongoing study reporting and analyzing mobile elements such as inversions in Nellore cattle, which still lacks information.

Using complementary bioinformatic tools and applying exigent filtering criteria within and between software and sample, we detected a total of 215,031high-confidence structural variants corresponding to 183,032 deletions-DEL, and 14,013 duplications-DUP, and 17,986 inversions (INV). Moreover, 3,752 non-redundant SVR that were frequent in more than 5% of bulls were identified, corresponding in more than 97% to regions of copy number variants (CNVR), and 3% to regions of inversions (INVR). From CNVR, 90% corresponded to regions of deletions (CNVR-DEL), 3% to regions of duplications (CNVR-DUP) and 7% to mixed regions, in which occur both duplication and deletion events (CNV-MIXED). Among the SVR, 532 were shared by more than 50% of the analyzed key ancestors, and harbored 1,164 positional candidate genes and QTLs with known functional implications in cattle.

Our results contribute to the unravelling of genetic mechanisms driving environmental adaptation of Nellore cattle, caused by structural changes and re arrangements of functionally important genes. As an example, the genes belonging to the highly preserved gene families, as the highlighted olfactory receptors-OR, the defensins, the secretory phospholipase A2 family, the adhesion G protein-coupled receptors, and the zinc fingers binding proteins, that overlapped genomic regions in which different structural variation events are occurring, and which have been mainly described in the literature for being potential candidates for feed efficiency indicator traits.

6. REFERENCES

Abyzov, A., Urban, A.E., Snyder, M., & Gerstein, M. (2011). CNVnator : An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research*, *21*, 974–984. https://doi.org/10.1101/gr.114876.110.974

Alkan, C., Coe, B. P., & Eichler, E. E. (2011b). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, *12*(5), 363–376. https://doi.org/10.1038/nrg2958

Almeida Santana, M. H., Junior, G. A. O., Cesar, A. S. M., Freua, M. C., da Costa Gomes, R., da Luz e Silva, S., Leme, P. R., Fukumasu, H., Carvalho, M. E., Ventura, R. V., Coutinho, L. L., Kadarmideen, H. N., & Ferraz, J. B. S. (2016). Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. *Journal of Applied Genetics*, *57*(4), 495–504. https://doi.org/10.1007/s13353-016-0344-7

Antunes de Lemos, M. V., Berton, M. P., Ferreira de Camargo, G. M., Peripolli, E., de Oliveira Silva, R. M., Ferreira Olivieri, B., Cesar, A. S. M., Pereira, A. S. C., de Albuquerque, L. G., de Oliveira, H. N., Tonhati, H., & Baldi, F. (2018). Copy number variation regions in Nellore cattle: Evidences of environment adaptation. *Livestock Science*, *207*, 51–58. https://doi.org/10.1016/j.livsci.2017.11.008

Arlt, M. F., Wilson, T. E., & Glover, T. W. (2012a). Replication stress and mechanisms of CNV formation. *Current Opinion in Genetics & Development*, *22*(3), 204–210. https://doi.org/10.1016/j.gde.2012.01.009

Arlt, M. F., Wilson, T. E., & Glover, T. W. (2012b). Replication stress and mechanisms of CNV formation. *Current Opinion in Genetics and Development*, *22*(3), 204–210. https://doi.org/10.1016/j.gde.2012.01.009

Berton, M. P., de Antunes Lemos, M. V., Chud, T. C. S., Stafuzza, N. B., Kluska, S., Amorim, S. T., ... & Baldi, F. (2021). Genome-wide association study between copy number variation regions and carcass-and meat-quality traits in Nellore cattle. *Animal Production Science*, *61*(8), 731-744

Bickhart, D. M., Hou, Y., Schroeder, S. G., Alkan, C., Cardone, M. F., Matukumalli, L. K., Song, J., Schnabel, R. D., Ventura, M., Taylor, J. F., Garcia, J. F., Tassell, C. P. Van, Sonstegard, T. S., Eichler, E. E., Liu, G. E., & Hughes, H. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Research*, *22*, 778–790. https://doi.org/10.1101/gr.133967.111.778

Bickhart, D. M., & Liu, G. E. (2014). The challenges and importance of structural variation detection in livestock. In *Frontiers in Genetics* (Vol. 5, Issue FEB). Frontiers Research Foundation. https://doi.org/10.3389/fgene.2014.00037

Chen, L., Chamberlain, A. J., Reich, C. M., Daetwyler, H. D., & Hayes, B. J. (2017). Detection and validation of structural variations in bovine whole-genome sequence data. *Genetics Selection Evolution*, *49*(1). https://doi.org/10.1186/s12711-017-0286-5

Connor, E. E., Zhou, Y., & Liu, G. E. (2018). The essence of appetite: Does olfactory receptor variation play a role? In *Journal of Animal Science* (Vol. 96, Issue 4, pp. 1551–1558). American Society of Animal Science. https://doi.org/10.1093/jas/sky068

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., Macarthur, D. G., Macdonald, J. R., Onyiah, I., Wing, A., Pang, C., … Hurles, M. E. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, *464*(7289), 704–712. https://doi.org/10.1038/nature08516

Couldrey, C., Keehan, M., Johnson, T., Tiplady, K., Winkelman, A., Littlejohn, M. D., & Scott, A. (2017). Detection and assessment of copy number variation using PacBio long-read and Illumina sequencing in New Zealand dairy cattle. *Journal of Dairy Science*, *100*(7), 5472–5478. https://doi.org/10.3168/jds.2016-12199

Couldrey, C., Keehan, M., Johnson, T., Tiplady, K., Winkelman, A., Littlejohn, M. D., Scott, A., Kemper, K. E., Hayes, B., Davis, S. R., & Spelman, R. J. (2017). Detection and assessment of copy number variation using PacBio long-read and Illumina sequencing in New Zealand dairy cattle. *Journal of Dairy Science*, *100*(7), 5472–5478. https://doi.org/10.3168/jds.2016-12199

Dobigny, G., Britton-Davidian, J., & Robinson, T. J. (2017). Chromosomal polymorphism in mammals: an evolutionary perspective. *Biological Reviews*, *92*(1), 1–21. https://doi.org/10.1111/brv.12213

Durán Aguilar, M., Román Ponce, S. I., Ruiz López, F. J., González Padilla, E., Vásquez Peláez, C. G., Bagnato, A., & Strillacci, M. G. (2017). Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *Journal of Animal Breeding and Genetics*, *134*, 49–59. https://doi.org/10.1111/jbg.12238

Escaramís, G., Docampo, E., & Rabionet, R. (2015). A decade of structural variants: description, history, and methods to detect structural variation. *Briefings in functional genomics*, *14*(5), 305-314.

Fernandes Júnior, G.A., de Oliveira, H.N., Carvalheiro, R. *et al.* Whole-genome sequencing provides new insights into genetic mechanisms of tropical adaptation in Nellore (*Bos primigenius indicus*). *Sci Rep* **10**, 9412 (2020). https://doi.org/10.1038/s41598-020-66272-7

Fleischer, J., R. Bumbalo, V. Bautze, J. Strotmann, and H. Breer. 2015. Expression of odorant receptor Olfr78 in enteroendocrine cells of the colon. Cell Tissue Res. 361:697–710. doi:10.1007/s00441-015-2165-0

Fonseca, P. A. S., Suárez-Vega, A., Marras, G., & Cánovas, Á. (2020). GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. *GigaScience*, *9*(12). https://doi.org/10.1093/gigascience/giaa149

Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., Mccarroll, S. A., Altshuler, D. M., Aburatani, H., Jones, K. W., Tyler-smith, C., Hurles, M. E., Carter, N. P., Scherer, S. W., & Lee, C. (2006). Copy number variation : New insights in genome diversity. *Genome Research*, *16*(617), 949–961. https://doi.org/10.1101/gr.3677206.16

Georges, M., Charlier, C., & Hayes, B. (2019). Harnessing genomic information. *Nature Reviews Genetics*, *20*, 135–156. https://doi.org/10.1038/s41576-018-0082-2

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, *10*(6), 381–391. https://doi.org/10.1038/nrg2575

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nature Reviews Genetics*, *10*(AuguST), 551–564. https://doi.org/10.1038/nrg2593

Hayes, B., & Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome*, *53*, 876–883. https://doi.org/10.1139/G10-076

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review : Genomic selection in dairy cattle : Progress and challenges. *Journal of Dairy Science*, *92*(2), 433–443. https://doi.org/10.3168/jds.2008-1646

Henrique, V., Correia, L., Regitano, D. A., Geistlinger, L., Morosini, S., Zimmer, R., & Coutinho, L. L. (2016). Genome-Wide Detection of CNVs and Their Association with Meat Tenderness in Nelore Cattle. *PLOS ONE*, *11*, 1–22. https://doi.org/10.1371/journal.pone.0157711

Higgins, M. G., Fitzsimons, C., McClure, M. C., McKenna, C., Conroy, S., Kenny, D. A., McGee, M., Waters, S. M., & Morris, D. W. (2018). GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and

GFRA2 expression in beef cattle. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-32374-6

Ho, S. S., Urban, A. E., & Mills, R. E. (2020). Structural variation in the sequencing era. In *Nature Reviews Genetics* (Vol. 21, Issue 3, pp. 171–189). Nature Research. https://doi.org/10.1038/s41576-019-0180-9

Hou, Y., Bickhart, D. M., Hvinden, M. L., Li, C., Song, J., Boichard, D. A., Fritz, S., Eggen, A., DeNise, S., Wiggans, G. R., Sonstegard, T. S., Van Tassell, C. P., & Liu, G. E. (2012). Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. *BMC Genomics*, *13*(1). https://doi.org/10.1186/1471-2164-13-376

Hu, Y., Xia, H., Li, M., Xu, C., Ye, X., Su, R., Zhang, M., Nash, O., Sonstegard, T. S., Yang, L., Liu, G. E., & Zhou, Y. (2020). Comparative analyses of copy number variations between Bos taurus and Bos indicus. *BMC Genomics*, *21*(1). https://doi.org/10.1186/s12864-020-07097-6

Islam, S., Akhand, M. R. N., & Hasan, M. (2023). Evolutionary trend of bovine β-defensin proteins toward functionality prediction: A domain-based bioinformatics study. *Heliyon*, *9*(3).

Julliard, A. K., D. Al Koborssy, D. A. Fadool, and B. Palouzier- Paulignan. 2017. Nutrient sensing: another chemosen- sitivity of the olfactory system. Front. Physiol. 8:468. doi:10.3389/fphys.2017.00468

Keel, B. N., Keele, J. W., & Snelling, W. M. (2017). Genome-wide copy number variation in the bovine genome detected using low coverage sequence of popular beef breeds. *Animal Genetics*, *48*(2), 141–150.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). *HISAT: a fast spliced aligner with low memory requirements*. *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317.HISAT

Lee, K., Nguyen, D. T., Choi, M., Cha, S. Y., Kim, J. H., Dadi, H., Seo, H. G., Seo, K., Chun, T., & Park, C. (2013). Analysis of cattle olfactory subgenome: The first detail study on the characteristics of the complete olfactory receptor repertoire of a ruminant. *BMC Genomics*, *14*(1). https://doi.org/10.1186/1471-2164-14-596

Lee, Y.-L., Costa, G., Moreira, M., Karim, L., Druet, T., Oget-Ebrad, C., Veerkamp, R. F., Groenen, M. A. M., Georges, M., & Bouwman, A. C. (2022). *High-resolution structural variation catalogue in a large-scale whole genome sequenced bovine family cohort data*. https://doi.org/10.21203/rs.3.rs-2007387/v1

Lemos, M. V., Peripolli, E., Berton, M. P., Loise, F., Feitosa, B., Olivieri, B. F., Stafuzza, N. B., Tonussi, R. L., Kluska, S., Lucas, H., Chiaia, J., Mueller, L.,

Ferrinho, A. M., Simone, A., Prereira, C., Oliveira, H. N. De, Albuquerque, L. G. De, & Baldi, F. (2018). Association study between copy number variation and beef fatty acid profile of Nellore cattle. *Animal Frontiers*, *59*, 203–223. https://doi.org/https://doi.org/10.1007/s13353-018-0436-7

Letaief, R., Rebours, E., Grohs, C., Meersseman, C., Fritz, S., Trouilh, L., Esquerré, D., Barbieri, J., Klopp, C., Philippe, R., Blanquet, V., Boichard, D., Rocha, D., & Boussaha, M. (2017a). Identification of copy number variation in French dairy and beef breeds using next - generation sequencing. *Genetics Selection Evolution*, *49*(77), 1–15. https://doi.org/10.1186/s12711-017-0352-z

Letaief, R., Rebours, E., Grohs, C., Meersseman, C., Fritz, S., Trouilh, L., Esquerré, D., Barbieri, J., Klopp, C., Philippe, R., Blanquet, V., Boichard, D., Rocha, D., & Boussaha, M. (2017b). Identification of copy number variation in French dairy and beef breeds using next - generation sequencing. *Genetics Selection Evolution*, *49*(77), 1–15. https://doi.org/10.1186/s12711-017-0352-z

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Mukiibi, R., Wang, Y., Plastow, G. S., & Li, C. (2021). Identification of candidate genes and enriched biological functions for feed efficiency traits by integrating plasma metabolites and imputed whole genome sequence variants in beef cattle. *BMC Genomics*, *22*(1). https://doi.org/10.1186/s12864-021-08064-5

Martínez-Fundichely, A., Casillas, S., Egea, R., Ràmia, M., Barbadilla, A., Pantano, L., Puig, M., & Cáceres, M. (2014). InvFEST, a database integrating information of polymorphic inversions in the human genome. *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1122

Melo, T. P., Camargo, G. M. F., Albuquerque, L. G., & Carvalheiro, R. (2017). Genome-wide association study provides strong evidence of genes affecting the reproductive performance of Nellore beef cows. *PLOS ONE*, *12*(5), 1–14. https://doi.org/10.1371/journal.pone.0178551

Mills, R. E., Walter, K., Stewart, C., Handsaker, R. E., Chen, K., Alkan, C., Abyzov, A., Yoon, S. C., Ye, K., Cheetham, R. K., Chinwalla, A., Conrad, D. F., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L. M., Iqbal, Z., Kang, S., … Peterson, J. L. (2011). Mapping copy number variation by population-scale genome sequencing. *Nature*, *470*(7332), 59–65. https://doi.org/10.1038/nature09708

Montaldo, H. H., & Meza-herrera, C. A. (1998). Use of molecular markers and major genes in the genetic improvement of livestock. *Electronic Journal of Biotechnology*, *1*(2), 1–7.

Niimura, Y., & Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS ONE*, *2*(8). https://doi.org/10.1371/journal.pone.0000708

Olender, T., Lancet, D., & Nebert, D. W. (2008). Update on the olfactory receptor (OR) gene superfamily. *Human Genomics*, *3*(1), 1479–7364. https://doi.org/10.1186/1479-7364-3-1-87

Olivieri, B. F., Mercadante, M. E. Z., Cyrillo, J. N. D. S. G., Branco, R. H., Bonilha, S. F. M., De Albuquerque, L. G., De Oliveira Silva, R. M., & Baldi, F. (2016). Genomic regions associated with feed efficiency indicator traits in an experimental nellore cattle population. *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0164390

Ornitz, D.M., Itoh, N. Fibroblast growth factors. *Genome Biol* **2**, reviews3005.1 (2001). https://doi.org/10.1186/gb-2001-2-3-reviews3005

Palouzier-Paulignan, B., M. C. Lacroix, P. Aimé, C. Baly, M. Caillol, P. Congar, A. K. Julliard, K. ucker, and D. A. Fadool. 2012. Olfaction under metabolic influences.

Peterson, B. G., Carl, P., Boudt, K., Bennett, R., Ulrich, J., Zivot, E., ... & Wuertz. D. (2014). PerformanceAnalytics: Econometric tools for performance and risk analysis. *R package version*. *1*(3).

Puig, M., Casillas, S., Villatoro, S., & Cáceres, M. (2015). Human inversions and their functional consequences. *Briefings in Functional Genomics*, *14*(5), 369–379. https://doi.org/10.1093/bfgp/elv020

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Radke, D. W., & Lee, C. (2015). Adaptive potential of genomic structural variation in human and mammalian evolution. *Briefings in functional genomics*, *14*(5), 358-368.

Rafter, P., Gormley, I. C., Purfield, D., Parnell, A. C., Naderi, S., & Berry, D. P. (2021). Genome-wide association analyses of carcass traits using copy number variants and raw intensity values of single nucleotide polymorphisms in cattle. *BMC Genomics*, *22*(1). https://doi.org/10.1186/s12864-021-08075-2

Rasekh, M. E., Chiatante, G., Miroballo, M., Tang, J., Ventura, M., Amemiya, C. T., Eichler, E. E., Antonacci, F., & Alkan, C. (2017). Discovery of large genomic

inversions using long range information. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-016-3444-1

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., Global variation in copy number in the human genome. Nature. 2006 Nov 23;444(7118):444-54. doi: 10.1038/nature05329. PMID: 17122850; PMCID: PMC2669898.

Salmon Hillbertz N.H.C., Isaksson M., Karlsson E.K., Hellmén E., Pielberg G.R., Savolainen P., et al. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nat Genet. 2007;39:1318–20.

Samuel, B., & Dinka, H. (2020). In silico analysis of the promoter region of olfactory receptors in cattle (Bos indicus) to understand its gene regulation. *Nucleosides, Nucleotides and Nucleic Acids*, *39*(6), 853–865. https://doi.org/10.1080/15257770.2020.1711524

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., Bhattarai, E. K., Molaei, M., Freetly, H. C., Hansen, S. L., Yampara-Iquise, H., Johnson, K. A., Kerley, M. S., Kim, J. W., Loy, D. D., Marques, E., Neibergs, H. L., Schnabel, R. D., Shike, D. W., … Taylor, J. F. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-3754-y

Silva, J. M., Giachetto, P. F., da Silva, L. O., Cintra, L. C., Paiva, S. R., Yamagishi, M. E. B., & Caetano, A. R. (2016). Genome-wide copy number variation (CNV) detection in Nelore cattle reveals highly frequent variants in genome regions harboring QTLs affecting production traits. *BMC Genomics*, *17*(1). https://doi.org/10.1186/s12864-016-2752-9

Singh, U., Deb, R., Rahman, R., Alex, R., Kumar, S., Chakraborty, S., Dhama, K., & Sharma, A. (2014). Molecular markers and their applications in cattle genetic research : A review. *Biomarkers and Genomic Medicine*, *6*(2), 49–58. https://doi.org/10.1016/j.bgm.2014.03.001

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C., Beazley, C., Thorne, N., Redon, R., Bird, C. P., Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S. W., Tavaré, S., Deloukas, P., Hurles, M., & Dermitzakis, E. T. (2014).

Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science*, *315*(2007), 848–853. https://doi.org/10.1126/science.1136678

Sun, T., Pei, S., Liu, Y., Hanif, Q., Xu, H., Chen, N., Lei, C., & Yue, X. (2023). Whole genome sequencing of simmental cattle for SNP and CNV discovery. *BMC Genomics*, *24*(1). https://doi.org/10.1186/s12864-023-09248-x

Syvänen, A. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, *2*(December), 930–942.

Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S., & Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. In *Bioinformatics* (Vol. 28, Issue 21, pp. 2711–2718). https://doi.org/10.1093/bioinformatics/bts535

Wellenreuther, M., & Bernatchez, L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. In *Trends in Ecology and Evolution* (Vol. 33, Issue 6, pp. 427–440). Elsevier Ltd. https://doi.org/10.1016/j.tree.2018.04.002

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, Hadley. (2009). *Ggplot2 : elegant graphics for data analysis*. Springer.

Yudav, A. K., Tomar, S. S., Jha, A. K., & Singh, J. (2017). Importance of Molecular Markers in Livestock Improvement : A Review. *International Journal of Agriculture Innovations and Research*, *5*(4).

Zarate, S., Carroll, A., Mahmoud, M., Krasheninina, O., Jun, G., Salerno, W. J., Schatz, M. C., Boerwinkle, E., Gibbs, R. A., & Sedlazeck, F. J. (2021). Parliament2: Accurate structural variant calling at scale. *GigaScience*, *9*(12), 1–9. https://doi.org/10.1093/gigascience/giaa145

Zhang, F., Gu, W., Hurles, M. E., & Lupski, J. R. (2009). Copy Number Variation in Human Health , Disease , and Evolution. *Annu Rev Genomics Hum Genet*, *10*, 451–481. https://doi.org/10.1146/annurev.genom.9.081307.164217

Zhao, M., Wang, Q., Wang, Q., Jia, P., & Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data : features and perspectives. *BMC Bioinformatics*, *14*. https://doi.org/doi:10.1186/1471-2105-14-S11-S1

Zhou, Y., Utsunomiya, Y. T., Xu, L., Hamidi, E., Bickhart, D. M., Alexandre, P. A., Rosen, B. D., Schroeder, S. G., Carvalheiro, R., Henrique, H., Neves, D. R.,

Sonstegard, T. S., Tassell, C. P. Van, Bento, J., Ferraz, S., Fukumasu, H., Garcia, J. F., & Liu, G. E. (2016). Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. *BMC Genomics*, *17*, 1–9. https://doi.org/10.1186/s12864-016-2461-4