UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CAMPUS DE JABOTICABAL

GENOME-WIDE ASSOCIATION STUDY BETWEEN COPY NUMBER VARIATION AND FEEDING BEHAVIOR, FEED EFFICIENCY, AND GROWTH TRAITS IN NELLORE CATTLE

Lorena Ferreira Benfica

Zootecnista

2023

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CAMPUS DE JABOTICABAL

GENOME-WIDE ASSOCIATION STUDY BETWEEN COPY NUMBER VARIATION AND FEEDING BEHAVIOR, FEED EFFICIENCY, AND GROWTH TRAITS IN NELLORE CATTLE

Lorena Ferreira Benfica

Orientadora: Dra. Maria Eugenia Zerlotti Mercadante

Coorientador: Dr. Luiz Fernando Brito

Trabalho apresentado à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de doutora em Genética e Melhoramento Animal

2023

B465g

Benfica, Lorena Ferreira
Genome-wide association study between copy number variation and feeding behavior, feed efficiency, and growth traits in Nellore cattle / Lorena Ferreira Benfica. -- Jaboticabal, 2023
95 p. : il., tabs., fotos

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientadora: Maria Eugênia Zerlotti Mercadante
Coorientadora: Luiz Fernando Brito

1. Zebu. 2. Genética animal. 3. Seleção genética. 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: GENOME-WIDE ASSOCIATION STUDY BETWEEN COPY NUMBER VARIATION AND FEEDING BEHAVIOR, FEED EFFICIENCY, AND GROWTH TRAITS IN NELLORE CATTLE

AUTORA: LORENA FERREIRA BENFICA
ORIENTADORA: MARIA EUGÊNIA ZERLOTTI MERCADANTE
COORIENTADOR: LUIZ FERNANDO BRITO

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

gov.br Documento assinado digitalmente
MARIA EUGENIA ZERLOTTI MERCADANTE
Data: 20/11/2023 18:30:45-0300
Verifique em https://validar.iti.gov.br

Pesquisadora Dra. MARIA EUGÊNIA ZERLOTTI MERCADANTE (Participaçao Virtual)
Instituto de Zootecnia / Sertaozinho/SP

gov.br Documento assinado digitalmente
LEONARDO SIQUEIRA GLORIA
Data: 20/11/2023 20:55:56-0300
Verifique em https://validar.iti.gov.br

Dr. LEONARDO SIQUEIRA GLÓRIA (Participaçao Virtual)
PIC Genetics / Hendersonville/TN/EUA

gov.br Documento assinado digitalmente
ALESSANDRA ALVES DA SILVA
Data: 21/11/2023 21:12:49-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. ALESSANDRA ALVES DA SILVA (Participaçao Virtual)
Departamento de Zootecnia / Universidade Federal de Viçosa (UFV) - Viçosa/MG

gov.br Documento assinado digitalmente
LENIRA EL FARO ZADRA
Data: 22/11/2023 14:37:05-0300
Verifique em https://validar.iti.gov.br

Pesquisadora Dra. LENIRA EL FARO ZADRA (Participaçao Virtual)
Instituto de Zootecnia / Sertãozinho/SP

gov.br Documento assinado digitalmente
LUCIO FLAVIO MACEDO MOTA
Data: 22/11/2023 15:19:53-0300
Verifique em https://validar.iti.gov.br

Pós doutorando LÚCIO FLÁVIO MACÊDO MOTA (Participaçao Virtual)
Departamento de Zootecnia / FCAV UNESP Jaboticabal

Jaboticabal, 17 de novembro de 2023

AUTHOR'S CURRICULAR INFORMATION

Lorena Ferreira Benfica – Born in Araxa, MG, on March 20th, 1992, she enrolled in the Animal Science course in February 2010 at the Federal Institute of Education, Science, and Technology of the Triangulo Mineiro, Uberaba-MG. During her undergraduate studies, she was a research fellow of the Minas Gerais Research Funding Foundation (FAPEMIG) and a fellow of the Coordination for the Improvement of Higher Education Personnel (CAPES) in the Science without Borders Program. Through this program, she spent three academic semesters at University College Dublin, Ireland, and obtained her Bachelor's degree in Animal Science in December 2015. She took her Master's degree at the Beef Cattle Research Center, Institute of Animal Science (IZ), in March 2018, under the supervision of Dr. Maria Eugenia Zerlotti Mercadante. She was a CAPES Master's fellow and obtained her Master's degree in Sustainable Animal Production in February 2020 with a dissertation titled "Genetic evaluation of feeding behavior traits and their correlation with growth and feed efficiency." In March 2020, she started her Ph.D. studies at São Paulo State University, Faculty of Agricultural and Veterinary Sciences, Jaboticabal-SP, under the supervision of Dr. Maria Eugenia Zerlotti Mercadante and co-supervision of Dr. Luiz Fernando Brito, with a CAPES scholarship. In July 2022, she started her Ph.D. research internship at Purdue University, West Lafayette, IN, United States, under the supervision of Dr. Luiz Fernando Brito.

Epigraph

"… A vida de quem inventa de voar é paradoxal, todo dia. É o peito eternamente divido. É chorar porque queria estar lá, sem deixar de querer estar aqui. É ver o céu e o inferno na partida, o pesadelo e o sonho na permanência. É se orgulhar da escolha que te ofereceu mil tesouros e se odiar pela mesma escolha que te subtraiu outras mil pedras preciosas.

E começamos a viver um roteiro clássico: deitar na cama, pensar no antigo-eterno lar, nos quilômetros de distância, pensar nas pessoas amadas, no que eles estão fazendo sem você, nos risos que você não riu, nos perrengues que você não estava lá para ajudar. É tentar, sem sucesso, conter um chorinho de canto e suspirar sabendo que é o único responsável pela própria escolha. No dia seguinte, ao acordar, já está tudo bem, a vida escolhida volta a fazer sentido. Mas você sabe que outras noites dessa virão.

Mas será que a gente aprende? A ficar doente sem colo, a sentir o cheiro da comida com os olhos, a transformar apartamentos vazios na nossa casa, transformar colegas em amigos, dores em resistência, saudades cortantes em faltas corriqueiras?

Será que a gente aprende? A ser filho de longe, a amar via Skype, a ver crianças crescerem por vídeos, a fingir que a mesa do bar pode ser substituída pelo grupo do whatsapp, a ser amigo através de caracteres e não de abraços, a rir alto com HAHAHAHA, a engolir o choro e tocar em frente?

Será que a vida será sempre esta sina, em qualquer dos lados em que a gente esteja? Será que estaremos aqui nos perguntando se deveríamos estar lá e vice versa? Será teste, será opção, será coragem ou será carma?

Será que um dia saberemos, afinal, se estamos no lugar certo? Será que há, enfim, algum lugar certo para viver essa vida que é um turbilhão de incertezas que a gente insiste em fingir que acredita controlar?..."

Ruth Manus

ACKNOWLEDGEMENTS

The philosophy that each individual who crosses our path has a special purpose has been a belief I carry with me. I believe that each encounter brings with it a unique exchange that contributes in various ways to our journey of transformation, and as each chapter unfolded, this belief solidified even further. It is with deep gratitude that I begin this acknowledgments section, expressing my sincere appreciation for all those I have had the privilege to share the last four years of this Ph.D. journey with. Each individual has contributed in a unique way to propel me toward this culmination point.

First, I would like to thank you my advisor, Dra. Maria Eugênia Mercadante, who has been with me since the beginning of my journey in graduate school. I am immensely grateful for all the support, knowledge, and assistance provided over the last six years. Your insights, critiques, and encouragements have been invaluable in refining my research and shaping my perspectives. I deeply appreciate your patience, trust, wisdom, and the countless hours you dedicated to nurturing my academic pursuits. Thank you for sparing no effort to assist me, for encouraging my dreams, and for celebrating my achievements with me. This achievement would not have been possible without your guidance, and I am truly privileged to have had you as my advisor.

Dr. Luiz Brito, what a stroke of luck and privilege it has been to have had the opportunity to be a part of your team and to call you as my co-advisor. Your boundless generosity, humility, and empathy are truly admirable. You are a fantastic professional and human, and you have become an example for me. I would like to express my deep gratitude for all the opportunities and knowledge you have shared with me. Furthermore, I thank you for making me feel welcome in your research group and in your home. Your support and guidance have been invaluable, and I am profoundly grateful for all the support and guidance. Thank you very much!

I cannot forget the “IZ gang”, all of you who have been with me for so long, and together we have formed a family. You cheer and celebrate every new achievement with me and also support me and stand by my side in all the difficult moments. Luana Gomes, Marcelo Sant’Ana, Ricardo do Bem, Marina Oliveira, Valesca Andrade, Ana

Carolina Paz, Karine Costa, David Pereira, Danielly Broleze, Catiucia Miranda, thank you for making graduate school fun, light, and lively. According to Miranda (2020), everyone should have the privilege and honor of knowing you. She couldn't be righter!! You are incredible! and in a mix of customs, personalities, habits, and thoughts, we understand each other, we complement each other. Now, even though each of us is on a different path, and following a different direction, when we meet, we are still us, with the same harmony and the same friendship!! Thank you for everything!

And what can I say about Brito's Lab?! You are fantastic guys!! You are an incredible group! Katiene Sousa, Leonardo Gloria, Artur Rocha, Leticia Oliveira, Victor Pedrosa, Mariah Castro, Henrique Mulim, Gabriel Campos, Taiana Cortez, Tatiana Cortez, Felipe Carvalho, Ali Haider Saleem, Anel Retief, Andre Araujo, Jason Graham, Jacob Maskal, Sharlenne Hartman, Hui Wen, Rick Hernandez, Caitlin Watts, Dr. Alan Schinkel, Dr. Hinayah Rojas, thank you for welcoming me so well into your group, for helping me whenever I needed, for embracing and adopting me as part of the family. Everyone should be lucky enough to have a work team as fantastic as you!!! I couldn't leave out thanking Rita and Renata, what a privilege and pleasant surprise it was to meet you, who soon became very special people. Purdue also brought Marion as a gift, who spared no effort to help me, and became a great friend. Thank you for all the "help" in times of despair, for the fun trips and conversations!

Thank you, Karla Alves, for the partnership that has been with me since our undergraduate days. I consider myself lucky to have your friendship. I appreciate the laughter, the meetings, and the reunions over the years.

In a period marked by COVID, quarantine, and online classes, I had the pleasure of crossing paths with Tadia. Even though we met in person only once, she accompanied me throughout this Ph.D. journey. We shared frustrations, anxieties, fears, and knowledge, and we cheered each other on. Thank you for assisting me even from a distance during this time.

And this incredible journey would not have been possible without having a solid foundation and a place to always turn to: my family! Without you, I wouldn't be capable of achieving anything. You are always there for me and the reason I keep moving forward and striving to be better. Thank you, Ronan (The best brother), for being my role model. Even from afar, you have always supported me, rooted for me, and

provided me with guidance and support. Thank you for the scolding's, the laughter, the shared moments, and the endless help. Thank you, Mom (Vilma) for giving me unwavering support and unconditional encouragement in all my choices. You allowed me to spread my wings, but you always made it clear that I could come back home. You are an example of strength, resilience, and determination.

I also want to thank Dr. Victor Pedrosa, Dr. Gabriel Campos, Dra. Hinayah Rojas, and Dr. Fernando Baldi for their availability to be part of my qualification committee. Dr. Leonardo Gloria, Dra. Alessandra Silva, Dr. Lucio Mota, and Dra. Lenira El Faro for their participation in my defense committee. Thank you for the suggestions and corrections that were essential for the improvement of the work.

To my "English tutor", Anthony, thank you for the partnership and support during the challenging moments over this past year. You are amazing!

To all the researchers and employees of the Institute of Animal Science who have always welcomed and assisted me. Dra. Joslaine, Dra. Lenira, Dra. Roberta, Dra. Sarah, Aline, Dona Irma, Estela, Leu, Sr. Braz, Batata, Marcos, Cará, Dona Clesia, Dona Tete, and countless others. Your support during both my master's and Ph.D. studies has been invaluable for the development and execution of my research projects. Thank you for being part of my academic experience.

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

To National Council for Scientific and Technological Development (CNPq, grant #434681/2018-0) and São Paulo Research Foundation (FAPESP, grant #2017/10630-2). To Unesp and Purdue University, which allowed me to go further and deepen my studies.

SUMMARY

CHAPTER 1 – General considerations 13
1 Introduction 13
2 Literature review 15
2.1 Growth, feed efficiency, and feeding behavior traits 15
2.2 Nellore Selection Program of the Institute of Animal Science 16
2.3 Copy number variation 18
REFERENCES 20
CHAPTER 2 - Genome-wide association study between copy number variation and feeding behavior, feed efficiency, and growth traits in Nellore cattle[&] 26
ABSTRACT 26
BACKGROUND 27
RESULTS 30
Copy number variation and CNVR detection 30
Gene annotation, enrichment analyses, and QTL identification 32
DISCUSSION 33
MATERIAL AND METHODS 41
Animals and phenotypic datasets 42
Copy number variation identification 44
Copy number variation regions identification 45
Association analyses 45
Gene annotation and functional analyses 46
DECLARATIONS 47
REFERENCES 48
CHAPTER 3 - Detection and characterization of copy number variation in three differentially-selected Nellore cattle populations 54
1. INTRODUCTION 55
2. MATERIAL AND METHODS 56
2.1 Animals and experimental breeding program design 56
2.2 Genomic datasets 57
2.3 Copy number variation identification 58
2.4 Copy number variation regions identification 59
2.5 Gene annotation and functional analyses 59
3 RESULTS 59
3.1 Copy number variation and CNVR detection for the Nellore population 60

3.2 Copy number variation and CNVR detection by selection line 61
3.3 Common and exclusive CNVRs in the Nellore lines and gene annotation 63
3.4 Gene ontology 66
4 DISCUSSION 67
4.1 Copy number variation and CNVR detection for the Nellore population 68
4.2 Copy number variation and CNVR detection by line 70
4.3 Gene annotation and gene ontology 71
5 CONCLUSIONS 74
REFERENCES 75
ADDITIONAL FILE 1 81
ADDITIONAL FILE 2 82
ADDITIONAL FILE 3 86
ADDITIONAL FILE 4 87
ADDITIONAL FILE 5 89
ADDITIONAL FILE 6 90

ESTUDO DE ASSOCIAÇÃO GENÔMICA ENTRE VARIAÇÃO DO NÚMERO DE CÓPIAS E CARACTERÍSTICAS DE CRESCIMENTO, EFICIÊNCIA ALIMENTAR E COMPORTAMENTO INGESTIVO EM BOVINOS NELORE

RESUMO - A variação no número de cópias (CNV) no genoma animal vem sendo cada vez mais estudada, uma vez que essas podem contribuir para a variação fenotípica entre os indivíduos. Os objetivos deste estudo foram: 1) Analisar a distribuição de CNVs e regiões de variação no número de cópias (CNVR) nos genomas de animais Nelore provenientes de três linhas de seleção para características de crescimento e eficiência alimentar; 2) Avaliar a associação de CNVs e CNVRs com características de crescimento (peso à desmama (P210) e peso pós-desmama (Psel)), eficiência alimentar (consumo de matéria seca (CMS), consumo alimentar residual (CAR) e ganho médio diário (GMD)) e comportamento alimentar (tempo de permanência no cocho (TP) e frequência de visitas ao cocho(FV)); 3) Realizar a anotação funcional das CNVRs significativas para cada característica; 4) Examinar a distribuição de CNVs e CNVRs entre as três linhas de seleção; 5) Realizar a anotação funcional das CNVRs identificadas nas linhas seleção. O estudo utilizou registros de animais da raça Nelore nascidos entre 1978 e 2020, pertencentes a três linhas de seleção para peso pós-desmama e CAR. Os fenotipos das sete características foram ajustados para os efeitos fixos e utilizados nas análises de associação. Um total de 2.266 animais foram genotipados, sendo 770 com o chip Illumina BovineHD BeadChip 770k, 1.338 com o GeneSeek Genomic Profiler HDi 75K e 158 com o GeneSeek Genomic Profiler Indicus 50k. A identificação de CNVs foi realizada usando o software PennCNV. As CNVRs foram determinadas agrupando as CNVs identificadas que se sobrepuseram por pelo menos 1 par de bases. A análise de associação foi realizada testando cada CNVR de maneira individual. As CNVRs significativamente associadas com as características fenotípicas foram utilizadas para a anotação de genes e QTLs, além disso, foi realizada a análise de *gene ontology* (GO). Posteriormente, os CNV's foram separados de acordo com cada linha de seleção e identificada as CNVRs exclusivas presentes em pelo menos 10% dos animais de cada linha. Um total de 3.161 CNVs e 561 CNVRs foram identifcadas na população. As CNVRs abrangeram 3,99% do genoma autossômico dos animais analisados. Dezessete CNVRs foram significativamente associadas a características estudadas. A linha NeT apresentou o maior numero de CNVRs (1,493), seguida pela NeS (823) e NeC (482). Seis, duas e quatro CNVRs foram identificadas exclusivamente nas linhas NeC, NeS e NeT, respectivamente. A anotação gênica e GO das CNVRs exclusivas de cada linha revelou genes e processos biológicos específicos que podem estar envovidos na expressão de características de crescimento e eficiência alimentar. Em conclusão, foram identificadas e caracterizadas mais de 3.100 CNVs e 550 CNVRs na população Nelore estudada, algumas CNVRs foram significativamente associadas a características de CMS e FV. Além disso, o estudo revelou variabilidade dos CNVs e CNVRs dentro das três linhas de seleção.

Palavras-chave: *Bos indicus*, genoma bovino, linhas de seleção, paineis de SNP, variação estrutural

GENOME-WIDE ASSOCIATION STUDY BETWEEN COPY NUMBER VARIATION AND GROWTH, FEED EFFICIENCY, AND FEEDING BEHAVIOR TRAITS IN NELLORE CATTLE

Abstract - The investigation of copy number variation (CNVs) in animal genomes has gained increasing attention due to their potential contribution to phenotypic variation among individuals. This study aimed to: 1) Analyze the distribution of CNVs and CNVRs (Copy Number Variation Regions) within the genomes of Nellore animals from three selection herds for growth and feed efficiency traits; 2) Assess the association of CNVs and CNVRs with growth traits (weaning weight (W210) and body weight measured at the time of selection (WSel)), feed efficiency traits (dry matter intake (DMI), residual feed intake (RFI), and average daily gain (ADG)), and feeding behavior traits (time spent at the feed bunk (TF) and frequency of visits to the feed bunk (FF)); 3) Perform functional annotation of significant CNVRs for each trait; 4) Examine the distribution of CNVs and CNVRs across the three selection herds; 5) Perform functional annotation of the identified CNVRs in the Nellore herds. The study utilized records from Nellore breed animals born between 1978 and 2020, belonging to three selection herds for post-weaning weight and residual feed intake. The phenotypes of the seven traits were adjusted for fixed effects and used in the association analyses. A total of 2,266 animals were genotyped, with 770 using the Illumina BovineHD BeadChip 770k, 1,338 using the GeneSeek Genomic Profiler HDi 75K, and 158 using the GeneSeek Genomic Profiler Indicus 50k. The identification of CNVs was performed using the PennCNV software. CNVRs were determined by grouping overlapping CNVs identified by at least 1 base pair. The association analysis was performed by testing each CNVR individually. The CNVRs significantly associated with the phenotypes were used for the gene and QTL annotation. Additionally, a Gene Ontology (GO) analysis was performed. Subsequently, the CNVs were separated according to each selection line, and exclusive CNVRs present in at least 10% of the animals in each line were identified. A total of 3,161 CNVs and 561 CNVRs were identified in the studied population, and the CNVRs covered up to 3.99% of the Nellore autosomal genome. Seventeen CNVR were significantly associated with growth, feed efficiency, and feeding behavior traits. The NeT line exhibited the highest number of CNVRs (1,493), followed by the NeS (823) and NeC lines (482). Gene annotation and GO of the exclusive CNVRs to each line revealed specific genes and biological processes involved in the expression of growth and feed efficiency traits. In conclusion, over 3,100 CNVs and 550 CNVRs were identified and characterized in the studied Nellore population. Some CNVRs were significantly associated with DMI and FF. Furthermore, the study unveiled variability in CNVs and CNVRs within the three selection lines.

Keywords: *Bos indicus*, bovine genome, SNP panel, selection herds, structural variants

CHAPTER 1 – General considerations

1 Introduction

Livestock farming is one of the main economic pillars of Brazil. The country stands out as the largest beef exporter and boasts the world's largest commercial cattle herd, consisting of approximately 80% Nellore (Bos taurus indicus) or Nellore composite breed animals (EMBRAPA, 2021). In 2019, the Brazilian bovine population totaled 213.5 million heads (IBGE, 2019), representing one-fifth of the 982.1 million cattle worldwide (USDA, 2020). The global market for beef is becoming increasingly demanding, putting pressure on producers to adopt more sustainable, efficient, and profitable systems. Due to this growing pressure on the sector, nutritional and genetic strategies have been implemented to ensure the competitiveness of the Brazilian production system in the global market.

Considering that animal feed represents the highest cost in meat production and is one of the main determinants of profitability for the producer (Greewood, 2021), increasing the feed efficiency of animals has been regarded as a way to reduce production costs. According to Basarab et al. (2003), the ability to identify animals that consume less feed without compromising performance and carcass quality could substantially enhance profitability and reduce the environmental impacts of production systems. Additionally, studies have shown that feeding behavior traits are correlated with feed efficiency and can serve as indicators of economically important traits (Nkrumah et al., 2007; Kelly et al., 2010; Benfica et al., 2020).

The utilization of genomic information as a strategy for improving phenotypes of interest, such as feed efficiency (Hayes et al., 2007), and the development of new technologies to understand the genetic architecture of traits, genetic variations, and heritable traits in living organisms have been studied and developed, generating new insights. The sequencing of the bovine genome has led to the discovery of thousands of genetic markers known as SNPs (Single Nucleotide Polymorphisms), along with the development of SNP subsets that can characterize the bovine genome with greater coverage and at a lower cost (Matukumalli et al., 2009). In addition to providing information on individual SNPs, SNP panel data can also be used for identifying a form

of genomic structural variation known as copy number variation (CNV, Matukumalli, et al., 2009).

Copy number variations (CNVs) are structural variations in an individual's genome, in the form of gain or loss of DNA fragments ranging from >1 Kb to several Mb (Henrichsen et al., 2009). The CNVs are considered as polymorphic genetic markers and can be inherited across generations (McCarroll and Altshuler, 2007), representing a significant source of genetic and phenotypic variation (Zhang et al., 2009). Compared to SNPs, CNVs cover broader chromosomal regions and can potentially lead to changes in gene structure, modifications in regulation, alterations in gene dosage, and unmasking of recessive alleles, and influence on gene expression resulting in substantial phenotypic effects (Stafuzza et al., 2019).

Studies involving CNVs have been conducted in many species with the aim of identifying genetic variations in copy number related to the expression of important traits. However, the understanding of how CNVs contribute to the phenotypic variation of growth, feed efficiency, and feeding behavior traits in cattle is still limited, particularly in indicine cattle such as Nellore. More research is needed to unravel the role of CNVs in shaping these economically important traits in bovines, especially in indicine breeds that are prevalent in certain regions and have unique adaptive traits. By gaining a better understanding of CNVs' influence on these traits, we can potentially enhance breeding strategies and management practices to improve productivity and sustainability in cattle production systems.

Therefore, the main objectives with this study were to: 1) identify CNV and CNV regions (CNVR) in the genome of a Nellore cattle population; 2) assess the association between the identified CNVR and growth (weaning weight (W210), body weight measured at the time of selection (WSel), average daily gain (ADG), and dry matter intake (DMI)), feed efficiency (residual feed intake (RFI)), and feeding behavior (time spent at feed bunk (TF) and frequency of visits to the feed bunk (FF)) traits; 3) perform functional annotation of the associated CNV regions; 4) identify and characterize CNV regions (CNVR) in Nellore cattle from three selection herds for growth and feed efficiency.

2 Literature review

2.1 Growth, feed efficiency, and feeding behavior traits

Selection is defined as the process of identifying superior individuals within a population for the production of the next generation, and it is one of the primary tools used by breeders for the improvement of their herds (Aaron et al., 1986). The key traits utilized in selection programs and genetic evaluation in beef cattle in Brazil are those associated with animal growth, such as birth weight, weaning weight, and postweaning weight gain. These traits are chosen due to their ease of measurement and their significance in assessing herd efficiency (Kamei et al., 2017).

Weaning weight is usually obtained at around 7 or 8 months of animal age and plays a crucial role in assessing the potential for future growth and weight gain. The yearling weight represents an individual growth potential and evaluates the potential for weight gain in the pre-weaning and post-weaning periods, enabling the selection of more precocious animals (Lôbo et al., 2000). The traits related to weight are easily measurable and can be performed by the producer on the farm and are directly linked to weight gain in finishing and the final weight of the animal. Moreover, it is highly accepted by producers since slaughterhouses plants pay for the same selected unit, i.e., per kilogram of product.

Another important trait that has been adopted in selection programs is feed efficiency. There are several measures to evaluate the feed efficiency of animals, each reflecting different biological and mathematical aspects of efficiency (Archer et al., 1999), with residual feed intake being one of the most commonly used measures nowadays.

Residual feed intake (RFI) is a concept introduced by Koch et al. (1963). This trait is independent of growth rate and body weight and is defined as the difference between the actual feed intake and the predicted feed intake, calculated through a linear regression equation of individual actual feed intake on the average metabolic body weight and average daily gain. The residual from this equation represents the RFI (Koch et al., 1963). Animals with more negative RFI values are considered more efficient (their observed feed intake is lower than predicted for their achieved weight gain and maintenance), while those with more positive RFI values are less efficient

(their observed feed intake is higher than predicted for weight gain and maintenance). Selection for lower RFI aims to choose animals with lower feed intake and reduced maintenance requirements, without altering the level of animal production (Koch et al., 1963; Carstens et al., 2002; Basarab et al., 2003).

Cattle that differ in terms of RFI also exhibit distinct feeding behaviors (Richardson and Herd, 2004; Nkrumah et al., 2006; Benfica et al., 2020). Records of feeding behavior are based on feeding events or visits to the feed trough, separated by intervals of varying durations (Mendes et al., 2011). In general, there is a range of feeding behavior traits that are typically investigated, including frequency and duration of trough visits, frequency and duration of feeding, average meal duration and size, and feeding rate (Green et al., 2013). Feeding behavior traits stand out as they are potentially associated with the energy costs of feeding (Montanholi et al., 2010), since the time spent at the feed trough and the feeding rate are key factors in determining the energetic cost of feed intake in cattle, and these traits are correlated with dry matter intake (Adam et al., 1984, Benfica et al., 2020). Furthermore, an individual animal's fedding behavior is typically consistent and repeatable, and it can be used to predict differences in performance and feed efficiency (Gibb et al., 1998).

2.2 Nellore Selection Program of the Institute of Animal Science

The Beef Cattle Research Center - Institute of Animal Science (IZ) is a research center is located in the São Paulo State and is primarily devoted to the genetic improvement of Nellore and Caracu breeds (Mercadante and Razook, 2010).

The formation of the Nellore herd at IZ began in 1937, and initially, the selection of sires followed the established criteria of the time, primarily based on racial characterization and conformation. In 1976, a group of researchers (including Prof. Irineu Umberto Packer, Dr. José Benedito de Freitas Trovo, Dr. Willian T. Magee, and Dr. Celso Boin) came together to outline the selection program for the IZ Nellore herd, which was intended to be similar to those that already existed in European beef breeds in temperate climate countries. Thus, the IZ selection project was implemented with a commitment to achieve practical results on the effects of selection within herds on the

key economic components of beef cattle, such as growth, reproduction, and carcass traits (Mercadante and Razook, 2010).

The establishment of the IZ selection program began with the reduction of the breeding season, redefinition of the methodology for weight gain testing, and the selection of founding sires used in 1980. At the same time, genetic material from the main representative families existing in Brazil at that time (Cantor, Everest, Karvardi, Kurupathi, and Nagpur) was introduced into the Nellore herd to expand the genetic base and ensure variability (Cyrillo et al., 2000; Mercadante and Razook, 2010).

In 1980, at the beginning of the program, out of a total of 350 breeding-eligible females, 180 of the youngest ones were randomly divided into two herds: Nellore Control Herd (NeC, 60 animals) and Nellore Selection Herd (NeS, 120 animals). The remaining 170 females were used to form a third herd, named Nellore Traditional Herd (NeT) (Cyrillo et al., 2000, Mercadante et al., 2003).

In the NeC herd, both males and females with selection differentials close to zero for P378 (males) and P550 (females) within herd and contemporary group were selected, following a stabilizing selection scheme. In the Selection (NeS) and Traditional (NeT) herds, animals with higher selection differentials for P378 (males) and P550 (females) were selected (Mercadante et al., 2003). In the year 2008, the RFI was introduced as selection criteria in the breeding program of the NeT herd. Since then, individuals with higher estimated breeding values for the P378 and lower estimated breeding values for RFI have been prioritized as breeding candidates (Mercadante et al., 2003; Cardoso et al., 2018; Benfica et al., 2020).

The Institute of Animal Science has established itself as a national reference in the field of beef cattle breeding. Since the establishment of the three selection herds, many research projects have been developed covering a wide range of topics, including those related to carcass traits (Pires et al., 2022; Paz et al., 2022), reproduction (Borges et al., 2023; Silva et al., 2022; Fernandes et al., 2022) parameters estimation for production traits (Valente et al., 2023; Benfica et al., 2020) and more recently, methane emissions (Sakamoto et al., 2021; Magnani et al., 2023). These studies involve the estimation of genetic parameters and genetic associations among traits, as well as investigations into enteric methane emissions and their correlation with performance and feed efficiency traits.

2.3 Copy number variation

Copy number variation are structural variations in an individual's genome in the form of losses or gains of DNA fragments range from 1 kb to several mega-bases in comparison to the reference genome of the species (Henrichsen et al., 2009). The CNVs are considered polymorphic genetic markers and are inherited across generations (McCarroll and Altshuler, 2007), representing an important source of genetic and phenotypic variation (Zhang et al., 2009). Compared to SNPs, CNVs cover larger chromosomal regions and can potentially be responsible for changes in gene structure, regulatory modifications, alterations in gene dosage, and exposure of recessive alleles, leading to significant influence in gene expression and phenotypic effects (Zhang et al., 2009, Stafuzza et al., 2019).

Studies have reported that CNVs capture between 18% to 30% of the genetic variation in humans and rats (Henrichsen et al., 2009). As a result, CNV microarrays enable the exploration of the genome for sources of variability that can explain the robust genetic component of many economically relevant traits, going beyond the explanatory power of SNPs.

The detection of CNVs can be performed using different molecular techniques, such as array comparative genome hybridization, single nucleotide polymorphism (SNP) panels, and next-generation sequencing (Bickhart et al., 2012; Liu et al., 2013). One of the most used techniques has been SNP panels, mainly because of the SNP panels data already generated for genomic selection purposes. The SNP panels have the advantage of providing normalized total intensities (Log R ratio - LRR) and allelic intensity ratios (B allele frequency - BAF) which represent overall copy numbers and allelic contrasts (Peiffer et al., 2006).

Studies involving CNVs have been conducted in many species with the main aim of identifying genetic variations in copy numbers associated with important traits. In humans, studies have identified associations between CNVs and several diseases, including Crohn's disease, psoriasis, schizophrenia, and autism (Fellermann et al. 2006; Sebat et al. 2007; Walsh et al. 2008; Moreno-De-Luca et al., 2010). In domestic animals, CNVs are also associated with the expression of different traits. In birds, studies have reported associations of CNVs with late feathering, dark brown feather

color, and dermal hyperpigmentation (Elferink et al., 2008). Stafuzza et al. (2019) conducted CNV analysis in pigs and reported that out of 2865 samples, 46118 CNVs and 425 CNV regions were identified and associated with reproductive traits. In sheep, Yang et al. (2018) identified 619 CNV regions associated with fetal muscle development and prostaglandin synthesis. Berton et al. (2017) identified several regions related to the immune system, inflammatory response, lymphocyte regulation, and leukocyte proliferation in Santa Inês breed sheep. Moreover, duplications overlapping the KIT (proto-oncogene tyrosine kinase receptor) and ASIP (agouti signaling protein) genes are considered responsible for white skin in pigs and sheep, respectively (Johansson et al., 1996; Norris et al., 2008).

In cattle, CNVs associated with production traits, body measurements, and parasite resistance have been identified. Xu et al. (2014) characterized and reported 34 CNVs significantly associated with milk production traits in Holstein cattle and found a deletion polymorphism associated with resistance to gastrointestinal nematodes. Butty et al. (2020) working with dairy cattle identified 23,256 CNVs that were associated with hoof health traits. Zhou et al. (2016), in a genome-wide association study (GWAS) for CNVs and body traits in Nellore cattle, identified 231 CNVs, with 17 of them significantly associated with growth traits. Lemos et al. (2018) studied animals of the same breed and identified 43 CNVRs associated with fatty acid profiles, indicating some genes that could influence fatty acid composition and metabolism. Berton (2017), while also studying Nellore cattle, identified genomic regions associated with fat deposition (FABP7) and muscle contraction (SLC34A3, TRDN), extracellular matrix (INS and COL10A1), as well as creatine transporter genes (SLC6A18 and SLC6A19), which could influence beef quality.

Thus, studies to identify and associate CNVs with production traits can improve the understanding of the genetic mechanisms that contribute to the expression of important productive traits.

REFERENCES

AARON, D.K.; FRAHM, R.R.; BUCHANAN, D.S. Direct and correlated responses to selection for increased weaning or yearling weight in Angus Cattle. I. Measurement of selection applied. Journal of Animal Science, v.62, p.54-65, 1986.

Adam, I., Young, B.A., Nicol, A.M. & Degen, A.A. (1984). Energy cost of eating in cattle given diets of different form. Animal Production 38, 53–56.

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., et al. (2018). Copy number variation regions in Nellore cattle: Evidences of environment adaptation. Livestock Science Elsevier B.V. 207, 51–58.

ARCHER, J.A.; RICHARDSON, J.A.; HERD, R.M.; ARTHUR, P.F. Potential for selection to improve efficiency of feed use in beef cattle: a review. Australian Journal of Agricultural Research, v.50, p.147-161, 1999.

Basarab, J.A., Price, M.A., Aalhus, J.L., Okine, E.K., Snelling, W.M. & Lyle, K.L. Residual feed intake and body composition in young growing cattle.

Benfica, L.F., Sakamoto, L.S., Magalhães, A.F.B., De Oliveira, M.H.V., De Albuquerque, L.G., Cavalheiro, R., Branco, R.H., Cyrillo, J.N.D.S.G. & Mercadante, M.E.Z. (2020). Genetic association among feeding behavior, feed efficiency, and growth traits in growing indicine cattle. Journal of Animal Science, 98.

Berton, M.P., de Lemos, M.V.A., Stafuzza, N.B., Simielli Fonseca, L.F., Silva, D.B. dos S., Peripolli, E., Pereira, A.S.C., Magalhães, A.F.B., et al. (2022). Integration analyses of structural variations and differential gene expression associated with beef fatty acid profile in Nellore cattle. Animal Genetics John Wiley and Sons Inc 53, 570–582.

Berton, M.P., Oliveira Silva, R.M., Peripolli, E., Stafuzza, N.B., Martin, J.F., Álvarez, M.S., Gavinã, B.V., Toro, M.A., et al. (2017). Genomic regions and pathways associated with gastrointestinal parasites resistance in Santa Inês breed adapted to tropical climate. Journal of Animal Science and Biotechnology BioMed Central Ltd. 8,.

Bickhart, D.M., Hou, Y., Schroeder, S.G., Alkan, C., Cardone, M.F., Matukumalli, L.K., Song, J., Schnabel, R.D., et al. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. Genome Research 22, 778–790.

Butty, A.M., Chud, T.C.S., Cardoso, D.F., Lopes, L.S.F., Miglior, F., Schenkel, F.S., Cánovas, A., Häfliger, I.M., et al. (2021). Genome-wide association study between

copy number variants and hoof health traits in Holstein dairy cattle. Journal of Dairy Science 104, 8050–8061.

Borges, M.S., Silva, M.O., Fernandes, L.G., Rodrigues, N.N., Rossi, G.F., Dellaqua, C.P.F., Bonilha, S.F.M.; Mercadante, M.E.Z., Monteiro, F.M. (2023). Relationship of residual feed intake with semen parameters and testicular ultrasound of Nellore bulls. Revista brasileira de zootecnia 52.

Cardoso, D.F., Albuquerque, L.G., Reimer, C., Qanbari, S., Erbe, M., Nascimento, A.V., Venturini, G.C., Scalez, D.C.B., Baldi, F., Camargo, G.M.F., Mercadante, M.E.Z., Cyrillo, J.N.S.G., Simianer, H., Tonhati. (2018). Genome-wide scan reveals population stratifcation and footprints of recent selection in Nelore cattle. Genetics Selection Evolution 50, 22.

Carstens, G.E., Theis, C.M., Kurz, M.W., Hollenbeck, R.J., White, M.B., Welsh, T.H., Warrington, B.G., Randel, R.O., et al. Relationship between net feed intake, performance traits and ultrasound measures of composition in beef steers.

Elferink, M.G., Vallée, A.A.A., Jungerius, A.P., Crooijmans, R.P.M.A. & Groenen, M.A.M. (2008). Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. BMC Genomics 9, 1–9.

Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). Brasil é o quarto maior produtor de grãos e o maior exportador de carne bovina do mundo, diz estudo. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/62619259/brasil-e-o-quarto-maior-produtor-de-graos-e-o-maior-exportador-de-carne-bovina-do-mundo-diz-estudo>. Acessado em: 07 Aug 2022.

Fellermann, K., Stange, D.E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C.L., Reinisch, W., Teml, A., et al. (2006). A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. American Journal of Human Genetics 79, 439–448.

Fernandes, L.G., Borges, M.S., Silva, M.O., Rodrigues, N.N., Vicentini, R.R., Crespilho, A.M., Mercadante, M.E.Z., Monteiro, F.M. (2022). Use of auxiliary techniques for assessing the reproductive capacity of natural service Nellore bulls. Animal reproduction science 247, 107090.

Green, T.C., Jago, J.G., Macdonald, K.A. & Waghorn, G.C. (2013). Relationships between residual feed intake, average daily gain, and feeding behavior in growing dairy heifers. Journal of Dairy Science 96, 3098–3107.

Greenwood, P.L. (2021). Review: An overview of beef production from pasture and feedlot globally, as demand for beef and the need for sustainable practices increase. Animal 15.

Hayes, B.J., Chamberlain, A.J., McPartlan, H., MacLeod, I., Sethuraman, L. & Goddard, M.E. (2007). Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. Genetical Research 89, 215–220.

Henrichsen, C.N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., Ruedi, M., Kaessmann, H. & Reymond, A. (2009). Segmental copy number variation shapes tissue transcriptomes. Nature Genetics 41, 424–429.

IBGE. Instituto Brasileiro de Geografia e Estatística (2020). Disponível em: . Acessado em: 07 Aug 2022.

KOCH, R.M.; SWIGER, L.A.; CHAMBERS, D. Efficiency of feed use in beef cattle. Journal of Animal Science, v.22, p.486-494, 1963.

Liu, J., Zhang, L., Xu, L., Ren, H., Lu, J., Zhang, X., Zhang, S., Zhou, X., et al. (2013). Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BMC Genomics 14,.

Lobo, R.N.B., Madalena, F.E., Vieira, A.R. (2000). Average estimates of genetics parameters for beef and dairy cattle in tropical regions. Animal Breeding Abstracts, 68.

Magnani, E., Silva, T.H., Sakamoto, L.S., Manella, M.Q., Fabio, M.G.N., Mercadante, M.E.Z., Darren, H., Marcatto, J.O.S., Paula, E.M., Branco, R.H. (2023). Phenotypic association among performance, feed efficiency and methane emission traits in Nellore cattle. PLoS One 7.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., et al. (2009). Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS ONE 4,.
McCarroll, S.A. & Altshuler, D.M. (2007). Copy-number variation and association studies of human disease. Nature Genetics 39, S37–S42.

Mendes, E.D.M., Carstens, G.E., Tedeschi, L.O., Pinchak, W.E. & Friend, T.H. (2011). Validation of a system for monitoring feeding behavior in beef cattle. Journal of Animal Science 89, 2904–2910.

Mercadante, M.E.Z., Packer, I.U., Razook, A.G., Cyrillo, J.N.S.G. & Figueiredo, L.A. (2003). Direct and correlated responses to selection for yearling weight on reproductive performance of Nelore cows 1. J. Anim. Sci. https://academic.oup.com/jas/article/81/2/376/4789734.

Mercadante, M.E.Z., Razook, A.G. (2010) Bovinocultura de Corte, p. 835-848.

Moiler, M.J., Chaudhary, R., Hellm~n, E., Hoyheim, B., Chowdhary, B. & Andersson, L. (1996). Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. Mammalian Genome.

Montanholi, Y.R., Swanson, K.C., Palme, R., Schenkel, F.S., McBride, B.W., Lu, D. & Miller, S.P. (2010). Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. Animal 4, 692–701.

Moreno-De-Luca, D., Mulle, J.G., Kaminsky, E.B., Sanders, S.J., Myers, S.M., Adam, M.P., Pakula, A.T., Eisenhauer, N.J., et al. (2010). Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. American Journal of Human Genetics 87, 618–630.

Nkrumah, J.D., Crews, D.H., Basarab, J.A., Price, M.A., Okine, E.K., Wang, Z., Li, C. & Moore, S.S. (2007). Genetic and phenotypic relationships of feeding behavior and temperament with performance, feed efficiency, ultrasound, and carcass merit of beef cattle. Journal of Animal Science 85, 2382–2390.

Noely, J., Gonçalves Cyrillo, S., Razook, A.G., Andrade De Figueiredo, L., Martins, L., Neto, B., Ruggieri, A.C. & Tonhati, H. Efeitos da Seleção para Peso Pós-desmame sobre Medidas Corporais e Perímetro Escrotal de Machos Nelore de Sertãozinho (SP) 1.

Norris, B.J. & Whan, V.A. (2008). A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Research 18, 1282–1293.

Peiffer, D.A., Le, J.M., Steemers, F.J., Chang, W., Jenniges, T., Garcia, F., Haden, K., Li, J., et al. (2006). High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Research 16, 1136–1148.

Paz, A.C.A.R., Ladeira, M.M., Teixeira, P.D., Canesin, R.C., Batalha, C.D.A., Mercadante, M.E.Z., Bonilha, S.F.M. (2022). Performance and muscle lipogenesis of calves born to Nellore cows with different residual feed intake classification. PLoS One 1, 1-18.

Pires, B.V., Reolon, H.G., Abduch, N.G., Souza, L.L., Sakamoto, L.S., Mercadante, M.E.Z., Silva, R.M.O., Fragomeni, B.O., Baldi, F., Paz, C.C.P., Stafuzza, N.B. (2022). Animals 12, 3196.

Richardson, E.C.; Herd, R.M. (2004). Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. Australian Journal of Experimental Agriculture, v.44, p.431-440.

Sakamoto, L.S., Souza, L.L., Gianvecchio, S.B., Oliveira, M.H.V., Silva, J.A.V., Canesin, R.C., Branco, R.H., Bacan, M., Albuquerque, L.G., Mercadante, M.E.Z. (2021). Phenotypic association among performance, feed efficiency and methane emission traits in Nellore cattle. PLoS One 16.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., et al. (2007). Strong Association of De Novo Copy Number Mutations with Autism Interdepartmental Program in the Neurosciences, Program in. Science 316, 445–449. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2993504/pdf/nihms239818.pdf.

Silva, M.O., Borges, M.S., Fernandes, L.G., Rodrigues, N.N., Watanabe, Y.F., Carlino, J.D., Oliveira, C.S., Feuchard, V.L.S., Cyrillo, J.N.S.G., Mercadante, M.E.Z., Monteiro, F.M. (2022). Effect of Nellore (Bos indicus) donor age on in vitro embryo production and pregnancy rate. Reproduction in domestic animals, 1-9.

Stafuzza, N.B., Silva, R.M.D.O., Fragomeni, B.D.O., Masuda, Y., Huang, Y., Gray, K. & Lourenco, D.A.L. (2019). A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive. BMC Genomics 20.
United States Department of Agriculture (USDA). Livestock and Poultry: World Markets and Trade. Disponível em: . Acessado em: 16 jun 2022.

Valente, J.P.S., Sousa, K.T., Tenffen, K., Deniz, M., Bonilha, S.F.M., Mercadante, M.E.Z., Dias, L.T. (2023). Heritability and genetic association of social organization traits with feeding behavior, feed efficiency and growth in Bos indicus. Applied animal behaviour science 265, 105976.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320, 539–543.

Xu, L., Cole, J.B., Bickhart, D.M., Hou, Y., Song, J., VanRaden, P.M., Sonstegard, T.S., Van Tassell, C.P. & Liu, G.E. (2014). Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics 15, 1–10.

Yang, L., Xu, L., Zhou, Y., Liu, M., Wang, L., Kijas, J.W., Zhang, H., Li, L. & Liu, G.E. (2018). Diversity of copy number variation in a worldwide population of sheep. Genomics 110, 143–148.

Zhang, F., Gu, W., Hurles, M.E. & Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. Annual Review of Genomics and Human Genetics.

Zhou, Y., Utsunomiya, Y.T., Xu, L., Hay, E.H. abdel., Bickhart, D.M., Alexandre, P.A., Rosen, B.D., Schroeder, S.G., et al. (2016). Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. BMC Genomics BMC Genomics 17, 1–9. http://dx.doi.org/10.1186/s12864-016-2461-4.

CHAPTER 2 - Genome-wide association study between copy number variation and feeding behavior, feed efficiency, and growth traits in Nellore cattle[&]

Lorena F. Benfica[1,2*], Luiz F. Brito[1], Ricardo D. do Bem[2], Henrique A. Mulim[1], Joseph Glessner[3,4], Larissa G. Braga[2], Leonardo S. Gloria[1], Joslaine N. S. G. Cyrillo[5], Sarah F.M. Bonilha[5], Maria E. Z. Mercadante[5*]

[1]Department of Animal Sciences, Purdue University, 270 S. Russell Street, West Lafayette, IN, 47907, USA

[2]Department of Animal Science, Faculty of Agricultural and Veterinary Sciences, Sao Paulo State University, Jaboticabal, SP, Brazil

[3]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[4]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[5]Institute of Animal Science, Sertaozinho, SP, Brazil

*Corresponding author: lorenafbenfica@gmail.com and mezmercadante@gmail.com

ABSTRACT

Background: Feeding costs represent the largest expenditures in beef production. Therefore, the animal efficiency in converting feed in high-quality protein for human consumption plays a major role in the environmental impact of the industry and in the beef producers' profitability. In this context, breeding animals for improved feed efficiency through genomic selection has been considered as a strategic action in modern breeding programs around the world. Copy number variation (CNV) is a less-studied source of genetic variation that can contribute to

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

phenotypic variability in complex traits. In this context, this study aimed to: 1) identify CNV and CNV region (CNVR) in the genome of Nellore cattle (*Bos taurus indicus*); 2) assess potential association between the identified CNVR and weaning weight (W210), body weight measured at the time of selection (WSel), average daily gain (ADG), dry matter intake (DMI), residual feed intake (RFI), time spent at the feed bunk (TF), and frequency of visits to the feed bunk (FF); and, 3) perform functional annotation analyses of the significant CNVR identified for each of the traits evaluated.

Results: A total of 3,161 CNVs and 561 CNVRs ranging from 4,973 bp to 3,215,394 bp were identified. The CNVRs covered up to 99,221,894 bp (3.99%) of the Nellore autosomal genome. Seventeen CNVR were significantly associated with dry matter intake and feeding frequency (number of daily visits to the feed bunk). The functional annotation of the associated CNVRs revealed important candidate genes related to metabolism that may be associated with the phenotypic expression of the evaluated traits. Furthermore, Gene Ontology (GO) analyses revealed 19 enrichment processes associated with FF.

Conclusions: A total of 3,161 CNVs and 561 CNVRs were identified and characterized in a Nellore cattle population. Various CNVRs were significantly associated with DMI and FF, indicating that CNVs play an important role in key biological pathways and in the phenotypic expression of feeding behavior and growth traits in Nellore cattle.

Key words: *Bos indicus*, bovine genome, SNP panel, structural variants

BACKGROUND

Brazil is one of the largest beef exporters in the world, with a cattle population composed of about 80% of Nellore (*Bos taurus indicus*) or Nellore composite breed animals

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

[1]. With a rapid increase in the world population and reduction in poverty, beef consumption is expected to increase from 60 to 130 million tons by 2050, and ~ 70% of this growth is projected to be provided by beef production systems from tropical and subtropical regions, including Brazil [2]. To meet the world's growing beef demand and reduce the environmental impact of the industry, especially in developing countries, there is an urgent need to develop more efficient breeding strategies for genetically improving tropically-adapted cattle raised in pasture-based production systems. Since feed represents the largest costs in beef production and is a major determinant of beef cattle producers' profitability [3], improving cattle feed efficiency has been considered as a strategic and major breeding goal in worldwide beef cattle breeding programs [4–7]. Additionally, feeding behavior traits are associated with feed efficiency and growth traits [8, 9], and could be used as auxiliary traits for further improving beef cattle feed efficiency.

The sequencing of the cattle genome has led to the discovery of thousands of single nucleotide polymorphism (SNP) markers [10], which are common variants of individual nucleotide sequences that are frequently observed in the population (> 1%). Following the sequence of the cattle genome, various SNP panels containing thousands of markers with great genome coverage were developed [10]. In addition to providing information on individual SNPs, SNP panel data can also be used for identifying a form of genomic structural variation known as copy number variation (CNV; [11]).

The CNVs are a less-studied source of genetic variation that can influence phenotypic variability in complex traits. They can be defined as structural variations in an individual's genome in the form of losses or gains of DNA fragments ranging from 1 kb to several mega-bases in comparison to the reference genome of the species [12–14]. Additionally, CNVs are polymorphic genetic markers that can be inherited across generations [15]. CNVs can be

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

detected using various platforms and sequencing tools, including array comparative genome hybridization (aCGH), single nucleotide polymorphism (SNP) panels, and next-generation sequencing (NGS) tools [16–18]. One of the most used data sources has been SNP panels as they are already generated for genomic selection purposes in commercial breeding programs [19].

Compared to individual SNPs, CNVs cover wider chromosomal regions [20], which contribute to changes in genome structure, alteration in regulation or gene dosage, exposure of recessive alleles, and alterations in gene expression, and consequently, phenotypic variability in complex traits [21–23]. CNVs can also have unique functional consequences not producible by SNPs. For instance, duplications can increase gene dosage while deletions can eliminate regulatory elements [24]. In addition, the lack of linkage disequilibrium between SNPs and 25% of the detected CNVs indicate that CNVs contain information not captured solely based on SNP information [25, 26]. Therefore, CNV is an additional source of information to explain the genetic variance of complex traits not accounted for by SNPs alone [26].

Many studies investigating CNVs have been carried out over the past few years. These studies have shown that these structural variations are major contributors to genetic diversity and phenotypic variability in many species, including humans [27–31], birds [32], pigs [23, 33], sheep [34–36], and cattle [25, 37–39]. However, there is limited information on how CNVs contribute to the phenotypic variation in traits related to feed efficiency, feeding behavior, and growth in cattle, especially in Zebu cattle (*Bos taurus indicus*) such as in the Nellore breed – the major Zebu breed in Brazil. Therefore, the main objectives of this study were to: 1) identify CNV and CNV regions (CNVR) in the genome of a Nellore cattle population; 2) assess potential associations between the identified CNVR and weaning weight adjusted to 210 days (W210), body weight measured at the time of selection (WSel), average daily gain (ADG), dry

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

matter intake (DMI), residual feed intake (RFI), time spent at the feed bunk (TF), and frequency of visits to the feed bunk (FF) traits; and, 3) perform functional genomic annotation of the associated CNV regions (CNVRs).

RESULTS

The descriptive statistics for the phenotype and the adjusted phenotypes used for the analyses are presented in Additional File 1.

Copy number variation and CNVR detection

Initially, 8,170 individual CNVs were detected in 1,222 samples. After the quality control, 3,161 CNVs located in the autosomal chromosomes from 620 samples remained for further analyses, with a mean of five CNVs per animal (range of 1 to 35). Out of the CNVs identified, 1,401 were deletions and 1,760 were duplications. The length of the CNVs ranged from 4,974 bp to 2,191,266 bp with an average length of 176,335±181,997 bp. No CNVs were detected on *Bos taurus* autosomes (BTA) BTA27 and BTA28. However, BTA6 exhibited the highest number of CNVs, with a maximum of 296 CNVs. On the other hand, BTA24 had the lowest number of CNVs, with only three CNVs identified.

The 3,161 CNVs remaining after quality control were used to infer CNVR by merging CNV with at least 1bp overlap. Thus, 561 CNVRs were identified, in which the average CNVRs length was 176,866±263,706 bp and they ranged from 4,973 bp to 3,215,394 bp. Among these CNVRs, 256 corresponded to genome deletions, 245 to duplications, and 60 to mixed (the same chromosomal segment was a deletion or duplication in different animals). The deletion-to-duplication ratio was 1.04. Thirty-nine CNVRs were identified in at least 1% of the population studied. The number and proportion of chromosomes covered by CNVRs varied considerably

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

and no CNVRs were identified in BTA27 and BTA28 (Fig. 1, Table 1). BTA1 had the largest number of CNVRs (n = 49), which covered 4.08% of the chromosome, and BTA12 presented the highest coverage of a chromosome sequence (10.2%) with 41 CNVRs. The CNVRs inferred in our study covered 99,221,894 bp of the autosomal genome sequence, which corresponds to 3.99% of the cattle genome size.

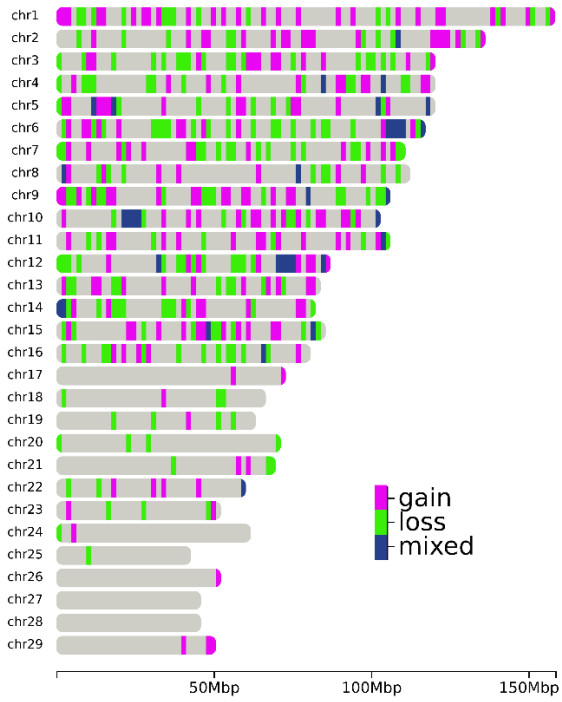

Fig 1. Distribution of copy number variation regions (deletions, duplications, and mixed type) by chromosome.

Table 1. Chromosome distribution of all 561 copy number variation regions (CNVRs) detected in the Nellore genome.

Chr[a]	Chr length (bp)	CNVR number	CNVR length (bp)	%[b]
BTA1	158,534,110	49	6,469,894	4.08
BTA2	136,231,102	35	5,392,406	3.96
BTA3	121,005,158	40	5,802,111	4.79
BTA4	120,000,601	33	4,625,409	3.85
BTA5	120,089,316	30	8,863,760	7.38
BTA6	117,806,340	43	9,353,904	7.94
BTA7	110,682,743	39	4,660,233	4.21
BTA8	113,319,770	21	3,178,109	2.80

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

BTA9	105,454,467	39	5,500,040	5.21
BTA10	103,308,737	23	5,660,780	5.48
BTA11	106,982,474	26	4,572,944	4.27
BTA12	87,216,183	41	8,878,934	10.2
BTA13	83,472,345	22	2,443,887	2.93
BTA14	82,403,003	26	4,704,971	5.71
BTA15	85,007,780	29	4,596,564	5.41
BTA16	81,013,979	20	2,809,390	3.47
BTA17	73,167,244	3	2,360,049	3.22
BTA18	65,820,629	4	579,684	0.88
BTA19	63,449,741	6	914,327	1.44
BTA20	71,974,595	4	1,039,708	1.44
BTA21	69,862,954	6	2,422,071	3.47
BTA22	60,773,035	8	1,785,240	2.94
BTA23	52,498,615	5	899,028	1.71
BTA24	62,317,253	2	127,751	0.20
BTA25	42,350,435	2	803,807	1.89
BTA26	51,992,305	2	454,688	0.87
BTA27	45,612,108	0	0	0
BTA28	45,940,150	0	0	0
BTA29	51,098,607	3	322,205	0.63
Total	**2,489,385,779**	**561**	**99,221,894**	**3.99**

[a]Chromosome
[b]The percentage of the chromosome covered by the copy number variation regions

Gene annotation, enrichment analyses, and QTL identification

Association analyses between the traits and the discovered CNVs of 620 animals led to the identification of 17 CNVRs significantly associated with at least 1 of the evaluated traits ($P < 0.0005$). The 17 regions represent 2 deletions, 9 duplications, and 6 mixed, distributed across 14 chromosomes with an average length of 273,250 ± 209,907 bp (Table 2). Out of the 17 CNVRs, 16 were associated with FF and 1 CNVR was associated with DMI and 12 were found in gap of the reference assembly ARS-UCD1.2 (Additional file 2). No CNVRs were significantly associated with the other traits.

Table 2. Copy number variation regions associated with growth, feed efficiency, and feeding behavior in Nellore cattle.

CNVR[a]	Chr[b]	Start (bp)	End (bp)	Type	Trait[c]
CNVR1	BTA2	135,110,420	135,653,313	Mixed	FF
CNVR2	BTA5	117,080,458	117,820,070	Deletion	FF
CNVR3	BTA6	116,755,758	117,164,372	Duplication	FF

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

CNVR4	BTA7	10,092,268	10,174,209	Duplication	FF
CNVR5	BTA7	42,951,015	43,292,715	Mixed	FF
CNVR6	BTA7	43,359,066	43,823,809	Mixed	FF
CNVR7	BTA8	15,562,312	15,781,720	Duplication	FF
CNVR8	BTA8	38,356,510	38,610,355	Duplication	FF
CNVR9	BTA8	85,996,187	86,508,867	Deletion	DMI
CNVR10	BTA9	2,637,837	2,700,411	Mixed	FF
CNVR11	BTA9	16,366,613	16,894,948	Duplication	FF
CNVR12	BTA9	15,312,685	15,469,154	Duplication	FF
CNVR13	BTA12	73,233,249	73,770,215	Mixed	FF
CNVR14	BTA12	74,302,958	74,578,587	Mixed	FF
CNVR15	BTA12	64,618,237	64,736,496	Duplication	FF
CNVR16	BTA13	12,552,408	12,829,168	Duplication	FF
CNVR17	BTA26	51,032,219	51,267,717	Duplication	FF

[a]Copy number variation region (CNVR) significantly ($P < 0.005$) associated with the traits
[b]Chromosome
[c]DMI: dry matter intake; FF: feed frequency

Functional enrichment analyses were performed to obtain broad functional insights into the set of genes significantly associated with the CNVRs for each trait. Gene Ontology enrichment analyses revealed 19 processes for FF, which are categorized as 10 biological processes, three cellular components, four molecular functions, and two metabolic pathways ($p < 0.05$, as shown in Additional file 3). No GO enrichment processes were identified for WSel, W210, ADG, DMI, RFI, TF.

In total, 73 previously-reported quantitative trait loci (QTL) overlapped with the genomic regions associated with DMI and FF (Additional file 4). The number of overlapping QTLs were 2 for DMI and 71 for FF. The QTL identified span a wide range of trait types, including meat and carcass, milk, reproduction, and production. These QTL are also associated with many important traits, including marbling score, fat cover carcass, body height, body weight, carcass weight, dry matter intake, and average daily gain.

DISCUSSION

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

All the growth, feed efficiency, and feeding behavior traits included in this study are heritable with heritability estimates ranging from 0.17 ± 0.03 (RFI) to 0.51 ± 0.06 (TF) [9]. Considering that CNVs are genomic alterations that can affect gene expression and, consequently, influence an individual's phenotype [22, 23] and that heritability indicates the proportion of phenotypic variation in a population attributable to genetic factors, a portion of the heritability of a given trait may be explained by genomic structural variations, such as CNVs. For instance, high heritability for specific traits implies that a substantial portion of the trait's phenotypic variability is attributable to genetic factors, which may include CNVs. This highlights the importance of considering these genomic alterations in the context of inheritance and phenotypic variability within populations.

Identification of CNV and CNVR

After applying quality control measures, approximately 60% of the initially detected CNVs were excluded from further analyses and many factors might be associated with this substantial reduction in the number of CNVs. For instance, the use of a rigorous quality control aiming to minimize false-positive CNV calls could result in the exclusion of true CNVs. Additionally, the design of the SNP panel itself could have played a key role in the reduction in the number of CNVs identified. The distribution of markers across the genome may not have fully captured certain CNVs, particularly those located in regions less represented by the SNP panel. Furthermore, the gap between markers could have also influenced the detection, as it might ignore smaller CNVs located within these gaps, or even the exclusion of larger CNVs when the gap between markers is too large.

After the quality control, 3,161 CNVs (1,401 deletions and 1,760 duplications) with an average length of 176,335 ± 181,997 bp remained for further analyses. Despite the number of

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

CNVs corroborating with values reported in the literature, there are often significant variability in the results of CNV analyses from different studies even within the same species. For instance, Butty et al. [39], while studying a population of 10,682 Holstein animals using SNP panels of different densities (BovineHD Beadchip HD, Genome Profiler Bovine 150K, Genome Profiler Bovine HD, BovineSNP50, and Genome Profiler Bovine 50K), identified an average of four CNVs per animal and a total of 23,256 CNVs with an average length of 168,520 bp. Peripolli et al. [40], studying whole-genome re-sequencing from 36 animals of different breeds, reported 7,285 CNVs in the population, with an average of 607.08 CNVs per animal, and an average length of 28,300 bp. Lemos et al. [38], studying 3,794 Nellore breed animals genotyped with a high density (HD) SNP panel and without adopting any quality control, identified 399,361 CNVs with an average length of 54,744 bp. Hou et al. [41], working with a population of 427 Angus reported 2,724 CNVs with an average of six CNVs per animal. These differences observed can be explained by differences in the data and methodologies used, including (1) the use of different platforms and methods (based on Comparative Genomic Hybridization, SNP-array, and Next Generation Sequencing); (2) the software utilized for the analyses; (3) quality control metrics and thresholds; (4) density of the genotyping SNP panels; and, (5) sample size [21, 42–44]. Despite the fact that all four studies mentioned above, and the current study focused on cattle, they all utilized different genotyping platforms, represented different populations, and differed on the sample sizes. These differences could justify the results obtained and make direct comparisons among studies unfair. It is important to acknowledge the impact of these differences when interpreting the findings. However, despite these discrepancies among studies, we observed some trends indicating that higher density panels tend to be associated with a greater number of CNVs and a shorter CNVs on average.

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

Deletion and duplication are genetic events that involve the number of copies of a particular DNA sequence. Deletion refers to the loss of a DNA segment from a chromosome and can impact phenotypic expression by interrupting genes and causing loss of biological functions [45]. Duplication refers to the process in which a segment of DNA is duplicated, resulting in additional copy of that sequence. Duplications are usually reported associated with digestion processes, lactation, reproduction, and immune system such as antigen processing and major histocompatibility genes in the cattle genome [46]. More duplications than deletions were identified in the present study, which is in agreement with Ladeira et al. [35] who reported 322 deletions and 835 duplications in sheep, and Liu et al. [47] who observed more duplication type (n = 11) CNVs as compared to deletion (n = 9) CNVs in Angus cattle. These findings deviate from the prevailing pattern observed in the literature regarding CNV surveys in animals, where deletions tend to be more frequent than duplications [23, 39, 40, 48].

The average length of the CNVs identified in the present study is in line with the distance between markers on the GGP 75K panel (107,700 bp; at least three SNPs were required to be considered as a CNV). The distribution of CNVs was not uniform across the cattle autosomes. This observation may be related with the different mechanisms of CNV formation, such as nonallelic homologous recombination (NAHR), fork stalling and template switching (Fostes, a mechanism based on DNA replication error), and nonhomologous end-joining (NHEJ), once each mechanism would take place more often in certain genomic regions than others [20].

The proportion of the genome covered by CNVRs (3.99%) was consistent with values reported in the literature, which range from 0.68% to 9.43% [20, 37, 39, 47]. The variation in genome coverage observed across studies can be attributed to the specific SNP panels used, which, similarly to CNV detection, can potentially affect the sensitivity of CNV detection.

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

When a low or medium-density SNP panel is used, there is a lower number of potential breakpoints compared to higher density SNP panels. As a result, the CNVs identified tend to be longer and cover a larger portion of the genome [39].

Association between CNVR and growth, feed efficiency, and feeding behavior traits

Copy number variation has the potential to modify gene expression, as deletions or duplications of gene segments, either partial or complete, can disrupt gene function and lead to phenotypic changes [49]. Consequently, identifying CNVRs that overlap with genes becomes a crucial step in assessing their functional impact. In this study, we further investigated the CNVRs identified based on the ARS-UCD1.2 cattle genome assembly. Remarkably, many CNVRs were associated with growth and feeding behavior, indicating that these CNVRs may play a role in influencing the phenotypic expression of these traits.

In this study, approximately 71% of the CNVRs significantly associated with the studied traits overlapped with Ensembl genes. This observation indicates that CNVRs often occur in gene-rich regions and suggests that these CNVRs could potentially have functional implications. This is because they might influence the expression or regulation of nearby genes and possibly contribute to the phenotypic variations associated with important traits. In total, 95 genes overlapped with these genomic regions and 83 of them were classified as protein-coding genes. A similar trend was observed in previous studies in pigs [50], beef cattle [46], and humans [51], where CNVRs were also found to be concentrated in protein-coding genes. Protein-coding genes are segments of DNA that serve as template for transcription of the DNA into RNA sequences and the complementary chain [52, 53]. The importance of this type of genes lies in their ability to direct the synthesis of specific proteins. They are also essential to the task of translating the information in the sequence of the genome into biologically relevant

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

knowledge and can affect dosage-sensitive genes [52]. Therefore, the presence of CNVRs in protein-coding genes may be relevant to explain the associations with the studied traits.

The lack of significantly identified CNVRs associated with WSel, W210, ADG, RFI, and TF traits may be linked to the limited number of genotyped animals with these phenotypes. Increasing the sample size for future studies could enhance the precision in identifying genetic variants associated with the phenotypes of interest and increase the accuracy and reliability of genetic variations associated with the traits under investigation. FF was the trait with the highest number of significant CNVRs and overlapped with important genes and QTLs. FF was also associated with some QTL related to dry matter intake, body weight and bovine respiratory disease (BRD) susceptibility in cattle. These are important traits, since BRD is one of the most common and costly disease of feedlot cattle, and has a negative impact on ADG, where calves diagnosed with BRD tend to have lower ADG compared with healthy animals [54, 55]. Therefore, the genetic associations between FF with BRD susceptibility, dry matter intake, and body weight QTLs highlight the interaction between genetic factors, feeding behavior, and overall cattle performance, with significant implications for the cattle industry. Additionally, significant biological processes associated with FF were identified, particularly related to the activation of GTPase activity (GO:0090630), GMP catabolic process (GO:0046038), and cellular response to mechanical stimulus (GO:0071260).

The CNVR12 was significantly associated with FF and overlapped with *MYO9A* gene. This is a gene member of the myosin superfamily that is related to ATPase activity. ATPase activity is essential for many cellular processes, plays a crucial role in cellular energy metabolism, and is involved in a wide range of physiological processes. The CNVR17, located in the chromosome 26 overlapped with five genes, including *INPP5A* and *MRPL13*. *INPP5A* is gene that code a protein responsible for mobilizing intracellular calcium and acts as a second

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

messenger mediating cell responses to several stimulations [56]. The *MRPL13* gene is a component of the mitochondrial ribosomal protein (MRP) family. The MRP are synthesized in the cytoplasm before being transported into the mitochondria for the purpose of mitochondrial ribosome assembly. MRP is vital for mitochondrial oxidative phosphorylation and plays a significant role in the regulation of apoptosis-inducing factors, and an alteration in the MRP expression could result in a range of disorders, including mitochondrial metabolic disorders and cellular dysfunction [57]. Although no studies have reported a direct relationship between the *MRPL13* gene and feeding behavior, it is important to highlight that the mitochondria are cellular organelles responsible for energy conversion and adenosine triphosphate (ATP) production in eukaryotic cells [58]. In addition to their function in energy metabolism, they play an important role in diverse cellular processes, such as apoptosis [59] and aging [60]. Therefore, it is plausible that the *MRPL13* gene, being involved in mitochondrial function and cellular metabolism, may have indirect implications for the organism's response to feeding behavior, and consequently, feed efficiency and growth.

Feeding behavior is a complex process that encompasses a multitude of psychological factors, neuronal mechanisms, and metabolic processes. Thus, while none of the identified genes have been directly associated with feeding behavior traits yet, the fact that many of them are linked to metabolic activities might suggest a relationship between these genes and the FF. These novel findings highlight the importance of developing a reference genome for Nellore cattle and performing detailed functional annotation of the Nellore cattle genome.

Despite the findings obtained, there are also inherent limitations in this study. These limitations provide a solid foundation for identifying research areas that require further investigation and pave the way for more comprehensive and in-depth studies in the future. One limitation of this study is that the current study lacks detailed information regarding the storage

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

and processing procedures of all utilized DNA samples. This stands as a crucial aspect as it can significantly impact the detection of CNVs. Variations in the DNA source have the potential to influence the genetic material's integrity, while the processing methods such as DNA extraction, storage, or amplification might affect the quality of the genomic data acquired [61], potentially impacting the accuracy of genetic variations identification. Furthermore, low quantity and/or quality DNA samples can lead to a higher number of genotyping errors [62].

Another important point is the relatively small number of animals, which, although larger than many previous studies, may still have impacted the results reported here. Additionally, the use of the reference genome ARS-UCD1.2 could also be a limitation that affected the results, as it is based on the genome of a *Bos taurus taurus* (Taurine) animal of the Hereford breed, while the animals in the present study are *Bos taurus indicus* (Zebu) of the Nellore breed. Therefore, future studies could take these potential factors into consideration for a more comprehensive understanding of the CNVs in Zebu cattle. Furthermore, validation of CNVs and CNVRs were not performed in this study. Validation studies are important to ensure the accuracy and reliability of the CNVs detected. One alternative is to use Whole Genome Sequencing (WGS) data that can provide a more comprehensive overview of the entire genome and tends to be more sensitivity for the detection of structural and complex variants in the genome [63]. Therefore, WGS could be a powerful tool for CNV validation in future studies.

The results of the present study have significant implications and can have several practical applications. Based on the findings, one possible application is the creation of SNP panels with a higher number of markers in the regions with large incidence of CNVRs. This would allow more comprehensive and refined genomic analyses, providing a more detailed understanding of the genetic variations in those regions. Additionally, an alternative approach to consider in genomic prediction is the differential weighting of SNPs located within CNVRs.

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

A potential method for this is the use of the Weighted Single-step Genomic BLUP procedure, as proposed by Wang et al. [64], that is an iterative process involving updates of SNP solutions with appropriate weights. Incorporating differential weighting of SNPs within genomic models could enable a more realistic representation of the actual distribution of SNP effects, with a particular emphasis on CNVRs with larger effects, and could potentially improve the accuracy of genomic prediction of breeding values [26].

CONCLUSIONS

This study aimed to investigate CNVs and CNVRs in the genome of a Nellore cattle population and explore their associations with growth, feed efficiency, and feeding behavior traits. A total of 3,161 CNVs and 561 CNVRs were identified and characterized within the Nellore cattle genome. The results revealed that some CNVRs are significantly associated with the traits analyzed, showing the potential influence of structural genome variations on economically relevant traits in Nellore cattle. The functional annotation of the associated CNVR revealed some important genes that may be related to the expression of the traits studied. Various QTLs overlapping with the CNVRs identified are related with growth, feed efficiency, and feeding behavior traits in Nellore cattle.

MATERIAL AND METHODS

No Animal Care Committee approval was necessary for the purposes of this study, as all information required was obtained from pre-existing databases.

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

Animals and phenotypic datasets

Performance records were obtained from 1,338 animals born from 1983 to 2020, which belong to the Nellore cattle herd of the Institute of Animal Science (IZ), Sertãozinho, SP, Brazil. The analyzed traits were weaning weight adjusted for 210 days of age (W210), body weight measured at the time of selection (WSel), average daily gain (ADG), dry matter intake (DMI), residual feed intake (RFI), time spent at the feed bunk (TF), and frequency of visits to the feed bunk (FF).

Weaning weight adjusted for 210 days of age was calculated based on the weight gain between birth and weaning, using the following equation:

$$\mathrm{W210} = \left(\frac{\mathrm{WW} - \mathrm{BW}}{\mathrm{AAW}}\right) * 210 + \mathrm{BW}$$

where W210 is the weaning weight adjusted for 210 days of age; WW is the weaning weight; BW is the birth weight; and AAW is the age of the animal at weaning. WSel is the postweaning weight adjusted to 378 days of age for males in feedlot performance tests and postweaning weight adjusted to 550 days of age for females on pasture.

Dry matter intake (DMI) and average daily gain (ADG) were obtained for males and females, which participated of 21 performance tests after weaning with a minimum of 21 days of adaptation to the diet and facilities and 86±13 test days. The animals started the test at the age of 293±42 days and remained in individual or collective pens equipped with the Vytelle LLC® (Vytelle LLC, Calgary, AB, Canada) or Intergado® (Contagem, Minas Gerais, Brazil) electronic monitoring system. The animals had *ad libitum* access to diet and water. The diet was formulated for an ADG of 1.1 kg. Body weights were recorded at a maximum interval of 28 days.

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

The DMI was obtained as the average of all valid days of intake multiplied by the dry matter content of the diet offered each week. ADG was calculated as the linear regression coefficient of weights as a function of days on test:

$$\mathbf{y_i = \alpha + \beta \times DOT_i + e_i},$$

where $\mathbf{y_i}$ is the animal's weight in the i^{th} observation; $\boldsymbol{\alpha}$ is the intercept corresponding to the initial weight; $\boldsymbol{\beta}$ is the linear regression coefficient corresponding to ADG; and $\mathbf{e_i}$ is the random error. RFI was estimated as the residual of the linear regression equation of DMI on ADG and $BW^{0.75}$ (Koch et al., 1963). Feeding behavior data was only available for males kept in collective pens equipped with Vytelle LLC® (Vytelle LLC, Calgary, AB, Canada) in 12 feeding tests. The electronic trough systems were configured to scan the electronic identification tags of animals entering the trough every 1.0 to 6.3 s. The start of a meal event is defined when the tag of an animal was identified by the system. The meal event ends when the time between the last two readings of the same tag was longer than 300 seconds, and can be made up of a single or several feeding events from different bunks, or when a new tag was detected in the same trough (Vytelle LLC®, Calgary, AB, Canada). Meal events with a feed intake lower than 1 kg and time at bunk lower than 3 seconds were discarded. The following feeding behavior traits were analyzed: time spent at the feed bunk (TF, average daily time the animal spent at the feed bunk during the test period, min per day) and frequency of visits to the feed bunk (FF, average sum of feeding events of the animal per day, number of feeding events per day).

The phenotypic records were adjusted for the fixed effects listed in Table 5 using the *lm() function* in R. The adjusted phenotypes were then used for the association analyses.

Table 6. Effects included for estimating genomic breeding values for growth, feed efficiency, and feeding behavior traits.

Trait[a]	Categorical fixed effects[b]	Covariates[c]

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

W210 (kg)	CGw	CA, CA^2 AW
WSel (kg)	CGw	CA, CA^2, ASW
ADG ($kg.day^{-1}$)	CGr, BM	AST
DMI ($kg.day^{-1}$)	CGr, BM	AST
RFI ($kg.day^{-1}$)	CGr, BM	AST
TF ($hour.day^{-1}$)	CGr, BM	AST
FF ($n.day^{-1}$)	CGr, BM	AST

[a]W210: weight at 210 days; WSel: body weight measured at the time of selection; ADG: average daily gain; DMI: dry matter intake; RFI: residual feed intake; TF: time spent at the feed bunk; FF: frequency of visits to the feed bunk.
[b]CGw: contemporary group for weight (birth year, birth month, line, and sex); CGr: contemporary group for RFI (test start year, test start month, installation, and sex); BM: birth month
[c]CA and CA^2: cow age (linear and quadratic effects); AW: age at weaning; ASW: age at selection weight; AST: age at the start of the feeding test.

Genomic datasets

A total of 1,338 animals were genotyped with the GeneSeek Genomic Profiler HDi 75K (GeneSeek Inc., Lincoln, NE, USA) panel containing 74,677 SNPs distributed along the genome, with a mean distance between markers equal to 32.3±10 kilobases (Kb). The genotyped animals included 817 males, 519 females, and 2 founders (unknown sex). Non-autosomal SNPs, SNPs with unknow genome position, and SNPs with a GenCall score below 0.15 were removed during the genomic quality control. After the quality control, 69,680 SNPs remained for further analyses.

Copy number variation identification

The CNV identification was performed using the PennCNV.1.0.5 software [74], which integrates Log R Ratio (LRR) and B Allele Frequency (BAF) on a per sample basis into a hidden Markov model to determine the number of copies and genotypes of each CNV. LRR measures the total signal intensity while BAF measures the proportion of the B allele in each sample. The population frequency of the B allele information was calculated using the BAF value of each SNP in all samples. To reduce false-positive results, the genomic waves were

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

adjusted using the -gcmodel option in PennCNV program. Genomic waves refer to a signal noise related to the GC content in the genome, which interferes with accurate CNV detection., and the cattle *gcmodel* file was generated by calculating the guanine-cytosine (GC) content of each marker. The LRR values of each SNP were adjusted for the genomic waves along the genomic regions, taking into account the expected GC content in the bovine genome in a region of 500 Kb around each SNP and based on a regression model [70].

After CNV calling, a sample-based quality control was performed, which removed CNVs with a BAF drift < 0.01, standard deviation of LRR > 0.30, minimum length of 1,000 bp, maximum length of 5,000,000 bp, and GC wave factor < 0.05 (after genomic waves were corrected by guanine-cytosine content) to generate raw CNV calls. In addition, CNVs with less than three consecutive SNPs were discarded. Finally, after the control quality, 620 individuals and 3,161 CNV remained for further analyses.

Copy number variation regions identification

The CNVR were determined by grouping the 3,161 CNVs that overlapped by at least 1 bp within each algorithm, using the *mergeBed* option of the BEDtools suite tool [76]. CNVRs were classified in deletions when the animal showed a region with loss of a chromosomal segment, duplication for repeated chromosomal regions, and mixed, when it was identified deletions and duplications in the same genomic region.

Association analyses

The association analyses were performed considering only the CNVR identified in the autosomal chromosomes. The CNVRs were coded as -1 (deletion), 0 (neutral), and 1

&Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

(duplication). To test the potential associations between CNVRs with the pre-adjusted phenotypes, the model fitted is:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},$$

where **y** is the vector of pre-adjusted phenotypes for each trait; **b** is the fixed effect of the CNVR tested for potential association with the phenotype, **X** is a vector containing the genotype score for the tested CNVR; **u** is the random vector of polygenic effect with $\mathbf{u} \sim N(0, \mathbf{A}\sigma_u^2)$, where **A** is the additive genetic relationship matrix obtained from the animal pedigree, σ_u^2 is the additive genetic variance for each trait; **Z** is the incidence matrix for **u**; and **e** is a random vector of residual effects with $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where **I** is an identity matrix and σ_e^2 is the residual variance. The variance components were previously calculated by fitting the above-described model excluding the CNVR as a fixed effect in the model. The estimation of variance components and the CNVR effect size estimation were performed using the BLUPF90+ program [69].

The model adjusted provided the effect size (β) and standard error (SE) for each CNVR, which were used to compute the t statistic ($t = \beta / SE$). Subsequently, p-values were computed by assuming that a t-distribution with "n – 2" degrees of freedom, where n is the sample size (i.e., the number of animals used to obtain the CNVR effect for each pre-adjusted phenotype). To adjust for multiple testing, a Bonferroni correction at $\alpha = 0.05$ genome-wise significance level was applied by dividing α by the total number of CNVRs.

Gene annotation and functional analyses

The CNVRs significantly associated with the phenotypes were used for the annotation. The gene and QTL annotation in these regions was performed using the GALLO package [79], utilizing annotated data for *Bos taurus* sourced from the Ensembl database (www.ensembl.org/Bos_taurus/Info/Index) and the reference genome ARS-UCD1.2 [80].

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

Additionally, the Cattle QTL database (www.animalgenome.org/cgi-bin/QTLdb/BT/index) was used as a resource for obtaining previously-reported QTL information. candidate regions for each QTL type is compared with the observed number of QTLs in the reference database. Subsequently, The Database for Annotation, Visualization and Integrated Discovery (DAVID; version 6.8) [75] was used for conducting Gene Ontology (GO) and KEGG pathway enrichment ($p<0.05$) analyses to identify biological processes, molecular functions, and cellular components associated with the positional candidate genes identified.

DECLARATIONS

Ethics approval and consent to participate

Animal Care and Use Committee approval was not obtained for this study because analyses were performed on pre-existing datasets.

Consent for publication

Not applicable.

Availability of data and materials

The data supporting this study's findings belongs to an experimental animal breeding program, and restrictions are applied to the availability of data. However, data are available by contacting the corresponding authors upon reasonable request and with permission of the program (contacting the researcher mezmercadante@gmail.com).

Competing interests

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

The authors declare that they have no competing interests.

Funding

This study was financially supported by the São Paulo Research Foundation (FAPESP, 2017/10630-2 and 2017/50339-5) and Coordination for the Improvement of Higher Education Personnel (CAPES, Brasilia, DF, Brazil; Finance Code 001).

Authors' contributions

LFB, LFB, and MEZM conceived and designed the study and conducted the data analyses. LFB, LFB, RDB, HAM, JG, LGB, LSG, JNSGC, and SFMB contributed to the data acquisition and interpretation of the results. LFB, LFB, and MEZM wrote and edited the manuscript. All authors reviewed and contributed to the editing of the manuscript and approved its final version.

REFERENCES

1. Goulart RS, Tedeschi LO, Silva SL, Leme PR, De Alencar MM, Lanna DPD. The energy requirement for maintenance of Nellore crossbreds in tropical conditions during the finishing period. J Anim Sci. 2022;100.

2. Cooke RF, Daigle CL, Moriel P, Smith SB, Tedeschi LO, Vendramini JMB. Cattle adapted to tropical and subtropical environments: Social, nutritional, and carcass quality considerations. In: Journal of Animal Science. Oxford University Press; 2020.

3. Kelly AK, McGee M, Crews DH, Fahey AG, Wylie AR, Kenny DA. Effect of divergence in residual feed intake on feeding behavior, blood metabolic variables, and body composition traits in growing beef heifers. J Anim Sci. 2010;88:109–23.

4. Basarab JA, Price MA, Aalhus JL, Okine EK, Snelling WM, Lyle KL. Residual feed intake and body composition in young growing cattle. Can. J. Anim.Sci. 2003;189-204.

5. Nkrumah JD, Crews DH, Basarab JA, Price MA, Okine EK, Wang Z, et al. Genetic and phenotypic relationships of feeding behavior and temperament with performance, feed efficiency, ultrasound, and carcass merit of beef cattle. J Anim Sci. 2007;85:2382–90.

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

6. Herd RM, Arthur PF. Physiological basis for residual feed intake. J Anim Sci. 2009;87.

7. Sakamoto LS, Souza LL, Gianvecchio SB, de Oliveira MHV, de Vasconcelos Silva JA, Canesin RC, et al. Phenotypic association among performance, feed efficiency and methane emission traits in Nellore cattle. PLoS One. 2021.

8. Montanholi YR, Swanson KC, Palme R, Schenkel FS, McBride BW, Lu D, et al. Assessing feed efficiency in beef steers through feeding behavior, infrared thermography and glucocorticoids. Animal. 2010;4:692–701.

9. Benfica LF, Sakamoto LS, Magalhães AFB, De Oliveira MHV, De Albuquerque LG, Cavalheiro R, et al. Genetic association among feeding behavior, feed efficiency, and growth traits in growing indicine cattle. J Anim Sci. 2020;98.

10. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS One. 2009;4.

11. Balagu-Dobón L, Cáceres A, González JR. Fully exploiting SNP arrays: A systematic review on the tools to extract underlying genomic structure. Briefings in Bioinformatics. 2022;23.

12. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526:75–81.

13. Abel HJ, Larson DE, Regier AA, Chiang C, Das I, Kanchi KL, et al. Mapping and characterization of structural variation in 17,795 human genomes. Nature. 2020;583:83–9.

14. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. Nature. 2020;581:444–51.

15. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. Nat Genet. 2007;39:37–42.

16. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, et al. Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res. 2012;22:778–90.

17. Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, et al. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BMC Genomics. 2013;14.

18. Schurink A, da Silva VH, Velie BD, Dibbits BW, Crooijmans RPMA, François L, et al. Copy number variations in Friesian horses and genetic risk factors for insect bite hypersensitivity. BMC Genet. 2018;19:1–13.

19. Solé M, Ablondi M, Binzer-Panchal A, Velie BD, Hollfelder N, Buys N, et al. Inter- A nd intra-breed genome-wide copy number diversity in a large cohort of European equine breeds. BMC Genomics. 2019;20:1–12.

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

20. Fadista J, Thomsen B, Holm L-E, Bendixen C. Copy number variation in the bovine genome. BMC Genomics. 2010;11.

21. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444:444–54.

22. Jiang L, Jiang J, Yang J, Liu X, Wang J, Wang H, et al. Genome-wide detection of copy number variations using high-density SNP genotyping platforms in Holsteins. BMC Genomics. 2013;14:23–6.

23. Stafuzza NB, Silva RMDO, Fragomeni BDO, Masuda Y, Huang Y, Gray K, et al. A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive. BMC Genomics. 2019;20:1–11.

24. Hujoel MLA, Sherman MA, Barton AR, Mukamel RE, Sankaran VG, Terao C, et al. Influences of rare copy-number variation on human complex traits. Cell. 2022;185:4233-4248.e27.

25. Xu L, Cole JB, Bickhart DM, Hou Y, Song J, VanRaden PM, et al. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics. 2014;15:1–10.

26. Hay EHA, Utsunomiya YT, Xu L, Zhou Y, Neves HHR, Carvalheiro R, et al. Genomic predictions combining SNP markers and copy number variations in Nellore cattle. BMC Genomics. 2018;19.

27. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet. 2006;79:439–48.

28. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong Association of De Novo Copy Number Mutations with Autism Interdepartmental Program in the Neurosciences, Program in. Science. 2007;316:445–9.

29. Walsh T, Mcclellan JM, Mccarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia. Science. 2008;320:539-43

30. Moreno-De-Luca D, Mulle JG, Kaminsky EB, Sanders SJ, Myers SM, Adam MP, et al. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. Am J Hum Genet. 2010;87:618–30.

31. Trost B, Thiruvahindrapuram B, Chan AJ, Engchuan W, Higginbotham EJ, Howe JL, et al. Genomic architecture of Autism Spectrum Disorder from comprehensive whole-genome sequence annotation. Cell. 2022;23:4409-27.

32. Elferink MG, Vallée AAA, Jungerius AP, Crooijmans RPMA, Groenen MAM. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. BMC Genomics. 2008;9:1–9.

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

33. Paudel Y, Madsen O, Megens HJ, Frantz LAF, Bosse M, Crooijmans RPMA, et al. Copy number variation in the speciation of pigs: A possible prominent role for olfactory receptors. BMC Genomics. 2015;16:1–14.

34. Yang L, Xu L, Zhou Y, Liu M, Wang L, Kijas JW, et al. Diversity of copy number variation in a worldwide population of sheep. Genomics. 2018;110:143–8.

35. Ladeira GC, Pilonetto F, Fernandes AC, Bóscollo PP, Dauria BD, Titto CG, et al. CNV detection and their association with growth, efficiency and carcass traits in Santa Inês sheep. Journal of Animal Breeding and Genetics. 2022;139:476–87.

36. Taghizadeh S, Gholizadeh M, rahimi-Mianji G, Moradi MH, Costilla R, Moore S, et al. Genome-wide identification of copy number variation and association with fat deposition in thin and fat-tailed sheep breeds. Sci Rep. 2022;12:1–12.

37. Zhou Y, Utsunomiya YT, Xu L, Hay EH abdel, Bickhart DM, Alexandre PA, et al. Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. BMC Genomics. 2016;17:1–9.

38. Antunes de Lemos MV, Berton MP, 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. 2018;207 November 2017:51–8.

39. Butty AM, Chud TCS, Cardoso DF, Lopes LSF, Miglior F, Schenkel FS, et al. Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle. J Dairy Sci. 2021;104:8050–61.

40. Peripolli E, Stafuzza NB, Machado MA, do Carmo Panetto JC, do Egito AA, Baldi F, et al. Assessment of copy number variants in three Brazilian locally adapted cattle breeds using whole-genome re-sequencing data. Anim Genet. 2023;54:254-70.

41. Hou Y, Bickhart DM, Chung H, Hutchison JL, Norman HD, Connor EE, et al. Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. Funct Integr Genomics. 2012;12:717–23.

42. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: Features and perspectives. BMC Bioinformatics. 2013;14.

43. Duan J, Zhang J-G, Deng H-W, Wang Y-P. CNV-TV: A robust method to discover copy number variation from short sequencing reads. BMC Bioinformatics. 2013;150.

44. Castellani CA, Melka MG, Wishart AE, Locke MEO, Awamleh Z, O'Reilly RL, et al. Biological relevance of CNV calling methods using familial relatedness including monozygotic twins. BMC Bioinformatics. 2014;15.

45. Liu B, Yang L, Huang B, Cheng M, Wang H, Li Y, et al. A functional copy-number variation in MAPKAPK2 predicts risk and prognosis of lung cancer. Am J Hum Genet. 2012;91:384–90.

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

46. Keel BN, Lindholm-Perry AK, Snelling WM. Evolutionary and functional features of copy number variation in the cattle genome. Front Genet. 2016;7.

47. Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, et al. Analysis of copy number variations among diverse cattle breeds. Genome Res. 2010;20:693–703.

48. Braga LG, Chud TCS, Watanabe RN, Savegnago RP, Sena TM, do Carmo AS, et al. Identification of copy number variations in the genome of Dairy Gir cattle. PLoS One. 2023;18:e0284085.

49. Lupski JR, Stankiewicz P. Genomic disorders: Molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genetics. 2005;1:0627–33.

50. Wang Z, Guo Y, Liu S, Meng Q. Genome-Wide Assessment Characteristics of Genes Overlapping Copy Number Variation Regions in Duroc Purebred Population. Front Genet. 2021;12.

51. Narasimhan A, Kumaran M, Gioulbasanis I, Skipworth RJE, Bathe OF, Kaasa S, et al. A genome wide association study to identify germline variants associated with cancer-associated cachexia-a preliminary analysis. 2023.

52. Harrow J, Nagy A, Reymond A, Alioto T, Patthy L, Antonarakis SE, et al. Identifying protein-coding genes in genomic sequences. Genome biology. 2009;10:201.

53. Amaral Haddad Editor L. Human Genome Structure, Function and Clinical Considerations.

54. Snowder GD, Van Vleck LD, Cundiff L V., Bennett GL. Bovine respiratory disease in feedlot cattle: Environmental, genetic, and economic factors. J Anim Sci. 2006;84:1999–2008.

55. Smith DR. Risk factors for bovine respiratory disease in beef cattle. Animal Health Research Reviews. 2020;21:149–52.

56. Mills SJ, Vandeput F, Trusselle MN, Safrany ST, Erneux C, Potter BVL. Benzene polyphosphates as tools for cell signalling: Inhibition of inositol 1,4,5-trisphosphate 5-phosphatase and interaction with the PH domain of protein kinase Bα. ChemBioChem. 2008;9:1757–66.

57. Dong T, Lu S, Li X, Yang J, Liu Y. Genetic association between ankylosing spondylitis and major depressive disorders: Shared pathways, protein networks and the key gene. Medicine. 2023;102:e33985.

58. Greber BJ, Ban N. Structure and Function of the Mitochondrial Ribosome. Annu Rev Biochem. 2016;85:103–32.

59. Tait SWG, Green DR. Mitochondria and cell death: Outer membrane permeabilization and beyond. Nature Reviews Molecular Cell Biology. 2010;11:621–32.

60. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153:1194.

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

61. Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P. How to track and assess genotyping errors in population genetics studies. Molecular Ecology. 2004;13:3261–73.

62. Bayés M, Gut IG. Overview of Genotyping. In: Molecular Analysis and Genome Discovery: Second Edition. wiley; 2012. p. 1–23.

63. Jones JL, Wang L, Ceric O, Nemser SM, Rotstein DS, Jurkovic DA, et al. Whole genome sequencing confirms source of pathogens associated with bacterial foodborne illness in pets fed raw pet food. Journal of Veterinary Diagnostic Investigation. 2019;31:235–40.

64. Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet Res (Camb). 2012;94:73–83.

65. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, et al. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience. 2020;9.

66. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17:1665–74.

67. Diskin SJ, Li M, Hou C, Yang S, Glessner J, Hakonarson H, et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 2008;36:1–12.

68. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.

69. Misztal I, Tsuruta S, Lourenco D, Masuda Y, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. University of Georgia. 2022.

70. Fonseca PAS, Suárez-Vega A, Marras G, Cánovas Á. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. Gigascience. 2020;9.

71. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane C, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. 2003.

[&]Este capítulo corresponde ao artigo científico submetido à revista BMC genomics e encontra-se em avaliação para publicação.

CHAPTER 3 - Detection and characterization of copy number variation in three differentially-selected Nellore cattle populations

ABSTRACT: Nellore cattle (Bos taurus indicus) is the main beef cattle breed raised in Brazil – one of the largest beef producers in the world. Nellore cattle are well adapted to tropical conditions and, more recently, have experienced intensive genetic selection for multiple performance traits. Over the past 43 years, an experimental breeding program has been developed at the Institute of Animal Science (IZ, Sertaozinho, SP, Brazil), which resulted in three differentially-selected lines known as Nellore Control (NeC), Nellore Selection (NeS), and Nellore Traditional (NeT). The primary goal of this selection experiment was to determine the response to selection for yearling weight (YW) and residual feed intake (RFI) on Nellore cattle. These populations are a great resource for studying the genetic mechanisms involved in the phenotypic variability on YW and RFI observed across the three lines. In this context, the primary objectives of this study were to: 1) identify copy number variations (CNVs) in Nellore cattle from three selection lines; 2) identify and characterize CNV regions (CNVR) on these three lines; and 3) to perform functional enrichment analyses of the CNVR identified. The study utilized records of Nellore animals born between 1978 and 2020, belonging to three selection lines for post-weaning weight and RFI. A total of 928 animals were genotyped, with 770 animals using the Illumina BovineHD BeadChip 770k and 158 animals using the GeneSeek Genomic Profiler Indicus 50k. The CNV identification was conducted separately for each panel using PennCNV 1.0.5. The CNVs were separated according to each selection line, and CNVRs were determined for each selection line by grouping identified CNVs that overlapped for at least 1 base pair. Subsequently, exclusive CNVRs present in at least 10% of animals in each line were identified. The exclusive CNVRs were used for gene and QTL annotation, and gene ontology (GO) analysis. A total of 14,914 unique CNVs and 1,884 unique CNVRs were identified when consider the whole population in a single analyze. The CNVRs were non-uniformly distributed across the chromosomes of the three selection lines under study. The NeT line had the highest number of CNVRs (1,493), followed by the NeS (823) and NeC lines (482). The CNVRs covered 23,449,890 bp (0.94%), 40,175,556 bp (1.61%), and 63,212,273 bp (2.54%) of the genome of the NeC, NeS, and NeT lines, respectively. Two CNVRs were found in common between the three lines, and six, two, and four exclusive regions were identified for NeC, NeS, and NeT, respectively. All the exclusive regions overlap with important genes, such as SMARCD3, SLC15A1, and MAPK1. Key biological processes associated with the candidate genes were identified, including pathways related to growth and metabolism. This study revealed large variability in CNVs and CNVRs across three Nellore lines differentially selected for YW and RFI. Gene annotation and gene ontology analyses of the exclusive CNVRs to each line revealed specific genes and biological processes involved in the expression of growth and feed efficiency traits. These finds contribute to the understanding of the genetic mechanisms underlying the phenotypic differences among the three selection lines.

KEYWORDS: Copy number variation, gene annotation, Residual feed intake, SNP panel

1. INTRODUCTION

Nellore cattle (Bos taurus indicus) is the main beef cattle breed raised in Brazil, i.e. one of the largest beef producers and exporters in the world (USDA, 2023). Nellore animals are well adapted to harsh climatic conditions and Brazilian herds have experienced major genetic progress for performance traits over the past decades (Fernandes Jr. et al., 2022). In addition to national Nellore breeding programs, an experimental breeding program was initiated in 1980 in the Institute of Animal Science (IZ; Sertãozinho, SP, Brazil), with the development of three selection lines. At the beginning of the breeding program, the primary goal of the experiment was to assess the response to selection for heavier weights in a tropical beef cattle population (Mercadante et al., 2003). Briefly, the three selection lines were established by randomly dividing the founder animals into three groups: Nellore Control (NeC), Nellore Selection (NeS), and Nellore Traditional (NeT). NeC was maintained under stabilizing selection, in which animals with a yearling weight (YW) close to the average of the contemporary group were selected for reproduction each year. NeS and NeT were selected for higher selection differentials for YW, and in 2008, residual feed intake (RFI) was also introduced as a selection criterion in the NeT line (Mercadante et al., 2003; Cardoso et al., 2018; Benfica et al., 2020).

After more than 40 years of selection, there are clear phenotypic differences among the lines subjected to stabilizing and directional selection. These differences include traits such as average body weight at various ages, body measurements, RFI, scrotal circumference, and carcass quality (Mercadante et al., 2003; Monteiro et al., 2013; Ceacero et al., 2016). Therefore, the three lines developed are a great resource for identifying genomic regions that have been influenced by selection, thus providing insights into the genes involved in the phenotypic expression of these traits. Numerous studies have been conducted to unravel the genetic mechanisms underpinning the phenotypic variations among the three Nellore lines. For instance, genome-wide association studies (GWAS) reported key genes associated with growth and feed efficiency traits, and population genetic stratification highlighted autosomal genomic regions with selection footprints (Ayres et al., 2010; Souza et al., 2011; Cardoso et al.,

2018).A new approach that could be further explored are copy number variations (CNV).

Copy Number Variations are structural variations within an individual's genome, involving the loss or gain of DNA fragments, which can range from 1 kilobase (kb) to several megabases in size when compared to the reference genome of the species (Henrichsen et al., 2009). CNVs span extensive chromosomal regions and can change gene structure, regulatory modifications, alterations in gene dosage, and exposure of recessive alleles, leading to significant impact on gene expression (Zhang et al., 2009; Stafuzza et al., 2019) and phenotypic variability in complex traits (Zhang et al., 2009). In this context, the study of CNVs serves as a valuable source of information to elucidate some of the mechanisms contributing to the differences among the three experimental selection lines and in the phenotypic variations observed in economically important traits. Hence, the primary objectives of this study were to: 1) identify and characterize CNVs and CNV regions (CNVR) in Nellore cattle from three differentially-selected lines; and, 2) perform functional enrichment analyses of the identified CNVRs.

2. MATERIAL AND METHODS

2.1 Animals and experimental breeding program design

Data was collected from 928 animals, including 114 from the NeC line, 245 from the NeS line, and 569 from the NeT line. These animals were born between 2004 and 2019 and are part of the Nellore cattle herd at the Institute of Animal Science (IZ) in Sertãozinho, SP, Brazil. These animals are part of an experimental breeding program initiated in 1980 and separated into three selection lines: Nellore Control (NeC), Nellore Selection (NeS), and Nellore Traditional (NeT). The NeC and NeS lines are considered closed lines, while at times, bulls from external sources have been introduced in the NeT line, including animals from commercial herds or those culled from the NeC and NeS lines, as described by Mercadante et al. (2003). Bulls were chosen from contemporary groups (defined by line and year) based on their YW adjusted to 378 days (W378) after a 168-day feedlot performance test. Replacement females, on the other hand, are selected based on their YW adjusted to 550 days (W550) while kept on pasture.

In the NeC line, males and females with a selection differential close to zero for YW were retained for breeding. Animals from the NeC line have maintained YW values that are close to the average observed at the outset of the breeding program in 1980. In contrast, for the selected NeS and NeT lines, both males and females with higher adjusted weights were selected over time. Starting in 2008, the sires from the NeT line have been selected based on higher genomic estimated breeding values (GEBV) for YW and lower GEBV for RFI (Mercadante et al., 2003; Cardoso et al., 2018; Benfica et al., 2020). RFI was estimated as the residual of the linear regression equation of dry matter intake (DMI) on average daily gain (ADG) and mid-test metabolic weight (BW0.75) (1v et al., n.d.) in each test group.

The sire selection strategy has been consistently applied to this day, involving the annual replacement of 50% of the three-year-old sires within each line. Furthermore, the annual culling rate for cows is approximately 20%. Fig. 1 illustrates the differentiation in the phenotypic performance of the lines achieved through selection.

Figure 1. Four-year-old sires from three differentially selected Nellore lines. A: NeS (left) and NeC (right); B: NeS (right) and NeC (left) (Institute of Animal Science, 2020)
.

2.2 Genomic datasets

A total of 928 Nellore animals, including 625 males and 303 females, were genotyped with the Illumina BovineHD BeadChip (HD, Illumina Inc., San Diego, CA, USA; n = 770) or GeneSeek Genomic Profiler 50K (50K, GeneSeek Inc., Lincoln, NE, USA; n = 158) SNP panels (Additional file 5). The HD and 50K panels contained 777,962 and 54,791 SNPs, respectively, distributed throughout the genome. The mean

distance between markers in the HD panel was approximately equal to 3.43±4.4 kilobases (Kb), while in the 50K panel, it was 49.2±99.1 Kb. To ensure genomic data quality, non-autosomal SNPs, SNPs with an unknown position, and SNPs with a GenCall score below 0.15 were removed during the quality control. After the quality control process, 734,593 and 51,613 SNPs remained in the HD and 50K panels, respectively, for subsequent analyses.

2.3 Copy number variation identification

The CNV identification was carried out separately for each SNP panel dataset using the PennCNV.1.0.5 software (Wang et al., 2007).This software integrates Log R Ratio (LRR) and B Allele Frequency (BAF) data on a per-sample basis into a hidden Markov model to determine the number of copies and genotypes of each CNV. LRR measures the total signal intensity, while BAF measures the proportion of the B allele in each sample. The population frequency of the B allele was calculated using the BAF value of each SNP in all samples. Furthermore, the LRR values were adjusted for the guanine-cytosine content at 500 kb upstream and downstream of each SNP based on a regression model (Diskin et al., 2008). This correction aims to reduce waviness that may result from the correlation between LRR and guanine-cytosine content in genomic regions, which could interfere with CNV detection.

Following CNV calling, a sample-based quality control process was implemented. This quality control step entailed the removal of CNVs with a BAF drift of less than 0.01, a standard deviation of LRR exceeding 0.30, a minimum length of 1,000 bp, a maximum length of 5,000,000 bp, and GC wave factor less than 0.05 (after genomic wave correction based on guanine-cytosine content). CNVs with less than three consecutive SNPs were also discarded. After this quality control, 883 animals and 14,914 CNVs (14,391 from the HD panel and 523 from the 50K panel) remained for further analyses. The CNVs identified were categorized and separated into the three distinct selection lines. This segregation led to the creation of distinct CNV datasets for each line, which were then utilized for conducting individual line-specific analyses. This approach enabled a thorough evaluation of CNVs within each selection line, providing valuable insights into the genetic diversity and potential functional significance of CNVs in these populations.

2.4 Copy number variation regions identification

The CNVR were defined by grouping CNVs that had at least 1 bp overlap using the mergeBed option of the BEDtools suite tool (Quinlan and Hall, 2010). This approach was applied in two contexts: across the entire population and within the specific selection lines being studied. CNVRs were classified as “loss” when an animal displayed a region with a loss of a chromosomal segment in comparison to the reference genome (deletions), “gain” for repeated chromosomal regions (duplications), and “mixed” when both loss and gain were identified within the same genomic region. Furthermore, CNVRs that were present in at least 10% of each line were identified. These identified CNVs were subsequently compared across the lines, enabling the distinction of regions that were exclusive to each line and regions that were shared among the three lines.

2.5 Gene annotation and functional analyses

The CNVRs exclusive to each line were used for annotation purposes. The gene and QTL annotation in these regions were performed using the GALLO package (Fonseca et al., 2020), utilizing annotated data for Bos taurus sourced from the Ensembl database (www.ensembl.org/Bos_taurus/Info/Index) and the reference genome ARS-UCD1.2 (Rosen et al., 2020). Additionally, the Cattle QTL database (www.animalgenome.org/cgi-bin/QTLdb/BT/index) was used as a resource for obtaining previously-reported QTL information. The Database for Annotation, Visualization and Integrated Discovery (DAVID; version 6.8, (Dennis et al., 2003)) was used for conducting Gene Ontology (GO) and KEGG pathway enrichment ($p<0.05$) analyses to identify biological processes, molecular functions, cellular components, and biological pathways associated with the positional candidate genes identified.

3 RESULTS

Table 1 presents descriptive statistics of all the animals from the three selection lines included in this study. The NeT line comprises the largest number of animals, followed by NeS and NeC. W378 ranged from 298 kg for NeC to 382 kg for NeS. In the case of W550, NeT had the highest average weight (363±28 kg). Furthermore, the

NeC line had the lowest average RFI (-0.112±0.53), followed by NeS (-0.032±0.61) and NeT (0.032±0.60).

Table 1. Descriptive statistics for Nellore Control, Nellore Selection, and Nellore Traditional

Selection Line[a]	Number of animals	Trait[b] W378 (kg)	W550 (kg)	RFI (kg dry matter/day)
NeC	114	298±41	266±22	-0.112±0.53
NeS	245	382±46	336±29	-0.032±0.61
NeT	569	370±46	363±28	0.032±0.60

[a]NeC: Nellore Control; NeS: Nellore Selection; NeT: Nellore Traditional

[b]W378: Yearling weight adjusted to 378 days; W550: Yearling weight adjusted to 550 days; RFI: Residual feed intake

3.1 Copy number variation and CNVR detection for the Nellore population

Initially, 20,259 CNVs were identified in 922 animals. After quality control, 14,914 CNVs located on autosomal chromosomes of 883 animals remained for further analyses, with an average of 16 CNVs per animal (range: 1 to 45). Among these identified CNVs, 3,680 were categorized as losses and 11,234 as gains. The length of the CNVs varied from 1,216 bp to 1,119,208 bp, with an average length of 75,632±100,827 bp. Notably, CNVs were detected on all autosomal chromosomes, non-uniformly distributed across the genome.

The 14,914 CNVs that remained after quality control were used to infer CNVRs by merging CNVs with at least a 1 bp overlap. This resulted in the identification of 1,884 CNVRs, with an average CNVR length of 40,887±104,812 bp (range: 1,215 to 1,807,286 bp). Among these CNVRs, 400 of them were associated with genome losses, 1,412 with gains, and 72 with a mixed pattern, where the same chromosomal segment exhibited both deletion and duplication in the population. The number and proportion of chromosomes covered by CNVRs varied considerably (Table 2). BTA1 had the highest number of CNVRs (n = 181), covering 4.03% of the chromosome, while BTA12 had the highest coverage of a chromosome sequence (7.94%) with 107 CNVRs. In contrast, BTA25 had the lowest number of CNVR (n = 23) and BTA24 had the lowest coverage of a chromosome sequence at 0.87%. In total, the CNVRs identified in this study covered 77,031,673 bp of the autosomal genome sequence, which corresponds to approximately 3.09% of the cattle genome size.

Table 2. Chromosome distribution of all 1,884 copy number variation regions (CNVRs) detected in the Nellore cattle genome.

Chr[a]	Chr length (bp)	CNVR number	CNVR length (bp)	%[b]
BTA1	158,534,110	181	4,812,669	3.03
BTA2	136,231,102	88	3,278,939	2.41
BTA3	121,005,158	69	2,877,284	2.38
BTA4	120,000,601	94	3,077,289	2.56
BTA5	120,089,316	78	3,499,701	2.91
BTA6	117,806,340	108	5,347,225	4.54
BTA7	110,682,743	98	4,258,382	3.85
BTA8	113,319,770	82	2,149,785	1.89
BTA9	105,454,467	83	2,801,410	2.66
BTA10	103,308,737	61	4,176,247	4.04
BTA11	106,982,474	53	3,057,729	2.86
BTA12	87,216,183	107	6,927,991	7.94
BTA13	83,472,345	43	1,438,060	1.72
BTA14	82,403,003	82	3,206,441	3.89
BTA15	85,007,780	82	4,634,544	5.45
BTA16	81,013,979	66	2,632,004	3.25
BTA17	73,167,244	58	2,707,483	3.70
BTA18	65,820,629	43	996,000	1.51
BTA19	63,449,741	46	1,420,441	2.24
BTA20	71,974,595	49	2,085,746	2.89
BTA21	69,862,954	55	1,898,712	2.72
BTA22	60,773,035	30	700,159	1.15
BTA23	52,498,615	27	1,104,129	2.10
BTA24	62,317,253	32	541,048	0.87
BTA25	42,350,435	23	1,986,333	4.69
BTA26	51,992,305	39	1,278,351	2.46
BTA27	45,612,108	34	740,213	1.62
BTA28	45,940,150	29	1,102,340	2.39
BTA29	51,098,607	44	2,295,018	4.49
Total	2,489,385,779	1,884	77,031,673	3.09

[a]Chromosome
[b]Percentage of the chromosome covered by the CNVRs

A noteworthy CNVR was identified in 847 animals, encompassing approximately 90% of the studied population (928 animals). This particular mixed type CNVR is located on BTA7, spanning a length of 1,133,904 bp. The gene content of this CNVR was thoroughly investigated, revealing an overlapped with a total of 62 known genes (Additional file 6).

3.2 Copy number variation and CNVR detection by selection line

The 14,914 identified CNVs were categorized based on their respective selection lines, resulting in 1,510 CNVs in NeC animals, 3,899 CNVs in NeS, and 9,448 CNVs in NeT. The average CNV length were similar across the three selection lines, ranging from 71,886±97,489 bp in NeC to 78,724±102,183 bp in NeS. In all three lines, the number of loss type CNVs exceed that of gain CNVs, and the average number of CNVs per animal were 13.9, 16.3, and 17.6 for NeC, NeS, and NeT, respectively. Detailed information about the CNVs per selection line after the quality control can be found in Table 3.

Table 3. Copy number variation description per selection line

Events	Number	Min (bp)	Mean (bp)	Max (bp)	Min SNPs	Mean SNPs	Max SNPSs	Mean CNVs per animal
NeC (n = 108 animals)								
Loss	321	1,359	53,322	1,029,591	3	7.74	125	
Gain	1,189	1,871	76,897	794,561	3	17.1	160	13.9
Total	1,510	1,359	71,886	1,029,591	3	15.1	160	
NeS (n = 239 animals)								
Loss	1,023	1,359	85,879	766,200	3	9.48	64	
Gain	2,876	1,454	76,179	883,985	3	17	175	16.3
Total	3,899	1,359	78,724	883,985	3	15	175	
NeT (n = 536 animals)								
Loss	2,313	1,274	73,236	739,884	3	8.78	90	
Gain	7,135	1,216	75,292	1,119,208	3	16.9	238	17.6
Total	9,448	1,216	74,789	1,119,208	3	14.9	238	

NeC: Nellore Control; NeS: Nellore Selection; NeT: Nellore Traditional

The CNVRs were randomly distributed across the chromosomes of the three cattle lines (Fig. 2). NeT had the highest number of CNVRs (n = 1,493), followed by NeS (n = 823) and NeC (n = 482). Among the three lines, BTA1 had the largest number of CNVRs, with 34 CNVRs identified in NeC, 81 in NeS, and 130 in NeT. On the other hand, BTA24 had the lowest CNVR count in both the NeC and NeS lines, with seven

CNVRs in each line. NeT's lowest CNVR count was observed on BTA25, with a total of 18 CNVRs. The CNVR coverage in the genomes of NeC, NeS, and NeT summed up to 23,449,890 bp, 40,175,556 bp, and 63,212,273 bp, respectively. This represents 0.94%, 1.61%, and 2.54% of the bovine autosomal genome for NeC, NeS, and NeT, respectively.

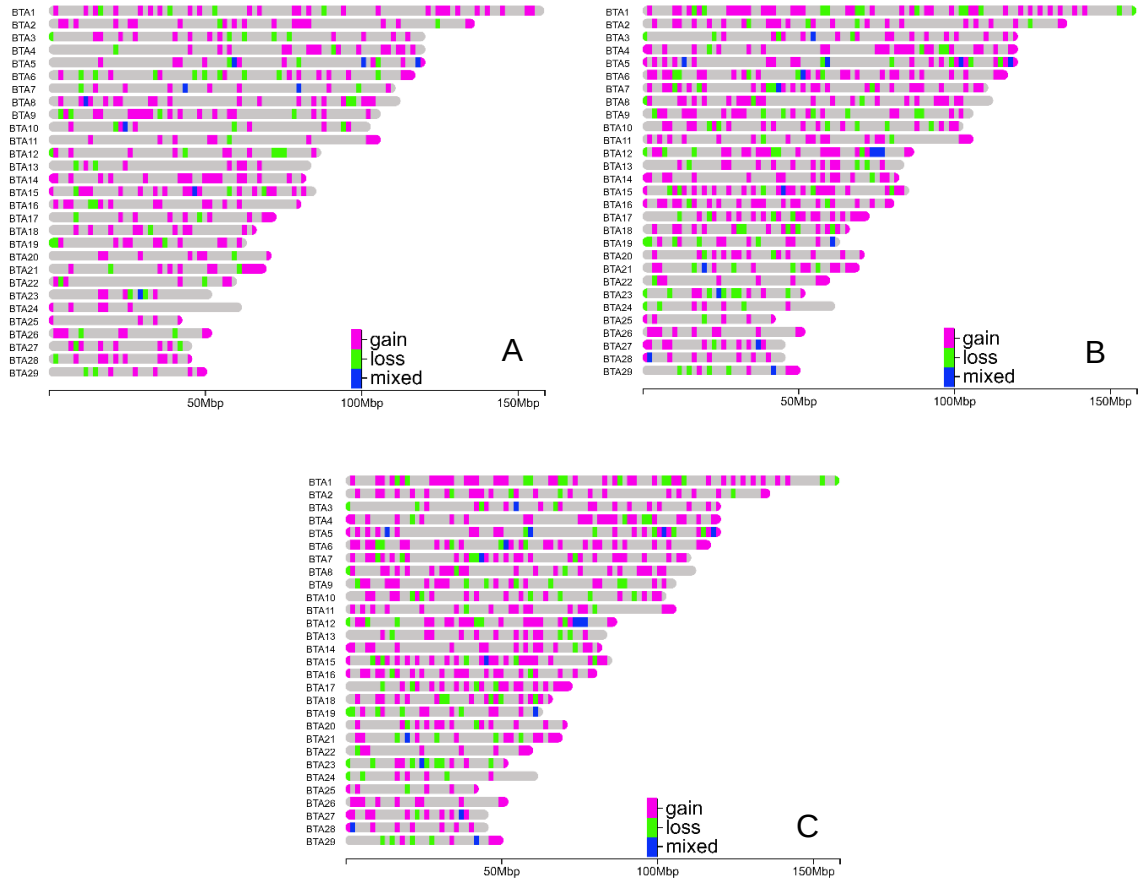

Figure 2: Distribution of copy number variation regions (deletions, duplications, and mixed type) by chromosome and selection line (A: Nellore Control (NeC); B: Nellore Selection (NeS); C: Nellore Traditional (NeT)).

3.3 Common and exclusive CNVRs in the Nellore lines and gene annotation

Twenty-five CNVRs, consisting of 6 losses, 4 gains, and 15 mixed CNVRs, were identified in at least 10% of the NeC animals. In the NeS line, 32 CNVRs were observed, including 3 losses, 17 gains, and 12 mixed CNVRs. In the NeT line, 33 CNVRs were identified, with 4 losses, 18 gains, and 11 mixed CNVRs. The average

length of these CNVRs was 283,307±283,739 bp for NeC, 355,917±290,815 bp for NeS, and 381,594±354,594 bp for NeT. Interestingly, two CNVRs were found to be common across all three lines. Additionally, there were 18 regions shared between NeC and NeS, 18 regions shared between NeC and NeT, and 29 regions shared between NeS and NeT, as illustrated in Fig. 3.

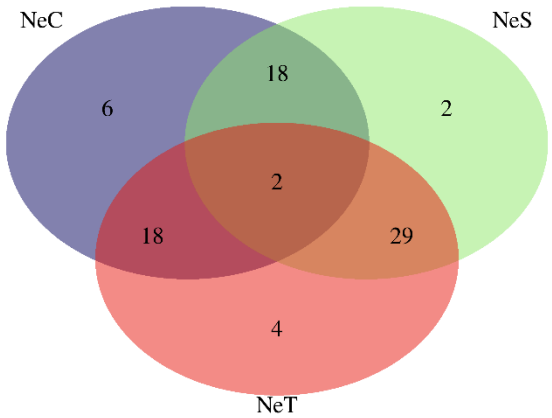

Fig. 3: Venn Diagram for the copy number variation region (CNVR) present in at least 10% of animals from Nellore Control (NeC), Nellore Selection (NeS), and Nellore Traditional (NeT).

The two regions that were identified as common to all three lines overlapped with a total of 11 genes (Table 4).

Table 4. Description of the copy number variation regions identified in common between the three selection lines

CNVR	BTA[a]	Start	End	Type	Gene Ensembl ID	Gene ID
CNVR1	5	59,421,039	59,650,955	Mixed	*ENSBTAG00000052799*	*OR9K2I*
					ENSBTAG00000054478	*OR9K2H*
					ENSBTAG00000038567	*OR9K2K*
					ENSBTAG00000045722	
					ENSBTAG00000050988	*OR9K2C*
					ENSBTAG00000054855	
CNVR2	6	113,462,586	113,603,960	Gain	*ENSBTAG00000005493*	*TBC1D14*
					ENSBTAG00000047810	*CCDC96*
					ENSBTAG00000010185	*TADA2B*
					ENSBTAG00000010181	*GRPEL1*
					ENSBTAG00000044374	*bta-mir-2453*

[a]Bos taurus autosomes chromosomes

Regarding the esclusive regions, there were 6 regions identified for the NeC line, 2 regions for the NeS line, and 4 regions for the NeT line. Among the 6 exclusive CNVRs in the NeC line, there were 3 loss regions and 3 gain regions, distributed across 6 chromosomes, with an average length of 91,745±119,203 bp. Out of the 6 CNVRs, 3 of them overlapped with a total of 16 genes (Table 5).

Table 5. Description of the copy number variation regions identified exclusively in Nellore Control line

CNVR	BTA[a]	Start	End	Type	Gene Ensembl ID	Gene ID
CNVR3	3	25,683	82,937	Loss	*ENSBTAG00000047047*	*OR5W25P*
					ENSBTAG00000011229	*TMUB1*
					ENSBTAG00000048379	
					ENSBTAG00000008468	*AGAP3*
					ENSBTAG00000016752	*ASB10*
					ENSBTAG00000011120	*GBX1*
CNVR4	4	113,657,341	113,819,352	Gain	*ENSBTAG00000018838*	*IQCA1L*
					ENSBTAG00000024204	*H2BK1*
					ENSBTAG00000000607	*ABCF2*
					ENSBTAG00000014371	*CHPF2*
					ENSBTAG00000036355	*bta-mir-671*
					ENSBTAG00000014372	*SMARCD3*
					ENSBTAG00000020815	*UHRF2*
CNVR5	8	38,362,326	38,537,300	Gain	*ENSBTAG00000011161*	
					ENSBTAG00000011160	*TPD52L3*
					ENSBTAG00000018347	*IL33*
CNVR6	9	4,374,671	4,386,831	Loss		
CNVR7	13	7,963,451	8,013,473	Loss		
CNVR8	18	18,822,010	18,843,543	Gain		

[a]Bos taurus autosomal chromosomes

In the case of the NeS line, there were two exclusive CNVRs, and both of these regions were classified as mixed, indicating both losses and gains. These CNVRs were found on BTA12, with an average length of approximately 812,093±147,962 bp each. Notably, both CNVRs were identified within genomic gaps in the reference genome assembly and overlapped with a total of 8 genes, as shown in Table 6.

Table 6. Description of the copy number variation regions identified exclusively in Nellore Selection line

CNVR	BTA[a]	Start	End	Type	Gene Ensembl ID	Gene ID
					ENSBTAG00000026070	
CNVR9	12	71,274,536	72,191,254	Mixed	*ENSBTAG00000052990*	
					ENSBTAG00000047181	
CNVR10	12	75,238,779	75,946,247	Mixed	*ENSBTAG00000000146*	*FARP1*

					ENSBTAG00000045262	*U6*
					ENSBTAG00000032234	*STK24*
					ENSBTAG00000004440	*SLC15A1*
					ENSBTAG00000010395	*DOCK9*

[a]*Bos taurus* autosomes chromosomes

In the NeT line, there were two exclusive loss regions and two exclusive gain regions, distributed across four different chromosomes. The average length of these exclusive CNVRs was approximately 233,107±279,300 bp. Among these regions, three overlapped with a total of 21 genes, as outlined in Table 7.

Table 7. Description of the copy number variation regions identified exclusively in Nellore Traditional line

CNVR	BTA[a]	Start	End	Type	Gene Ensembl ID	Gene ID
CNVR11	1	15,986,400	15,996,851	Loss		
CNVR12	6	114,086,816	114,716,539	Gain	*ENSBTAG00000027434*	*SORCS2*
					ENSBTAG00000012464	*AFAP1*
					ENSBTAG00000015926	*ABLIM2*
					ENSBTAG00000036422	*bta-mir-95*
					ENSBTAG00000047876	*SH3TC1*
					ENSBTAG00000047613	*HTRA3*
					ENSBTAG00000055218	
					ENSBTAG00000004893	*ACOX3*
					ENSBTAG00000017277	*RAB36*
					ENSBTAG00000036113	*RSPH14*
					ENSBTAG00000047664	*GNAZ*
CNVR13	17	71,806,249	72,029,884	Gain	*ENSBTAG00000023007*	
					ENSBTAG00000055118	*VPREB1*
					ENSBTAG00000052191	*VPREB1*
					ENSBTAG00000053372	*TOP3B*
					ENSBTAG00000006938	*PPM1F*
					ENSBTAG00000010312	*MAPK1*
					ENSBTAG00000042240	*U6*
CNVR14	21	20,062,038	20,130,658	Loss	*ENSBTAG00000001526*	
					ENSBTAG00000053783	
					ENSBTAG00000050647	

[a]*Bos taurus* autosomal chromosomes

3.4 Gene ontology

The genes that overlapped with exclusive CNVRs from each selection line were included in the gene ontology analyses. While the functional analysis of genes

conducted using the DAVID software did not yield significant results for the NeC and NeT cattle lines, a closer investigation of the functions of biological processes associated with these genes revealed their involvement in specific pathways. These genes were involved in pathways such as thermogenesis (NeC), metabolism (NeC), and protein digestion and 67ellore67on (NeS). For the NeT line, functional enrichment was observed in the cellular component category, specifically for the term GO:0016020 – Integral component of membrane. Genes within the exclusive regions of NeT also contribute to various biological processes, including positive regulation of growth (GO:0045927), positive regulation of gene expression (GO:0010628), and insulin-like growth factor receptor signaling pathway (GO:0048009). Furthermore, these genes play an important role in metabolic pathways related to growth hormone synthesis and secretion.

4 DISCUSSION

The Nellore experimental breeding program from IZ has gained national recognition and makes substantial contributions to the field of beef cattle breeding and genetics. The differential selection among the three selection lines has enabled in-depth studies of weight-related traits and feed efficiency, providing essential insights into the genetic information of livestock (e.g. Noely et al., n.d.; Ayres et al., 2010; Cardoso et al., 2014, 2018).

The NeC line had the lowest average for W378 and W550, Ih was expected since this line is characterized as stabilizing selection with an average YW close to the weight at the start of the breeding program. The NeS line exhibited the highest mean for W378, which aligns with this line’s selection focus on increased post-weaning

weight, highlighting the success of the breeding program in attaining its specific breeding objective. Considering the substantial difference in the average of W378 and W550 between lines, the three lines provide a great opportunity to identify genomic regions altered by selection. The NeC animals can be used as a reference point to compare the lines and understand the genetic progress achieved over time and the mechanisms involved in the phenotypic expression of the selected traits. NeC exhibited the lowest phenotypic average for RFI (more efficient), followed by NeS and NeT. However, it is important to highlight that the standard deviations (SD) were high for these averages, and these values are representative of only a small subset of Nellore animals, thus not accurately reflecting the population mean of each line.

4.1 Copy number variation and CNVR detection for the Nellore population

Numerous studies have previously investigated the distribution and characterization of CNVs and CNVRs within the cattle genome (e.g., Fadista et al., 2010; Liu et al., 2010; Hou et al., 2012; Peripolli et al., 2023), each yielding diverse findings and insights about the presence and the function of these variants in the cattle genome. For instance, Silva et al. (2016) identified 68,007 CNVs and 7,319 CNVRs in a population of 1,509 Nellore animals. Additionally, Upadhyay et al. (2017) reported 9,944 CNVs and 923 CNVRs in 149 European cattle, while Lemos et al. (2018) identified 195,873 CNVs and 9,805 CNVRs in 3,794 Nellore animals. In a study of Holstein cattle, Butty et al. (2021) found 23,256 CNVs and 1,645 CNVRs. There is a clear notable discrepancy in the number of CNVs and CNVRs between the previously reported study and our current findings. However, each study utilized different SNP panel densities, quality control thresholds, and sample sizes, which may contribute to

these differences (Fadista et al., 2010; Hou et al., 2012). Furthermore, the implementation of quality control measures, accounting for batch effects, addressing population stratification, managing experimental variations, and the robustness of statistical models can all impact the detection and accuracy of CNVs (Dellinger et al., 2010). Therefore, any comparisons between studies should be made cautiously, considering all these factors described above. The proportion of the genome covered by CNVRs (3.09%) falls within the range reported in the literature. Previous studies have reported values ranging from 0.68% to 13.0% in cattle populations (Fadista et al., 2010; Zhou et al., 2016; Lemos et al., 2018).

The distribution of CNVRs across chromosomes did not follow any clear pattern, and BTA1 exhibited the highest number of CNVRs (n = 181), a trend also noted by Silva et al. (2016). Although no particular pattern or correlation was discerned, this result may be associated with the fact that BTA1 is the largest chromosome in the cattle genome. Another interesting finding in the present study was the identification of a CNVR present in 90% of the individuals included in the study. This observation suggests the existence of a region that has remained conserved within this Nellore population over time, highlighting potential genetic stability or selection pressure within this genomic region. This might also reflect the fact that the reference genome used was based on a taurine (Bos taurus taurus) animal while Nellore is a different subspecies (Bos taurus indicus). This highlights the need of developing pangenomes in cattle (e.g., (Zhou et al., 2022).

This common region observed in 90% of the studied population is a gene-rich region, overlapping with 62 annotated genes. Several genes associated with male and female reproductive traits were identified, including THEG (Nayernia et al., 1999;

Mannan et al., 2003), FGF22 (Castilho et al., 2017, 2019), KISS1R (D'Occhio et al., 2020; Singh et al., 2020), and ARID3A (Yang et al., 2018). Furthermore, genes linked to the immune system such as AZU1 (Xu et al., 2018; Verardo et al., 2021) and ELANE (Cassatella et al., 2019; Verardo et al., 2021) were also identified. The CFD gene was also previously associated with fat accumulation (Wang et al., 2023) and overlapped with the region cited above.

4.2 Copy number variation and CNVR detection by line

While previous studies have identified CNVs within and between cattle populations, our study is one of the first endeavors to investigate the population-genetic properties in three closed Nellore lines that were differentially selected for high post-weaning weight and RFI. Substantial differences in CNV counts were identified among the three lines studied. NeS and NeT exhibited a relatively high number of CNVs and CNVs per individual compared to NeC, along with a high chromosome coverage by CNVRs. The results in this study are based on a population of 928 animals with an uneven distribution among the lines. However, for the purpose of comparison and confirmation of the results, CNVs and CNVRs were also identified considering a reduced number of animals with an equal number of samples per line (n = 114). Remarkably, the results remained consistent with the same pattern (results not shown), where animals from the NeS and NeT lines exhibited a higher number of CNVs and CNVRs.

The results obtained align with the previous expectations and are supported by the findings from Upadhyay et al. (2017), who reported that the population size, gene flow, and the selection process in a population can contribute to differential CNV

abundance among populations. Selection for a specific trait can indeed lead to changes in allele frequencies within the population, resulting in alterations within the cattle genome and giving rise to significant phenotypic and genetic variability (Bickhart et al., 2016).Furthermore, the present findings are consistent with the results of Strillacci et al. (2018), who reported CNVs and CNVRs within the genome of Valdostana Red Pied cattle, an Italian dual-purpose cattle population that did not undergo strong artificial selection for production traits. Following the CNV identification, the authors conducted a comparative analysis of the CNVs detected in their study with those available from published research in the Italian Brown Swiss and Mexican Holstein populations. Their findings revealed the presence of unique and highly differentiated CNVs, leading to the conclusion that directional selection occurring within a population exerts a significant impact on the genome in terms of CNVs.

Despite differences in the numbers of CNVs identified, all three selection lines exhibited a higher frequency of duplications than deletions. This observation aligns with findings from previous studies, such as Laseca et al. (2022) in horses, Ladeira et al. (2022) in sheep, and Liu et al. (2010) in cattle. While there is no clear pattern of duplication and deletion distribution across the genome, duplications are more likely to occur in CNVs with greater lengths (Locke et al., 2006). Furthermore, according to Amos et al. (2003) and Conrad et al. (2006), deletion events may go unnoticed using SNP genotyping methods.

4.3 Gene annotation and gene ontology

The deletion or duplication of genomic regions can have various consequences. The deletion of a genomic region that contains important genes can lead to the loss of

gene function, potentially being associated with diseases, genetic disorders, and reduced fitness (Stenson et al., 2017). Moreover, the duplication of gene-rich regions may also be associated with adaptation (Meredith et al., 2014; Sharma et al., 2018). On the other hand, the duplication of gene-rich regions is typically linked to genetic diversity. Gene duplication is believed to play an important role in evolution and adaptation and may be involved in the development of new gene functions (Zhang, 2003; Magadum et al., 2013; Lallemand et al., 2020). Thus, we identified genes present in exclusive regions for each selection line, which may help elucidate differences between lines and the expression of traits in a selection process.

In the NeC line, an important gene was identified within the CNVR4, located on BTA4. This region was characterized as a gain region and overlapped with the SMARCD3 gene. The SMARCD3 plays a crucial role as a subunit of the SWI/SNF family of proteins, which are known for their helicase and ATPase activities and their capacity to modulate the transcription of specific genes by modifying the chromatin structure surrounding those genes. Additionally, in the NeC line, another important region is the CNVR5 on BTA8, which contains the UHF2 gene. This gene encodes a nuclear protein involved in cell-cycle regulation (Lu and Hallstrom, 2013). The UHF2 gene has been reported to be involved in the regulation of many biological processes, including metabolic pathways, growth, and reproduction (Magoro et al., 2022). These findings suggest a potential link between the SMARCD3 and UHRF2 genes, and critical pathways related to metabolism, growth, and health, warranting further investigation into their roles in cattle populations.

In the NeS line, a region classified as “mixed” was identified on BTA12, which overlapped with the SLC15A1 gene. This gene encodes an intestinal hydrogen peptide

cotransporter and belongs to the solute carrier family 15. It plays an important role in the uptake and digestion of dietary proteins (Liang et al., 1995). Additionally, SLC15A1 has also been associated with the small intestine weight and embryo development in chickens (Zeng et al., 2011; Li et al., 2013). The presence of this gene within the identified CNVR suggests its potential significance in regulating critical processes related to nutrient absorption, intestinal development, and overall growth in cattle.

In NeT line, several exclusive regions that overlapped with important genes were also identified. One of these regions, CNVR13, stood out as a gain region located on BTA17. This region encompasses 10 genes, with particular emphasis on MAPK1. MAPK1 encodes a member of the MAP kinase family. MAP kinases, also known as extracellular signal-regulated kinases, serve as a central hub for integrating multiple biochemical signals and play integral roles in a wide array of cellular processes, including proliferation, differentiation, transcription regulation, and development (Jiang et al., 2011; Liu et al., 2016). Moreover, studies have reported that MAPK10 gene is linked to cell growth in phosphorylation and protein modification process, which are needed for the production of meat in the muscle mechanism (Shin et al., 2014).

Gene ontology analysis is an essential tool for elucidating the functional landscape of genetic elements, as it helps to comprehend and interpret the functions of genes. In the current study, no enrichment of biological processes was observed for the genes identified. This suggests that collectively, they do not participate in any similar biological process, potentially indicating a diverse array of gene functions. However, even though enriched processes were not identified, the genes individually participate in crucial biological processes and pathways. In NeT line, specific terms related to growth were detected, such as GO:0010628, defined as positive regulation

of gene expression, and GO:0045927, defined as positive regulation of growth. These findings suggest that while there may not be overall enriched processes, the individual genes within these regions may collectively contribute to the regulation of vital biological processes associated with growth and gene expression.

5 CONCLUSIONS

We described a variability of CNVs and CNVRs within three Nellore lines differentially selected for YW and RFI. Through the gene annotation and gene ontology analyses of the exclusive CNVRs identified in each line, specific genes and biological processes involved in the expression of growth and feed efficiency traits were found. These results not only show the structural differences present in the genomes of animals from the three studied selection lines but also indicate that these variations may account for a portion of the observed differences among them. These findings provide valuable insights for future research and breeding strategies to enhance these important traits in Nellore cattle populations.

REFERENCES

Amos, C. I., Shete, S., Chen, J., and Yu, R. K. (2003). Positional identification of microdeletions with genetic markers. In Human Heredity, 107–118. Doi: 10.1159/000073738.

Antunes de Lemos, M. V., Berton, M. P., Ferreira de Camargo, G. M., Peripolli, E., de Oliveira Silva, R. M., Ferreira Olivieri, B., et al. (2018). Copy number variation regions in Nellore cattle: Evidences of environment adaptation. Livest Sci 207, 51–58. Doi: 10.1016/j.livsci.2017.11.008.

Ayres, D. R., Souza, F. R. P., Mercadante, M. E. Z., Fonseca, L. F. S., Tonhati, H., Cyrillo, J. N. S. G., et al. (2010). Evaluation of tfam and fabp4 gene polymorphisms in three lines of 75ellore cattle selected for growth. Genetics and Molecular Research 9, 2050–2059. Doi: 10.4238/vol9-4gmr850.

Benfica, L. F., Sakamoto, L. S., Magalhães, A. F. B., De Oliveira, M. H. V., De Albuquerque, L. G., Cavalheiro, R., et al. (2020). Genetic association among feeding behavior, feed efficiency, and growth traits in growing indicine cattle. J Anim Sci 98. Doi: 10.1093/JAS/SKAA350.

Bickhart, D. M., Xu, L., Hutchison, J. L., Cole, J. B., Null, D. J., Schroeder, S. G., et al. (2016). Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. DNA Research 23, 253–262. Doi: 10.1093/dnares/dsw013.

Butty, A. M., Chud, T. C. S., Cardoso, D. F., Lopes, L. S. F., Miglior, F., Schenkel, F. S., et al. (2021). Genome-wide association study between copy number variants and hoof health traits in Holstein dairy cattle. J Dairy Sci 104, 8050–8061. Doi: 10.3168/jds.2020-19879.

Cardoso, D. F., De Albuquerque, L. G., Reimer, C., Qanbari, S., Erbe, M., Do Nascimento, A. V., et al. (2018). Genome-wide scan reveals population stratification and footprints of recent selection in Nelore cattle. Genetics Selection Evolution 50. Doi: 10.1186/s12711-018-0381-2.

Cardoso, D. F., de Souza, F. R. P., de Camargo, G. M. F., Fonseca, P. D. da S., Fonseca, L. F. S., Braz, C. U., et al. (2014). Polymorphism analysis in genes of the somatotropic axis in Nellore cattle selected for growth. Gene 545, 215–219. Doi: 10.1016/j.gene.2014.05.033.

Cassatella, M. A., Östberg, N. K., Tamassia, N., and Soehnlein, O. (2019). Biological Roles of Neutrophil-Derived Granule Proteins and Cytokines. Trends Immunol 40, 648–664. Doi: 10.1016/j.it.2019.05.003.

Castilho, A. C. S., Dalanezi, F. M., Franchi, F. F., Price, C. A., Ferreira, J. C. P., Trevisol, E., et al. (2019). Expression of fibroblast growth factor 22 (FGF22) and its

receptor, FGFR1B, during development and regression of bovine corpus luteum. Theriogenology 125, 1–5. Doi: 10.1016/j.theriogenology.2018.09.024.

Castilho, A. C. S., Price, C. A., Dalanezi, F., Ereno, R. L., Machado, M. F., Barros, C. M., et al. (2017). Evidence that fibroblast growth factor 10 plays a role in follicle selection in cattle. Reprod Fertil Dev 29, 234–243. Doi: 10.1071/RD15017.

Ceacero, T. M., Mercadante, M. E. Z., Cyrillo, J. N. D. S. G., Canesin, R. C., Bonilha, S. F. M., and De Albuquerque, L. G. (2016). Phenotypic and genetic correlations of feed efficiency traits with growth and carcass traits in 76ellore cattle selected for postweaning weight. PloS One 11. Doi: 10.1371/journal.pone.0161366.

Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E., and Pritchard, J. K. (2006). A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38, 75–81. Doi: 10.1038/ng1697.

Da Silva, J. M., Giachetto, P. F., da Silva, L. O., Cintra, L. C., Paiva, S. R., Yamagishi, M. E. B., et al. (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. Doi: 10.1186/s12864-016-2752-9.

Dellinger, A. E., Saw, S. M., Goh, L. K., Seielstad, M., Young, T. L., and Li, Y. J. (2010). Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. Nucleic Acids Res 38. Doi: 10.1093/nar/gkq040.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, C., et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Available at: http://dot.ped.med.umich.edu:2000/.

D'Occhio, M. J., Campanile, G., and Baruselli, P. S. (2020). Peripheral action of kisspeptin at reproductive tissues-role in ovarian function and embryo implantation and relevance to assisted reproductive technology in livestock: A review. Biol Reprod 103, 1157–1170. Doi: 10.1093/biolre/ioaa135.

Fadista, J., Thomsen, B., Holm, L.-E., and Bendixen, C. (2010). Copy number variation in the bovine genome. Available at: http://www.biomedcentral.com/1471-2164/11/284.

Fernandes Júnior, G. A., Peripolli, E., Schmidt, P. I., Campos, G. S., Mota, L. F. M., Mercadante, M. E. Z., et al. (2022). Current applications and perspectives of genomic selection in Bos indicus (Nellore) cattle. Livest Sci 263. Doi: 10.1016/j.livsci.2022.105001.

Fonseca, P. A. S., Suárez-Veja, A., Marras, G., and Cánovas, Á. (2020). GALLO: An R package for genomic annotation and integration of multiple data sources

in livestock for positional candidate loci. Gigascience 9. Doi: 10.1093/gigascience/giaa149.

Henrichsen, C. N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., et al. (2009). Segmental copy number variation shapes tissue transcriptomes. Nat Genet 41, 424–429. Doi: 10.1038/ng.345.

Hou, Y., Bickhart, D. M., Chung, H., Hutchison, J. L., Norman, H. D., Connor, E. E., et al. (2012). Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. Funct Integr Genomics 12, 717–723. Doi: 10.1007/s10142-012-0295-y.

Jiang, Q., Ho, Y. Y., Hao, L., Berrios, C. N., and Chakravarti, A. (2011). Copy number variants in candidate genes are genetic modifiers of Hirschsprung disease. PloS One 6. Doi: 10.1371/journal.pone.0021219.

Ladeira, G. C., Pilonetto, F., Fernandes, A. C., Bóscollo, P. P., Dauria, B. D., Titto, C. G., et al. (2022). CNV detection and their association with growth, efficiency and carcass traits in Santa Inês sheep. Journal of Animal Breeding and Genetics 139, 476–487. Doi: 10.1111/jbg.12671.

Lallemand, T., Leduc, M., Landès, C., Rizzon, C., and Lerat, E. (2020). An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. Genes (Basel) 11, 1–40. Doi: 10.3390/genes11091046.

Laseca, N., Molina, A., Valera, M., Antonini, A., and Demyda-Peyrás, S. (2022). Copy Number Variation (CNV): A New Genomic Insight in Horses. Animals 12. Doi: 10.3390/ani12111435.

Li, X. guang, Chen, X. ling, and Wang, X. qi (2013). Changes in relative organ weights and intestinal transporter gene expression in embryos from white Plymouth Rock and WENS Yellow Feather Chickens. Comparative Biochemistry and Physiology – A Molecular and Integrative Physiology 164, 368–375. Doi: 10.1016/j.cbpa.2012.11.016.

Liang, R., Fei, Y. J., Prasad, P. D., Ramamoorthy, S., Han, H., Yang-Feng, T. L., et al. (1995). Human intestinal H+/peptide cotransporter. Cloning, functional expression, and chromosomal localization. Journal of Biological Chemistry 270, 6456–6463. Doi: 10.1074/jbc.270.12.6456.

Liu, G. E., Hou, Y., Zhu, B., Cardone, M. F., Jiang, L., Cellamare, A., et al. (2010). Analysis of copy number variations among diverse cattle breeds. Genome Res 20, 693–703. Doi: 10.1101/gr.105403.110.

Liu, M., Li, B., Huang, Y., Yang, M., Lan, X., Lei, C., et al. (2016). Copy number variation of bovine MAPK10 modulates the transcriptional activity and affects growth traits. Livest Sci 194, 44–50. Doi: 10.1016/j.livsci.2016.09.014.

Livestock and Poultry: World Markets and Trade EU Share of Pork, Beef, and Chicken Global Exports Continues Decline (n.d.). Available at: https://public.govdelivery.com/accounts/USDAFAS/subscriber/new.

Locke, D. P., Sharp, A. J., McCarroll, S. A., McGrath, S. D., Newman, T. L., Cheng, Z., et al. (2006). Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet 79, 275–290. Doi: 10.1086/505653.

Lu, H., and Hallstrom, T. C. (2013). The nuclear protein UHRF2 is a direct target of the transcription factor E2F1 in the induction of apoptosis. Journal of Biological Chemistry 288, 23833–23843. Doi: 10.1074/jbc.M112.447276.

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., and And Ravikesavan, R. (2013). Gene duplication as a major force in evolution.

Magoro, A. M., Mtileni, B., Hadebe, K., and Zwane, A. (2022). Assessment of Genetic Diversity and Conservation in South African Indigenous Goat Ecotypes: A Review. Animals 12. Doi: 10.3390/ani12233353.

Mannan, A. U., Nayernia, K., Mueller, C., Burfeind, P., Adham, I. M., and Enge, W. (2003). Male mice lacking the Theg (testicular haploid expressed gene) protein undergo normal spermatogenesis and are fertile. Biol Reprod 69, 788–796. Doi: 10.1095/biolreprod.103.017400.

Mercadante, M. E. Z., Packer, I. U., Razook, A. G., Cyrillo, J. N. S. G., and Figueiredo, L. A. (2003). Direct and correlated responses to selection for yearling weight on reproductive performance of Nelore cows 1. Available at: https://academic.oup.com/jas/article/81/2/376/4789734.

Meredith, R. W., Zhang, G., Thomas, M., Gilbert, P., Jarvis, E. D., and Springer, M. S. (n.d.). Evidence for a single loss of mineralized teeth in the common avian ancestor.

Monteiro, F. M., Mercadante, M. E. Z., Barros, C. M., Satrapa, R. A., Silva, J. A. V., Oliveira, L. Z., et al. (2013). Reproductive tract development and puberty in two lines of Nellore heifers selected for postweaning weight. Theriogenology 80, 10–17. Doi: 10.1016/j.theriogenology.2013.02.013.

Nayernia, K., Von Mering, M. H. P., Kraszucka, K., Burfeind, P., Wehrend, A., Kö, M., et al. (1999). A Novel Testicular Haploid Expressed Gene (THEG) Involved in Mouse Spermatid-Sertoli Cell Interaction 1. Available at: https://academic.oup.com/biolreprod/article/60/6/1488/2741135.

Noely, J., Gonçalves Cyrillo, S., Razook, A. G., Andrade De Figueiredo, L., Martins, L., Neto, B., et al. (n.d.). Efeitos da Seleção para Peso Pós-desmame sobre Medidas Corporais e Perímetro Escrotal de Machos Nelore de Sertãozinho (SP) 1.

Peripolli, E., Stafuzza, N. B., Machado, M. A., do Carmo Panetto, J. C., do Egito, A. A., Baldi, F., et al. (2023). Assessment of copy number variants in three Brazilian locally adapted cattle breeds using whole-genome re-sequencing data. Anim Genet. Doi: 10.1111/age.13298.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. Doi: 10.1093/bioinformatics/btq033.

Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., et al. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience 9. Doi: 10.1093/gigascience/giaa021.

Sharma, V., Hecker, N., Roscito, J. G., Foerster, L., Langer, B. E., and Hiller, M. (2018). A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. Nat Commun 9. Doi: 10.1038/s41467-018-03667-1.

Shin, D. H., Lee, H. J., Cho, S., Kim, H. J., Hwang, J. Y., Lee, C. K., et al. (2014). Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. BMC Genomics 15. Doi: 10.1186/1471-2164-15-240.

Singh, U., Alex, R., Kumar, S., Deb, R., Venkatesan Raja, T., Singhal, S., et al. (2020). Association of bovine KISS1 single nucleotide polymorphisms with reproductive traits in Indian Cattle. Reproduction in Domestic Animals 55, 922–930. Doi: 10.1111/rda.13704.

Stafuzza, N. B., Silva, R. M. D. O., Fragomeni, B. D. O., Masuda, Y., Huang, Y., Gray, K., et al. (2019). A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive. BMC Genomics 20. Doi: 10.1186/s12864-019-5687-0.

Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., et al. (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet 136, 665–677. Doi: 10.1007/s00439-017-1779-6.

Strillacci, M. G., Gorla, E., Cozzi, M. C., Vevey, M., Genova, F., Scienski, K., et al. (2018). A copy number variant scan in the autochthonous Valdostana Red Pied cattle breed and comparison with specialized dairy populations. PloS One 13. Doi: 10.1371/journal.pone.0204669.

Upadhyay, M., da Silva, V. H., Megens, H. J., Visker, M. H. P. W., Ajmone-Marsan, P., Bâlteanu, V. A., et al. (2017). Distribution and functionality of copy number variation across European cattle populations. Front Genet 8. Doi: 10.3389/fgene.2017.00108.

Verardo, L. L., e Silva, F. F., Machado, M. A., do Carmo Panetto, J. C., de Lima Reis Faza, D. R., Otto, P. I., et al. (2021). Genome-Wide Analyses Reveal the Genetic Architecture and Candidate Genes of Indicine, Taurine, Synthetic Crossbreds, and Locally Adapted Cattle in Brazil. Front Genet 12. Doi: 10.3389/fgene.2021.702822.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F. A., et al. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 17, 1665–1674. Doi: 10.1101/gr.6861907.

Wang, L., Gao, P., Li, C., Liu, Q., Yao, Z., Li, Y., et al. (2023). A single-cell atlas of bovine skeletal muscle reveals mechanisms regulating intramuscular adipogenesis and fibrogenesis. J Cachexia Sarcopenia Muscle 14, 2152–2167. Doi: 10.1002/jcsm.13292.

Xu, L., Zhang, W. G., Shen, H. X., Zhang, Y., Zhao, Y. M., Jia, Y. T., et al. (2018). Genome-wide scanning reveals genetic diversity and signatures of selection in Chinese indigenous cattle breeds. Livest Sci 216, 100–108. Doi: 10.1016/j.livsci.2018.08.005.

Yang, C., Wang, J., Liu, J., Sun, Y., Guo, Y., Jiang, Q., et al. (2018). Functional haplotypes of ARID4A affect promoter activity and semen quality of bulls. Anim Reprod Sci 197, 257–267. Doi: 10.1016/j.anireprosci.2018.08.038.

Zeng, P. L., Li, X. G., Wang, X. Q., Zhang, D. X., Shu, G., and Luo, Q. B. (2011). The relationship between gene expression of cationic and neutral amino acid transporters in the small intestine of chick embryos and chick breed, development, sex, and egg amino acid concentration. Poult Sci 90, 2548–2556. Doi: 10.3382/ps.2011-01458.

Zhang, F., Gu, W., Hurles, M. E., and Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10, 451–481. Doi: 10.1146/annurev.genom.9.081307.164217.

Zhang, J. (2003). Evolution by gene duplication: An update. Trends Ecol Evol 18, 292–298. Doi: 10.1016/S0169-5347(03)00033-8.

Zhou, Y., Utsunomiya, Y. T., Xu, L., Hay, E. H. abdel, Bickhart, D. M., Alexandre, P. A., et al. (2016). Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. BMC Genomics 17, 1–9. Doi: 10.1186/s12864-016-2461-4.

Zhou, Y., Yang, L., Han, X., Han, J., Hu, Y., Li, F., et al. (2022). Assembly of a pangenome for global cattle reveals missing sequences and novel structural variations, providing new insights into their diversity and evolutionary history. Genome Res 32, 1585–1601. Doi: 10.1101/gr.276550.122.

ADDITIONAL FILE 1

Table S1. Descriptive statistics of raw and adjusted phenotypes of growth, feed efficiency, and feeding behavior traits

Trait[a]	Nº of animals with phenotype	Phenotype			Adjusted Phenotype		
		Minimum	Mean	Maximum	Minimum	Mean	Maximum
W210 (kg)	1,314	90.1	199.5	320.5	-232	-0.04	72.3
Wsel (kg)	1,306	137	331	512	-163	2.47	139
ADG ($kg.d^{-1}$)	734	0.286	1.054	1.69	-0.689	0.001	0.618
DMI ($kg.d^{-1}$)	734	3.68	8.48	12.6	-3.53	0.002	3.54
RFI ($kg.d^{-1}$)	734	-2.19	-0.007	3.17	-2.21	0.002	2.99
TF ($hour.d^{-1}$)	623	1.012	2.69	8.009	-2.069	0.288	5.56
FF ($visit.d^{-1}$)	623	8.081	24.6	91.8	-21.8	6.018	60.95

[a]W210: weight at 210 days; Wsel: body weight measured at the time of selection; ADG: average daily gain; DMI: dry matter intake; RFI: residual feed intake; TF: time spent on the feed bunk; FF: feed frequency.

ADDITIONAL FILE 2

Table S2. Copy number variation regions of deletion type associated with growth, feed efficiency, and feeding behavior.

CNVR[a]	Chr[b]	Start (bp)	End (bp)	Type	Trait[c]	p_value	Ensembl ID	Gene ID
CNVR1	2	135,110,420	135,653,313	Mixed	FF	$2.743x10^{-6}$	*ENSBTAG00000000684*	*ARHGEF10L*
							ENSBTAG00000048830	
							ENSBTAG00000008579	*RCC2*
							ENSBTAG00000045172	*bta-mir-2358*
							ENSBTAG00000038945	*PADI6*
							ENSBTAG00000012052	*PADI4*
							ENSBTAG00000012043	*PADI3*
							ENSBTAG00000002138	*PADI1*
							ENSBTAG00000043637	*5S_rRNA*
							ENSBTAG00000003403	*PADI2*
							ENSBTAG00000008314	*SDHB*
							ENSBTAG00000008309	*ATP13A2*
							ENSBTAG00000003832	*MFAP2*
							ENSBTAG00000003822	*CROCC*
CNVR2	5	117,080,458	117,820,070	Deletion	FF	$1.134x10^{-9}$	*ENSBTAG00000012291*	*TBC1D22A*
							ENSBTAG00000044449	*bta-mir-2285o-5*
CNVR3	6	116,755,758	117,164,372	Duplication	FF	$3.232x10^{-9}$	*ENSBTAG00000020108*	*LETM1*
							ENSBTAG00000007164	*FGFR3*
							ENSBTAG00000051271	
							ENSBTAG00000011044	*TACC3*
							ENSBTAG00000011042	*TMEM129*
							ENSBTAG00000001328	*SLBP*
							ENSBTAG00000048974	
							ENSBTAG00000048789	*NKX1-1*
							ENSBTAG00000012577	*UVSSA*
							ENSBTAG00000012575	*MAEA*
CNVR4	7	10,092,268	10,174,209	Duplication	FF	$1.092x10^{-10}$	*ENSBTAG00000053498*	

CNVR5	7	42,951,015	43,292,715	Mixed	FF	$8.2x10^{-7}$	*ENSBTAG00000000717*	*PLPP2*
							ENSBTAG00000019767	*MIER2*
							ENSBTAG00000048393	*THEG*
							ENSBTAG00000047560	*C2CD4C*
							ENSBTAG00000016072	*SHC2*
							ENSBTAG00000039372	*ODF3L2*
							ENSBTAG00000050765	
							ENSBTAG00000039355	*MADCAM1*
							ENSBTAG00000010772	*TPGS1*
							ENSBTAG00000002098	*CDC34*
							ENSBTAG00000002100	*GZMM*
							ENSBTAG00000016648	*BSG*
							ENSBTAG00000015050	*HCN2*
							ENSBTAG00000015051	*POLRMT*
							ENSBTAG00000027357	*FGF22*
							ENSBTAG00000014349	*RNF126*
							ENSBTAG00000003018	*FSTL3*
							ENSBTAG00000021616	*PRSS57*
CNVR6	7	43,359,066	43,823,809	Mixed	FF	$1.821x10^{-9}$	*ENSBTAG00000045828*	*PTBP1*
							ENSBTAG00000051609	*PLPPR3*
							ENSBTAG00000045829	*AZU1*
							ENSBTAG00000049418	*U6*
							ENSBTAG00000046105	*PRTN3*
							ENSBTAG00000046188	*ELANE*
							ENSBTAG00000048122	*CFD*
							ENSBTAG00000047217	*MED16*
							ENSBTAG00000053281	*U6*
							ENSBTAG00000046406	*R3HDM4*
							ENSBTAG00000037418	*KISS1R*
							ENSBTAG00000008607	*ARID3A*

							ENSBTAG00000002434	*WDR18*
							ENSBTAG00000025597	*GRIN3B*
							ENSBTAG00000011351	*TMEM259*
							ENSBTAG00000020764	*CNN2*
							ENSBTAG00000020766	*ABCA7*
							ENSBTAG00000020772	*ARHGAP45*
							ENSBTAG00000020776	*POLR2E*
							ENSBTAG00000053003	*GPX4*
							ENSBTAG00000020780	*SBNO2*
							ENSBTAG00000011639	*STK11*
							ENSBTAG00000025233	*CBARP*
							ENSBTAG00000000550	*ATP5F1D*
							ENSBTAG00000012172	*MIDN*
							ENSBTAG00000007480	*CIRBP*
							ENSBTAG00000046542	*FAM174C*
							ENSBTAG00000038221	*EFNA2*
							ENSBTAG00000030839	*PWWP3A*
							ENSBTAG00000019419	*NDUFS7*
							ENSBTAG00000004112	*GAMT*
CNVR7	8	15,562,312	15,781,720	Duplication	FF	$2.007x10^{-7}$		
CNVR8	8	38,356,510	38,610,355	Duplication	FF	$3.292x10^{-6}$	*ENSBTAG00000020815*	*UHRF2*
							ENSBTAG00000011161	
							ENSBTAG00000011160	*TPD52L3*
							ENSBTAG00000018347	*IL33*
CNVR9	8	85,996,187	86,508,867	Deletion	DMI	$4.10x10^{-5}$	*ENSBTAG00000005092*	*ROR2*
							ENSBTAG00000054391	
							ENSBTAG00000047649	*AUH*
CNVR10	9	2,637,837	2,700,411	Mixed	FF	$9.307x10^{-6}$		
CNVR11	9	16,366,613	16,894,948	Duplication	FF	$1.096x10^{-10}$		
CNVR12	9	15,312,685	15,469,154	Duplication	FF	$4.583x10^{-6}$	*ENSBTAG00000005869*	*SENP6*

							ENSBTAG00000051278	
							ENSBTAG00000016751	*MYO6*
CNVR13	12	73,233,249	73,770,215	Mixed	FF	$2.861x10^{-9}$	*ENSBTAG00000004401*	*UGGT2*
							ENSBTAG00000039065	*HS6ST3*
CNVR14	12	74,302,958	74,578,587	Mixed	FF	$8.424x10^{-7}$		
CNVR15	12	64,618,237	64,736,496	Duplication	FF	$2.611x10^{-11}$		
CNVR16	13	12,552,408	12,829,168	Duplication	FF	$4.023x10^{-11}$	ENSBTAG00000019488	USP6NL
CNVR17	26	51,032,219	51,267,717	Duplication	FF	$5.725x10^{-6}$	*ENSBTAG00000009181*	*INPP5A*
							ENSBTAG00000054967	
							ENSBTAG00000051139	
							ENSBTAG00000017804	*BNIP3*
							ENSBTAG00000050527	

[a]Copy number variation region (CNVR) significantly ($P < 0.005$) associated with the traits
[b]Chromosome
[c]DMI: dry matter intake; FF: feed frequency

ADDITIONAL FILE 3

Table S3. Significant Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses.

Category[a]	Term	p-value[b]	Trait[c]
BP	GO:0036414 – Histone citrullination	5.71×10^{-9}	FF
BP	GO:0043951 – Negative regulation of cAMP-mediated signaling	0.00183	FF
BP	GO:0009060 – Aerobic respiration	0.00443	FF
BP	GO:0036413 – Histone H3-R26 citrullination	0.01178	FF
BP	GO:0090630 – Activation of GTPase activity	0.01670	FF
BP	GO:0046038 – GMP catabolic process	0.02919	FF
BP	GO:0071260 – Cellular response to mechanical stimulus	0.03143	FF
BP	GO:0071321 – Cellular response to cGMP	0.03492	FF
BP	GO:0006909 – Phagocytosis	0.03783	FF
BP	GO:0006955 – Immune response	0.00704	FF
CC	GO:0005634 – Nucleus	0.00278	FF
CC	GO:0030133 – Transport vesicle	0.02346	FF
CC	GO:0005743 – Mitochondrial inner membrane	0.04617	FF
MF	GO:0004668 – Protein-arginine deiminase activity	6.69×10^{-5}	FF
MF	GO:0004252 – Serine-type endopeptidase activity	0.00691	FF
MF	GO:0030246 – Carbohydrate binding	0.01088	FF
MF	GO:0003899 – DNA-directed 5'-3' RNA polymerase activity	0.01185	FF
PATHWAY	bta04740 – Olfactory transduction	0.02403	FF
PATHWAY	bta03420 – Nucleotide excision repair	0.04908	FF

[a]BP: biological process; CC: cellular process; MF: molecular function; Pathway: metabolic pathaway
[b]Significance level: $p < 0.05$
[c]FF: feed frequency

ADDITIONAL FILE 4

Table S4. Enriched quantitative trait loci (QTL) that overlapped with the genomic regions associated with DMI and FF

QTL	p_value	QTL type	Trait
Lean meat yield	0.0115112	Meat and Carcass	DMI
Milk fat percentage	0.0131573	Milk	DMI
Average daily gain	0.0056228	Production	FF
Body depth	0.1363043	Production	FF
Body height	0.1107805	Production	FF
Body weight	0.0387307	Production	FF
Body weight gain	0.0261365	Production	FF
Bovine respiratory disease susceptibility	0.073254	Health	FF
Bovine tuberculosis susceptibility	0.4126641	Health	FF
Calving ease	0.8477273	Reproduction	FF
Calving interval	0.0998353	Reproduction	FF
Coat color	1.046E-05	Exterior	FF
Conception rate	0.4500531	Reproduction	FF
Connective tissue amount	0.1884542	Meat and Carcass	FF
Dairy form	0.1309514	Exterior	FF
Dopamine level	0.0301066	Health	FF
Dry matter intake	0.9321903	Production	FF
Fat cover	0.0312919	Meat and Carcass	FF
Feet and leg conformation	0.5361104	Exterior	FF
Fertility index	0.12374	Reproduction	FF
First service conception	0.4755485	Reproduction	FF
Foot angle	0.561042	Exterior	FF
Hair length	0.0001148	Exterior	FF
Heel horn erosion	0.0241583	Exterior	FF
Inseminations per conception	0.6202659	Reproduction	FF
Interval to first estrus after calving	0.3688475	Reproduction	FF
Lactation persistency	0.0029695	Milk	FF
LDL cholesterol level	0.0265419	Health	FF
Lean meat yield	0.5326815	Meat and Carcass	FF
Length of productive life	0.9150415	Production	FF
M. paratuberculosis susceptibility	0.12201	Health	FF
Marbling score	0.1826957	Meat and Carcass	FF
Maturity rate	0.2572032	Production	FF
Milk alpha-S2-casein percentage	0.250809	Milk	FF
Milk C14 index	0.9731294	Milk	FF
Milk C16 index	0.7036025	Milk	FF
Milk capric acid content	0.6731459	Milk	FF
Milk casein percentage	0.2716153	Milk	FF

Milk fat percentage	0.9998846	Milk	FF
Milk fat yield	0.99131	Milk	FF
Milk lactose content	0.1205179	Milk	FF
Milk linoleic acid content	0.0215906	Milk	FF
Milk margaric acid content	0.0696093	Milk	FF
Milk mid-infrared spectra	0.0650463	Milk	FF
Milk palmitoleic acid content	0.9493991	Milk	FF
Milk protein percentage	7.133E-10	Milk	FF
Milk protein yield	0.9780389	Milk	FF
Milk tridecylic acid content	0.0006781	Milk	FF
Milk unglycosylated kappa-casein percentage	0.549324	Milk	FF
Milk urea nitrogen content	0.0430661	Milk	FF
Milk yield	0.9960463	Milk	FF
Milk zinc content	2.67E-12	Milk	FF
Milking speed	0.001637	Milk	FF
Multiple birth	0.1107805	Reproduction	FF
Muscle carnosine content	0.0786689	Meat and Carcass	FF
Muscle iron content	0.0707466	Meat and Carcass	FF
Omega-3 unsaturated fatty acid content	0.0085223	Meat and Carcass	FF
Pregnancy rate	0.0283827	Reproduction	FF
PTA type	0.1788791	Production	FF
Rear leg placement - rear view	0.1212048	Exterior	FF
Rump width	0.1276824	Production	FF
Shear force	0.8777472	Meat and Carcass	FF
Sole ulcer	0.0786689	Exterior	FF
Stature	0.1056326	Exterior	FF
Stayability	0.0133598	Reproduction	FF
Stillbirth	0.233264	Reproduction	FF
Strength	0.1951127	Exterior	FF
Teat placement - rear	0.3055899	Exterior	FF
Tenderness score	0.800387	Meat and Carcass	FF
Tick resistance	0.1031329	Health	FF
Udder attachment	0.5512335	Exterior	FF
Udder cleft	0.1160044	Exterior	FF
Udder height	0.1264616	Exterior	FF

ADDITIONAL FILE 5

Table S5: Number of genotyped animals per line and SNP panel

	Panel[b]	
Line[a]	50K	HD
NeC	25	89
NeS	51	194
NeT	82	487

[a]NeC: Nellore Control; NeS: Nellore Selection; NeT: Nellore Traditional
[b]50K: GeneSeek Genomic Profiler 50K; HD: Illumina BovineHD BeadChip

ADDITIONAL FILE 6

Table S6. Copy number variation region and the genes present in at least 90% of the studied population

CNVR	BTA	Start	End	Type	Gene Ensembl ID	Gene ID
CNVR	7	42,718,008	43,851,912	Mixed	*ENSBTAG00000019814*	*PGBD2*
					ENSBTAG00000046200	*OR2AV1*
					ENSBTAG00000046953	*OR2AV14*
					ENSBTAG00000009273	*OR2AV2*
					ENSBTAG00000045589	*OR2AV11*
					ENSBTAG00000049046	*OR2AZ3*
					ENSBTAG00000052753	*OR2AZ3B*
					ENSBTAG00000052768	
					ENSBTAG00000048841	*OR2AZ1*
					ENSBTAG00000000717	*PLPP2*
					ENSBTAG00000019767	*MIER2*
					ENSBTAG00000048393	*THEG*
					ENSBTAG00000047560	*C2CD4C*
					ENSBTAG00000016072	*SHC2*
					ENSBTAG00000039372	*ODF3L2*
					ENSBTAG00000050765	
					ENSBTAG00000039355	*MADCAM1*
					ENSBTAG00000010772	*TPGS1*
					ENSBTAG00000002098	*CDC34*
					ENSBTAG00000002100	*GZMM*
					ENSBTAG00000016648	*BSG*
					ENSBTAG00000015050	*HCN2*
					ENSBTAG00000015051	*POLRMT*
					ENSBTAG00000027357	*FGF22*
					ENSBTAG00000014349	*RNF126*
					ENSBTAG00000003018	*FSTL3*
					ENSBTAG00000021616	*PRSS57*
					ENSBTAG00000019771	*PALM*
					ENSBTAG00000020269	*MISP*
					ENSBTAG00000045828	*PTBP1*
					ENSBTAG00000051609	*PLPPR3*
					ENSBTAG00000045829	*AZU1*
					ENSBTAG00000049418	*U6*
					ENSBTAG00000046105	*PRTN3*
					ENSBTAG00000046188	*ELANE*
					ENSBTAG00000048122	*CFD*
					ENSBTAG00000047217	*MED16*
					ENSBTAG00000053281	*U6*
					ENSBTAG00000046406	*R3HDM4*
					ENSBTAG00000037418	*KISS1R*
					ENSBTAG00000008607	*ARID3A*
					ENSBTAG00000002434	*WDR18*
					ENSBTAG00000025597	*GRIN3B*

	ENSBTAG00000011351	*TMEM259*
	ENSBTAG00000020764	*CNN2*
	ENSBTAG00000020766	*ABCA7*
	ENSBTAG00000020772	*ARHGAP45*
	ENSBTAG00000020776	*POLR2E*
	ENSBTAG00000053003	*GPX4*
	ENSBTAG00000020780	*SBNO2*
	ENSBTAG00000011639	*STK11*
	ENSBTAG00000025233	*CBARP*
	ENSBTAG00000000550	*ATP5F1D*
	ENSBTAG00000012172	*MIDN*
	ENSBTAG00000007480	*CIRBP*
	ENSBTAG00000046542	*FAM174C*
	ENSBTAG00000038221	*EFNA2*
	ENSBTAG00000030839	*PWWP3A*
	ENSBTAG00000019419	*NDUFS7*
	ENSBTAG00000004112	*GAMT*
	ENSBTAG00000019714	*DAZAP1*
	ENSBTAG00000019718	*RPS15*