UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ZOOTECNIA E ENGENHARIA DE ALIMENTOS

LAÍS GRIGOLETTO

Genomic studies in Montana Tropical Composite cattle

Pirassununga
2020

LAIS GRIGOLETTO

Genomic studies in Montana Tropical Composite cattle

Versão Corrigida

Thesis submitted to the College of Animal Science and Food Engineering, University of São Paulo in partial fulfillment of the requirements for the degree of Doctor in Science from the Animal Biosciences program.

Concentration area: Genetics, Molecular and Cellular Biology

Supervisor: Prof. Dr. José Bento Sterman Ferraz

Co-supervisor: Prof. Dr. Fernando Baldi

Pirassununga

2020

Ficha catalográfica elaborada pelo
Serviço de Biblioteca e Informação, FZEA/USP,
com os dados fornecidos pelo(a) autor(a)

Grigoletto, Laís
G857g Genomic studies in Montana Tropical Composite cattle / Laís Grigoletto ; orientador José Bento Sterman Ferraz ; coorientador Fernando Baldi. -- Pirassununga, 2020.
183 f.

Tese (Doutorado - Programa de Pós-Graduação em Biociência Animal) -- Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo.

1. beef cattle. 2. composite. 3. genomics. 4. imputation. 5. genetic progress. I. Ferraz, José Bento Sterman, orient. II. Baldi, Fernando, coorient. III. Título.

DISPENSA DE ANÁLISE ÉTICA

Comunicamos que o projeto de pesquisa abaixo identificado está dispensado da análise ética por utilizar animais oriundos de coleções biológicas formadas anteriormente ao ano de 2008, ano da promulgação da Lei nº 11.794/2008 – lei que estabelece procedimentos para o uso científico de animais.

Ressaltamos que atividades realizadas na vigência da referida lei, ou que resulte em incremento do acervo biológico, devem ser submetidas à análise desta CEUA conforme disposto pelo *Conselho Nacional de Controle de Experimentação Animal* (CONCEA).

Projeto: Genomic Studies in Montana tropical composite cattle.

Finalidade: Pesquisa Acadêmica – Tese de Doutorado

Pesquisador: Laís Grigoletto

Pesquisador responsável: Prof. Dr. José Bento Sterman Ferraz (Dep. ZMV)

Instituição: Faculdade de Zootecnia e Engenharia de Alimentos

Informações adicionais: - .

Pirassununga, 18 de agosto de 2020.

Profa. Dra. CRISTIANE GONÇALVES TITTO
Presidente da Comissão de Ética no Uso de Animais

Profa. Dra. DANIELE DOS SANTOS MARTINS
Vice-Presidente da Comissão de Ética no Uso de Animais

Universidade de São Paulo

ATA DE DEFESA

Ata de defesa de Tese do(a) Senhor(a) Laís Grigoletto no Programa: Biociência Animal, do(a) Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo.

Aos 10 dias do mês de julho de 2020, no(a) Sala do docente (GMAB) realizou-se a Defesa da Tese do(a) Senhor(a) Laís Grigoletto, apresentada para a obtenção do título de Doutora intitulada:

"Estudos genômicos em bovinos de corte composto Montana tropical"

Após declarada aberta a sessão, o(a) Sr(a) Presidente passa a palavra ao candidato para exposição e a seguir aos examinadores para as devidas arguições que se desenvolvem nos termos regimentais. Em seguida, a Comissão Julgadora proclama o resultado:

Nome dos Participantes da Banca	Função	Sigla da CPG	Resultado
José Bento Sterman Ferraz	Presidente	FZEA - USP	_____
Heidge Fukumasu	Titular	FZEA - USP	_____
Juliana Petrini	Titular	UNIFAL - Externo	_____
Ricardo Vieira Ventura	Titular	FMVZ - USP	_____
Luiz Fernando Brito	Titular	PU - Externo	_____

Resultado Final: _____

Parecer da Comissão Julgadora *

Eu, Erica Cristina Mello Ferraz _____ , lavrei a presente ata, que assino juntamente com os(as) Senhores(as). Pirassununga, aos 10 dias do mês de julho de 2020.

Heidge Fukumasu

Juliana Petrini

Ricardo Vieira Ventura

Luiz Fernando Brito

José Bento Sterman Ferraz
Presidente da Comissão Julgadora

* Obs: Se o candidato for reprovado por algum dos membros, o preenchimento do parecer é obrigatório.

A defesa foi homologada pela Comissão de Pós-Graduação em _____ e, portanto, o(a) aluno(a) _____ jus ao título de Doutora em Ciências obtido no Programa Biociência Animal.

_____
Presidente da Comissão de Pós-Graduação

To my lovely and supportive family.
In memory of my grandmother's Maria Luiza and Cacilda (“Cacá”).

I dedicate.

ACKNOWLEDGEMENTS

I truly believe that every person, situation, and the lesson come in our life for a reason. I am very blessed to have and met amazing people that I have learned with and can call then family and friends. I am very glad to no longer be the same person as I was in the past year. I've learned with each one and situation in my life. Also, I feel so realized for each decision I made in life until now.

I was very lucky during my Ph.D. I met (and re-met) very special people, which did not only teach me about genetics and genomics but also taught me a new perspective and how to be a better person and chase my dreams. They helped me to not give up, even when I was scared and did not believe in myself. I could have never done half of what I did in my Ph.D. if I have not met them at some point in my path. There is a theory called six degrees of separation originated from a scientific study d that describes, in the world, a maximum of six bonds of friendship is necessary for two people to be connected, so on in the past five years, I may have more connections built. Thus, I would like to show here my thanks to some of the people that were part of these five (2 of masters and 3 of Ph.D.) tremendous years of my life.

First of all, I would like to thank my family. My parents Renato and Angela to give me the support, care, and they love to go through this intense cycle of my life. I do not have words to describe how I am thankful. To my brother, Dr. Renan, who shared the whole life with me, also be very protective for me, and of course, we divide the Ph.D. life together. Even more, I am so thankful to Renan and my sister in law Carolina, to my blessed and perfect nephew Bento. To all my uncles and aunts, and cousins. To my little cousin Gabi ("Pri") to all the support! If I had the chance of choosing my family, I would choose all of you again!

I would like to say a very big thank you to my advisor, Dr. José Bento Sterman Ferraz, for all support that he has given me since when I sent an email asking to be your Master student. That decision to give up my job and start this journey worth each second. Bento, you gave me a lot of opportunities that I could have never imagined. You were always there ready to help me with everything that I needed. You always helped me to achieve my dreams as if they were yours, and you believed in my potential even when I did not. You are the best advisor that I could have asked for, and I am very proud of being your student and called you as a friend (sometimes more like my

second father). You have a special place in my life, and we need to have our barbecue as soon as possible. Thank you very much for everything that you have done for me. I admire you a lot for everything that you have been doing for the improvement of the beef cattle industry in Brazil.
Dr. Luiz Fernando Brito, you have been the best person I have met in life. It was a pleasure to have the opportunity to stay close with you last year (2019-2020) doing part of my Ph.D. at Purdue University. I am very thankful for having trusted and believe in me. Thanks for being by my side helping, teaching, and cheering for me. You proved to me how far we can go if we work hard, and how life brings us good surprises when we do our best. But especially, our discussions/partnerships/projects/travels were fundamental in making me who I am today. I remember that the first thing you said to me was: "Lais, my job here, is to make you think". And, Luiz every day that I need to do something, I always think why should I need to do this When? How? I am sure that we will continue to work together. Thanks for everything. Dra. Hinayah, you are an example of a professional and a person for me. Thanks for all your help, discussion, dinners, and friendship. I miss sharing my day and all my stories with you. I can have a piece of what it is to have a sister in my life.

Elisangela, our kindness "Li", you've been all my supporting and friend in this important cycle of my life. I could have never imagined that someone would be so nice and kind. You were the most sweetness surprise from life, and I have sure that you will always be present in my life. Thank you for all these years, coffee times, advisers, care, listening, and other more special moments.

Dr. Fernando Baldi, you helped me a lot in my early beginning in the academic field. Thanks for all these years. I've learned a lot from you. Dr. Joanir Pereira Eler, thanks for sharing all the data and your knowledge with me. Dr. Miguel Henrique Almeida Santana, I would like to thank you for always have the right words and trust. And, of course, for your friendship. You are like a brother to me. Dr. Ricardo Ventura, thank you very much for your support and knowledge in those years. Moreover, thank you for all your great suggestions and comments in our manuscripts, especially in the last chapters of my thesis. I also would like to say thank you to Dr. Heidge Fukumasu, Dra. Juliana Petrini, Dr. Rafael Spigolan, Dr. Victor Pedrosa and Dr. Fabyano da Fonseca for accepting being part of my defense committee, and for all their contributions to this thesis.

To GMAB family, I made some very special friends that were very important at the begging of my Master's and Ph.D. I appreciated the fact that I could learn more about different things, and I enjoyed the time that we spent together. Thus, I would like to thank you to Barbara, Fernando (for all conversations and friendship), Pâmela, Felipe, Gerson, Alisson, Gerardo, Isabela, Lucas, Hugo, Rosiane, and Mariana. Also, thank you Sra. Vera. To Purdue family, I would like to manifest here my sincere thank you to Dr. Allan Schinkel (especially for the papers, and good behaving), Amanda (my U.S and Brazilian sister for all conversations, joy, and friendship), Hanna, Pedro, Laís, Fabiana, Isis, Scott, Marcela, Thaís, Guilherme, Débora, Augusto, André, and Sirlene. I had a super fun time with you, guys! I hope we can repeat it in the future.

Of course, I also have to thanks all my good friends from the Araras, where all dreams start. Over these 31 years is impossible do not think about a lot of special people that are/were very important in my life. Thus, to represent all my "old friends", I would like to say a special thank you to Paulinha (my half soul), Isa, Fer, Tammy (who always listen to my anxious and happy moments) and Nathy (We'd passed through difficult times, learned together and support each other from almost 15 years). Especially to Fer Orpi, who always shared life moments and Ph.D. life (all the afflictions and joy) as well, thank you "miga" to always be by my side. It does not matter how physically far we are, our friendship and affection for each other is always the same.

And last but not least, I would like to say thanks to the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)" and "Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)" for the financial support that I received during my Ph.D. If I did not have the FAPESP scholarship I would not be able to achieve this dream. Also, I would like to say thank you to the "Department of Animal Science at Purdue University" that allowed me to go further and show my study in different parts of the world.

Some people arrive without any intension and end up becoming a fundamental part of our history.

“Put the things you can control in order. Repair what is in disorder and make what is already good better.”

Jordan B. Peterson

ABSTRACT

GRIGOLETTO, L. **Genomic studies in Montana Tropical Composite cattle.** 2020, 183p. Thesis – College of Animal Science and Food Engineering, University of Sao Paulo, Pirassununga (FZEA/USP).

Beef cattle production is a major component of the Brazilian economy and places the country amongst the top worldwide beef producers and exporters. The development and use of composite populations represent a promising alternative to increase production efficiency and quality in beef cattle. The strategic combination of different breeds poses a great opportunity to maximize genetic variability and exploit heterosis and breed complementarity. In order to successfully implement genomic evaluations for beef cattle, various genomic approaches and methodologies need to be investigated. In addition, the majority of genomic models and methods currently used for genomic evaluations were developed based on purebred animals and the knowledge on the use of genomic information in synthetic or composite cattle breeds is still very limited. Therefore, studies investigating more sophisticated methodologies and genomic approaches will enable a more efficient use of the genomic information available for composite breeds and consequently, more accurate breeding values. The overall objectives of this thesis are to: 1) identify genomic regions and potential candidate genes associated with various economically important traits in composite beef cattle populations; 2) investigate genomic selection methodologies and imputation strategies to improve the accuracy of genomic prediction of breeding values in the genetically diverse population of Montana Tropical Composite® cattle. Approximately 4,000 genotyped animals from multiple breeds (Montana, Nellore, Aberdeen Angus, Red Angus, Senepol, and Simmental) and SNP chip panels were available for this research. In summary, this project contributed to the application of genomic approaches to fastener genetic progress for a variety of economically important traits in composite beef cattle and increase the profitability of beef producers.

Keywords: beef cattle, composite, genomics, imputation, genetic progress

RESUMO

GRIGOLETTO, L. **Estudos genômicos em bovinos de corte Composto Montana Tropical.** 2020, 183p. Tese (Doutorado) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga (FZEA/USP).

A pecuária é um componente importante da economia brasileira e coloca o país entre os principais produtores e exportadores mundiais de carne bovina. O desenvolvimento e uso de populações de compostos representam uma alternativa promissora para aumentar a eficiência e a qualidade da produção em bovinos de corte. A combinação estratégica de diferentes raças representa uma grande oportunidade para maximizar a variabilidade genética e explorar a heterose e complementaridade entre raças. Para implementar com sucesso avaliações genômicas para bovinos de corte, várias abordagens e metodologias genéticas e genômicas precisam ser investigadas. Além disso, a maioria dos modelos e métodos genômicos atualmente utilizados para avaliações genômicas foram desenvolvidos com base em animais de raça pura e o conhecimento sobre o uso de informações genômicas em raças sintéticas ou compostas de gado ainda é muito limitado. Portanto, estudos que investigam metodologias mais sofisticadas e abordagens genômicas permitirão um uso mais eficiente das informações genômicas disponíveis para raças de raça pura e composta e, consequentemente, predição de valores genéticos genômicos mais acurados. Os objetivos gerais são: 1) identificar regiões genômicas e potenciais genes candidatos associados a várias características econômicas importantes em animais compostos; 2) investigar metodologias de seleção genômica e estratégias de imputação para melhorar a acurácia de predição de valores genéticos genómicos na população geneticamente diversa do composto Montana Tropical®. Aproximadamente, 4.000 genótipos de animais de várias raças (Montana, Nelore, Aberdeen Angus, Red Angus, Senepol e Simmental) em diversas densidades foram utilizados nesta pesquisa. Dessa forma, este trabalho contribuiriu para a viabilidade da aplicação de métodos e abordagens genômicas para o progresso genético para uma variedade de características economicamente importantes em bovinos de corte compostos e a lucratividade dos produtores de carne bovina.

Palavras-chave: bovinos de corte, compostos, genômica, imputação, progresso genético.

FIGURES

TABLES

TABLE OF CONTENTS

CHAPTER 1. GENERAL INTRODUCTION

1.1 BACKGROUND AND JUSTIFICATION

The Montana Tropical® composite population was developed based on crossbreeding schemes between multiple *Bos taurus taurus* and *Bos taurus indicus* breeds, and therefore, has improved productivity, adaptation, and profitability due to the combination of four different breed groups, as indicine and taurine breeds that are adapted to tropical environments were used in the development of this composite (FERRAZ; ELER; GOLDEN, 1999). In terms of genetic improvement, the main advantages of using composite animals are performance gains due to heterosis and its retention in successive generations. In addition, the complementarity among the breeds, under the advantageous combination of traits, present in each parental breed, is another great benefit, as reported by geneticists who developed the concept at the USDA's Meat Animal Research Center, located in Clay Center, Nebraska, USA (CUNDIFF; GREGORY, 1999). As a consequence of heterosis, the crossbred progeny has an adaptive and productive value greater than the average of its parents.

The efficiency in the selection process of genetically superior animals in breeding programs, whether purebred, crossbreds or composite animals, depends on the accuracy of estimation of the genetic value of the animals at a young age. The methodological procedures used in the genomic evaluation for multi-breed animals allow more accurate predictions of genetic values through direct comparison between animals of different breed compositions and for the additive and non-additive genetic components (MOURÃO et al., 2007; BUENO et al. 2011,

BOCCHI et al., 2016; DUENK et al., 2020; GRIGOLETTO et al., 2019; 2020). However, as reported by Elzo and Borjas (2004), Bueno et al. (2012), and Misztal et al. (2013), these statistical genetic models are complex, providing mismatch between genetic components and inclusion of non-additive genetic effects, dominance and epistasis.

The incorporation of genomic information (MEUWISSEN; HAYES; GODDARD, 2001; GODDARD; HAYES, 2007) has been intensively explored in livestock production, especially in pure breeds (HAYES et al., 2009; GARRICK et al., 2011; MEUWISSEN; HAYES; GODDARD, 2001) and also to evaluate the genetic value of animals with greater accuracy at a young age and therefore, fastener genetic progress (ELZO et al., 2012; LUND et al., 2014). The use of genomic selection is more advantageous for traits that are difficult or expensive to measure such as: meat quality, feed efficiency, and resistance to disease. Furthermore, the use of genomic selection plays an important role on reducing the generation interval and, consequently, increasing the genetic gain per year, through more accurate selection of young animals, without own performance or progeny evaluated (SCHAEFFER, 2006). On this regard, different statistical genomic methodologies have been proposed to incorporate pedigree, phenotype and genomic information, as multi-step methods using genomic and pedigree information separately (HAYES et al., 2009; VANRADEN et al., 2009) and single-step Genomic Best Linear Unbiased Prediction (ssGBLUP; LEGARRA et al., 2014), combining phenotypic, genotypic and pedigree records in a single analysis. This enables inclusion of genotyped and ungenotyped animals in a single step (LOURENÇO et al., 2015; MISZTAL, 2017).

There is still a great lack of research investigating the implementation of genomic selection in composite beef cattle populations, such as the Montana Tropical® Composite. By investigating alternative genomic prediction methods and imputation approaches may reveal that the integration

of multiple sources of information in the Montana composite population, including genomic data, may improve understanding of the biological processes and breeding schemes of composite beef cattle (PORTO-NETO et al., 2014; RAMAYO-CALDAS et al., 2016; HAY; ROBERTS, 2017; 2018; GRIGOLETTO et al., 2019; 2020).

1.2 OUTLINE

Chapters 2 and 3 were elaborated to identify genomic regions and potential candidate genes associated with growth, carcass, and meat quality traits for composite beef cattle through genome-wide association studies (GWAS) using the weighted single-step GBLUP methodology (WANG et al., 2012; 2014). By investigating the biological pathways in which the candidate genes are involved, we can have a better understanding of the biological processes that are (or might be) affected by direct selection for these traits.

Chapter 4 aimed to the characterize the population structure, linkage disequilibrium, consistency of gametic phase, and admixture levels in Tropical Composite beef cattle and perform a genotype imputation analysis also adding purebred animals, such as Aberdeen Angus, Nellore, Senepol, and a crossbreed population of Angus x Simmental.

Chapter 5 was designed to investigate the accuracy of genomic predictions for growth, reproductive, carcass, and meat quality traits in Montana Tropical Composite using imputed data and alternative methodologies such as GBLUP (VANRADEN, 2008), single-step GBLUP (MISZTAL et al., 2009; AGUILAR et al., 2010; CHRISTENSEN; LUND, 2010), and BayesR (ERBE et al., 2012) to indicate the method that yield genomic breeding values with the highest accuracy and lowest less bias.

1.3 OBJECTIVES

1.3.1 General Objectives

The overall objective of this thesis was to evaluate the best strategies and methods to genomically characterize the Montana Tropical® Composite population and provide the supporting information to enable implementation of genomic selection for growth, reproductive, carcass and meat quality traits in this composite population.

1.3.2 Specific Objectives

a) To identify genomic regions and potential candidate genes associated with various growth, carcass, and meat quality traits in composite population using the weighted single-step Genomic Best Linear Unbiased Prediction (WssGBLUP) method;
b) To estimate genetic parameters for various growth, carcass and meat quality traits in Montana Tropical® Composite cattle;
c) To assess genetic diversity levels and population structure based on genomic information;
d) To investigate different strategies for genotype imputation from low- to high-density SNP panels in Montana Tropical® cattle and purebred animals;
e) To evaluate the predictive ability of genomic breeding values using various statistical methods [GBLUP (VANRADEN, 2008), single-step GBLUP (MISZTAL et al., 2009; AGUILAR et al., 2010; CHRISTENSEN; LUND, 2010), and BayesR (ERBE et al., 2012)].

1.4 REFERENCES

AGUILAR, I. et al. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743–752, 2010. Disponível em: <http://dx.doi.org/10.3168/jds.2009-2730>.

BOCCHI, A. L.; OLIVEIRA, H. N. D., FERRAZ, J. B. S., & ELER, J. P. Multibreed genetic evaluation in bovines using simulated data employing a composite population. **Genetics and Molecular Research**, v. 15, n. 4, p. 1–17, 2016.

BUENO, R. S.; TORRES, R. D. A.; FERRAZ, J. B. S. et al. Inclusão da epistasia em modelo de avaliação genética de bovinos de corte compostos. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 63, n. 4, p. 948-953, 2011.

BUENO, R. S.; TORRES, R. D. A.; FERRAZ, J. B. S. et al. Métodos de estimação de efeitos genéticos não-aditivos para características de peso e perímetro escrotal em bovinos de corte mestiços. **Revista Brasileira de Zootecnia**, v. 41, n. 5, p. 1140–1145, 2012.

CHRISTENSEN, O. F.; LUND, M. S. Genomic relationship matrix when some animals are not genotyped Genomic prediction models. **Genetics Selection Evolution**, v. 42, n. 2, 2010.

CUNDIFF, L. V.; GREGORY, K. E. What is systematic crossbreeding. In: **Cattlemen's College, 1999 Cattle Industry Annual Meeting and Trade Show, National Cattlemen's Beef Association. Charlotte, NC**. 1999.

DUENK, P.; BIJMA, P.; CALUS, M. P.; WIENTJES, Y. C.; & VAN DER WERF, J. H. The impact of non-additive effects on the genetic correlation between populations. **G3: Genes| Genomes| Genetics**, v. 10, n. 2, p. 783, 2020.

ELZO, M. A.; LAMB, G. C.; JOHNSON, D. D. et al. Genomic-polygenic evaluation of Angus-Brahman multibreed cattle for feed efficiency and postweaning growth using the Illumina 3K chip. **Journal of animal science**, v. 90, n. 8, p. 2488-2497, 2012.

ELZO, M. A.; BORJAS, A. R. Perspectivas da avaliação genética multirracial em bovinos no Brasil. **Ciência Animal Brasileira**, v. 5, p. 171–185, 2004.

ERBE, M. et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. **Journal of dairy science**, v. 95, n. 7, p. 4114-4129, 2012.

FERRAZ, J. B. S.; ELER, J. P.; GOLDEN, B. L. A formação do composto Montana Tropical. **Revista Brasileira de Reprodução Animal**, v. 23, n. 2, p. 115-117, 1999.

GARRICK, D. J. The nature, scope and impact of genomic prediction in beef cattle in the United States. **Genetics Selection Evolution**, v. 43, n. 1, p. 17, 2011.

GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal of Animal breeding and Genetics**, v. 124, n. 6, p. 323-330, 2007.

GRIGOLETTO, L. FERRAZ, J. B. S.; OLIVEIRA, H. R.; ELER, J. P.; BUSSIMAN, F. O.; ABREU SILVA, B. C.; BALDI, F.; BRITO, L. F. Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical® Composite Beef Cattle. **Frontiers in genetics**, v. 11, p. 123, 2020.

GRIGOLETTO, L.; BRITO, L. F.; MATTOS, E. C.; ELER, J. P.; BUSSIMAN, F. O.; ABREU SILVA, B. C.; SILVA, R. P.; CARVALHO, F. E.; BERTON, M. P.; BALDI, F.; FERRAZ, J. B. S. Genome-wide associations and detection of candidate genes for direct and maternal

genetic effects influencing growth traits in the Montana Tropical® Composite population. **Livestock Science**, v. 229, p. 64-76, 2019.

HAY, E. H.; ROBERTS, A. Genotype× prenatal and post-weaning nutritional environment interaction in a composite beef cattle breed using reaction norms and a multi-trait model. **Journal of animal science**, v. 96, n. 2, p. 444-453, 2018.

HAY, E. H.; ROBERTS, A. Genomic prediction and genome-wide association analysis of female longevity in a composite beef cattle breed. **Journal of animal science**, v. 95, n. 4, p. 1467-1471, 2017.

HAYES, B. J.; VISSCHER, P. M.; GODDARD, M. E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics research**, v. 91, n. 1, p. 47-60, 2009.

LEGARRA, A.; CHRISTENSEN, O. F.; AGUILAR, I.; & MISZTAL, I. Single Step, a general approach for genomic selection. **Livestock Science**, v. 166, p. 54-65, 2014.

LOURENCO, D. A. L.; TSURUTA, S.; FRAGOMENI, B. O. et al. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. **Journal of animal science**, v. 93, n. 6, p. 2653-2662, 2015.

LUND, M. S.; SU, G.; JANSS, L.; GULDBRANDTSEN, B.; & BRØNDUM, R. F. Genomic evaluation of cattle in a multi-breed context. **Livestock Science**, v.166, p. 101-110, 2014.

MEUWISSEN, T.; HAYES, B.; GODDARD, M. Genomic selection: A paradigm shift in animal breeding. **Animal frontiers**, v. 6, n. 1, p. 6-14, 2016.

MISZTAL, I. et al. BLUPF90 and related programs (BGF90). In: **Proceedings of the 7th world congress on genetics applied to livestock production**. 2002. p. 743-744.

MISZTAL, I.; TSURUTA, S.; LOURENCO, D.; AGUILAR, I.; LEGARRA, A.; VITEZICA,

Z. Manual for BLUPF90 family of programs. **Athens: University of Georgia**, 2014. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

MISZTAL, I.; TSURUTA, S.; AGUILAR, I.; LEGARRA, A.; VANRADEN, P. M.; & LAWLOR, T. J. Methods to approximate reliabilities in single-step genomic evaluation. **Journal of dairy science**, v. 96, n. 1, p. 647-654, 2013.

MISZTAL, I.; LEGARRA, A. Invited review: efficient computation strategies in genomic selection. **animal**, v. 11, n. 5, p. 731-736, 2017.

MOURÃO, G. B.; FERRAZ, J. B. S.; ELER, J. P.; BALIEIRO, J. C. D. C.; BUENO, R. S.; MATTOS, E. C.; FIGUEIREDO, L. G. G. Genetic parameters for growth traits of a Brazilian Bos taurus x Bos indicus beef composite. **Genetics and Molecular Research**, v. 6, n. 4, p. 1190-1200, 2007.

PORTO-NETO, L. R.; KIJAS, J. W.; REVERTER, A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. **Genetics Selection Evolution**, v. 46, n. 1, p. 22, 2014.

RAMAYO-CALDAS, Y.; MACH, N.; LEPAGE, P. et al. Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. **The ISME journal**, v. 10, n. 12, p. 2973-2977, 2016.

SCHAEFFER, L. R. Strategy for applying genome-wide selection in dairy cattle. **Journal of animal Breeding and genetics**, v. 123, n. 4, p. 218-223, 2006.

VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 2008. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0022030208709901>.

VANRADEN, P. M.; VAN TASSELL, C. P., WIGGANS, G. R et al. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of dairy science**, v. 92, n. 1, p. 16-24, 2009.

WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes. **Genetics Research**, v. 94, n. 2, p. 73–83, 2012.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; FERNANDO, R. L.; VITEZICA, Z.; OKIMOTO, R.; WING, T.; HAWKEN, R.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in genetics**, v. 5, p. 134, 2014.

CHAPTER 2. GENOME-WIDE ASSOCIATIONS AND DETECTION OF CANDIDATE GENES FOR DIRECT AND MATERNAL GENETIC EFFECTS INFLUENCING GROWTH TRAITS IN THE MONTANA TROPICAL® COMPOSITE POPULATION

Laís Grigoletto[a,b,] Luiz F. Brito[b], Elisângela C. Mattos[a], Joanir Pereira Eler[a], Fernando Oliveira Bussiman[a], Barbara da Conceição Abreu Silva[a], Rosiane Pereira da Silva[a], Felipe Eguti Carvalho[a], Mariana Piatto Berton[c], Fernando Baldi[c], José Bento Sterman Ferraz[a]

[a]Department of Veterinary Medicine, Faculdade de Zootecnia e Engenharia de Alimentos, University of São Paulo, Pirassununga, São Paulo, Brazil

[b]Department of Animal Sciences, Purdue University, West Lafayette, IN, USA

[c]Department of Animal Sciences, Faculdade de Ciências Agrárias e Veterinárias, São Paulo State University, Jaboticabal, São Paulo, Brazil

Livestock Science, 2019
(Original Article, Published)

ABSTRACT

The Montana Tropical® Composite beef cattle have been developed in Brazil to serve as a genetic resource to meet the consumers' needs for higher-quality meat while animals are raised in extensive production systems under tropical conditions. In order to optimize the selection process for economically important traits in this population, various genetic and genomic studies are still lacking. In this regard, the aim of this study was to assess genetic parameters and identify genomic regions and potential candidate genes associated with various growth traits, using the single-step Genomic Best Linear Unbiased Predictor (ssGBLUP) method. Approximately 400,000 cows, bulls and progeny had measurements for birth weight (BW), weaning weight (WW), yearling weight (YW) and post-weaning weight gain (WG). A total of 1,394 animals were genotyped for 27,199 SNPs (after the quality control) to enable implementation of weighted single-step genome-wide association studies. The traits included in this study were shown to be moderately heritable (i.e. heritability estimates ranging from 0.16±0.01 to 0.33±0.04) and the genetic correlations ranged from 0.60±0.07 (between WW and WG) to 0.88±0.08 (between BW and WW). Single-trait weighted genome-wide association studies enabled the identification of 83 genomic regions for direct genetic effects (all traits) and 29 genomic regions associated with maternal genetic effects on BW and WW traits. Furthermore, biological processes and pathways associated with survival to adult age, calf behavior, fatty acid metabolism, muscle development, fertility, and immune system were identified. The findings of this study greatly contribute to a better understanding of the genetic architecture of growth traits in the Montana Tropical® Composite population. Furthermore, the genomic regions identified can be given more importance (weight) when implementing genomic selection for these traits, by using weighted ssGBLUP or Bayesian approaches.

Keywords: biological pathways, composite, crossbreeding, growth, genome-wide association, tropical beef cattle.

2.1 INTRODUCTION

Composite populations are formed by crossing individuals from two or more breeds with the purpose of exploiting heterosis and complementarity between them (GREGORY; CUNDIFF, 1999). This has been done in various species, including sheep (e.g., BRITO et al., 2017), dairy cattle (e.g., COLE; SILVA, 2016), and beef cattle (e.g., DODENHOFF et al., 1999; GREGORY et al., 1994; LUND et al., 2014; HAY; ROBERTS, 2018). In this regard, the Montana Tropical® Composite (Figure 2.1) was developed by crossing animals from four different biological types (i.e., breed groups) termed NABC (FERRAZ; ELER; GOLDEN, 1999). The genetic variability obtained by combining these four complementary biological types results in animals with better carcass yield and meat quality while still retaining important adaptation and robustness characteristics. These traits are paramount to obtain high performance in challenging environments, particularly in tropical regions with harsh environmental conditions (FERRAZ; ELER; GOLDEN, 1999; HANSEN, 2004; PORTO-NETO et al., 2014). Growth traits are key variables associated with the profitability of beef cattle production (e.g., SULLIVAN et al., 1999; GROSSI et al., 2008). Thus, they have a great economic importance to the Montana Tropical® Composite breeding program and account for approximately 70% of the selection index weights (SANTANA et al., 2014).

Identifying genomic regions influencing economically important traits through genome-wide association studies (GWAS) has a great value to better understand the genetic architecture of complex traits. This has been done in other composite populations (e.g., SCHMID; BENNEWITZ, 2017; HAY; ROBERTS, 2018). The weighted single-step GWAS method (WssGWAS; WANG

et al., 2012; 2014) enables the simultaneous integration of pedigree, genomic and phenotypic information (also for ungenotyped animals) to improve the estimation accuracy of Single Nucleotide Polymorphism (SNP) effects.

Previous GWAS in crossbred animals and other composite populations reported very interesting genes and metabolic pathways (SNELLING et al., 2010; BUZANSKAS et al., 2014; CRISPIM et al., 2015; MARTÍNEZ et al., 2017; HAY; ROBERTS, 2018), which contributed to better understand the nature of the genetic effects associated with the traits under study. For instance, these authors identified relevant genomic regions affecting growth traits, especially in a genomic window located on *Bos taurus* autosome (BTA) 14 harboring genes such as *TOX, NSMAF*, *MRPL15, RGS20,* and *SOX17.* Furthermore, those genes are neighboring the *PLAG1* gene, which has been previously associated with beef cattle growth rates (UTSUNOMIYA et al., 2013; 2017; TAKASUGA, 2016).

The main objectives of this study were: 1) to estimate variance components and genetic parameters for growth traits using both pedigree and genomic information; and, 2) to perform WssGWAS and functional enrichment analyses aiming to identify genomic regions associated with direct and maternal genetic effects for growth traits in the Montana Tropical® Composite population.

2.2 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.

2.2.1 Phenotypic and Pedigree Datasets

The datasets available for the analyses included pedigree and phenotypic records for 396,761 animals born between 1994 and 2016. These animals were raised in 58 farms distributed in the Brazilian states of São Paulo, Minas Gerais, Goiás, Mato Grosso do Sul and Pará, and also in Paraguay and Uruguay. The pedigree file contained 525,806 individuals, in which 4,172 were sires and 195,524 were dams. The datasets were provided by the Animal Breeding and Biotechnology Group of the College of Animal Science and Food Engineering (Pirassununga, Sao Paulo, Brazil). The traits analyzed included birth weight (BW, kg), weaning weight (WW, kg; measured around 205 days of age), yearling weight (YW, kg; measured around 420 days of age) and post-weaning weight gain (WG, kg; calculated as YW minus WW).

All animals were raised in extensive production systems, in pastures with predominance of grasses from the *Brachiaria* and *Panicum* genera. In some farms, nutritional supplementation was also provided in the dry period. Births were concentrated between September and December (spring season), in which there was greater abundance of better-quality forage and a favorable environment for calf development. The data editing procedure removed phenotypic records deviating from the mean ± 3.5 standard deviations and contemporary groups (CG) with less than five records. CG was defined based on farm, year and season of birth, sex and management group. Descriptive statistics of the records used for the analyses are presented in Table 2.1.

Figure 2.1 Illustration of Montana Tropical® Composite animals (bulls in the top and cows and calves in the bottom picture). Photo credits: Montana Tropical® Composite breeding program.

Table 2.1 Number of phenotypic records (N), contemporary groups (N_{GC}), mean and standard deviation (SD) for birth weight (BW), weaning weight (WW), yearling weight (YW) and post weaning weight gain (WG) traits.

Traits	N	N_{GC}	Mean	SD
BW (kg)	352,732	2,050	33.07	4.44
WW (kg)	381,269	2,088	194.37	37.42
YW (kg)	146,795	1,416	271.49	51.01
WG (kg)	94,162	690	92.14	29.93

The breed composition of the Montana Tropical® Composite cattle was not the same for all animals, as crossbreeding is still being performed according to the NABC system (FERRAZ; ELER; GOLDEN, 1999). Thus, the genetic composition of each animal derives from different breed proportions, following the criteria described in Table 2.2. This system classifies breeds based on specific biological types as follow: type N = composed by *Bos taurus indicus* breeds, i.e., Zebu breeds, mainly represented by the Nellore breed; type A = composed by *Bos taurus taurus* breeds adapted to the tropics, such as Senepol and Bonsmara; type B: animals from British origin (*Bos taurus taurus)* such as Angus, Devon, Hereford and Red Angus; and, type C: characterized by continental European breeds, including Charolais, Limousin, and Simmental.

Table 2.2 Descriptions of the total number of animals in the pedigree file according to their breed and biological type compositions by the NABC system.

Biological type[a]	Number of animals
Montana Tropical Composite[b]	315,483
Purebred animals[c]	57,838
Crossbred animals[d]	152,485
Total	**525,806**

[a]Biological types = N: Zebu breeds, A: breeds adapted to the Tropical regions, B: British breeds and C: Continental European breeds; [b]Montana Tropical Composite = 12.5% ≤N≥ 37.5%; 12.5% ≤A≥ 62.5%; 0% ≤B≥ 50%; 0% ≤C≥ 50%; 50% ≤NxA≥75%; 25% ≤BxC≥50%;
[c]Purebred animals = N, A, B or C ≥90%; [d]Combination of NABC types (NxA, NxB, NxC, AxB, AxC and BxC) = 50%.

2.2.2 Genotypes

A total of 1,436 animals (1,435 bulls and 1 cow) were genotyped using a medium-density SNP panel (GeneSeek® Genomic Profiler Bovine LD v4, Illumina Inc., San Diego, CA) containing 30,105 SNPs. The genotype quality control consisted of removing a total of 2,906 SNP markers: with minor allele frequency (MAF) lower than 0.05, call rate less than 90%, Hardy-Weinberg equilibrium defined by the maximum difference between observed and expected frequency of heterozygosis greater than 0.15 (WIGGANS et al., 2009), and located in non-autosomal chromosomes. In addition, 42 animals were excluded due to call rate lower than 0.90. A total of 27,199 SNPs and 1,394 animals remained for further analyses. Out of 1,394 genotyped animals 1,357, 1,392, 1,203, and 1,164 had measurements for BW, WW, YW and WG, respectively.

2.2.3 Genetic Parameter Estimation

The Restricted Maximum Likelihood (REML) approach, implemented in the AIREMLF90 software (MISZTAL et al., 2002; 2014), was used to estimate variance components. The linear model implemented for all growth traits (BW, WW, YW and WG) can be described as:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 a} + \boldsymbol{Z_2 m} + \boldsymbol{Wc} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is a vector of observations; $\boldsymbol{\beta}$ is the vector associated with the fixed effect of CG, corresponding to the biological type compositions for individual (A, B or C) and maternal (A_M, B_M or C_M) effects, also including the direct heterozygosity (N_xA, N_xB, N_xC, A_xB, A_xC and B_xC), the maternal total heterozygosity (H_M), and the age of the dam at first calving (for BW and WW) and age at weaning (for WW); $\boldsymbol{a}$ is a vector of random additive genetic effects; $\boldsymbol{m}$ is a vector of random maternal effects; $\boldsymbol{c}$ is a vector of random effects of maternal permanent environment and $\boldsymbol{e}$ is the vector of residual effects. The associated incidence matrices are $\mathbf{X}$, $\mathbf{Z_1}$, $\mathbf{Z_2}$, and $\mathbf{W}$. It was

assumed that E[**y**] = **Xb**; the direct additive genetic, maternal genetic, maternal permanent environmental, and residual effects were normally distributed with mean of zero and Var(**a**) = **H** $\otimes$ $\mathbf{S_g}$; Var(**m**)=**H** $\otimes$ $\mathbf{S_m}$; Var(**c**)=**H** $\otimes$ $\mathbf{S_c}$; and Var(**e**)=**I** $\otimes$ $\mathbf{S_e}$; in which $\mathbf{S_g}$ is the direct additive genetic (co)variance matrix; $\mathbf{S_m}$ is the maternal genetic (co)variance matrix, only for BW and WW; $\mathbf{S_c}$ is the maternal permanent environmental (co)variance matrix, only for BW and WW; $\mathbf{S_e}$ is the residual (co)variance matrix and **I** is an identity matrix. The covariance between direct and maternal genetic effects was set to zero. In order to predict the additive genetic effects corresponding to the biological types, each one was considered as a proportion in the genetic composition of the calf and its dam. The coefficients of direct (H_D) and maternal total heterozygosity (H_M) for each individual were derived using the linear relationship obtained by the subsequent equations:

$$\boldsymbol{H_D = 1 - \sum_{i=1}^{4} S_i \; x \; D_i;} \qquad\qquad \boldsymbol{H_M = 1 - \sum_{i=1}^{4} MGS_i \; x \; MGD_i,}$$

where ***4*** is the number of biological types (N, A, B, C); $\boldsymbol{S_i}$, $\boldsymbol{D_i}$, $\boldsymbol{MGS_i}$ and $\boldsymbol{MGD_i}$ are the fractions of the i^{th} biological type of sire, dam, maternal grandsire and maternal granddam, respectively.

The maternal direct effect was not included in the model for YW and WG. It is important to highlight that the biological type "N" was excluded from the statistical models for direct and maternal genetic effects to avoid multicollinearity issues, as the sum of the biological type fractions would be equal to one. Therefore, the effects of the biological types A, B, and C were estimated as deviations of the additive effects of N (DIAS et al., 2011). Direct heritability (h^2), maternal heritability (h^2_m), and total heritability (h_t^2) as defined by Willham (1972) were obtained by fitting a single-trait animal model, while genetic correlations were estimated using a multi-trait animal model. The single-step Genomic BLUP procedure (ssGBLUP; MISZTAL et al., 2009; AGUILAR

et al., 2010; CHRISTENSEN; LUND, 2010) was used for all analyses. The inverse of a hybrid genomic relationship matrix, $\mathbf{H^{-1}}$ (LEGARRA et al., 2014), was created as:

$$\boldsymbol{H^{-1}} = \boldsymbol{A^{-1}} + \begin{bmatrix} 0 & 0 \\ 0 & (\mathbf{0.95}\boldsymbol{G} + \mathbf{0.05}\boldsymbol{A_{22}} + \boldsymbol{\alpha^{-1}}) - \boldsymbol{\omega A_{22}^{-1}} \end{bmatrix}$$

where $\boldsymbol{A_{22}^{-1}}$ is the inverse of the numerator relationship matrix ($\boldsymbol{A^{-1}}$, based on pedigree) for genotyped animals and **G** is the genomic relationship matrix as described by VanRaden (2008):

$$\boldsymbol{G} = \frac{\boldsymbol{ZDZ'}}{\boldsymbol{q}}$$

Where, $\boldsymbol{q} = \mathbf{2}\sum \boldsymbol{p_i}(\mathbf{1} - \boldsymbol{p_i})$, in which $\boldsymbol{p_i}$ is the MAF of SNP$_i$, $\boldsymbol{Z}$ is a matrix of centered genotypes (-1, 0, 1); **D** is a matrix containing the SNP weights (initially $\mathbf{D} = \mathbf{I}$). Moreover, the scaling factors $\boldsymbol{\tau}$ and $\boldsymbol{\omega}$ were adjusted to reduce the inflation of predictions (KOIVULA et al., 2015) and also ensure the convergence of iterative approaches as mixed models (MISZTAL et al., 2010). These parameters were defined as τ = 1 and $\omega = 0.7$, based on preliminary analyses using the same datasets.

2.2.4 Estimation of Genetic Diversity Metrics

Genetic diversity metrics were estimated to better understand the genetic background of this population, which can affect genomic analyses such as GWAS. Linkage disequilibrium (LD) is a metric that indicates the degree of non-random association between markers or the inheritance and correlation between alleles within a population (HILL ; ROBERTESON, 1966, 1968; BUSH; MOORE, 2012). This is an important parameter for the interpretation of GWAS findings and implementation of genomic selection. The LD metric for each pair of SNPs was calculated using

r^2 (BOHMANOVA; SARGOLZAEI; SCHENKEL, 2010), as implemented in the PLINK v1.07 software (PURCELL et al., 2007).

To further analyze the relationship among all genotyped animals and population structure we performed a principal component analysis (PCA) using the PREGSF90 package (AGUILAR et al., 2014) based on the **G** matrix. Genotyped animals were highlighted in the plot based on the maximum proportion of each biological types (N, A, B, and C ≥ 50%). Traditional pedigree and genomic inbreeding (based on the **G** matrix) were also calculated for the genotyped animals.

2.2.5 Weighted Single-Step Genome-Wide Association Studies

Single-trait GWAS analyses were performed based on the Weighted single-step GWAS methodology (WssGWAS; WANG et al., 2012). The same statistical models described before were used. Analyses were performed using the BLUPF90 family software (MISZTAL et al., 2002; 2014). The "SNP moving average" option from the POSTGSF90 package (AGUILAR; MISZTAL; TSURUTA, 2014) was used to back solve the genomic breeding values and calculate SNP effects. Genomic windows of 10 consecutive SNPs were defined based on the average LD decay in the population. The SNP effects for additive and maternal genetic effects (a_g and m_g, respectively) were calculated based on the genomic breeding values for genotyped animals, following Wang et al. (2012):

$$a_g = \boldsymbol{Z_1} u_d,$$

$$m_g = \boldsymbol{Z_2} u_m,$$

where $\boldsymbol{Z}$ and $\boldsymbol{D}$ are matrices that relate genotypes of each locus, and u_d and u_m are vectors of effects of direct genetic and maternal genetic effects for each trait. The variances for direct and maternal genetic effects were assumed as:

$$Var(a_g) = Var(\mathbf{Z_1}u_d) = \mathbf{Z_1 D_1 Z_1'}\sigma^2_{u_d} = \mathbf{G}^*\sigma^2_a,$$

$$Var(m_g) = Var(\mathbf{Z_2}u_m) = \mathbf{Z_2 D_2 Z_2'}\sigma^2_{u_m} = \mathbf{G}^*\sigma^2_m,$$

in which $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$ are the diagonal matrix of weights for variances of markers for direct additive genetic and maternal genetic effects, respectively. $\sigma^2_{u_d}$ and $\sigma^2_{u_m}$ are the direct and maternal genetic variances captured by each SNP marker when there is similar weight for all markers and $\boldsymbol{G}^*$ is the weighted **G** matrix. (Co)variance for direct and maternal genetic effects (a_g and m_g) and SNPs (u_d and u_m) were calculated as:

$$Var\begin{bmatrix} a_g \\ u_d \end{bmatrix} = \begin{bmatrix} \mathbf{Z_1 D_1 Z_1'} & \mathbf{Z_1 D_1'} \\ \mathbf{D_1 Z_1'} & \mathbf{D_1} \end{bmatrix}\sigma^2_{u_d}$$

$$Var\begin{bmatrix} m_g \\ u_m \end{bmatrix} = \begin{bmatrix} \mathbf{Z_2 D_2 Z_2'} & \mathbf{Z_2 D_2'} \\ \mathbf{D_2 Z_2'} & \mathbf{D_2} \end{bmatrix}\sigma^2_{u_m}$$

Where:

$$\boldsymbol{G}^* = \frac{Var(a_g)}{\sigma^2_a} = \frac{Var(\mathbf{Z_{u_d}})}{\sigma^2_a} = \mathbf{Z_1 D_1 Z_1'}\lambda$$

$$\boldsymbol{G}^* = \frac{Var(m_g)}{\sigma^2_m} = \frac{Var(\mathbf{Z_{u_m}})}{\sigma^2_m} = \mathbf{Z_2 D_2 Z_2'}\lambda$$

The λ was considered as:

$$\lambda_1 = \frac{\sigma^2_{u_d}}{\sigma^2_a} = \frac{1}{\sum_{i-1}^{M_d} 2p_i(1-p_i)},$$

$$\lambda_2 = \frac{\sigma^2_{u_m}}{\sigma^2_m} = \frac{1}{\sum_{i-1}^{M_m} 2p_i(1-p_i)},$$

In which λ_1 (genetic direct effect) and λ_2 (maternal genetic effect) are the variance ratio or normalization constant (VANRADEN et al., 2009). $\boldsymbol{M}$ is the total number of SNPs and p_i is the allele frequency of the alternative allele of the i^{th} SNP. The observed allele frequencies (based on

the genotyped animals) were used for the calculations. The marker effects were calculated according to Strandén and Garrick (2009):

$$\hat{u}_d = \frac{\sigma^2_{u_d}}{\sigma^2_a} \boldsymbol{D_1 Z_1' G}^{*-1} \hat{a}_g = \boldsymbol{D_1 Z_1' [Z_1 D_1 Z_1']^{-1}} \hat{a}_g$$

$$\hat{u}_m = \frac{\sigma^2_{u_m}}{\sigma^2_m} \boldsymbol{D_2 Z_2' G}^{*-1} \hat{m}_g = \boldsymbol{D_2 Z_2' [Z_2 D_2 Z_2']^{-1}} \hat{m}_g$$

Estimates of SNP effects can be used to calculate the individual variance of each SNP effect (ZHANG et al., 2010) and to apply a different weighting for each marker:

$$\hat{\sigma}^2_{u_d,i} = \hat{u}_{d_i}{}^2 2p_i(1-p_i),$$

$$\hat{\sigma}^2_{u_m,i} = \hat{u}_{m_i}{}^2 2p_i(1-p_i),$$

The iteration process used is described below considering $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$ to estimate the effect of SNPs and weights (WANG et al., 2012):

1. As a first step **D=I**;
2. Calculation of the matrix $\boldsymbol{G = Z_1 D_1 Z_1' q}$; $\boldsymbol{G = Z_2 D_2 Z_2' q}$;
3. Calculate GEBVs (predicted genomic value) for all animals in the dataset using ssGBLUP;
4. Calculate the effect of SNPs: $\hat{u}_d = \lambda \boldsymbol{D_1 Z_1' G^{*-1}} \hat{a}_g$; $\hat{u}_m = \lambda \boldsymbol{D_2 Z_2' G^{*-1}} \hat{m}_g$;
5. Calculate the variance of each SNP: $d_i = \hat{u}_i^2 2p_i(1-p_i)$, where *i is the i*th SNP;
6. Normalize the SNP weights to maintain the direct and maternal additive genetic variance constant;
7. Return to step 2.

The effects of the markers were obtained by three iterations (ZHANG et al., 2016) from step 2 to 7. The percentage of genetic variance explained by the i[th] genomic region was calculated for direct effect and maternal effects, as described by Wang et al. (2014):

$$\frac{Var(a_i)}{\sigma_a^2} =\times 100 = \frac{Var\left(\sum_{j=1}^{10} \boldsymbol{Z_j}\hat{u}_{d\,j}\right)}{\sigma_a^2} \times 100,$$

$$\frac{Var(m_i)}{\sigma_m^2} =\times 100 = \frac{Var(\sum_{j=1}^{10} \boldsymbol{Z_j}\hat{u}_{m\,j})}{\sigma_m^2} \times 100,$$

where a_i and m_i correspond to the genetic values of the i[th] genomic region consisting of 10 consecutive SNPs, σ_a^2 and σ_m^2 are the direct and maternal genetic variances, respectively, $\boldsymbol{Z_j}$ is a vector of genetic content of the SNP j[th] for all individuals and $\hat{u}_{\boldsymbol{j}}$ considers the j[th] marker effect within the i[th] genomic region.

2.2.6 Gene Prospection and Functional Enrichment Analyses

The Ensembl Genome Browser tool (www.ensembl.org/index.html) was used to search for the list of genes located in the genomic regions (± 0.25 Mb in each direction) that explained more than 1.0% of the total additive direct or maternal genetic variances. The new version of the cattle genome (ARS-UCD1.2) was used. Subsequently, the GeneCards tool (www.genecards.org/) was used to describe the functions of the annotated genes. The classification of genes regarding to their biological functions was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool v6.8 (https://david.ncifcrf.gov/; HUANG et al., 2009). The functional enrichment analysis for gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were also investigated and considered *p*-value ≤ 0.05 and False Discovery Rate (FDR) ≤ 5 as the significance threshold.

2.3 RESULTS AND DISCUSSION

2.3.1 Genetic Parameter Estimates

As shown in Table 2.1, the average (± SD) BW, WW, YW, and WG was 33.07 ± 4.44, 194.37 ± 37.42, 271.49 ± 51.01, and 92.14 ± 29.93, respectively. This indicates that there is large phenotypic variability for these traits in the Montana Tropical® Composite population. The variance components and genetic parameter estimates for BW, WW, YG and WG are presented in Table 2.3. All the traits were moderately heritable: 0.29 (BW), 0.30 (WW), 0.33 (YW) and 0.16 (WG). Mourão et al. (2007), using a subset of the Montana Tropical® Composite data born between 1994 to 2003 and performing a pedigree-based analysis, reported heritability estimates of 0.30, 0.26, 0.26 and 0.16 for BW, WW, YW, and WG, respectively. These estimates are similar to the ones obtained in our study using genomic information (**H** matrix) and indicate that direct selection will enable fast genetic progress for growth traits in the Montana Tropical® Composite population.

Table 2.3 Variance component and genetic parameters estimates for growth traits (Birth weight; BW, Weaning weight; WW, Yearling weight; YW, and Weight gain; WG) in the Montana Tropical® Composite population.

Trait	σ^2_a (SE)	σ^2_m (SE)	r_{am} (SE)	σ^2_c (SE)	σ^2_e (SE)	h^2 (SE)	h^2_m (SE)	h^2_t (SE)	c^2 (SE)
BW	5.14 (0.08)	0.48 (0.03)	-0.37 (0.02)	0.36 (0.03)	9.56 (0.06)	0.29 (0.01)	0.06 (0.01)	0.25 (0.03)	0.02 (0.01)
WW	220.74 (4.02)	48.79 (2.29)	-0.60 (0.04)	72.78 (2.11)	446.03 (2.77)	0.30 (0.01)	0.12 (0.01)	0.26 (0.02)	0.09 (0.01)
YW	172.03 (3.95)	-	-	-	515.15 (6.28)	0.33 (0.04)	-	-	-
WG	74.89 (4.10)	-	-	-	391.00 (3.62)	0.16 (0.01)	-	-	-

σ^2_a: Additive genetic variance; σ^2_m: Maternal genetic variance; r_{am}: correlation between additive and maternal genetic effects; σ^2_c: Permanent environmental variance; σ^2_e: Residual variance; h^2: Direct heritability; h^2_m: Maternal heritability; h^2_t: Total heritability; c^2: Permanent environmental effect; SE: standard error.

The total heritability (h^2_t) had a lower magnitude compared to the direct heritability for BW and WW, mainly due to the negative correlation between direct and maternal additive effect (Table 2.3). This genetic antagonism has also been reported in various studies in purebred animals from the breeds used in the development of the Montana Tropical® Composite (e.g., Brangus, -0.29; Gelbvieh, -0.21; Hereford, -0.28; Simmental, -0.32; BIF, 1996). Various studies have debated the cause of this antagonism, from the biological to the statistical aspects involved in the modeling of these traits and effects (LEE; POLLAK, 1997; 2002). These authors also tested models considering the sire versus year interactions. They concluded that ignoring this effect could inflate the correlation between direct additive and maternal effect, but the negative correlation was still obtained. In agreement with that, Grigoletto et al. (2016) working with Nellore cattle, recommended to use models that consider the covariance between these genetic effects.

The genetic relationship among all 4 traits was also investigated. The highest genetic correlation was obtained between BW and WW (0.88±0.08), followed by BW and YW (0.87±0.08); WW and YW (0.87±0.08); YW and WG (0.73±0.07); BW and WG (0.64±0.07); and, WW and WG (0.60±0.06). These findings are also in agreement with Santana et al. (2013), who reported a genetic correlation of 0.94±0.01 for YW and WG in a Montana Tropical® Composite population. The positive and favorable genetic correlations observed indicate that indirect responses are expected by including any of the traits in a selection index. However, as the genetic correlations differ from the unit, there is a value by directly selecting for all of them simultaneously.

2.3.2 Genetic Diversity Metrics

Figure 2.2 shows the LD decay in the Montana Tropical® Composite up to 1,000 kb. The average r^2 across all marker pairs and chromosomes was 0.16±0.01 (Table S2.1). Meuwissen et al. (2001) indicated that an r^2 equal or higher than 0.2 would be desirable for accurate genomic selection. However, populations with similar LD values have still been successfully implemented genomic selection, considering that a large enough training population is assembled.

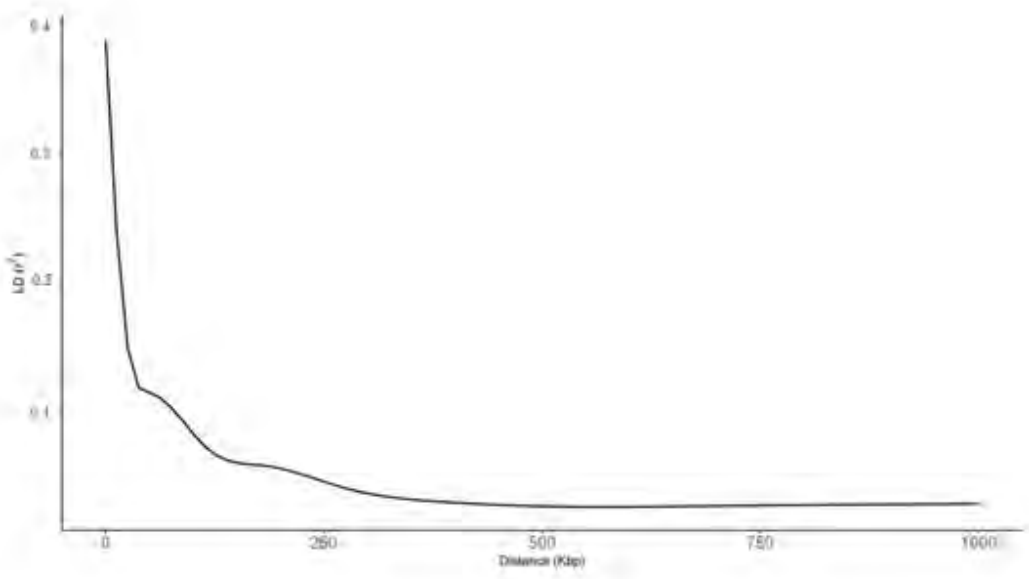


Figure 2.2 Linkage disequilibrium (r^2) decay on Montana Tropical® Composite population in autosomal chromosomes.

Porto-Neto et al. (2014) reported that at short distances (<10 kb), Composite populations (Tropical Composite, Santa Gertrudis and Belmont Red) had intermediate (r^2=0.32) LD levels when compared with pure breeds of *Bos taurus taurus* (r^2 = 0.45) and *Bos taurus indicus* (r^2=0.25). These LD estimates are slightly higher compared to the Montana Tropical Composite population

(Figure 2.2) and thus, a larger number of markers for GWAS analyses would be desirable and lower accuracy of genomic predictions would be expected.

Figure 2.3 shows a plot of the first two principal components (PC1 and PC2) of the **G** matrix, which explained only 3.4% and 2.6% of the total genetic variability, respectively. This indicates that there is not very clear population stratification. Furthermore, individual and maternal breed composition were considered by the statistical models to correct for any potential breed effects on the traits. It is worth noting that the largest proportion of animals were from the biological type A and no animals had more than 50% of the biological type C. The pedigree level of inbreeding was 0.001 for all the animals and 0.01 for the genotyped animals. The average genomic inbreeding (based on the **H** matrix) for the genotyped animals was ~0.0001. As expected, inbreeding is not an issue for this composite population.

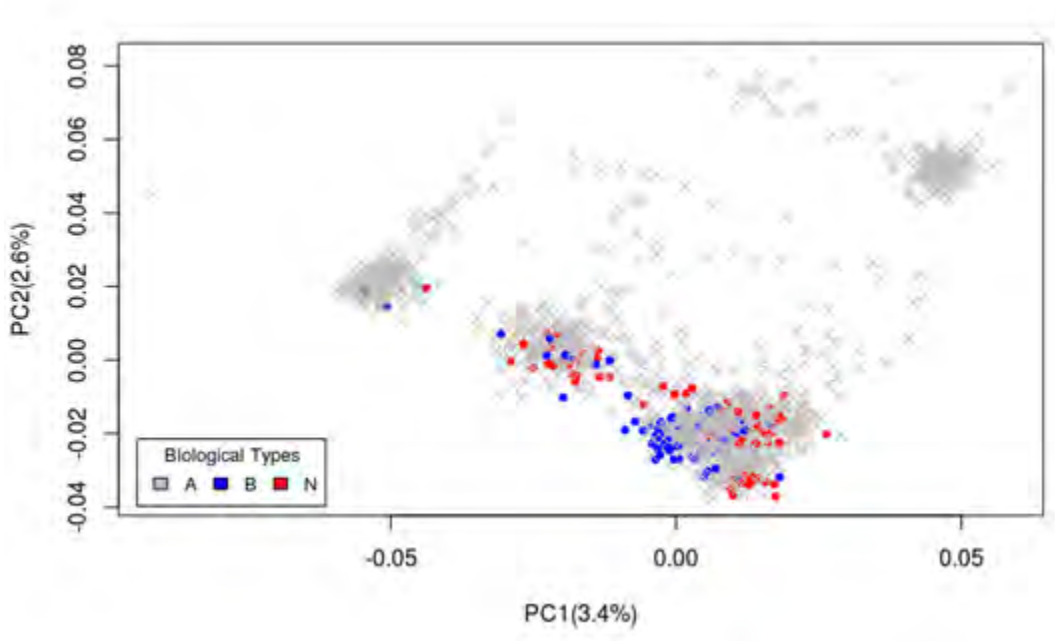


Figure 2.3 Animals clustered on the basis of the first two principal components (PC1 and PC2) of the genomic relationship matrix.

2.3.3 Genomic Regions and Gene Prospection

The genomic windows of 10 SNPs were defined based on the LD decay in the population and average distance between SNPs. Other studies (e.g. BRAZ et al., 2019) reported advantages in using windows with more than 6 SNPs to better capture QTL effects. This support our strategy to use a moving window containing 10 adjacent SNPs. A total of 21, 20, 24 and 18 genomic windows explaining more than 1 of the total genetic variances were identified for additive direct effects in BW, WW, YW and WG (Tables S2.2 to S2.5). For maternal genetic effect a total of 18 and 11 were obtained for BW and WW, respectively (Tables S2.2 and S2.3). The proportion of the total variance explained by each genomic window ranged from 1.0% to 10.79%. Overlapping regions were found at chromosomes BTA1, BTA3, BTA6, BTA7, BTA9, BTA10, BTA11, BTA16, BTA19, BTA21, BTA22, BTA23 and BTA29.

2.3.4 Birth Weight – Direct Genetic and Maternal Effects

A total of 21 and 18 genomic windows were identified for direct and maternal effects of BW, respectively (Figure 2.4). The regions associated with direct and maternal effects contained 24 genes in common: *GRB10, DDC, FIGNL1, IKZF1, SPATA48, ZPBP*, and *VWC2* located on BTA4; *DIAPH3, IRS2, COL4A1, COL4A2, RAB20, NAXDCA, RS2ING1,* and *ANKRD10* located at BTA12; *BPIFA2, BBPIFA3, BPIFA1, BPIFB1, BPIFB5, CDK5, RAP1*, and *SNTA1* located at BTA13; *DUSP8,* and *MOB2* located at BTA29. These are all important candidate genes to the traits investigated here. For instance, the Substrate Insulin Receptor 2 gene (*IRS2*) has been linked to milk production and milk solid composition in Gyr cattle (i.e. Zebu breed; PERIPOLLI et al., 2018), and thus was linked to maternal genetic effects for BW. Furthermore, this gene mediates

effects of insulin which has been shown to have depressing effect on the levels of blood glucose and milk production in dairy cows (SCHMIDT, 1966).

Another gene related with insulin receptors (MAGEE et al., 2010) and insulin-like growth factor receptors is the Growth Factor Receptor-bound Protein 10 gene (*GRB10*). Maternally Expressed gene 1 (*MEG1*) has also been shown to be involved with dynamically regulated by lactation and energy balance in dairy cows (RHOADS et al., 2008). The Dual Specificity Phosphatase 8 gene (*DUSP8)* belong to the DUSP protein family, which has been suggested to control and modulate the Mitogen-activated Protein Kinase (MAPK) signaling and associated with innate and adaptive immune effector functions. MAPK has been reported to play a function in oocyte quality and developmental (SUDIMAN et al., 2014; LIN et al., 2014; MOUSSA et al., 2015). As described by Noordman et al. (2006), MAPK has an important role in growth factors involved in cellular proliferation and differentiation processes. Furthermore, genes in the MAPK signaling have also been associated to feed efficiency in beef cattle (SERÃO et al., 2013).

The Insulin-like Growth Factor Binding Protein 3 (*IGFBP-3*) gene, one of the six binding proteins, mapped to BTA4 is involved in adipogenesis process, cell replication, proliferation, differentiation, protein synthesis, carbohydrate homeostasis and bone metabolism. It also plays a crucial role as a carrier of 80% of circulating *IGF-1,* and this action of *IGFBP-3* is regulated by growth hormone (ELLOUMI et al., 2005). Nevertheless, *IGFBP-3* has been associated with backfat thickness in Canchim beef cattle (i.e. a composite beef cattle breed developed in Brazil; VENERONI et al., 2010), and marbling in Wagyu beef cattle (MIZOGUCHI et al., 2005). In goats, the *IGFBP-3* has been reported to have a direct role on litter size and body weight traits (LAN et al., 2007).

The chromosomal region identified on BTA14 harbors various important genes associated with growth and other productive traits in cattle. This include *TOX, PLAG1*, *MRPL15*, *RGS20*, *FAM110B*, *RP1*, *TCEA1*, *UBXN2B*, and *RPL39*. For instance, the Pleiomorphic Adenoma 1 gene (*PLAG1*) has been extensively associated with cattle growth rates and body size (TAKASUGA, 2016; UTSUNOMIYA et al., 2017; XU et al., 2018). In addition, the *TOX* gene has been associated with carcass weight (LEE et al., 2013) and identified as a key transcription factor responsible for the molecular regulation of puberty in Brahman cattle (FORTES et al., 2012).

An interesting biological process linked to direct genetic effects of BW is long-term memory (GO:0007616; *p*-value=0.001; FDR=2.7), which is related to learning, associative learning, visual and olfactory learning. Moreover, there are biological processes linked to behavioral traits (GO:0007610; *p*-value=0.039; FDR=51), including feeding, visual ability, immune response, locomotion, and chemosensory behaviors. These processes are crucial to pre-weaning development (HULBERT et al., 2011).

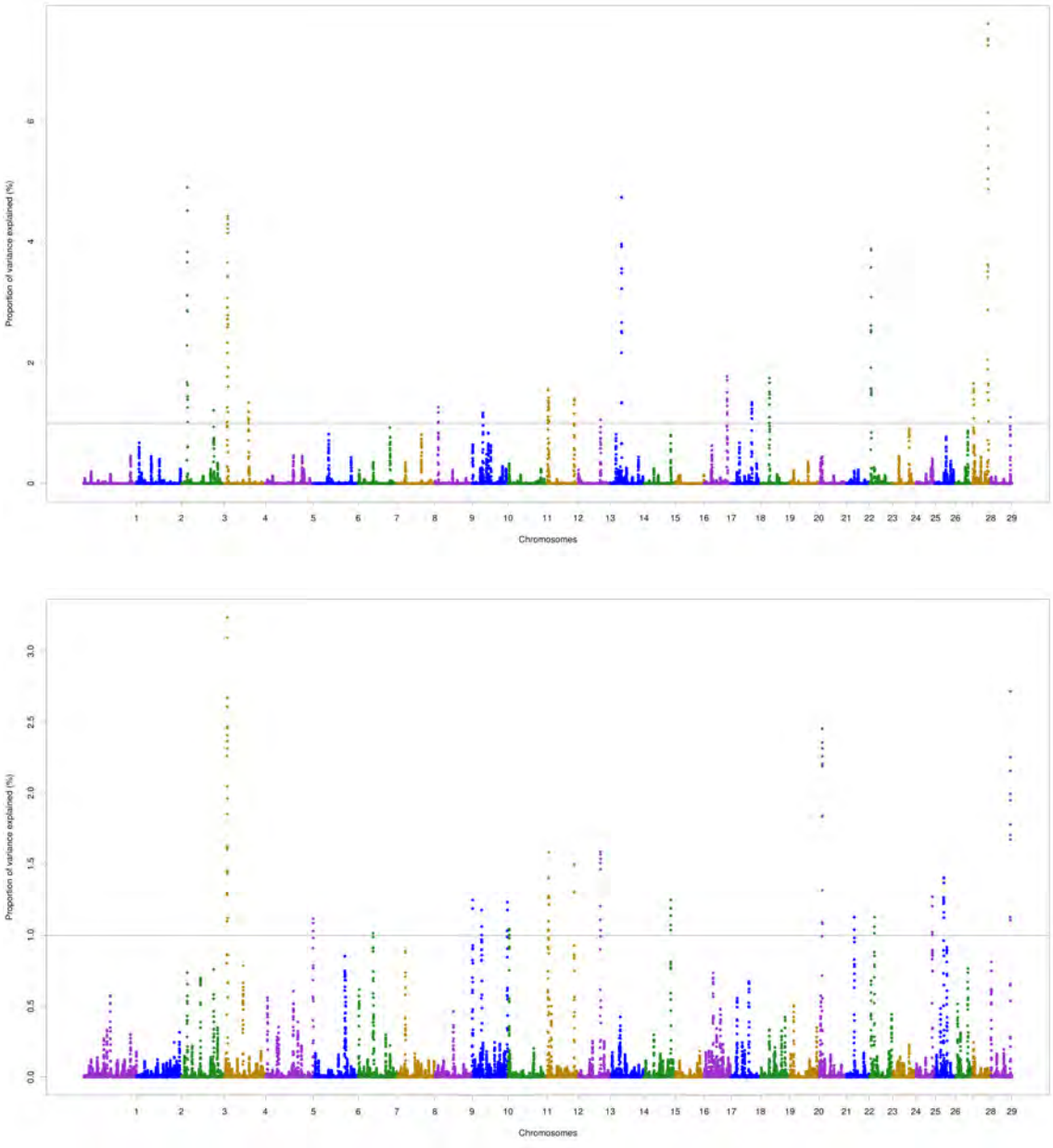


Figure 2.4 Manhattan plots of moving average windows of 10 adjacent SNP explained by the additive genetic variance for Birth weight (BW) on additive direct animal (top) and maternal effects (bottom). The gray line represents the genetic variance proportion threshold 1.0%.

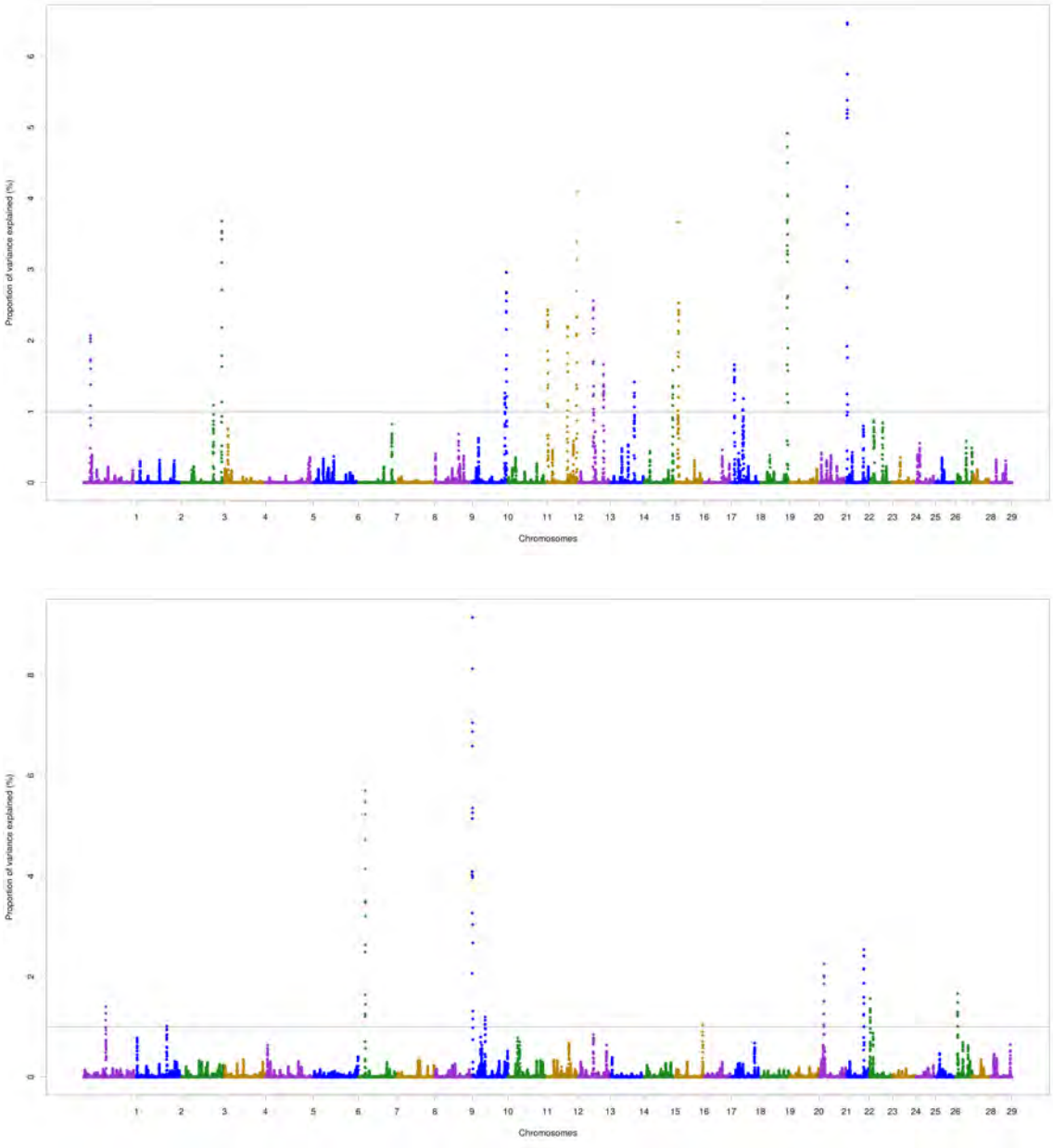


Figure 2.5 Manhattan plots of moving average windows of 10 adjacent SNP explained by the additive genetic variance for Weaning weight (WW) on additive direct animal (top) and maternal (bottom) effects. The gray line represents the genetic variance proportion threshold 1.0%.

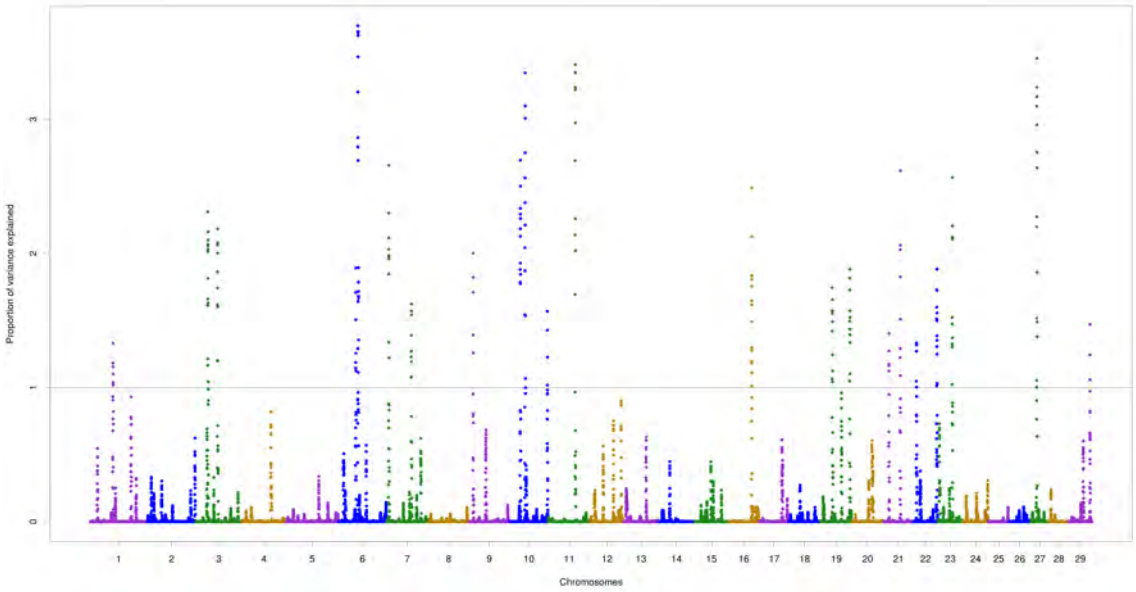


Figure 2.6 Manhattan plots of moving average windows of 10 adjacent SNP explained by the additive genetic variance for Yearling weight (YW) on additive direct animal effects. The gray line represents the genetic variance proportion threshold 1.0%.

2.3.5 Weaning Weight – Direct Genetic and Maternal Effects

Most of the genes linked to growth traits are associated with muscle, bone and lipid metabolism. A total of 20 and 11 genomic regions located at BTA1 to BTA27 were associated with WW for direct and maternal effects, respectively (Figure 2.5; Table S2.3). Genes located at BTA18, such as the *WWOX* gene (RAMAYO-CALDAS et al., 2016), have been associated with postnatal survival and maintenance of proper cell growth and metabolism, and meat quality traits in beef cattle (LEE et al., 2018). Another positional gene is *PRKAA2*, which has been shown to be involved in glucose and lipid homeostasis and protein synthesis, as well as in the regulation of energy intake and body weight (ZHANG et al., 2011). Differences in growth efficiency have been attributed to body composition. For example, differences in muscle and fat deposition influence the rate and efficiency of body weight gain (BULLE et al., 2007).

For WW maternal genetic effects, we identified various genes from the Phospholipase A family (*PLA2G4F*, *PLA2G4B*, *PLA2G4E*, *PLA2G4D*), which were previously associated with a number of physiological activities. This includes: lipid metabolism (BEER; WEBB, 2006), cell proliferation (STARKL et al., 2013), muscle contraction (JOHANSSON; THESLEFF, 1968) and inflammation process (CLARK et al., 1995). Interestingly, the *PLA2G4E* gene has also been associated with economically important traits in livestock species, including mastitis resistance, post-partum uterine disease, feed efficiency, embryonic development and survival to adult age (FONSECA et al., 2018; SEROUSSI et al., 2013; SHELDON et al., 2009; MERCADANTE et al., 2015). The *PLA2G4E* gene also plays a role on thyroid hormones metabolism (SHARMA et al., 2014) and thus, in the differentiation, growth, metabolism, and physiological functions involved in body development (KLIMIEN et al., 2008). The *PRG3* gene was previously identified in a selective sweep region for tropical environmental adaptation traits in indicine cattle (TAYE et al., 2017). This gene (*PRG3*) was also identified as a Copy Number Variation, associated with animal resilience traits, in a study including several taurine and indicine cattle breeds (BICKHART et al., 2016).

The genomic window located on BTA15, contains olfactory receptor genes (*OR5B17* and *OR6Q1*). These genes are hypothesized to play a role as the olfactory receptors associated with feeding behavior (LINDSTEDT, 1971) and immune system through major histocompatibility complex (MHC) genes (SPEHR et al., 2006). These genes could be closely associated with physiological and anatomical traits that have evolved as a result of natural selection. Adaptation has a great impact on their ability to grow, reproduce and be productive (BONSMA, 1949). Important pathways were enriched for maternal genetic effects (Figure 2.7). For instance, the GnRH signaling pathway plays a role in the characterization of the differential stages of growing

cattle until puberty (WIDMANN et al., 2013). Thus, this is considered a crucial factor at the weaning stage due to its association with male sexual maturation in cattle (OJEDA et al., 2006; 2010; RAWLINGS et al., 2008).

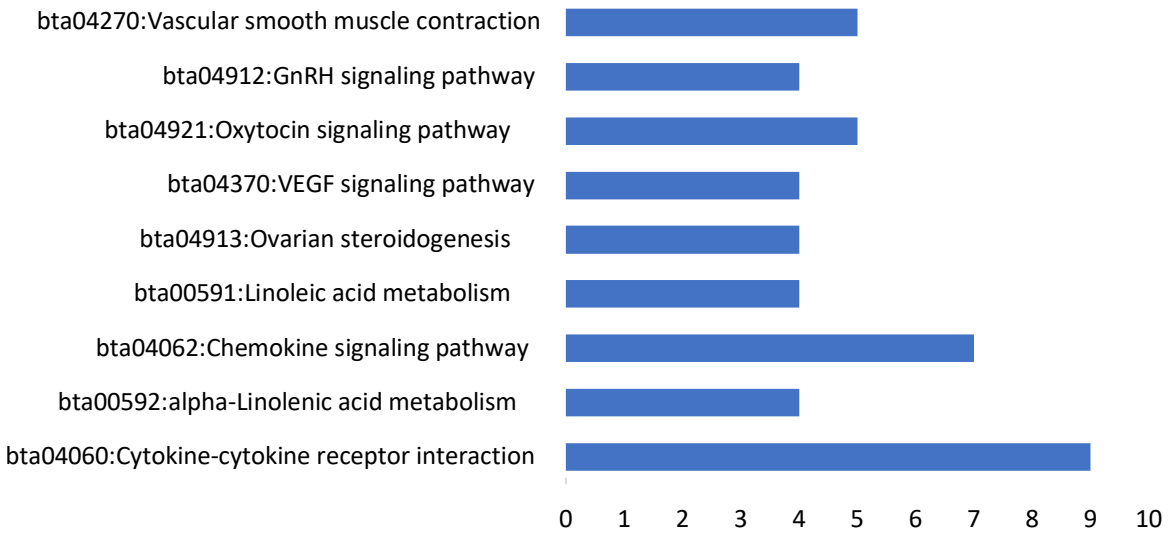


Figure 2.7 Kyoto Encyclopedia of Genes and Genomes pathways (blue) considered as threshold p-value ≤ 0.05 and False Discovery Rate as FDR ≤ 5 for maternal effects identified to WW in Montana Tropical® Composite.

2.3.6 Yearling Weight – Direct Genetic Effect

We identified 24 genomic regions associated with YW in BTA1, BTA3, BTA6, BTA7, BTA9, BTA10, BTA11, BTA16, BTA19, BTA21, BTA22, BTA23, BTA27 and BTA29 (Figure 2.6; Table S2.4). Mcclure et al. (2010) reported QTLs highly associated with YW, which were also identified in this study (BTA7:67.3-69.0 Mb). The genomic region located at BTA27 (15.0-15.2 Mb; Figure 2.6) contains the Long-chain Acyl-CoA Synthetase 1 (*ACSL1*) gene, previously associated with fatty acid metabolism, especially in the bovine skeletal muscle (WIDMANN et al., 2011). Similar findings were reported in pigs (LI et al., 2012).

The *ABCA5, ABCA6,* and *ABCA10* genes (part of the ATP-binding cassette transporters) were included on significant biological processes associated with weight gain and carcass fat deposition (BALDWIN et al., 2012). Important biological processes (P-value ≤ 0.05) were found to be associated with lipid transport, fatty acid metabolic process, and lipoxin metabolic process (Table S2.5). At the yearling stage, fat deposition has a greater increase compared to the growing/weaning stage, where the development of muscle tissue is occurring at a higher rate (ROUSE et al., 2003).

2.3.7 Post-Weaning Weight Gain – Direct Genetic Effect

A total of 18 genomic regions located on BTA1, BTA5, BTA6, BTA7, BTA10, BTA12, BTA16, BTA17, BTA19, BTA22, and BTA25 (Table S2.6; Figure 2.8) were found to be associated with WG. The genomic region located at BTA17 (72.8-72.9Mb) harbors the *GNB1L* and *COMT* genes, previously associated with social behavior (QAYYUM et al., 2015). The genomic region on BTA16 harbors the gene *USH2A,* which is known to be involved in the nervous system, especially with vision and hearing functions (FRIEDMAN et al., 2007). The sensory perception of sound is critical at birth and weaning phase to the survival of calves and their behavior. Vocalization is crucial for calves recognizing their mothers (WATTS; STOOKEY, 2000; MORA-MEDINA et al., 2016). As BW and WW are genetically correlated with WG, the identification of this gene might be associated with cumulative effects of better feeding (milk-based).

The gene Wnt inhibitory factor 1 (*WIF1)* located at BTA5 has been described as having a crucial role on the extent of fat cell differentiation. Results showed by Wu et al. (2015) suggested imprinting on the expression of *WIF1* with muscle and subcutaneous fat tissues in cattle. Fat deposition in the WG phase is crucial for satisfactory performance, as precocity of the animal is

associated with earlier fat deposition. This highlights the complementary of *Bos taurus taurus* breeds in the development of the Montana Tropical® Composite for improving meat quality.

The highest peak observed in this study was located at BTA22 spanning 54.8 to 55.6 Mb and accounting for 10.79% of the total genetic variance. This region contains the genes: *TGFBR2*, *GADL1*, *STT3B*, *OSBPL10*, *ATP2B2*, *ATG7*, *VGLL4*, and *TAMM41*. The *ATP2B2* has been previously associated with lactation phenotypes (e.g. milk yield and mastitis) in dairy cattle (OGOREVC et al., 2009). The Oxysterol Binding Protein-like 10 (*OSBPL10*) gene has been associated with lipid binding since is expressed in subcutaneous adipocytes (SHENG; NI; LIU, 2014). In addition, the Autophagy Related 7 (*ATG7*) has been linked to skeletal muscle growth in beef cattle (NAKANISHI et al., 2019).

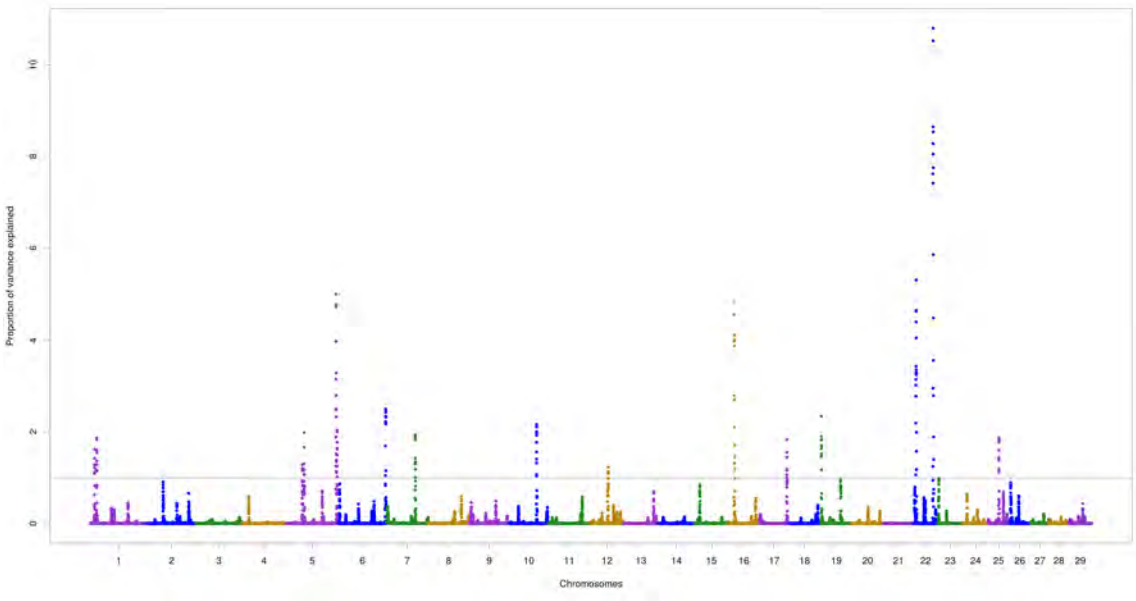


Figure 2.8 Manhattan plots of moving average windows of 10 adjacent SNP explained by the additive genetic variance for Post weaning weight gain (WG) on additive direct animal effect. The gray line represents the genetic variance proportion threshold 1.0%.

In particular, we can highlight some functional genes identified in WW and WG involved on social behavior such as: *OR5B17*, *OR6Q1, COMT, and USH2A* (LEE et al., 2013; QAYYUM

et al., 2015). The olfactory receptors (ORs) genes could influence on animal behavior. In mice, these genes have been reported to promote, via olfactory system, the attraction or aversion of differences odor preferences (HORIO et al., 2019). The maternal genetic effects for BW and WW had seven genes in common (*MCC*, *DCP2*, *REEP5*, *SRP19*, *APC*, *PNKD*, and *CHD2*) located at BTA2, BTA10 and BTA21. These genes have been previously associated with carcass traits (YOON; KO, 2016); reproduction (WALSH et al., 2011). Another important point is the difference in maternal ability between cows from *Bos taurus indicus* and *Bos taurus taurus* breeds. As mentioned by Gregory et al. (1985), cows sired by Angus (i.e. *Bos taurus taurus*) and Red Poll (i.e. *Bos taurus taurus*) bulls significantly outperformed cows sired by Boran (i.e. *Bos taurus indicus*) bulls, with regards to WW of the progeny and cow weight at calving and weaning. Thus, there is strong evidence that both the Angus and Red Poll breeds may have slightly higher additive maternal genetic effects.

2.4 GENERAL DISCUSSION AND IMPLICATIONS

In this study, we reported that there is enough phenotypic variability to implement genetic and genomic selection for growth traits in the Montana Tropical® Composite. All the traits were moderately heritable and thus, genetic progress is expected through direct and indirect selection. Most of the genetic correlations were favorable, except those with BW, in which there is no interest on further increasing BW of calves. As the genetic correlations between all traits were moderate, if balanced in a selection index, gains can be simultaneously obtained in all of them. As expected, the levels of inbreeding in this population are low. This is in part because of the incomplete pedigree (only 74 % of animals have both parents known) and also due to the high levels of crossbreeding in the population.

The levels of LD are comparable to other composite populations. Despite the fact that the values are in a lower range of the desirable values for implementing genomic selection and GWAS, they are still within acceptable ranges (average > 0.15). With large enough training populations, accurate genomic predictions can be obtained. With regards to GWAS studies, denser panels could yield better results (e.g. markers closer to the true QTLs). In this study, various genomic regions associated with growth traits were identified. Some of these regions had been described before, while novel ones were also reported. By investigating the biological pathways in which the candidate genes are involved, we have a better understanding of the biological processes that are potentially been affected by implementing genetic selection for these traits. It is worth noting that the number of genotyped animals and SNP markers included in this study is reduced compared to other GWAS studies in the literature (e.g., TWOMEY et al., 2019). However, this is the first GWAS report in the Montana Tropical® Composite and our results will serve as a basis for future studies including larger sample sizes.

The reduced number of genotyped animals could have resulted in spurious associations. However, many genomic regions identified in this study had been previously reported in other beef cattle studies and are biologically reasonable. Population structure is another factor that can affect GWAS studies (ASTLE; BALDING, 2009). Despite the diverse genetic composition of the population under study, the principal component analyses did not clearly cluster the population in separate groups. Furthermore, breed effects were accounted for by including individual and maternal breed covariables in the model and fitting the genomic relationship matrix.

The genomic regions identified here can be incorporated in online databases (e.g AnimalQTLdb, HU et al., 2006) and this information can be useful when implementing genomic selection approaches considering more important regions. Some examples of such methods are

weighted ssGBLUP, Bayesian approaches and inclusion of biological information in genomic prediction methods (NEVES et al., 2014; EDWARDS et al., 2016; MACLEOD et al., 2016; FANG et al., 2017; ABDOLLAHI-ARPANAHI et al., 2017; FIKERE et al., 2018).

2.5 CONCLUSIONS

Birth, weaning and yearling weights and post-weaning weight gain are moderately heritable and thus, genetic progress is expected through direct and indirect selection. Most of the genetic correlations were favorable, except those with BW. The levels of inbreeding in this population are very low, due to incomplete pedigree and high levels of crossbreeding. The levels of linkage disequilibrium are comparable to other composite populations in the lower end (average r^2: 0.16). Various genomic regions associated with growth traits were identified. Some of these regions had been described before, while novel ones were also reported. The genomic regions identified here can be incorporated in online databases and this information can be useful when implementing genomic selection.

2.6 REFERENCES

ABDOLLAHI-ARPANAHI, R.; MOROTA, G.; PEÑAGARICANO, F. Predicting bull fertility using genomic data and biological information. Journal of dairy science, v. 100, n. 12, p. 9656-9666, 2017. Disponível em: <https://doi.org/10.3168/jds.2017-13288

AGUILAR, I. et al. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science**, v. 93, n. 2, p. 743–752, 2010. Disponível em: <http://dx.doi.org/10.3168/jds.2009-2730>.

AGUILAR, I.; MISZTAL, I.; TSURUTA, S. PREGSF90–POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped. **10th World Congress on Genetics Applied to Livestock Production**, n. August, p. 90–92, 2014. Disponível em: <https://www.asas.org/docs/default-source/wcgalp-posters/680_paper_9756_manuscript_1666_0.pdf?sfvrsn=2>.

ANTONIO, R. et al. Multicollinearity in genetic effects for weaning weight in a beef cattle composite population. **Livestock Science**, v. 142, n. 1–3, p. 188–194, 2011. Disponível em: <http://dx.doi.org/10.1016/j.livsci.2011.07.016>.

BALDWIN, R. L. et al. Characterization of the longissimus lumborum transcriptome response to adding propionate to the diet of growing Angus beef steers. **Physiological Genomics**, v. 44, n. 10, p. 543–550, 2012.

BEER, F. C. De; WEBB, N. R. Inflammation and Atherosclerosis Group IIa and Group V sPLA 2 Are Not Redundant. **Arterioscler Thromb Vasc Biol.**, v. 2, p. 1421–1422, 2006.

BICKHART, D. M. et al. Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. v. 23, n. April, p. 253–262, 2016.

BEEF IMPROVEMENT FEDERATION. Guidelines for uniform beef improvement programs. Beef Improvement Federation, 2010.

BOHMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. S. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**, v. 11, n. 1, 2010.

BRAZ, C. U. et al. Sliding window haplotype approaches overcome single SNP analysis limitations in identifying genes for meat tenderness in Nelore cattle. **BMC Genetics**, v. 20, n. 8, p. 1–12, 2019.

BREED, A. A.; SCHMID, M.; BENNEWITZ, J. Invited review : Genome-wide association analysis for quantitative traits in livestock – a selective review of statistical models and experimental designs. n. 2016, p. 335–346, 2017.

BRITO, L. F. et al. Genetic diversity of a New Zealand multi-breed sheep population and composite breeds ' history revealed by a high-density SNP chip. p. 1–11, 2017.

BULLE, F. C. P. C. et al. Growth , carcass quality , and protein and energy metabolism in beef cattle with different growth potentials and residual feed intakes Growth , carcass quality , and protein and energy metabolism in beef cattle with different growth potentials and residua. n. May, 2007.

BUSH, W. S.; MOORE, J. H. Chapter 11 : Genome-Wide Association Studies. v. 8, n. 12, 2012.

BUZANSKAS, M. E. et al. Genome-Wide Association for Growth Traits in Canchim Beef Cattle. **PLoS ONE**, v. 9, n. 4, 2014.

CHRISTENSEN, O. F.; LUND, M. S. Genomic relationship matrix when some animals are not genotyped Genomic prediction models. **Genetics Selection Evolution**, v. 42, n. 2, 2010.

CLARK, J. D. et al. Cytosolic phospholipase A 2. **J. Lipid Mediators Cell Signalling**, v. 12, p.

83–117, 1995.

COLE, J. B.; SILVA, M. V. G. B. Genomic selection in multi-breed dairy cattle populations. **Revista Brasileira de Zootecnia**, v. 45, n. 4, p. 195-202, 2016.

CRISPIM, A. C. et al. Multi-Trait GWAS and New Candidate Genes Annotation for Growth Curve Parameters in Brahman Cattle. n. 167, p. 1–19, 2015.

FONSECA, P. et al. Functional Characterization of Genes Mapped in Transmission Ratio Distortion Regions of the Bovine Genome Affecting Reproduction . n. April, p. 2016–2018, 2018.

FORTES, M. R. S. et al. Finding genes for economically important traits : Brahman cattle puberty. v. 15, p. 143–150, 2012.

GREGORY, K. E. et al. Characterization of breeds of bos indicus and bos taurus cattle for maternal and individual traits. **Journal of Animal Science**, v. 60, n. 5, p. 1165–1174, 1985.

GRIGOLETTO, L. et al. Inclusion of cytoplasmic lineage effect and direct-maternal genetic covariance for genetic evaluation of growth traits in Nellore cattle. **Genetics and Molecular Research**, v. 15, n. 3, p. 1–9, 2016.

GROSSI, D. A. et al. Genetic associations between accumulated productivity , and reproductive and growth traits in Nelore cattle. v. 117, p. 139–146, 2008.

HAY, E. H.; ROBERTS, A. Genomic prediction and genome-wide association analysis of female longevity in a composite beef cattle breed 1 , 2. n. May, p. 1467–1471, 2018.

HUANG, D. W.; SHERMAN, B. T.; LEMPICKI, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. **Nature Protocols**, v. 4, n. 1, p. 44–57, 2009.

HULBERT, L. E. et al. Innate immune responses of temperamental and calm cattle after transportation. **Veterinary Immunology and Immunopathology**, v. 143, n. 1–2, p. 66–74, 2011.

JOHANSSON, P.; THESLEFF, S. A comparison of the effects of phospholipase c and tetrodotoxin on spike generation in muscle. **European Journal of Pharmacology**, v. 4, p. 347–348, 1968.

JR, M. L. S. et al. Genetic associations among average annual productivity , growth traits , and stayability : A parallel between Nelore and composite beef cattle 1. n. 1996, p. 2566–2574, 2014.

KLIMIEN, I. et al. Metabolic changes of thyroid hormones in cattle. Review. **Veterinarija ir zootechnika**, v. 42, n. 64, p. 56115, 2008.

KOIVULA, M. et al. Single-step genomic evaluation using multitrait random regression model and test-day data. **Journal of Dairy Science**, v. 98, n. 4, p. 2775–2784, 2015.

LAN, X. Y. et al. The novel SNPs of the. **Arc. Tierz**, v. 50, n. 2, p. 223–224, 2007.

LEE, C.; POLLAK, E. J. Relationship between sire x year interactions and direct-maternal genetic correlation for weaning weight of Simmental cattle The online version of this article , along with updated information and services , is located on the World Wide Web at : Relations. **Journal of Animal Science**, v. 75, p. 68–75, 1997.

LEE, C.; POLLAK, E. J. Genetic antagonism between body weight and milk production in beef cattle. **Journal of Animal Science**, v. 80, n. 2, p. 316–321, 2002.

LEE, S. et al. Genome-wide association analysis to identify QTL for carcass traits in Korean native cattle. **Journal of Animal Science**, v. 96, p. 516, 2018.

LEE, S. H. et al. Genome-Wide Association Study Identifies Major Loci for Carcass Weight on

BTA14 in Hanwoo (Korean Cattle). v. 8, n. 10, 2013.

LEGARRA, A. et al. Single Step , a general approach for genomic selection $. **Livestock Science**, v. 166, p. 54–65, 2014. Disponível em: <http://dx.doi.org/10.1016/j.livsci.2014.04.029>.

LI, Q.; TAO, Z.; SHI, L. Expression and genome polymorphism of ACSL1 gene in different pig breeds. **Mol Biol Rep**, n. 40, p. 8787–8792, 2012.

LIN, Z. L. et al. Effects of growth differentiation factor 9 and bone morphogenetic protein 15 on the in vitro maturation of porcine oocytes. **Reproduction in Domestic Animals**, v. 49, n. 2, p. 219–227, 2014.

LINDSTEDT, K. J. Chemical control of feeding behavior. **Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics**, v. 39, p. 553–581, 1971.

MAGEE, D. A. et al. DNA sequence polymorphisms in a panel of eight candidate bovine imprinted genes and their association with performance traits in Irish Holstein-Friesian cattle. p. 1–15, 2010.

MARTÍNEZ, R. et al. Genome-wide association study for birth , weaning and yearling weight in Colombian Brahman cattle. v. 459, p. 453–459, 2017.

MCCLURE, M. C. et al. A genome scan for quantitative trait loci influencing carcass , post-natal growth and reproductive traits in commercial Angus cattle. p. 597–607, 2010.

MERCADANTE, V. R. G. et al. Effects of anti-phospholipase A 2 antibody supplementation on dry matter intake feed efficiency , acute phase response , and blood differentials of steers fed forage- and grain-based diets 1. **J Animal Science**, v. 93, p. 776–785, 2015.

MISZTAL, I. et al. Choice of parameters for single-step genomic evaluation for type. **Journal of Dairy Science**, v. 93, n. August, p. 166, 2010.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of Dairy Science**, v. 92, n. 9, p. 4648–4655, set. 2009.

MISZTAL, I. et al. BLUPF90 and related programs (BGF90). In: **Proceedings of the 7th world congress on genetics applied to livestock production**. 2002. p. 743-744.

MISZTAL, I.; TSURUTA, S.; LOURENCO, D.; AGUILAR, I.; LEGARRA, A.; VITEZICA, Z. Manual for BLUPF90 family of programs. **Athens: University of Georgia**, 2014. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

MIZOGUCHI, Y. et al. Mapping of quantitative trait loci for carcass traits in a Japanese Black (Wagyu) cattle population. **Animal Genetics**, v. 37, n. 1994, p. 51–54, 2005.

MORA-MEDINA, P. et al. Sensory factors involved in mother-young bonding in sheep: A review. **Veterinarni Medicina**, v. 61, n. 11, p. 595–611, 2016.

MOURÃO, G. B.; FERRAZ, J. B. S.; ELER, J. P.; BALIEIRO, J. C. D. C.; BUENO, R. S.; MATTOS, E. C.; FIGUEIREDO, L. G. G. Genetic parameters for growth traits of a Brazilian Bos taurus x Bos indicus beef composite. **Genetics and Molecular Research**, v. 6, n. 4, p. 1190-1200, 2007.

MOUSSA, M. et al. Maternal control of oocyte quality in cattle “a review”. **Animal Reproduction Science**, v. 155, p. 11–27, 2015.

NOORDMAN, Y. E.; JANSEN, P. A. M.; HENDRIKS, W. J. A. J. Tyrosine-specific MAPK phosphatases and the control of ERK signaling in PC12 cells. **Journal of Molecular Signaling**, v. 1, p. 1–11, 2006.

OGOREVC, J. et al. Database of cattle candidate genes and genetic markers for milk production and mastitis. n. 2006, p. 832–851, 2009.

OJEDA, S. R. et al. Minireview: The Neuroendocrine Regulation of Puberty: Is the Time Ripe for a Systems Biology Approach? **Endocrinology**, v. 147, n. 3, p. 1166–1174, 2006.

PERIPOLLI, E. et al. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. p. 1–13, 2018.

PORTO-NETO, L. R.; KIJAS, J. W.; REVERTER, A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. **Genetics Selection Evolution**, v. 46, n. 1, p. 22, 2014. Disponível em: <http://www.gsejournal.org/content/46/1/22>.

PURCELL, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

QAYYUM, A. et al. The Role of the Catechol-o-methyltransferase (COMT) Gene Val158Met in Aggressive Behavior , A Review of Genetic Studies. **Current Neurophamarcology**, v. 13, p. 802–814, 2015.

RAMAYO-CALDAS, Y. et al. Multi-breed and multi-trait co-association analysis of meat tenderness and other meat quality traits in three French beef cattle breeds. **Genetics Selection Evolution**, v. 48, n. 1, p. 37, 2016.

RAWLINGS, N. et al. Sexual Maturation in the Bull. **Reproduction in Domestic Animals**, v. 43, n. SUPPL.2, p. 295–301, 2008.

RHOADS, M. L. et al. Growth Hormone Receptor , Insulin-Like Growth Factor (IGF) -1 , and IGF-Binding Protein-2 Expression in the Reproductive Tissues of Early Postpartum Dairy Cows. **Journal of Dairy Science**, v. 91, n. 5, p. 1802–1813, 2008.

ROUSE, G. H. et al. Body Composition Changes in Bulls from Weaning to Yearling Part I — Muscle , Waste Fat and Taste Fat Deposition Body Composition Changes in Bulls from Weaning to Yearling Part I . **Beef Research Report**, 2003.

SCHMIDT, G. H. Effect of Insulin on Yield and Composition of Milk of Dairy Cows. **Journal of Dairy Science**, v. 49, n. 4, p. 381–385, 1966.

SERÃO, N. V. et al. Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. **BMC genetics**, v. 14, p. 94, 2013.

SERGIO R. OJEDA et al. Gene Networks and the Neuroendocrine Regulation of Puberty. **Mol Cell Endocrinol**, v. 2153, n. 07, p. 2–6, 2010.

SEROUSSI, E. et al. Nonbactericidal secreted phospholipase A2s are potential anti-inflammatory factors in the mammary gland. **Immunogenetics**, v. 65, p. 861–871, 2013.

SHARMA, A. et al. Validation of genetic polymorphisms on BTA14 associated with carcass trait in a commercial Hanwoo population. n. 2013, p. 863–867, 2014.

SHELDON, I. M. et al. Minireview in the Female Reproductive Tract in Cattle 1. **Biology of reproduction**, v. 1032, n. April, p. 1025–1032, 2009.

SHENG, X.; NI, H.; LIU, Y. RNA-seq analysis of bovine intramuscular , subcutaneous and perirenal adipose tissues. p. 1631–1637, 2014.

SPEHR, M. et al. Essential Role of the Main Olfactory System in Social Recognition of Major Histocompatibility Complex Peptide Ligands. v. 26, n. 7, p. 1961–1970, 2006.

STARKL, P.; MARICHAL, T.; GALLI, S. J. news and views PLA2G3 promotes mast cell maturation and function. **Nature Publishing Group**, v. 14, n. 6, p. 527–529, 2013.

STRANDÉN, I.; GARRICK, D. J. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. **Journal of Dairy Science**, v. 92, n. 6, p. 2971–2975, 2009.

SUDIMAN, J. et al. Effects of differing oocyte-secreted factors during mouse in vitro maturation on subsequent embryo and fetal development. **Journal of Assisted Reproduction and Genetics**, v. 31, n. 3, p. 295–306, 2014.

TAKASUGA, A. PLAG1 and NCAPG-LCORL in livestock. n. December 2014, p. 159–167, 2016.

TAYE, M. et al. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. **Mammalian Genome**, v. 28, n. 11, p. 528–541, 2017.

UTSUNOMIYA, Y. T. et al. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. 2013.

UTSUNOMIYA, Y. T. et al. OPEN A PLAG1 mutation contributed to stature recovery in modern cattle. n. April, p. 1–15, 2017.

VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 2008. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0022030208709901>.

VANRADEN, P. M. et al. Invited review : Reliability of genomic predictions for North American Holstein bulls. **journal dairy science**, v. 92, p. 16–24, 2009.

VENERONI, G. B. et al. Prospecting candidate SNPs for backfat in Canchim beef cattle. v. 9, n. 4, p. 1997–2003, 2010.

WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without

genotypes. **Genetics Research**, v. 94, n. 2, p. 73–83, 2012.

WANG, H. et al. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v. 5, n. May, p. 1–10, 2014.

WATTS, J. M.; STOOKEY, J. M. Vocal behaviour in cattle : the animal ' s commentary on its biological processes and welfare. p. 15–33, 2000.

WIDMANN, P. et al. Association of an ACSL1 gene variant with polyunsaturated fatty acids in bovine skeletal muscle. **BMC Genetics**, v. 12, n. 96, p. 1–13, 2011.

WIDMANN, P. et al. A systems biology approach using metabolomic data reveals genes and pathways interacting to modulate divergent growth in cattle. **BMC Genomics**, v. 14, n. 1, p. 798, 2013.

WIGGANS, G. R. et al. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. **Journal of Dairy Science**, v. 92, n. 7, p. 3431–3436, 2009. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0022030209706617>.

WILLHAM, R. L. The Role of Maternal Effects in Animal Breeding: III. Biometrical Aspects of Maternal Effects in Animals. **Journal of animal science**, v. 35, n. 6, p. 1288–1293, 1972.

WILTON, J. W. Genetic Trends and Breed Overlap Derived from Multiple-Breed Genetic Evaluations of Beef Cattle for Growth Traits. n. September 1999, 2014.

XU, W. et al. Detection of 19-bp deletion within PLAG1 gene and its effect on growth traits in cattle. **Gene**, v. 675, n. March, p. 144–149, 2018.

ZHANG, Q. et al. SNP discovery and haplotype analysis in the bovine PRKAA2 gene. v. 1, p.

1551–1556, 2011.

ZHANG, Z. et al. Mixed linear model approach adapted for genome-wide association studies. **Nature Genetics**, v. 42, n. 4, p. 355–360, 2010.

CHAPTER 3. GENETIC ARCHITECTURE OF CARCASS AND MEAT QUALITY TRAITS IN MONTANA TROPICAL® COMPOSITE BEEF CATTLE

Laís Grigoletto[a,b], José Bento Sterman Ferraz[a], Hinayah R. Oliveira[b], Joanir P. Eler[a], Fernando O. Bussiman[a], Barbara C. Abreu Silva[a], Fernando Baldi[c] and Luiz F. Brito[b]

[a]Department of Animal Sciences, College of Animal Sciences and Food Engineering, University of São Paulo, Pirassununga, Sao Paulo, Brazil

[b]Department of Animal Sciences, Purdue University, West Lafayette, Indiana, USA

[c]Department of Animal Sciences, College of Agricultural and Veterinary Sciences, São Paulo State University, Jaboticabal, Sao Paulo, Brazil

Frontiers in Genetics, 2020
Research topic: Advances in Genomics of Crossbred Farm Animals
(Original Article, Published)

ABSTRACT

The Montana Tropical® Composite is a recently developed beef cattle population that is rapidly expanding in Brazil and other Tropical countries. This is mainly due to its improved meat quality and adaptation to tropical climate conditions compared to Zebu and Taurine cattle breeds, respectively. This study aimed to investigate the genetic architecture of ultrasound-based carcass and meat quality traits in Montana Tropical® Composite beef cattle. Therefore, we estimated variance components, genetic parameters and performed genome-wide association studies using the weighted single-step GBLUP approach. A pedigree dataset containing 28,480 animals was used, in which 1,436 were genotyped using a moderate-density Single Nucleotide Polymorphism panel (30K; 30,105 SNPs). A total of 9,358, 5,768, 7,996 and 1,972 phenotypic records for the traits *Longissimus* muscle area (LMA), backfat thickness (BFT), rump fat thickness (RFT) and for marbling score (MARB), respectively, were used for the analyses. Moderate to high heritability estimates were obtained and ranged from 0.16±0.03 (RFT) to 0.33±0.05 (MARB). A high genetic correlation was observed between BFT and RFT (0.97±0.02), suggesting that a similar set of genes affects both traits. The most relevant genomic regions associated with LMA, BFT, RFT, and MARB were found on BTA10 (5.4-5.8 Mb), BTA27 (25.2-25.5 Mb), BTA18 (60.6-61.0 Mb), and BTA21 (14.8-15.4 Mb). Two overlapping genomic regions were identified for RFT and MARB (BTA13:47.9-48.1 Mb), and for BFT and RFT (BTA13:61.5-62.3 Mb). Candidate genes identified in this study, including *PLAG1*, *LYN, WWOX,* and *PLAGL2*, were previously reported to be associated with growth, stature, skeletal muscle growth, fat thickness, and fatty acid composition. Our results indicate that ultrasound-based carcass and meat quality traits in the Montana Tropical® Composite beef cattle are heritable, and therefore, can be improved through selective breeding. In addition, various novel and already known genomic regions related to these traits were identified,

which contributes to a better understanding of the underlying genetic background of LMA, BFT, RFT, and MARB in the Montana Tropical® Composite population.

Keywords: candidate genes, composite cattle, crossbreeding, genomic regions, ssGWAS, single-step GBLUP, tropical beef cattle

3.1 INTRODUCTION

Both carcass and meat quality traits are paramount for optimizing the profitability of the beef cattle industry. These traits are influenced by diet and feeding practices, pre- and post-slaughter management, and meat processing and storage methods (ADZITEY, 2011; GUERRERO et al., 2013; NJISANE; MUCHENJE, 2017). Despite the clear effectiveness of these alternatives, genetic selection is a complementary approach in which the gains achieved are permanent and cumulative over generations. In this context, carcass and meat quality traits have been measured and incorporated in worldwide beef cattle breeding programs (REVERTER et al., 2000; YOKOO et al., 2010; BERRY et al., 2017; GORDO et al., 2018). Carcass and meat quality traits can be measured in live animals using ultrasound technology, which is a non-invasive technique (PATHAK et al., 2011; SCHOLZ et al., 2015). Ultrasound-based traits that are indicators of carcass and meat quality include *Longissimus* muscle area (LMA), backfat thickness (BFT), rump fat thickness (RFT), and marbling score (MARB) (PATHAK et al., 2011; FONT-I-FURNOLS et al., 2014; GORDO et al., 2018).

Brazil is one of the largest beef cattle producers in the world, with a population of over 230 million animals (USDA, 2018). More than 80% of the beef cattle animals currently raised in Brazil are from the Nellore breed (*Bos taurus indicus*; Zebu), which are well adapted to Tropical conditions (FERRAZ; FELÍCIO, 2010). However, Zebu breeds are also well known for poorer meat quality (CROUSE et al., 1989; BRESSAN et al., 2016; RODRIGUES et al., 2017) when

compared to Taurine (*Bos taurus taurus*) breeds (e.g., Aberdeen Angus, Red Angus, Senepol, Charolais). An alternative to improve carcass and meat quality traits, while keeping the adaptation characteristics of Zebu cattle, is through the development of composite populations (i.e., crossbreeding between Taurine and Zebu animals; e.g. PICCOLI et al., 2020).

The Montana Tropical® Composite population was firstly developed in 1994 following studies conducted by the U.S. Meat Animal Research Center at Clay Center, United States Department of Agriculture (USDA; GREGORY et al*.,* 1993; 1994). This composite population was developed by crossing animals from four different biological types or breed groups (FERRAZ; ELER; GOLDEN, 1999): 1) Zebu breeds (*Bos taurus indicus*), 2) Adapted Taurine breeds (*Bos taurus taurus*), 3) British breeds (*Bos taurus taurus*), and 4) Continental European breeds (*Bos taurus taurus*).

Over the past few years, there has been a great interest on genetically improving this composite population and better understanding its genetic background underlying phenotypic variation of economic importance to the breeders. In this context, genome-wide association studies (GWAS) can be performed to identify Quantitative Trait Loci (QTL) associated with key traits (e.g. carcass and meat quality). Recent GWAS have successfully revealed significant genomic regions in beef cattle composite populations (e.g., WENG et al*.,* 2016; HAY; ROBERTS, 2018; GRIGOLETTO et al*.,* 2019). Wang et al. (2012) proposed a GWAS method based on the single-step Genomic Best Linear Unbiased Predictor (ssGBLUP; AGUILAR et al*.,* 2010; LEGARRA et al*.,* 2009; 2014), which has become the gold-standard method for GWAS (also termed ssGWAS). A variation of this method, the weighted single-step GBLUP (WssGBLUP; WANG et al*.,* 2012) usually yields more accurate SNP effects (e.g. ZHANG et al., 2016), and consequently, a greater power to identify QTLs and functional genes. In this context, the main goals of this study were to:

1) estimate variance components and genetic parameters for four ultrasound-based carcass and meat quality traits (i.e., LMA, BFT, RFT, and MARB) in Montana Tropical® Composite beef cattle; and 2) identify relevant genomic regions, candidate genes, and metabolic pathways associated with these traits, using the WssGBLUP method.

3.2 MATERIALS AND METHODS

Animal Care Committee approval was not obtained for this study as all the analyses were performed using pre-existing databases.

3.2.1 Animals and Phenotypic Data

The descriptive statistics of the pedigree file, including the breed composition of the animals is shown in Table 3.1. The animals were classified within each biological group (NABC) as: 1) N: Zebu breeds, mainly represented by Nellore; 2) A: Taurine breeds adapted to Tropical conditions (Senepol, Belmont Red, Bonsmara, and Caracu); 3) B: Taurine breeds of British origin (mainly Angus, Devon, and Hereford); and, 4) C: Continental European breeds (mainly Charolais, Limousin, and Simmental).

To be considered as a Montana Tropical® Composite (Figure 3.1), the animals had to have at least three breeds in its genetic composition. In addition, the minimum percentage of the biological types (breed groups) required to be considered a Montana Tropical® Composite was 12.5% for group A; and 25% for groups N and A together. The maximum proportion of each group allowed was 37.5% for group N; 87.5% for group A; and 75% for groups B and C (Santana et al., 2013). The main contributing breeds to the development of this composite population were Aberdeen Angus, Red Angus, Nellore, Senepol, Limousin, Simmental, Hereford, and Bonsmara.

Table 3.1 Descriptive statistics of the pedigree dataset according to the breed and biological type composition of the animals.

[1]Biological Type		Number of animals
Montana Tropical® Composite	4444	7,136
	4480	4,693
	4804	3,125
	4840	3,127
Pure breeds	N ≥ 90%	3,730
	A ≥ 90%	1,461
	B ≥ 90%	1,630
	C ≥ 90%	181
Crossbreed	NxA	116
	NxB	2,230
	NxC	842
	AxB	153
	AxC	8
	BxC	48
Total		**28,480**

[1]Biological type (NABC system; FERRAZ; ELER; GOLDEN, 1999): Zebu breeds (N), Adapted Taurine breeds (A), British Taurine breeds (B), and Continental Taurine breeds (C). Breed composition of the Montana Tropical Composite animals: 4444 = 25% N, A, B and C; 4480 = 25% N, 25% A, 50% B and < 6.25% C; 4804 = 25% N, 50% A, < 6.25% B and 25% C; 4840 = 25% N, 50% A, 25% B and < 6.25% C. NxA, NxB, NxC, AxB, AxC and BxC = the combination of each breed group equals to 50%.

Figure 3.1 Illustration of a Montana Tropical® Composite bull (left) and location of the farms (right) participating in the Montana Tropical® Composite breeding program. The map regions in dark red indicate Brazilian states and the light red areas represent Paraguay and Uruguay. Photo Credits: Montana Tropical® Composite.

Four ultrasound-based carcass and meat quality traits (LMA, BFT, RFT and MARB), recorded on animals born between 2008 and 2016, were included in this study. Animals were raised in 18 farms located at different Brazilian states, Paraguay and Uruguay (Figure 3.1). In general, the animals were raised on pastures composed basically of *Brachiaria brizantha*. All farms provided feed supplements in the dry season (from May to August). With regards to the reproductive breeding scheme, around 60% of cows were artificially inseminated and 40% were kept in multiple-sire lots with a cows-to-bull' ratio of 30:1 or 25:1. The majority of calves were born between September and December (Spring season in South America and the beginning of the rainy period) and weaned at 7 months of age. Weight recording was obtained at birth and weaning. Further records of yearling weight, scrotal circumference, and other productive traits were collected between 14 and 18 months. More details are presented in Santana et al. (2012), and in a previous GWAS study from the same population (GRIGOLETTO et al., 2019).

The average (±standard deviation; SD) age of the animals at the ultrasound measurement was 580.27 (±75.08) days. *Longissimus* muscle area (LMA) was measured in cm^2, between the 12^{th} and 13^{th} ribs. Backfat thickness (BFT) was measured in mm, at a point three-fourths of transverse orientation over the LMA (BRETHOUR, 2004). Rump fat thickness (RFT) was also measured in mm, at the junction of the *biceps femoris* and *gluteus medius* between the ischium and ilium (GREINER et al., 2003; GORDO et al., 2012). Marbling score (MARB) was measured as an indicator of percentage of intramuscular fat, using a subjective scale ranging from 1 to 12, based on the U.S. Department of Agriculture (USDA) quality grades (www.uspremiumbeef.com/DocumentItem.aspx?ID=21). All traits were evaluated by ultrasonography using the ALOKA 500 V device, with a 3.5 MHz linear probe. These images were analyzed using the LINCE® software (GABÍN et al., 2012). Phenotypic quality control removed records deviating 3.5 SD from the overall mean within contemporary group (CG). The CG was defined based on farm, year and season of birth, sex, and management group. The CGs with less than five records were excluded from subsequent analyzes. Descriptive statistics for the ultrasound-based carcass and meat quality traits after the data editing are shown in Table 2.

3.2.2 Genotypic Quality Control

A total of 1,436 bulls were genotyped using a moderate-density SNP panel containing 30,105 SNPs (GeneSeek Genomic Profiler™ LDv4 - GGP Bovine LDv4; Illumina, San Diego, CA). Genotype quality control was performed using the PREGSF90 program (AGUILAR et al., 2014; 2019). In general, SNPs with minor allele frequency lower than 0.05, call rate lower than 90%, extreme deviation from Hardy-Weinberg equilibrium (defined as the maximum difference between observed and expected heterozygosity) greater than 0.15 (WIGGANS et al., 2009), and

SNPs located in non-autosomal chromosomes were excluded. A total of 27,196 SNPs distributed on 29 autosomal chromosomes, and 1,394 genotyped animals (42 animals were excluded due to call rate lower than 90%) remained for further analyzes. BTA1 is the largest chromosome, with 158.72 Megabase pairs (Mb) covered by 1,602 SNPs, while the BTA27 is the shortest one, with 42.33 Mb covered by 512 SNPs.

3.2.3 Statistical Analyses

Variance components and breeding value prediction. Single-trait linear animal models and the average-information restricted maximum likelihood (AI-REML) procedure were used to estimate heritability and variance components, using the AIREMLF90 package from the BLUPF90 family programs (MISZTAL et al., 2002; 2014). Genomic breeding values for all traits were directly predicted using the ssGBLUP procedure (MISZTAL et al., 2009; AGUILAR et al., 2010; CHRISTENSEN; LUND, 2010). The ssGBLUP is a modified version of the traditional BLUP, in which the inverse of the pedigree-based relationship matrix ($\mathbf{A}^{-1}$) is replaced by the $\mathbf{H}^{-1}$ matrix. The $\mathbf{H}^{-1}$ is defined as follow (LEGARRA et al., 2009; AGUILAR et al., 2010):

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\tau}\mathbf{G}^{-1} - \boldsymbol{\omega}\mathbf{A}_{\mathbf{22}}^{-1} \end{bmatrix},$$

where $\mathbf{A}^{-1}$ was previously defined, τ and ω are the scaling factors used to combine $\mathbf{G}$ and $\mathbf{A_{22}}$ (assumed as τ=1.0 and ω=0.7 in order to reduce bias; MISZTAL et al., 2010; TSURUTA et al., 2011), $\mathbf{A}_{\mathbf{22}}^{-1}$ is the inverse of pedigree-based relationship matrix for the genotyped animals, and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix ($\mathbf{G}$), which was calculated as (VANRADEN, 2008):

$$\mathbf{G} = \mathbf{ZZ}'/\mathbf{k},$$

where **Z** is the matrix containing the centered genotypes (-1, 0, 1) accounting for the observed allelic frequencies; and k is a scaling parameter, defined as $2\sum p(1-p)$, in which p is the observed allele frequency of each marker. The weighting factor can be derived either based on SNP frequencies (VANRADEN, 2008), or by ensuring that the average diagonal of **G** is close to one as in $\mathbf{A_{22}}$ (VITEZICA et al., 2011). In order to minimize issues with **G** inversion, 0.05 of **A** was added to 0.95 of the **G** matrices.

The single-trait animal models used in this study included the direct additive genetic and residual as random effects. CG, heterozygosity (described below), and age of the animal at the measurement were included as fixed effects in the model. Thus, the statistical model used in this study can be described as**:**

$$\mathbf{y}_{ijkl} = \mathbf{CG}_i + \mathbf{b_1}(\mathbf{Age}_j - \overline{\mathbf{Age}}) + \mathbf{b_2}\left(\mathbf{H_D}_k - \overline{\mathbf{H_D}}\right) + \boldsymbol{\alpha}_l + \boldsymbol{\varepsilon}_{ijkl},$$

where $\mathbf{y_{ijkl}}$ is the phenotypic record for each trait (LMA, BFT, RFT or MARB) recorded on the animal *l*, belonging to the **CG*i***, at age *j*, and heterozygosity (H_D) k. $\mathbf{b_1}$ and $\mathbf{b_2}$ are the linear regression coefficients related to the Age and $\mathbf{H_D}$ effects, respectively, which were considered as deviations from the mean ($\overline{\mathbf{Age}}$ and $\overline{\mathbf{H_D}}$). The $\boldsymbol{\alpha}_\mathbf{l}$ is the direct additive genetic random effect for the animal *l*, and $\boldsymbol{\varepsilon}_\mathbf{ijkl}$ is the residual random effect associated with the animal *l*, heterozygosity *k*, age *j* and CG*i*. Assuming a matrix notation, the previous model can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where, $\mathbf{y}$ is the vector of phenotypic observations for each trait; $\boldsymbol{\beta}$ is the vector of solutions for fixed effects; $\boldsymbol{\alpha}$ is the vector of predictions for random additive genetic animal effect; $\boldsymbol{\varepsilon}$ is the vector of random residual terms; $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices of fixed and random effects, respectively. It was assumed that: $\alpha \sim N(0, \mathbf{H}\sigma_\alpha^2)$ and $\varepsilon \sim N(0, \mathbf{I}\sigma_\varepsilon^2)$; where σ_α^2 is the additive genetic variance; σ_ε^2 is the residual variance; and $\mathbf{I}$ is an identity matrix. Thus, the (co)variance matrix ($\mathbf{V}$) of the random effects can be expressed as:

$$\mathbf{V} = \begin{bmatrix} \mathbf{H}\sigma_\alpha^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_\varepsilon^2 \end{bmatrix},$$

where $\mathbf{H}$ is the relationship matrix used in the ssGBLUP method. The non-additive effects of heterozygosity were obtained by a linear regression to the coefficients of direct heterozygosity (H_D), which were calculated as (DIAS et al*.,* 2011):

$$\mathbf{H_D} = \mathbf{1} - \sum_{\mathbf{i=1}}^{\mathbf{4}} \mathbf{S_i}\,\mathbf{D_i}$$

in which ***i*** represents the biological type (i.e., i = 1, 2, 3 or 4, indicating the proportion of N, A, B, C, respectively); $\mathbf{S_i}$ and $\mathbf{D_i}$ are the fractions of the i^{th} biological type of sire and dam, respectively. Coefficients for biological types (N, A, B, and C) were equal to the proportion of each biological type in the breed composition, and it was assumed that the sum of all proportions of biological types in one animal was equal to one. To avoid multicollinearity, direct additive effects of the biological type N were excluded from the statistical models, i.e., the effects for A, B, and C were estimated as deviations of the additive effects of N (DIAS et al*.,* 2011; PETRINI et al*.,* 2012).

Genetic correlations. A multiple-trait linear animal model was used to estimate the genetic and phenotypic correlation between all traits (LMA, BFT, RFT, and MARB) using pedigree and

genomic information. Genetic and phenotypic correlations were calculated using the AIREMLF90 package from the BLUPF90 family programs (MISZTAL et al., 2002; 2014). The multiple-trait model included the same fixed and random effects described above. However, it was assumed that: $\boldsymbol{\alpha} \sim N(0, \mathbf{G} \otimes \mathbf{H})$; $\boldsymbol{\varepsilon} \sim N(0, \mathbf{R} \otimes \mathbf{I})$; where $\boldsymbol{\alpha}$, $\mathbf{H}$, and $\mathbf{I}$ are the same as above; $\mathbf{G}$ is the additive genetic (co)variance matrix; $\mathbf{R}$ is the residual (co)variance matrix. In this reasoning, the (co)variance matrix for random effects was:

$$\mathbf{V} = \begin{bmatrix} \mathbf{G} \otimes \mathbf{H} & \mathbf{O} \\ \mathbf{0} & \mathbf{R} \otimes \mathbf{I} \end{bmatrix}$$

Genome-wide association studies (GWAS). The GWAS for each trait was carried out using the weighted ssGBLUP method (WssGBLUP; WANG et al., 2012). The same statistical models described to estimate the variance components and breeding values were used to identify genomic windows associated with the traits, as described by Wang et al. (2014) using the BLUPF90 family programs (MISZTAL et al., 2002; 2014). The PREGSF90 software (AGUILAR et al., 2014) was used as an interface to the genomic module to process the genomic information. Also, the POSTGSF90 software (AGUILAR et al., 2014) was used to back-solve the GEBVs for each trait. To calculate the SNP effects and weights, we followed the steps proposed by Wang et al. (2014). This method uses an iterative process, which was repeated three times in this study, to increase the weight of SNPs with larger effects and decrease the weight of those markers with smaller (close to zero) effects (WANG et al., 2014). The GWAS results are reported as the proportion of variance explained by a moving genomic window of five adjacent SNPs. Genomic windows that explained more than 1% of the total genetic variance were considered as relevant, i.e. associated with the trait being analyzed.

3.2.4 Functional Analyses

Positional candidate genes were annotated considering an upstream and downstream interval of 100 kb (threshold defined based on the level of linkage disequilibrium in the population) using the Ensembl Genome Browser (www.ensembl.org/index.html) and the ARS-UCD1.2 version of the cattle genome (ZERBINO et al., 2017). Furthermore, important SNPs (from the key genomic windows) were further explored using the Animal QTL Database (AnimalQTLdb; HU et al., 2019). Functional analyses were carried out to characterize the gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the Database for Annotation, Visualization and Integrated Discovery (DAVID; HUANG et al., 2009). In order to increase the statistical power of the study, all candidate genes identified for the four traits were considered in the same functional analysis, as they are all correlated traits. The significance thresholds used were p-value < 0.05 and False Discovery Rate (FDR) < 5.

3.3 RESULTS

3.3.1 Genetic Parameter Estimates

The variance components and heritability (h^2) estimates for LMA, BFT, RFT, and MARB are presented in Table 3.2. All traits had moderate to high estimates, which ranged from 0.16±0.03 to 0.33±0.05. The genetic and phenotypic correlation are shown in Table 3.3. The highest genetic correlation was obtained between BFT and RFT (0.97±0.02), followed by an unfavorable correlation between BFT and MARB (0.66±0.01). The heritability estimates from the single-trait and averaged bivariate model analyses were similar and therefore, only the heritability estimates from the single-trait models are reported and discussed here.

Table 3.2 Descriptive statistics, variance components and genetic parameter estimate for ultrasound carcass traits in the Montana Tropical® Composite cattle population.

[1]Trait	N	Mean	SD	σ_a^2(SE)	σ_e^2(SE)	h^2(SE)
LMA (cm^2)	9,358	58.40	12.79	13.99 (1.73)	33.35 (1.44)	0.29 (0.03)
BFT (mm)	5,768	2.84	0.71	0.25 (0.04)	0.68 (0.03)	0.26 (0.03)
RFT (mm)	7,996	3.16	1.37	0.09 (0.04)	0.45 (0.03)	0.16 (0.03)
MARB (score)	1,972	3.29	1.20	0.18 (0.04)	0.36 (0.04)	0.33 (0.05)

[1]Traits: Longissimus muscle area (LMA); backfat thickness (BFT); rump fat thickness (RFT); marbling score (MARB).N = number of animals; SD = standard deviation; SE = standard error; σ_a^2 = additive genetic variance; σ_e^2 = residual variance; h^2 = heritability.

Table 3.3 Genetic (below) and phenotypic (above) correlation (± standard error) for ultrasound carcass and meat quality traits in the Montana Tropical Composite beef cattle population.

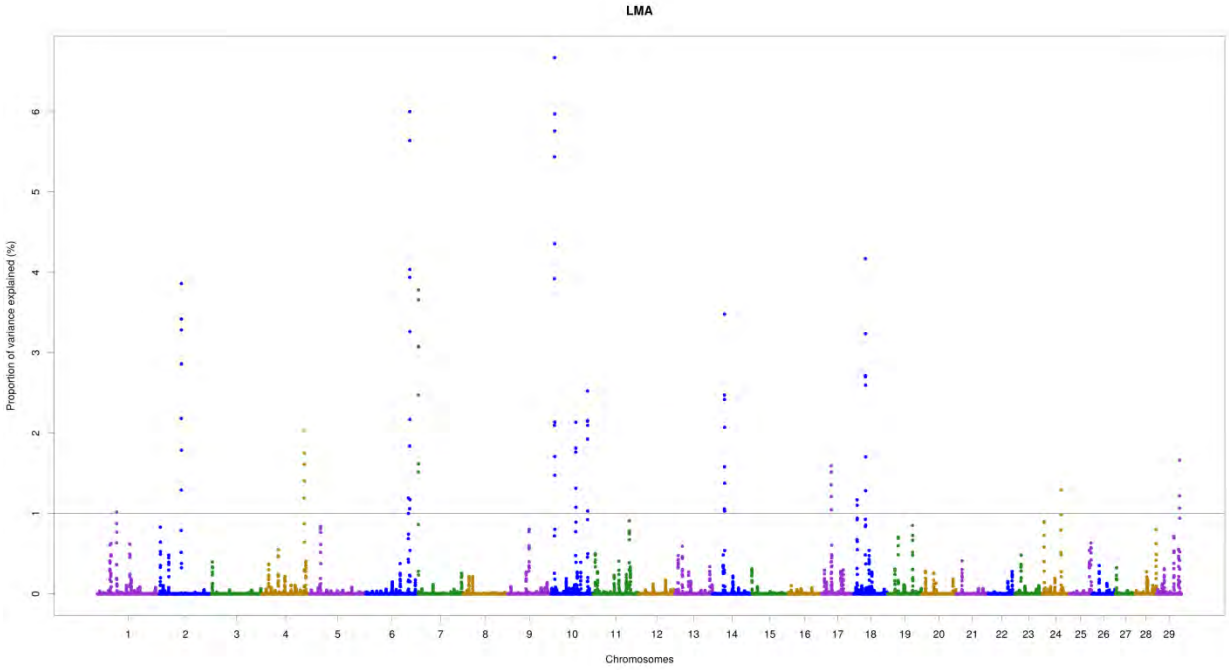
Traits[1]	LMA	BFT	RFT	MARB
LMA		0.46±0.05	0.29±0.08	0.27±0.05
BFT	0.53±0.08		0.64±0.03	0.50±0.02
RFT	0.39±0.12	0.97±0.02		0.47±0.03
MARB	0.23±0.01	0.66±0.01	0.55±0.02	

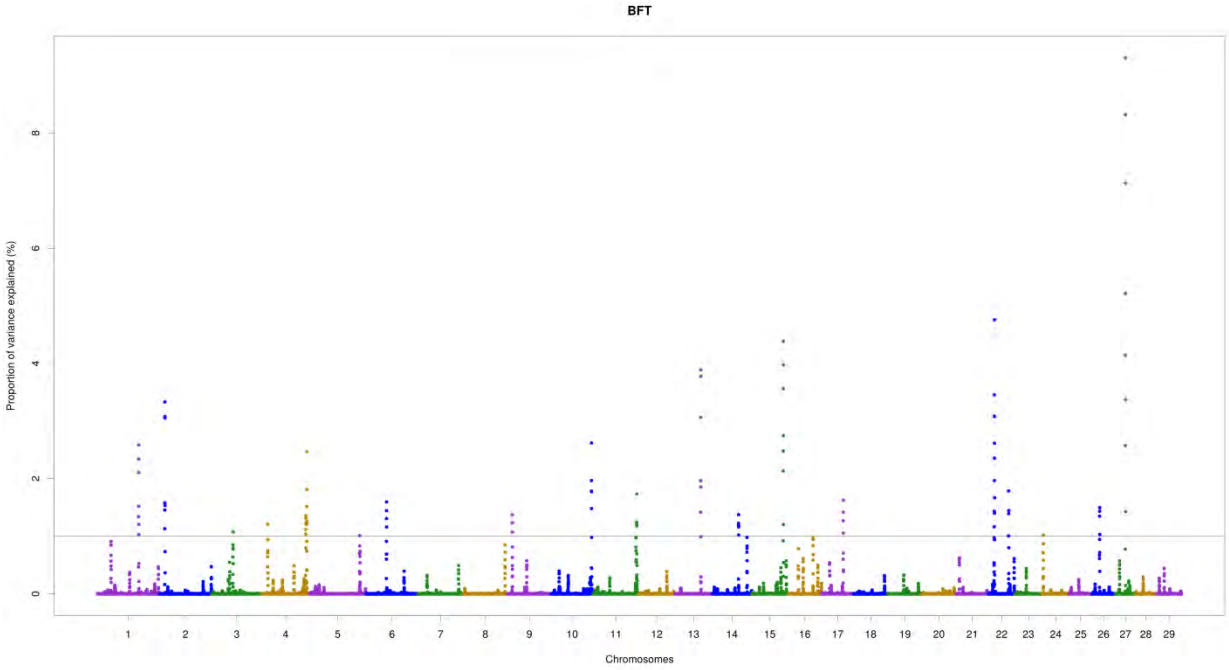
[1]Traits: *Longissimus* muscle area (LMA), backfat thickness (BFT), rump fat thickness (RFT) and marbling score (MARB).

3.3.2 GWAS and Functional Analyses

A total of 18, 22, 9, and 11 genomic windows explaining more than 1% of the total genetic variance were identified for LMA, BFT, RFT, and MARB, respectively. These regions are harboring or overlap with 241 positional genes. The main candidate genes are shown in Table 3.4 and the complete list is presented in the "Supplementary Material" section. The genomic windows identified are spread across all autosomal chromosomes, with exception of BTA8, BTA16,

BTA19, BTA20, and BTA25 (Tables S3.1-S3.4). The Manhattan plots for all traits are presented in Figure 3.2.





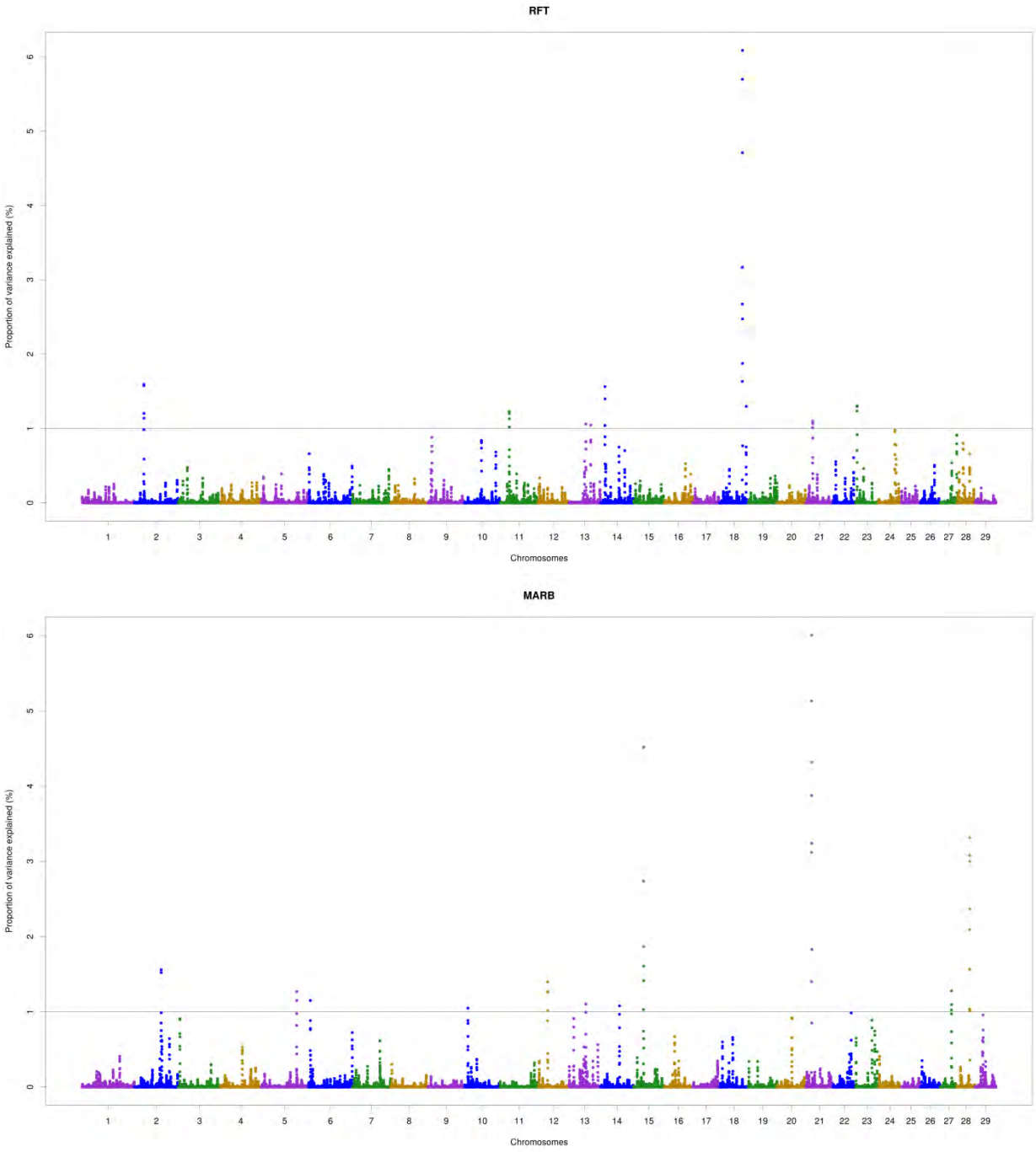


Figure 3.2 Manhattan plots of the genome-wide association analyses for *Longissimus* muscle area (A; LMA), backfat thickness (B; BFT), rump fat thickness (C; RFT) and marbling score (D; MARB) traits. The 29 autosomal chromosomes are shown in different colors. The x-axis represents the chromosome number whereas the y-axis shows the proportion of genetic variance explained by five adjacent SNPs. The gray line corresponds to the genome-wide threshold of each window that explained more than 1% of genetic variance.

Table 3.4 The main genomic regions explaining more than 1% of total genetic variance (%var) of ultrasound-based carcass traits in the Montana Tropical® Composite beef cattle.

[1]Trait	[2]BTA	Position (start-end, in bp)	%var	Candidate genes
LMA	2	64,808,388 – 65,069,037	3.86	*NCKAP5, LYPD1, GPR39*
	6	102,264,376 – 102,500,758	6.00	*HSD17B13, HSD17B11, NUDT9, SPARCL1, DSPP, DMP1, PPP2R2C, WFS1, JAKMIP1*
	10	5,392,944 – 5,807,684	6.67	*HRH2, SFXN1, DRD1*
	14	22,875,603 – 23,252,097	3.48	*XKR4, TMEM68, TGS1, LYN, RPS20, MOS, PLAG1*
	18	25,832,665 – 26,209,903	4.17	*KIFC3, CNGB1, TEPP, ZNF319, USB1, MMP15, CFAP20, CSNK2A2, CCDC113, PRSS54, GINS3, NDRG4, SETD6, CNOT1, SLC38A7, GOT2*
BFT	2	12,138,830 – 12,823,369	3.33	-
	13	61,566,683 – 62,224,699	3.88	*PLAGL2, POFUT1, KIF3B, ASXL1, NOL4L, COMMD7, DNMT3B, MAPRE1, EFCAB8, SUN5, BPIFB2*
	15	75,727,954 – 76,192,434	4.38	*MAPK8IP1, C15H11orf94, PEX16, LARGE2, PHF21A, CREB3L1, CHST1, SLC35C1, CRY2*
	22	12,826,540 – 13,203,551	4.75	*SCN10A, SCN11A, WDR48, GORASP1, TTC21A, CSRNP1, XIRP1, CX3CR1, CCR8, SLC25A38, RPSA, MOBP, MYRIP, EIF1B, ENTPD3, RPL14, ZNF619, ZNF621, HSPD1*
	27	25,252,766 – 25,558,906	9.31	*PPP1R3B, TNKS*
RFT	2	29,948,707 – 30,390,796	1.59	*SCN7A, SCN9A, SCN1A, TTC21B*
	14	7,844,432 – 8,107,746	1.56	*ST3GAL1, NDRG1, CCN4*
	18	60,682,096 – 61,018,825	6.08	*ZNF331, MGC139164, NLRP12, MGC157082*
	18	65,340,963 – 65,356,544	1.29	*ZNF814*
	23	3,159,017 – 3,581,582	1.30	*ZNF451, BEND6*

MARB	2	96,082,524 – 96,725,242	1.55	*PLEKHM3, CRYGD, CRYGC, CRYGB, CRYGA, C2H2orf80, IDH1, PIKFYVE, PTH2R*
	12	22,901,497 – 23,236,520	1.39	*LHFPL6, NHLRC3, PROSER1, STOML3, FREM2*
	15	24,216,319 – 24,219,946	4.51	*ZW10*
	21	14,800,548 – 15,428,801	6.00	*SLCO3A1*
	28	34,157,181 – 34,514,922	3.31	-

[1]Traits: *Longissimus* muscle area (LMA), backfat thickness (BFT), rump fat thickness (RFT) and marbling score (MARB). [2]*Bos taurus* autosome (BTA).

Two overlapping regions were identified on BTA13: 1) at 47.7-48.5 Mb for BFT and MARB, and 2) at 61.5-63.5 Mb for BFT and RFT. The highest peaks associated with LMA, BFT, RFT, and MARB were located on BTA10 (5.4-5.8 Mb; 6.6% of the genetic variance), BTA27 (25.2-25.5 Mb; 9.3% of the genetic variance), BTA18 (60.6-61.0 Mb; 6.0% of the genetic variance), and BTA21 (14.8-15.4 Mb; 6.0% of the genetic variance), respectively (Figure 3.2). For LMA, a single genomic window was identified on BTA14 (22.8 to 23.2 Mb) explaining close to 4% of the total additive genetic variation. Another region explaining 1.17% of the total additive genetic variance was identified on BTA18 (5.4 to 5.6 Mb) and contains the *WWOX* gene which plays a role on the composition of intramuscular fatty acid associated with cholesterol homeostasis and triglyceride biosynthesis (IATAN et al., 2014). A region located on BTA13 (61.6–62.5 Mb) identified to be associated with both BFT and RFT harbors the candidate genes *PLAGL2*, *ASXL1*, and *BPIFB2.* This suggests that these genes might have pleiotropic effects on BFT and RFT. The genomic region located at BTA22 and harboring the *SCAP* and *ENTPD3* genes accounted for 7.58% of the total genetic variance for BFT.

It is worth noting that we highlighted selected genes related with BFT and RFT, however, a total of 13 mutual genes (*HCK, TM9SF4, PLAGL2, POFUT1, KIF3B, ASXL1, NOL4L, COMMD7, DNMT3B, MAPRE1, EFCAB8, SUN5, BPIFB2*) were identified for this common

genomic region. A total of 12 biological processes (BP) and two pathways were significantly enriched (Table 3.5). Four biological processes involving visual behavior, associative learning, muscle tissue morphogenesis, and regulation of fatty acid biosynthetic processes were highlighted for further discussion.

Table 3.5 Enriched Gene Ontology (GO) and KEGG terms obtained from the DAVID database (https://david.ncifcrf.gov; HUANG et al., 2009).

Category GO	Term	*p*-value	FDR	Genes
Biological Process	GO:0008306~associative learning	4.97E-04	0.84	*DDHD2, NDRG4, DRD1, HRH2, LRRN4*
	GO:0007632~visual behavior	6.93E-04	1.17	*DDHD2, NDRG4, DRD1, HRH2, LRRN4*
	GO:0008542~visual learning	0.01	1.53	*DDHD2, NDRG4, DRD1, HRH2, LRRN4*
	GO:0060415~muscle tissue morphogenesis	0.008	2.57	*CCM2L, MYL3, RXRA, ZFPM2*
	GO:0007612~learning	0.014	3.6	*DDHD2, DRD1, LRRN4, HRH2, NDRG4*
	GO:0030817~regulation of cAMP biosynthetic process	0.04	5.2	*DRD1, GALR1, WFS1, GALR3*
	GO:0042304~regulation of fatty acid biosynthetic process	0.01	5.61	*SCAP, INSIG1, PDK4*
	GO:0044060~regulation of endocrine process	0.04	5.4	*FGFR1, CRY2, GALR1*
	GO:0060986~endocrine hormone secretion	0.04	5.9	*FGFR1, CRY2, GALR1*
	GO:0033002~muscle cell proliferation	0.01	7.3	*FGFR1, NDRG4, RXRA, ZFPM2*
	GO:0007611~learning or memory	0.01	8.1	*DDHD2, DRD1, LRRN4, HRH2, NDRG4*
	GO:0001501~skeletal system development	0.02	8.2	*WDR48, FGFR1, CSRNP1, PTH1R, ASXL1, INSIG1, PKDCC, SETD2, WWOX*
KEGG pathway	bta00270: Cysteine and methionine metabolism	0.04	14.63	*DNMT3B, GOT2*

	bta04151: PI3K-Akt signaling pathway	0.05	40.64	*CREB3L1, COL5A1, FGFR1, FLT4, PIK3AP1, PPP2R3C, RXRA*

3.4 DISCUSSION

3.4.1 Genetic Parameters

The genetic parameters obtained for ultrasound carcass and meat quality traits in the Montana Tropical® Composite population (Table 3.2) are similar to literature reports. For instance, Meirelles et al. (2010) estimated h^2 of 0.24 ± 0.09 and 0.33 ± 0.09 for BFT and LMA, respectively, in Canchim beef cattle (a synthetic population based on crossing between Charolais and Zebu breeds). Silva et al. (2019) also reported moderate to high h^2 estimates for BFT (0.17±0.06), RFT (0.27±0.07), and LMA (0.32±0.02) in Nellore beef cattle. Hay and Roberts (2018) also reported a high h^2 estimate for LMA (0.32 ± 0.08) in a composite population of 50% Red Angus, 25% Charolais, and 25% Tarentaise beef cattle. The moderate to high heritability estimates indicate that genetic progress can be achieved for these traits through selective breeding.

A high and favorable genetic correlation was estimated between BFT and RFT (r = 0.97±0.02), indicating that these traits are controlled by a similar set of genes. Furthermore, this high genetic correlation suggests that indirect genetic progress can be attained by including only one of these two traits in a breeding program. Positive but unfavorable genetic correlations were estimated between BFT and MARB (r = 0.66±0.01), RFT and MARB (r = 0.55±0.02), LMA and BFT (r = 0.53±0.08), and LMA and RFT (r = 0.39±0.02). This is because the industry aims to increase MARB and LMA while keeping BFT and RFT at a constant level. However, as these correlations are far from the unit, genetic progress for all the traits can be achieved by including

and properly weighting them in a selection index. A favorable correlation was observed between LMA and MARB (r = 0.23±0.01). Gordo et al. (2018) also obtained a moderate and positive correlation between LMA and MARB in Zebu cattle. These findings indicate that selection for carcass traits might indirectly improve meat quality.

3.4.2 GWAS and functional analyses

To our best knowledge, this is the first study reporting genomic regions and genetic parameters for carcass and meat quality traits in the Montana Tropical® Composite. The WssGBLUP method enables the inclusion of phenotypes of ungenotyped animals, which improves the accuracy of marker effect estimation (WANG et al., 2012; AGUILAR et al., 2019). The genomic regions presented in Table 3.4 are harboring candidate genes related to several biological mechanisms associated with carcass and meat quality traits. For instance, the *PPP1R3B* (*protein phosphatase 1, regulatory subunit 3B*) gene was identified to play a role in the expression of all the traits included in this study. *PPP1R3B* has been reported to be associated with meat quality traits in cattle, including pH, meat color, and shear force (EDWARDS et al., 2003; KAYAN, 2011; CINAR et al., 2012), and skeletal muscle development in humans (MUNRO et al., 2002). In addition, this gene is associated with glucose and glycogen metabolism. Therefore, it may affect the energy availability in skeletal muscle and consequently, contribute to greater muscle growth (ZHAO; HUANG; DU, 2019). Also, it regulates deposition of intramuscular fat relative to subcutaneous fat deposition (CHOAT et al., 2003).

The *PLAGL2*, *CALCR*, *ASXL1*, and *BP1FB*2 genes, identified to be associated with BFT and RFT, play a role in lipid metabolism (VAN DYCK et al., 2007). More specifically, *PLAGL2* is part of a subfamily of zinc finger (PLAG) gene family proteins (KAS et al., 1998). The *PLAG1* gene, also identified in this study, has a great impact on carcass weight in cattle (LITTLEJOHN et

al., 2012). Moreover, many studies have shown that the *PLAG* gene family is a key regulator of mammalian growth and body weight (LITTLEJOHN et al., 2012; FORTES et al., 2013; UTSUNOMIYA et al., 2017; MURAMATSU, 2018; ZHANG et al., 2019). The *CALCR* gene, located on BTA4 and identified to be associated with BFT, was previously reported to be associated with angularity, body condition score and body depth in Holstein cattle (MAGEE et al., 2010). The gene *INSIG1* (*Insulin induced gene 1*) has also been associated with growth and carcass traits, including body weight, hip width and withers height (LIU et al., 2012), residual feed intake (KARISA et al., 2013) and milk fatty acids (RINCON et al., 2012). Furthermore, a group of genes (*PLAG1*, *RPS20*, *ATP6V1H*, *RGS20*, *LYN*, *TCEA1*, *MRPL15*, *SOX17*, *RP1, CHCHD7, SDR16C5, SDR16C6, PENK, FAM110B, CYP7A1, SDCBP)* located on a conserved region on BTA14, previously reported as a selective sweep region in dairy and beef cattle breeds (ZHAO et al., 2015), might play a crucial role on carcass and meat quality traits. This region seems to be the most relevant association with carcass traits in beef cattle (MAGALHÃES et al., 2016; HAY; ROBERTS, 2018; ZHANG et al., 2019). Furthermore, *LYN, XKR4,* and *TGS1* genes have already been associated with hip height (AN et al., 2019), insulin-like growth factor 1 level (Fortes et al., 2012), and carcass traits (including RFT) in Blonde d'Aquitaine, Charolaise, Limousine, Belmont Red, Santa Gertrudis, and Nellore cattle (MAGALHÃES et al., 2016; PORTO-NETO et al., 2012; RAMAYO-CALDAS et al., 2014).

The considerable number of common candidate genes (i.e. 114 genes) identified for multiple carcass traits suggests that there are important pleiotropic effects regulating phenotypic expression of these traits. This is also supported by the moderate to high genetic correlation observed here and in other studies (e.g. TONUSSI et al., 2015; HERD et al., 2018). Recently, Silva et al. (2017) and Hay and Robert (2018) reported several significant regions on BTA14 associated

with BFT and other carcass traits in Zebu and composite beef cattle populations. The genomic region identified on BTA22 (harboring the *SCAP* and *ENTPD3* genes) was also reported by Hay and Robert (2018) to be associated with BFT in tropical composite cattle. The gene *DNMT3B* (*DNA cytosine-5-methyltransferase 3 beta*), associated with BFT in this study, was previously associated with marbling score, subcutaneous fat, *Longissimus* muscle area, body weight, carcass weight, dressing percentage in offspring of Wagyu and F1 crossbred cows of Limousin with Fuzhou Yellow cattle (LIU et al., 2012). *LCORL* has also been previously associated with carcass weight and fat thickness at the 12th rib in crossbred beef cattle (LINDHOLM-PERRY et al., 2011).

The *WWOX* gene, located on BTA18, has been previously associated with meat color in Korean native cattle (LEE et al., 2018). Meat color is one of the main parameters that influence consumers' preference (FONT-I-FURNOLS; GUERRERO, 2014). Additionally, meat color has currently been described to be related to cholesterol homeostasis and fatty acid biosynthesis, which is likely associated with lipid metabolism (IATAN et al., 2014). Furthermore, lipid metabolism in mammals is hypothesized to be associated with immune response and inflammatory processes. This consequently impacts lean deposition and subcutaneous fat deposition, as well as growth rate in cattle (SILVA-VIGNATO et al., 2019).

The number of genotyped animals with phenotypes for the traits of interest and the density of the panel used (number of SNPs after the quality control) are two key factors that influence the identification of important genomic regions, especially those with located in regions with low levels of linkage disequilibrium or with small effect on the trait. These two factors might have limited the genomic regions that were identified in this study. However, the SNP panel used in this study contains informative SNPs identified in several breeds, which were also used to develop the Montana Tropical® Composite population (Angus, Red Angus, Nellore, Brahman, Charolais,

Gelbvieh, Hereford, Limousin, Simmental, Holstein, Jersey, Brown Swiss, Ayrshire, Guernsey, Gyr, Girolando, Brangus, Beefmaster, and Braford), which might have minimized these effects. In view of the limitations described here, further studies using larger datasets and denser SNP panels should be performed to validate the results reported in this study.

3.4.3 Functional enrichment analyses

The moderate estimates of genetic correlation obtained, and the common genomic regions and candidate genes identified indicates that muscle development and fat deposition are directly correlated processes. Berg and Butterfield (1976) described that as soon as the animal reaches mature age, changes in the proportions of specific tissues are observed. This includes a decrease in muscle-bone growth rates, and an increase in fat deposition rate. The two main biological processes identified are: 1) "muscle tissue morphogenesis" (GO:006415); and 2) "regulation of fatty acid biosynthesis" (GO:0042304). A key gene of the muscle tissue morphogenesis is *RXRA* (*Retinoid X receptor, alpha*), which has been associated with weaning weight and yearling weight in Charolais and Brahman cattle (PAREDES-SANCHEZ et al*.,* 2015), and with BFT and meat fatty acids in an Angus-Hereford-Limousin crossbred population (GOSZCZYNSKI et al*.,* 2016). Fatty acid composition is directly linked with intramuscular fat content, and its major regulation is located in the skeletal muscle in mammals (MUOIO et al*.,* 2002). Meat fatty acid content is a crucial parameter of consumers acceptability and might become a key breeding goal in Nellore cattle (e.g., LEMOS et al*.,* 2016; FEITOSA et al*.,* 2017; 2019), one of the most influential breeds in the development of the Montana Composite population. In general, meat fatty acid content is related to meat quality and flavor, and complex interactions occurring during the animals' life and post-mortem period (MULLEN et al*.,* 2006).

Two of the highlighted processes are related to behavior indicator traits: 1) visual behavior; and 2) associative learning. The associative learning is defined as the capacity of an individual learning a behavior based on the association of two or more events (ABRAMSON; KIESON, 2016). In general, animals recognize events related to environmental factors through this process. For example, the animal's temperament from previous handling experiences produce an active learning process to determine how it will react in a next handling event. Furthermore, mounting behavior can result in carcass bruising and thus, reduce carcass quality especially depending on the level of BFT (HOFFMAN; LÜHL, 2012). This is a very important finding, as cattle temperament is significantly associated with handling stress and consequently, carcass damage and reduction in meat quality (YANG et al., 2019). The association between visual behavior and associative learning processes can also be related with feeding behavior which is a relevant process associated with feed efficiency, growth rate, and carcass composition.

The KEGG pathway PI3K-AKT is associated with stimulation of cell growth and proliferation, and simultaneously inhibits apoptosis. In this regard, PI3Ks plays a major role in insulin metabolism (MA et al., 2017), which is the major hormone controlling glucose and lipid metabolism (DIMITRIADIS et al., 2011). In this context, Shingu et al. (2001) suggested that insulin secretion may contribute to the difference in growth patterns and meat quality properties among beef cattle breeds. Another pathway enriched was “bta00270: Cysteine and methionine metabolism”, which is associated with meat flavor development in several species (MECCHI et al., 1964; MINOR et al., 1965; PEPPER; PEARSON, 1969; PIPPEN; MECCHI, 1969), and likely associated with intramuscular fat (or MARB). Cysteine and methionine are considered the largest components of meat flavor (WERKHOFF et al., 1990; KHAN et al., 2015) one of the most important and required of meat quality factors. Uncooked meat has little or no aroma and only a

blood-like taste, thus, the meat flavor is thermally derived by reactions between carbohydrates and amino acids (MOTTRAM, 1998).

3.5 CONCLUSIONS

Our findings indicate that ultrasound-based carcass and meat quality traits are heritable and therefore can be improved through selective breeding. The high genetic correlation between BFT and RFT indicate that indirect genetic response can be obtained by selecting for only one of them. The WssGBLUP method used to perform GWAS enabled the identification of various novel or already known candidate genes associated with the carcass and meat quality traits in the Montana Tropical® Composite population, but the traits studied have a polygenic nature. Some of the genes identified were previously associated with traits such as growth, carcass, body condition score, skeletal muscle growth, carcass fatness, and meat fatty acid composition. The main biological processes and pathways identified were “muscle tissue morphogenesis” and “regulation of fatty acid biosynthetic”, which biologically validate the ultrasound-based measurements. Further studies using larger datasets (ideally in independent populations) and denser SNP panels (> 30K) should be performed in order to validate the results reported in this study.

3.6 REFERENCES

ABRAMSON, C. I.; KIESON, E. Conditioning methods for animals in agriculture: A review. **Ciência Animal Brasileira,** v. 17, p. 359-375, 2016. <doi: 10.1590/1089-6891v17i341981>.

ADZITEY, F. Effect of pre-slaughter animal handling on carcass and meat quality. **International Food Research Journal,** v. 18, n. 2, 2011.

AGUILAR, I.; LEGARRA, A.; CARDOSO, F.; MASUDA, Y.; LOURENCO, D.; MISZTAL, I. (2019). Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. **Genetics Selection Evolution**, v. 51, n. 1, p. 28, 2019. <doi: 10.1186/s12711-019-0469-3>.

AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science,** v. 93, p.743-752, 2010. <doi: 10.3168/jds.2009-2730>.

AGUILAR, I.; MISZTAL, I.; TSURUTA, S.; LEGARRA, A.; WANG, H. PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. **In: Proceedings of the 10th world congress of genetics applied to livestock production.** 2014.

AN, B.; XIA, J.; CHANG, T.; WANG, X.; XU, L.; ZHANG, L.; GAO, X.; CHEN, Y.; LI, J. GAO, H. Genome-wide association study reveals candidate genes associated with body measurement traits in Chinese Wagyu beef cattle. **Animal genetics**, v. 50, n. 4, p. 386-390, 2019.doi: 10.1111/age.12805.

BERG, R. T.; BUTTERFIELD, R. M. **New concepts of cattle growth**. Sydney University Press, University of Sydney, 1976.

BERRY, D. P.; CONROY, S.; PABIOU, T.; CROMIE, A. R. Animal breeding strategies can improve meat quality attributes within entire populations. **Meat science**, v. 132, p. 6-18, 2017.

BRESSAN, M. C.; RODRIGUES, E. C.; PAULA, M. D. L. D.; RAMOS, E. M.; PORTUGAL, P. V.; SILVA, J. S.; BESSA, R. B.; GAMA, L.T. Differences in intramuscular fatty acid profiles among Bos indicus and crossbred Bos taurus× Bos indicus bulls finished on pasture or with concentrate feed in Brazil. **Italian Journal of Animal Science**, v. 15, n. 1, p. 10-21, 2016.

BRETHOUR, J. R. The relationship of average backfat thickness of feedlot steers to performance and relative efficiency of fat and protein retention. **Journal of animal science**, v. 82, n. 11, p. 3366-3372, 2004.

CHOAT, W. T.; KREHBIEL, C. R.; DUFF, G. C.; KIRKSEY, R. E.; LAURIAULT, L. M.; RIVERA, J. D.; CAPITAN, B. M.; WALKER, D. A.; DONART, G. B.; GOAD, C. L. Influence of grazing dormant native range or winter wheat pasture on subsequent finishing cattle performance, carcass characteristics, and ruminal metabolism. **Journal of animal science**, v. 81, n. 12, p. 3191-3201, 2003.

CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v. 42, n. 1, p. 2, 2010.

CINAR, M. U.; KAYAN, A.; UDDIN, M. J.; JONAS, E.; TESFAYE, D.; PHATSARA, C.; PONSUKSILI, S.; WIMMERS, K.; THOLEN, E.; LOOFT, C.; JÜNGST, H.; SCHELLANDER, K. Association and expression quantitative trait loci (eQTL) analysis of porcine AMBP, GC and PPP1R3B genes with meat quality traits. **Molecular biology reports**, v. 39, n. 4, p. 4809-4821, 2012.

CROUSE, J. D.; CUNDIFF, L. V.; KOCH, R. M.; KOOHMARAIE, M.; SEIDEMAN, S.

C. Comparisons of Bos indicus and Bos taurus inheritance for carcass beef characteristics and meat palatability. **Journal of Animal Science**, v. 67, n. 10, p. 2661-2668, 1989.

DIAS, R. A. P.; PETRINI, J.; FERRAZ, J. B. S.; ELER, J. P.; BUENO, R. S.; DA COSTA, A. L. L.; MOURÃO, G. B. Multicollinearity in genetic effects for weaning weight in a beef cattle composite population. **Livestock Science**, v. 142, n. 1-3, p. 188-194, 2011.

DIMITRIADIS, G.; MITROU, P.; LAMBADIARI, V.; MARATOU, E.; RAPTIS, S. A. Insulin effects in muscle and adipose tissue. **Diabetes research and clinical practice**, v. 93, p. S52-S59, 2011.

EDWARDS, D. B.; BATES, R. O.; OSBURN, W. N. Evaluation of Duroc-vs. Pietrain-sired pigs for carcass and meat quality measures. **Journal of Animal Science**, v. 81, n. 8, p. 1895-1899, 2003.

FEITOSA, F. L. B.; OLIVIERI, B. F.; ABOUJAOUDE, C.; PEREIRA, A. S. C. et al. Genetic correlation estimates between beef fatty acid profile with meat and carcass traits in Nellore cattle finished in feedlot. **Journal of applied genetics**, v. 58, n. 1, p. 123-132, 2017.

FEITOSA, F. L. B.; PEREIRA, A. S. C.; AMORIM, S. T.; PERIPOLLI, E.; DE OLIVEIRA SILVA, R. M.; BRAZ, C. U.; FERRINHO, A. M.; SCHENKEL, F. S.; BRITO, L. F.; ESPIGOLAN, R.; DE ALBUQUERQUE, L. G.; BALDI, F. Comparison between haplotype-based and individual snp-based genomic predictions for beef fatty acid profile in Nelore cattle. **Journal of Animal Breeding and Genetics**, 2019.

FERRAZ, J. B. S.; FELÍCIO, P. E. Production systems–An example from Brazil. **Meat science**, v. 84, n. 2, p. 238-243, 2010.

FERRAZ, J. B. S.; ELER, J. P.; GOLDEN, B. L. A formação do composto Montana Tropical. **Revista Brasileira de Reprodução Animal**, v. 23, p. 115-117, 1999.

FONT-I-FURNOLS, M.; GUERRERO, L. Consumer preference, behavior and perception about meat and meat products: An overview. **Meat science**, v. 98, n. 3, p. 361-371, 2014.

FORTES, M.R.; REVERTER, A.; HAWKEN, R.J.; BOLORMAA, S.; LEHNERT, S.A. Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. **Biology of reproduction**, v. 87, n. 3, p. 58, 1-8, 2012.

FORTES, M. R. S.; KEMPER, K.; SASAZAKI, S.; REVERTER, A.; PRYCE, J. E.; BARENDSE, W.; BUNCH, R.; MCCULLOCH, R.; HARRISON, B.; BOLORMAA, S.; ZHANG, Y. D.; HAWKEN, R. J.; GODDARD, M. E.; LEHNERT, S. A. Evidence for pleiotropism and recent selection in the PLAG 1 region in Australian Beef cattle. **Animal genetics**, v. 44, n. 6, p. 636-647, 2013.

GABÍN, B.; CAMERINO, O.; CASTAÑER, M.; ANGUERA, M. T. LINCE: new software to integrate registers and analysis on behavior observation. **Procedia Computer Science Technology**, 2012.

GORDO, D. G. M.; BALDI, F.; LÔBO, R. B.; FILHO, W. K.; SAINZ, R. D.; ALBUQUERQUE, L. G. D. Genetic association between body composition measured by ultrasound and visual scores in Brazilian Nelore cattle. **Journal of Animal Science**, v. 90, n. 12, p. 4223-4229, 2012.

GORDO, D. G. M.; ESPIGOLAN, R.; BRESOLIN, T.; FERNANDES JÚNIOR, G. A.; MAGALHÃES, A. F.; BRAZ, C. U., et al. Genetic analysis of carcass and meat quality traits in Nelore cattle. **Journal of animal science**, v. 96, n. 9, p. 3558-3564, 2018.

GOSZCZYNSKI, D. E.; MAZZUCCO, J. P.; RIPOLI, M. V.; VILLARREAL, E. L.;

ROGBERG-MUÑOZ, A.; MEZZADRA, C. A.; MELUCCI, L. M.; GIOVAMBATTISTA, G. Genetic characterisation of PPARG, CEBPA and RXRA, and their influence on meat quality traits in cattle. **Journal of animal science and technology**, v. 58, n. 1, p. 14, 2016.

GREGORY, K. E.; CUNDIFF, L. V.; KOCH, R. M.; LUNSTRA, D. D. Germplasm utilization in beef cattle. **Beef Res. Prog. Progress Report**, v. 4, p. 7-19, 1993. http://digitalcommons.unl.edu/hruskareports/140

GREGORY, K. E.; CUNDIFF, L. V.; KOCH, R. M.; DIKEMAN, M. E.; KOOHMARAIE, M. Breed effects, retained heterosis, and estimates of genetic and phenotypic parameters for carcass and meat traits of beef cattle. **Journal of animal science**, v. 72, n. 5, p. 1174-1183, 1994.

GREINER, S. P.; ROUSE, G. H.; WILSON, D. E.; CUNDIFF, L. V.; WHEELER, T. L. Prediction of retail product weight and percentage using ultrasound and carcass measurements in beef cattle. **Journal of animal science**, v. 81, n. 7, p. 1736-1742, 2003.

GRIGOLETTO, L.; BRITO, L. F.; MATTOS, E. C.; ELER, J. P.; BUSSIMAN, F. O.; SILVA, B. C. A.; SILVA, R. P.; CARVALHO, F. E.; BERTON, M.P.; BALDI, F.; FERRAZ, J. B. S. Genome-wide associations and detection of candidate genes for direct and maternal genetic effects influencing growth traits in the Montana Tropical® Composite population. **Livestock Science**, v. 229, p. 64-76, 2019.

GUERRERO, A.; VELANDIA VALERO, M.; CAMPO, M. M.; SAÑUDO, C. Some factors that affect ruminant meat quality: from the farm to the fork. Review. **Acta Scientiarum. Animal Sciences**, v. 35, n. 4, p. 335-347, 2013.

HAY, E. H.; ROBERTS, A. Genome-wide association study for carcass traits in a composite beef cattle breed. **Livestock Science**, v. 213, p. 35-43, 2018.

HERD, R. M.; ARTHUR, P. F.; BOTTEMA, C. D. K.; EGARR, A. R.; GEESINK, G. H.; LINES, D. S.; PIPER, S.; SIDDELL, J.P.; THOMPSON, J. M.; PITCHFORD, W.S. Genetic divergence in residual feed intake affects growth, feed efficiency, carcass and meat quality characteristics of Angus steers in a large commercial feedlot. **Animal Production Science**, v. 58, n. 1, p. 164-174, 2018.

HOFFMAN, L. C.; LÜHL, J. Causes of cattle bruising during handling and transport in Namibia. **Meat Science**, v. 92, n. 2, p. 115-124, 2012.

HUANG, D. W.; SHERMAN, B. T.; ZHENG, X.; YANG, J.; IMAMICHI, T.; STEPHENS, R.; LEMPICKI, R. A. Extracting biological meaning from large gene lists with DAVID. **Current protocols in bioinformatics**, v. 27, n. 1, p. 13.11. 1-13.11. 13, 2009.

HU, Z. L.; PARK C. A.; REECY, J. M. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. **Nucleic acids research**, v. 47, n. 1, p. 701-710, 2019.

IATAN, I.; CHOI, H. Y.; RUEL, I., et al. The WWOX gene modulates high-density lipoprotein and lipid metabolism. **Circulation: Cardiovascular Genetics**, v. 7, n. 4, p. 491-504, 2014.

KARISA, B. K.; THOMSON, J.; WANG, Z.; STOTHARD, P.; MOORE, S. S.; PLASTOW, G. S. Candidate genes and single nucleotide polymorphisms associated with variation in residual feed intake in beef cattle. **Journal of animal science**, v. 91, n. 8, p. 3502-3513, 2013.

KAS, K.; VOZ, M. L.; HENSEN, K.; MEYEN, E.; VAN DE VEN, W. J. Transcriptional activation capacity of the novel PLAG family of zinc finger proteins. **Journal of Biological Chemistry**, v. 273, n. 36, p. 23026-23032, 1998.

KAYAN, A. Identification of positional and functional candidate genes for meat and carcass quality in F2 Duroc x Pietrain resource population. Thesis. University of Bonn, Germany, 2011.

KHAN, M. I.; JO, C.; TARIQ, M. R. Meat flavor precursors and factors influencing flavor precursors—A systematic review. **Meat science**, v. 110, p. 278-284, 2015.

LEE, S.; MIN-WOOK, H.; SO-YOUNG, C.; KIM, J. Genome-wide association study identifies major loci for carcass weight on BTA14 in Hanwoo (Korean cattle). **PLoS One**, v. 8, n. 10, 2013.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of dairy science**, v. 92, n. 9, p. 4656-4663, 2009.

LEGARRA, A.; CHRISTENSEN, O. F.; AGUILAR, I.; MISZTAL, I. Single Step, a general approach for genomic selection. **Livestock Science**, v. 166, p. 54-65, 2014.

LEMOS, M. V.; CHIAIA, H. L. J.; BERTON, M. P.; FEITOSA, F. L.; ABOUJAOUD, C.; CAMARGO, G. M., et al. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. **BMC genomics**, v. 17, n. 1, p. 213, 2016.

LINDHOLM-PERRY, A. K.; SEXTEN, A. K.; KUEHN, L. A.; SMITH, T. P.; KING, D. A.; SHACKELFORD, S. D.; WHEELER, T. L.; FERRELL, C. L.; JENKINS, T. G.; SNELLING, W. M. FREETLY; H. C. Association, effects and validation of polymorphisms within the NCAPG-LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. **BMC genetics**, v. 12, n. 1, p. 103, 2011.

LITTLEJOHN, M.; GRALA, T.; SANDERS, K.; WALKER, C.; WAGHORN, G.;

MACDONALD, K.; COPPIETERS, W.; GEORGES, M.; SPELMAN, R.; HILLERTON, E.; DAVIS, S.; SNELL, R. Genetic variation in PLAG1 associates with early life body weight and peripubertal weight and growth in Bos taurus. **Animal Genetics**, v. 43, n. 5, p. 591-594, 2012.

LIU, Y.; LIU, X. L.; HE, H.; GU, Y. L. Four SNPs of insulin-induced gene 1 associated with growth and carcass traits in Qinchuan cattle in China. **Genet Mol Res**, v. 11, n. 2, p. 1209-1216, 2012.

LIU, X.; GUO, X.Y.; XU, X.Z.; WU, M.; ZHANG, X.; LI, Q.; MA, P.P.; ZHANG, Y.; WANG, C.Y.; GENG, F.J.; QIN, C.H. Novel single nucleotide polymorphisms of the bovine methyltransferase 3b gene and their association with meat quality traits in beef cattle. **Genet. Mol. Res**, v. 11, p. 2569-2577, 2012.

MA, M.; WANG, X.; CHEN, X.; CAI, R.; CHEN, F.; DONG, W.; YANG, G.; PANG, W. MicroRNA-432 targeting E2F3 and P55PIK inhibits myogenesis through PI3K/AKT/mTOR signaling pathway. **RNA biology**, v. 14, n. 3, p. 347-360, 2017.

MAGALHÃES, A. F.; DE CAMARGO, G. M.; JUNIOR, G. A. F.; GORDO, D. G.; TONUSSI, R. L.; COSTA, R. B., et al. Genome-wide association study of meat quality traits in Nellore cattle. **PLoS ONE**, v. 11, n. 6, p. e0157845, 2016.

MAGEE, D. A.; SIKORA, K. M.; BERKOWICZ, E. W.; BERRY, D. P.; HOWARD, D. J.; MULLEN, M. P.; EVANS, R. D.; SPILLANE, C.; MACHUGH, D. E. DNA sequence polymorphisms in a panel of eight candidate bovine imprinted genes and their association with performance traits in Irish Holstein-Friesian cattle. **BMC genetics**, v. 11, n. 1, p. 93, 2010.

MECCHI, E. P.; PIPPEN, E. L.; LINEWEAVER, H. Origin of hydrogen sulfide in heated chicken muscle. **Journal of Food Science**, v. 29, n. 4, p. 393-399, 1964.

MEIRELLES, S. L.; ALENCAR, M. M. D.; OLIVEIRA, H. N. D.; REGITANO, L. C. D. A. Efeitos de ambiente e estimativas de parâmetros genéticos para características de carcaça em bovinos da raça Canchim criados em pastagem. **Revista Brasileira de Zootecnia**, v. 39, n. 7, p. 1437-1442, 2010.

MINOR, L. J.; PEARSON, A. M.; DAWSON, L. E.; SCHWEIGERT, B. S. Chicken flavor: The identification of some chemical components and the importance of sulfur compounds in the cooked volatile fraction. **Journal of Food Science**, v. 30, n. 4, p. 686-696, 1965.

MISZTAL, I. et al. BLUPF90 and related programs (BGF90). In: **Proceedings of the 7th world congress on genetics applied to livestock production**. 2002. p. 743-744.

MISZTAL, I.; LEGARRA, A.; AGUILAR, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **Journal of dairy science**, v. 92, n. 9, p. 4648-4655, 2009.

MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; LAWLOR, T. J. Choice of parameters for single-step genomic evaluation for type. **J. Dairy Sci**, v. 93, n. Suppl 1, p. 533, 2010.

MISZTAL, I.; TSURUTA, S.; LOURENCO, D.; AGUILAR, I.; LEGARRA, A.; VITEZICA, Z. Manual for BLUPF90 family of programs. **Athens: University of Georgia**, 2014. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

MOTTRAM, D. S. Flavour formation in meat and meat products: a review. **Food chemistry**, v. 62, n. 4, p. 415-424, 1998.

MOURÃO, G. B.; FERRAZ, J. B. S.; ELER, J. P.; BALIEIRO, J. C. D. C.; BUENO, R. S.;

MATTOS, E. C.; FIGUEIREDO, L. G. G. Genetic parameters for growth traits of a Brazilian Bos taurus x Bos indicus beef composite. **Genetics and Molecular Research**, v. 6, n. 4, p. 1190-1200, 2007.

MULLEN, A. M.; STAPLETON, P. C.; CORCORAN, D.; HAMILL, R. M.; WHITE, A. Understanding meat quality through the application of genomic and proteomic approaches. **Meat science**, v. 74, n. 1, p. 3-16, 2006.

MUNRO, S.; CUTHBERTSON, D. J.; CUNNINGHAM, J.; SALES, M.; COHEN, P. T. Human skeletal muscle expresses a glycogen-targeting subunit of PP1 that is identical to the insulin-sensitive glycogen-targeting subunit GL of liver. **Diabetes**, v. 51, n. 3, p. 591-598, 2002.

MURAMATSU, Y. Multiple Marker Effects of Single Nucleotide Polymorphisms in Two Genes, NCAPG and PLAG1, for Carcass Weight in Japanese Black Cattle. **Open Journal of Animal Sciences**, v. 9, n. 1, p. 129-134, 2018.

MUOIO, D. M.; WAY, J. M.; TANNER, C. J.; WINEGAR, D. A.; KLIEWER, S. A.; HOUMARD, J. A., et al. Peroxisome proliferator-activated receptor-α regulates fatty acid utilization in primary human skeletal muscle cells. **Diabetes**, v. 51, n. 4, p. 901-909, 2002.

NJISANE, Y. Z.; MUCHENJE, V. Farm to abattoir conditions, animal factors and their subsequent effects on cattle behavioural responses and beef quality—A review. **Asian-Australasian journal of animal sciences**, v. 30, n. 6, p. 755, 2017.

PAREDES-SÁNCHEZ, F. A.; SIFUENTES-RINCÓN, A. M.; CABRERA, A. S.; PÉREZ, C. A. G.; BRACAMONTE, G. M. P.; MORALES, P. A. Associations of SNPs located at candidate genes to bovine growth traits, prioritized with an interaction networks construction approach. **BMC genetics**, v. 16, n. 1, p. 91, 2015.

PATHAK, V.; SINGH, V. P.; SANJAV, Y. Ultrasound as a modern tool for carcass evaluation and meat processing: a review. **Int J Meat Sci**, v. 1, n. 1, p. 83-92, 2011.

PEPPER, F. H.; PEARSON, A. M. Changes in hydrogen sulfide and sulfhydryl content of heated beef adipose tissue. **Journal of Food Science**, v. 34, n. 1, p. 10-12, 1969.

PETRINI, J.; DIAS, R. A. P.; PERTILE, S. F. N.; ELER, J. P.; FERRAZ, J. B. S.; MOURÃO, G. B. Degree of multicollinearity and variables involved in linear dependence in additive-dominant models. **Pesquisa Agropecuária Brasileira**, v. 47, n. 12, p. 1743-1750, 2012.

PICCOLI, M. L.; BRITO, L. F.; BRACCINI, J.; OLIVEIRA, H. R.; CARDOSO, F. F.; ROSO, V. M.; SARGOLZAEI, M.; SCHENKEL, F. S. Comparison of genomic prediction methods for evaluation of adaptation and productive efficiency traits in Braford and Hereford cattle. **Livestock Science**, v. 231, p. 103864, 2020.

PIPPEN, E. L.; MECCHI, E. P.; NONAKA, M. Origin and nature of aroma in fat of cooked poultry. **Journal of Food Science**, v. 34, n. 5, p. 436-442, 1969.

PORTO-NETO, L. R.; BUNCH, R. J.; HARRISON, B. E.; BARENDSE, W. Variation in the XKR4 gene was significantly associated with subcutaneous rump fat thickness in indicine and composite cattle. **Animal genetics**, v. 43, n. 6, p. 785-789, 2012.

RAMAYO-CALDAS, Y.; FORTES, M. R. S.; HUDSON, N. J.; PORTO-NETO, L. R.; BOLORMAA, S.; BARENDSE, W., et al. A marker-derived gene network reveals the regulatory role of PPARGC1A, HNF4G, and FOXP3 in intramuscular fat deposition of beef cattle. **Journal of Animal Science**, v. 92, n. 7, p. 2832-2845, 2014.

REVERTER, A.; JOHNSTON, D. J.; GRASER, H. U.; WOLCOTT, M. L.; UPTON, W. H. Genetic analyses of live-animal ultrasound and abattoir carcass traits in Australian Angus and Hereford cattle. **Journal of Animal Science**, v. 78, n. 7, p. 1786-1795, 2000.

RINCON, G.; ISLAS-TREJO, A.; CASTILLO, A. R.; BAUMAN, D. E.; GERMAN, B. J.; MEDRANO, J. F. Polymorphisms in genes in the SREBP1 signalling pathway and SCD are associated with milk fatty acid composition in Holstein cattle. **Journal of Dairy Research**, v. 79, n. 1, p. 66-75, 2012.

RODRIGUES, R. T. S.; CHIZZOTTI, M. L.; VITAL, C. E.; BARACAT-PEREIRA, M. C.; BARROS, E.; BUSATO, K. C.; GOMES, R. A.; LADEIRA, M. M.; MARTINS, T. S. Differences in beef quality between Angus (Bos taurus taurus) and Nellore (Bos taurus indicus) cattle through a proteomic and phosphoproteomic approach. **PloS one**, v. 12, n. 1, 2017.

SANTANA, M. L.; ELER, J. P.; CARDOSO, F. F.; ALBUQUERQUE, L. G. D.; BIGNARDI, A. B.; FERRAZ, J. B. S. Genotype by environment interaction for birth and weaning weights of composite beef cattle in different regions of Brazil. **Livestock science**, v. 149, n. 3, p. 242-249, 2012.

SANTANA, M. L.; ELER, J. P.; CARDOSO, F. F.; ALBUQUERQUE, L. G. D.; FERRAZ, J. B. S. Phenotypic plasticity of composite beef cattle performance using reaction norms model with unknown covariate. **Animal**, v. 7, n. 2, p. 202-210, 2013.

SCHOLZ, A. M.; BÜNGER, L.; KONGSRO, J.; BAULAIN, U.; MITCHELL, A. D. Non-invasive methods for the determination of body and carcass composition in livestock: dual-energy X-ray absorptiometry, computed tomography, magnetic resonance imaging and ultrasound: invited review. **Animal**, v. 9, n. 7, p. 1250-1264, 2015.

SILVA-VIGNATO, B.; COUTINHO, L. L.; POLETI, M. D.; CESAR, A. S.; MONCAU, C. T.; REGITANO, L. C.; BALIEIRO, J. C. Gene co-expression networks associated with carcass traits reveal new pathways for muscle and fat deposition in Nelore cattle. **BMC genomics**, v. 20, n. 1, p. 32, 2019.

SILVA, R. M. O.; STAFUZZA, N. B.; DE OLIVEIRA FRAGOMENI, B.; DE CAMARGO, G. M. F.; CEACERO, T. M.; CYRILLO, J. N. D. S. G.; BALDI, F.; BOLIGON, A. A.; MERCANDANTE, M.E.Z.; LOURENCO, D. L.; MISZTAL, I.; ALBUQUERQUE L. G. Genome-wide association study for carcass traits in an experimental Nelore cattle population. **PLoS ONE**, v. 12, n. 1, 2017.

SILVA, R. P.; BERTON, M. P.; GRIGOLETTO, L.; CARVALHO, F. E.; SILVA, R. M.; PERIPOLLI, E.; CASTRO, L. M.; FERRAZ, J. B. S.; ELER, J. P.; LÔBO, R. B.; BALDI, F. Genomic regions and enrichment analyses associated with carcass composition indicator traits in Nellore cattle. **Journal of animal breeding and genetics**, v. 136, n. 2, p. 118-133, 2019.

SHINGU, H.; HODATE, K.; KUSHIBIKI, S.; UEDA, Y.; WATANABE, A.; SHINODA, M.; MATSUMOTO, M. Profiles of growth hormone and insulin secretion, and glucose response to insulin in growing Japanese Black heifers (beef type): comparison with Holstein heifers (dairy type). **Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology**, v. 130, n. 2, p. 259-270, 2001.

TONUSSI, R. L.; ESPIGOLAN, R.; GORDO, D. G. M.; MAGALHÃES, A. F. B.; VENTURINI, G. C.; BALDI, F.; OLIVEIRA, H. N.; CHARDULO, L. A.; TONHATI, H.; ALBUQUERQUE, L. G. Genetic association of growth traits with carcass and meat traits in Nellore cattle. **Genet. Mol. Res**, v. 14, n. 4, p. 18713-18719, 2015.

TSURUTA, S.; MISZTAL, I.; AGUILAR, I.; LAWLOR, T. J. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. **Journal of Dairy Science**, v. 94, n. 8, p. 4198-4204, 2011.

UNITED STATES DEPARTMENT OF AGRICULTURE (USDA). National Agricultural Statistics Service (NASS), AGRICULTURAL STATISTICS BOARD, 2018.

UTSUNOMIYA, Y. T., et al. A PLAG1 mutation contributed to stature recovery in modern cattle. **Scientific reports**, v. 7, n. 1, p. 1-15, 2017.

VAN DYCK, F.; BRAEM, C. V.; CHEN, Z.; DECLERCQ, J., et al. Loss of the PlagL2 transcription factor affects lacteal uptake of chylomicrons. **Cell metabolism**, v. 6, n. 5, p. 406-413, 2007.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of dairy science**, v. 91, n. 11, p. 4414-4423, 2008.

VITEZICA, Z. G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. Bias in genomic predictions for populations under selection. **Genetics Research**, v. 93, n. 5, p. 357-366, 2011.

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; FERNANDO, R. L.; VITEZICA, Z.; OKIMOTO, R.; WING, T.; HAWKEN, R.; MUIR, W. M. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in genetics**, v. 5, p. 134, 2014.

WENG, Z.; SU, H.; SAATCHI, M.; LEE, J.; THOMAS, M. G.; DUNKELBERGER, J. R.;

GARRICK, D. J. Genome-wide association study of growth and body composition traits in Brangus beef cattle. **Livestock Science**, v. 183, p. 4-11, 2016.

WERKHOFF, P.; BRUENING, J.; EMBERGER, R.; GUENTERT, M.; KOEPSEL, M.; KUHN, W.; SURBURG, H. Isolation and characterization of volatile sulfur-containing meat flavor

components in model systems. **Journal of Agricultural and Food Chemistry**, v. 38, n. 3, p. 777-791, 1990.

WIGGANS, G. R.; SONSTEGARD, T. S.; VANRADEN, P. M.; MATUKUMALLI, L. K.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S.; VAN TASSELL, C. P. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. **Journal of dairy science**, v. 92, n. 7, p. 3431-3436, 2009.

YANG, F. L.; ANSCHUTZ, K. S.; BALL, J. J.; HORNSBY, P.; REYNOLDS, J. L.; POHLMAN, F. W. Evaluating the relationship of animal temperament to carcass characteristics and meat quality. **Meat and Muscle Biology**, v. 3, n. 1, p. 70-75, 2019.

YOKOO, M. J.; LOBO, R. B.; ARAUJO, F. R. C.; BEZERRA, L. A. F.; SAINZ, R. D.; ALBUQUERQUE, L. G. D. Genetic associations between carcass traits measured by real-time ultrasound and scrotal circumference and growth traits in Nelore cattle. **Journal of Animal Science**, v. 88, n. 1, p. 52-58, 2010.

ZERBINO, D. R.; ACHUTHAN, P.; AKANNI, W.; AMODE, M. R. et al. Ensembl 2018. **Nucleic acids research**, v. 46, n. 1, p. 754-761, 2018.

ZHANG, X.; LOURENCO, D.; AGUILAR, I.; LEGARRA, A.; MISZTAL, I. Weighting strategies for single-step genomic BLUP: an iterative approach for accurate calculation of GEBV and GWAS. **Frontiers in genetics**, v. 7, p. 151, 2016.

ZHANG, R.; MIAO, J.; SONG, Y.; ZHANG, W.; XU, L.; CHEN, Y.; ZHANG, L.; GAO, H.; ZHU, B.; LI, J.; GAO X. Genome-wide association study identifies the PLAG1-OXR1 region on BTA14 for carcass meat yield in cattle. **Physiological genomics**, v. 51, n. 5, p. 137-144, 2019.

ZHAO, F.; MCPARLAND, S.; KEARNEY, F.; DU, L.; BERRY, D. P. Detection of selection signatures in dairy and beef cattle using high-density genomic information. **Genetics Selection Evolution**, v. 47, n. 1, p. 49, 2015.

ZHAO, L.; HUANG, Y.; DU, M Farm animals for studying muscle development and metabolism: dual purposes for animal production and human health. **Animal Frontiers**, v. 9, n. 3, p. 21-27, 2019.

CHAPTER 4. GENOTYPE IMPUTATION AND GENETIC DIVERSITY OF TROPICAL COMPOSITE BEEF CATTLE

Laís Grigoletto[a,b], Hinayah R. Oliveira[b], Joanir Pereira Eler[a], Fernando Oliveira Bussiman[a], José Bento Sterman Ferraz[a] and Luiz F. Brito[b]

[a]Department of Animal Sciences, College of Animal Sciences and Food Engineering, University of São Paulo, Pirassununga, Sao Paulo, Brazil

[b]Department of Animal Sciences, Purdue University, West Lafayette, Indiana, USA

ABSTRACT

Tropical composite beef cattle have been developed to improve Brazilian beef production. This is a common strategy used in tropical and subtropical regions. This population became a powerful genetic source as the population was formed by mating taurine (*Bos taurus taurus*) and indicine (*Bos taurus indicus*) cattle breeds, exploiting the heterosis and complementarity. The evaluation of genetic diversity metrics, such as linkage disequilibrium (LD) and population structure, as well as the consistency of the gametic phase across breeds for the successful implementation of genomic selection, is crucial. Therefore, the objectives of this study were: (1) to design a moderate density SNP chip panel and perform genotype imputation through different scenarios including purebred and crossbreed genotypes; and (2) to characterize the genetic diversity of Montana Tropical® Composite, Angus, Nellore, Senepol, and crossbred Angus x Simmental. The pedigree included more than 550,000 animals and the genomic database consisted of a total of 4,092 samples from five different breeds. All marker panels used were from the Illumina Bovine BeadChip array from a diversity of densities ranging from a high (777,962 SNPs) to moderate (52,598 to 30,105 SNPs) number of SNPs. The level of linkage disequilibrium (r^2) decayed rapidly with greater distances (> 0.10 Mb) in composite animals. The highest consistency of phase was found for Montana Tropical® composite and crossbreed Angus x Simmental, suggesting a greater level of relatedness between these populations. This indicates an opportunity for joint genomic analysis. Our study provided a great deal of information regarding the population structure and genetic diversity using the pedigree and genomic information of tropical composite cattle breeds. These results contribute to optimizing the implementation of genomic selection in these populations.

Keywords: composite, consistency of gametic phase, linkage disequilibrium, genomic, SNP panel

4.1 INTRODUCTION

The Montana Tropical® composite beef cattle have been developed to be raised in tropical and subtropical conditions, such as Brazil, based on the Clay Center methodology (CUNDIFF; GREGORY, 1999) to develop a crossbreeding population to rely on high genetic performing animals (FERRAZ; ELER; GOLDEN, 1999). This population became a powerful genetic resource as the animals were formed by mating taurine (*Bos taurus taurus*) and indicine (*Bos taurus indicus*) cattle subspecies. To exploit the genetic phenomena of heterosis and complementarity, the higher meat quality and precocity reproductive traits in taurine breeds were combined with the adaption of heat tolerance and resistance for disease and parasites present in indicine breeds, thus, improving the expression of productive, and hybrid vigor in animals.

In the literature, many key studies were performed for this composite population considering different models to estimate genetic parameters (MOURÃO et al., 2007; BUENO et al., 2012; DIAS et al., 2011), the use of different genetic groups (SANTANA et al., 2012; 2014), and the interaction of the genetic by environment (PETRINI et al., 2012). However, only a limited number of studies including genomic information has been conducted (PERIPOLLI et al., 2019; GRIGOLETTO et al., 2019; 2020). In this regard, the complexity of studying composite populations is due to the characterization of those animals with the aspects of population history due to the possible occurrence of population stratification, providing a better understanding of the genetic background in order to avoid adding noise on further genomic analyses (i.e., genomic prediction and genome- wide association studies).

Genomic evaluations have been successfully implemented in livestock, particularly in purebred cattle populations, and this allowed substantial increases in the rates of genetic gain (HAYES et al., 2009; WIGGANS et al., 2011; OLSON et al., 2012; GARCÍA-RUIZ et al., 2016). A large number of phenotyped individuals with genotypes for high-density SNP panels are required to increase the accuracy of genomic breeding values, however this can be very expensive. Most beef cattle populations (e.g., Nellore, Montana) have only a small number of young bulls as selection candidates available to construct the reference population (COLE; SILVA, 2016). Thus, genotype imputation is a well-established powerful statistical tool combining genotyped information from different density panels (e.g., high, moderate, and low-density) in order to infer unknown marker information, provided enough overlap exists between panels. To date, several studies using different strategies and methodologies for genomic imputation have been performed for different species, including cattle, pigs, sheep, and poultry. In this sense, Ventura et al. (2014), Chud et al. (2015), and Mokry et al. (2014) have shown suitable results in crossbred cattle. Otherwise, the implementation of genomic methods, such as genotype imputation, are still challenging in composite populations.

In this way, it is relevant to estimate the metrics to characterize the genetic diversity for composite populations. Additionally, in order to better understand the population structure of a composite population, single nucleotide polymorphisms (SNPs) are a useful tool to capture accurate information compared with pedigree analyses. The extent of linkage disequilibrium (LD) can be defined as a central parameter for the success implementation of genomic selection, since this metric is defined as a non-random association of alleles at two or more loci and is highly influenced by the population history, which can vary between breeds (BRITO et al., 2015). In addition, the principal component analysis (PCA), admixture, and other parameters, such as the

level of heterozygosis and the consistency of the gametic phases, are crucial to better understanding the genetic background across this composite population and other breeds. Therefore, the objectives of this study were: (1) to design a moderate density SNP chip and perform the genotype imputation through different scenarios, including purebreds and crossbreed strategies; (2) to characterize the genetic diversity of beef breeds using genomic information from Montana Tropical Composite, Angus, Nellore, Senepol, and the crossbred Angus x Simmental, concerning the population history of this population.

4.2 MATERIAL AND METHODS

4.2.1 Pedigree Structure

The Montana Tropical® population (Figure 4.1A) contains around 550,000 animals, from which almost 400,000 were measured for economically important traits, such as growth, reproduction, meat quality, and carcass (Table 4.1). The breed composition of the Montana Tropical® Composite cattle is not the same for all animals, as crossbreeding is still being performed according to the NABC system: N: Zebu breeds (*Bos taurus indicus*), A: Adapted Taurine breeds (*Bos taurus taurus*), B: British breeds (*Bos taurus taurus*), and C: Continental European breeds (*Bos taurus taurus*) (FERRAZ; ELER; GOLDEN, 1999). Therefore, the genetic composition of each animal derives from different breed proportions, but follows the criteria that determine the minimum percentage of each biological type accepted to be considered a Montana Tropical® Composite: 12.5% for group A, and 25% for groups N plus A. The maximum proportions considered was 37.5% for group N (Zebu breeds); 87.5% for group A (Adapted breeds); and 75% for groups B (British breeds) and C (Continental breeds; SANTANA et al., 2013; Table 4.1).

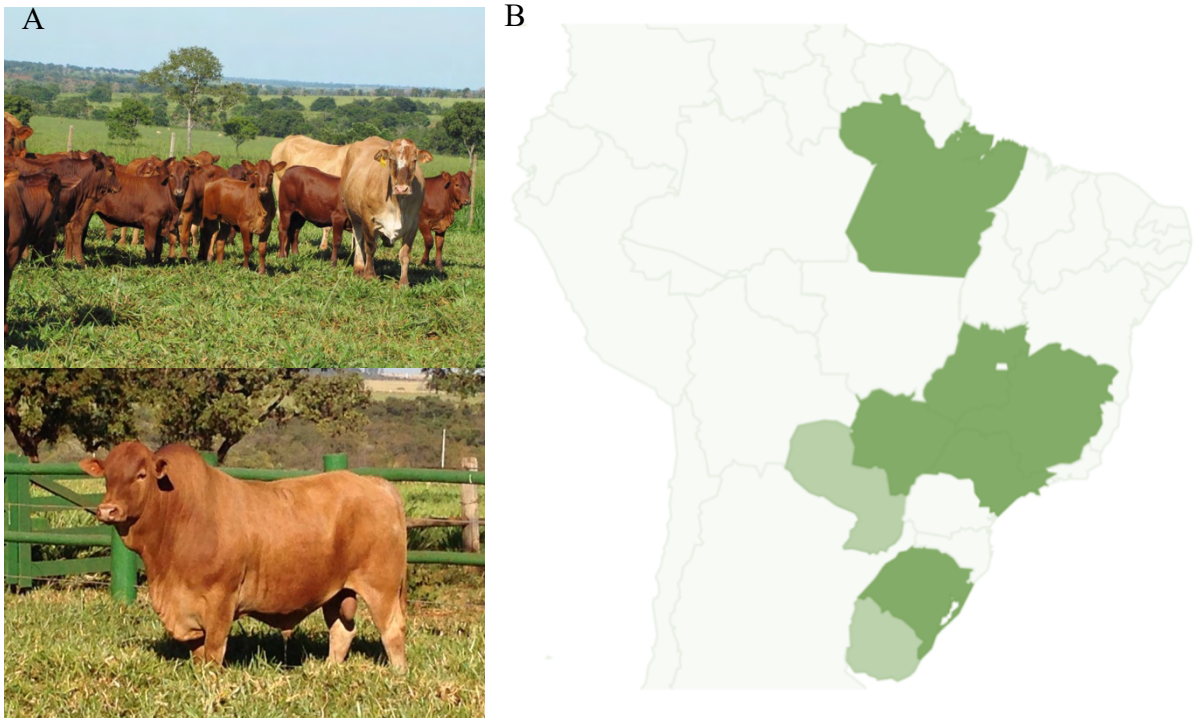


Figure 4.1 Illustration of Montana Tropical® Composite animals (A). Location of farms participating in the Montana Tropical® breeding program (B). Brazilian states are showed in dark green, and Paraguay e Uruguay are showed in light green. Photo Credits: Montana Composto Tropical® website (compostomontana.com.br/criadores-montana/).

Table 4.1 Summary of the NABC* system criteria used in the Montana Tropical® composite breeding program.

Biological Type	Minimum	Maximum
Number of breeds	3	No limits
N	0	37.5%
A	12.5%	87.5%
N + A	25%	100%
B	0	75%
C	0	75%
B + C	0	75%

*N: Zebu breeds (Nellore); A: Adapted taurine breeds; B: British breeds; C: Continental breeds.

The CFC program (SARGOLZAEI et al., 2006) was used to calculate a summary of the pedigree statistics and genetic structure analysis of the population. In addition, this program was used to check the pedigree for possible errors.

4.2.2 Breeds and genomic dataset

The genomic dataset was provided by the Montana Tropical® composite breeding program stored and analyzed by the Animal Breeding and Biotechnology Group of the College of Animal Science and Food Engineering (Pirassununga, Sao Paulo, Brazil; GMAB/FZEA), and the Department of Animal Science of the Purdue University (West Lafayette, Indiana, USA). The dataset consisted of a total of 4,092 animals from five different breeds and sources, as described in Table 4.2. All SNP chip panels used were from the Illumina Bovine BeadChip (Illumina, San Diego, CA) array, but from different densities. The SNP densities ranged from a high (777,962 SNPs) to moderate (52,598 to 30,105 SNPs) number of SNPs. Thus, this study was conducted using two genotype sources: (i) A total of 3,605 animals from Montana Tropical composite (N = 1,900), Nellore (indicine breed; N = 1,590), Angus (taurine breed; N = 65), and Senepol (taurine breed; N = 50), made available from the Animal Breeding and Biotechnology Group (GMAB/FZEA); and (ii) 487 crossbred animals from Angus and Simmental (taurine breeds) from Purdue University (West Lafayette, IN, USA).

Table 4.2 Number of genotyped animals, name of SNP panel, number of SNPs and source for each breed.

Breed	Number of Animals	SNP panel	Number of SNPs	Source
Montana Tropical® composite	1,436	GGP® Low-Density (GGP-LD)	30,105	GMAB/FZEA
	448	GGP® Low-Density (GGP LD)	35,339	
	16	Illumina® Bovine HD	777,962	
Nellore	1,590	Illumina® Bovine HD	777,962	GMAB/FZEA
Purdue (Angus + Simmental)	487	Custom SNP panel*	52,598	Purdue University
Angus	65	GGP® Bovine 50K	47,843	GMAB/FZEA
Senepol	50	GGP® Bovine 50K	47,843	GMAB/FZEA
Total	4,092			

GGP = GeneSeek® Genomic Profile
*Panel designed particularly for the use of this information by the Purdue University.

4.2.3 Genotype Imputation

For this study, a custom SNP array was developed based on imputation, in order to increase the number of markers used to combine both moderate-density SNP panels (30K plus 35K). This custom SNP array contained approximately 55,000 SNPs, which was also used as target to reduce the HD genotypes in the second step. Genotypes were codified as 0, 1, and 2 to denote homozygous (AA), heterozygous (Aa), and homozygous (aa), respectively. Missing markers were coded as 5. The number of SNPs present in each SNP array density are described in Table 4.3, as well the number of overlapping SNPs between them.

To combine the different SNP panels, all overlapping markers from each breed and panel were considered by merging files in the PLINK software (PURCELL et al., 2007), using the flag *–merge*. Thereafter, the imputation was performed using the FImpute software (SARGOLZAEI et al., 2014), including all animals in the reference population. The software used is available for research purposes at http://www.aps.uoguelph.ca/~msargol/fimpute. An illustration of the genotyping imputation strategy is shown in Figure 4.2.

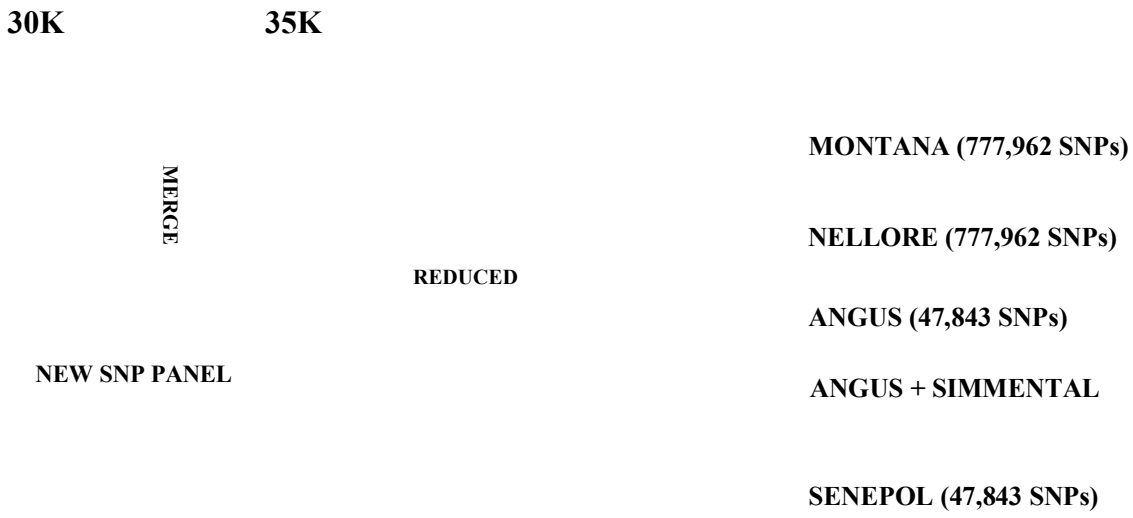


Figure 4.2 Illustration of the genotyping imputation strategy.

4.2.4 Imputation scenarios

Regarding the scenarios used, different '*pools*' of genotypes available including the Montana, crossbred, and purebred animals. At first, one exclusive scenario was tested to assess the imputation accuracy from the Montana animals between both moderate SNP densities (30K and 35K), as described below. The scenarios used to assess genotype imputation accuracy using the reference group consisted of all individuals, considered as (A) Montana Tropical®, (B) Montana Tropical®+Nelore+Senepol+Purdue+Angus, and (C) same as scenario B, but excluding the Purdue

samples. The imputation was performed by FImpute (SARGOLZAEI et al., 2014) using the population method.

The scenario test was performed to access the accuracy of imputation considering moderate SNP array (30K and 35K). First, the animals were divided into a training and validation population, based on their birth year. For the 30K panel, the population was split with 1,149 animals in training and 287 on validation. For the 35K panel, the animals were split with 358 and 90 for training and validation, respectively. Only the overlapping SNPs between both panels were considered (N=9,830 SNPs) for imputation in the validation population.

4.2.5 Genotype Quality Control

Genotype quality control was performed in PLINK using all SNPs with known positions. SNPs with a minor allele frequency (MAF) lower than 0.01, a call rate lower than 95%, SNPs located on the sexual and mitochondrial chromosome, and those that did not follow the Hardy Weinberg Equilibrium (HWE; considering as threshold 1e-15) were excluded. After the genotype quality control, 51,709 SNPs remained in this study.

Table 4.3 The number of SNP on each panel (diagonals), and number of overlapping SNPs between panels (off diagonals).

Panel	GGP® Low-Density (GGP-LD)	GGP® *indicus*	Illumina® Bovine HD	Custom SNP panel	GGP® Bovine 50K
GGP® Low-Density (GGP-LD)	**30,105**				
GGP® *indicus*	9,830	**35,339**			
Illumina® Bovine HD	29,141	35,269	**777,962**		
Custom SNP panel	10,229	2,803	46,653	**52,598**	
GGP® Bovine 50K	10,263	3,552	41,848	15,133	**47,843**

4.3 GENETIC DIVERSITY METRICS

4.3.1 Minor allele frequency

The average minor allele frequency (MAF) was estimated by applying the *–freq* flag in PLINK (PURCELL et al., 2007) software to estimate the frequency of the least common allele per breed.

4.3.2 Linkage disequilibrium

Pair-wise LD was calculated by the genotype squared correlation coefficient (r^2) between alleles at two loci, as proposed by Hill and Robertson (1968):

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

where D = *f*(AB) – *f*(A) *f*(B) and *f* (AB), *f* (A), *f* (a), *f* (B), and *f* (b), are the observed frequencies of haplotype AB and alleles A, a, B, and b, respectively. The r^2 was estimated for each pair of loci on each chromosome to determine the LD between adjacent SNPs and the LD decay over different distances, by using the *–r2 –ld-window* 99999 *–ld-window-r2* 0 commands in PLINK v1.9 (PURCELL et al., 2007). To estimate the LD decay over different distances, SNP pairs on the autosomes were sorted into five distance bins based on their pair-wise marker distance and the average of each bin was calculated following the distances: lower than 0.05 Mb, until 0.1 Mb, 0.25 Mb, 0.50, and greater than 0.75 Mb.

4.3.3 Heterozygosity

The observed heterozygosity (HO) per animal, within breeds, was calculated using the SNP markers that passed the quality control and was compared to the expected heterozygosity (HE) under the Hardy Weinberg Equilibrium (HWE). Therefore, the HO was calculated as the number of heterozygotes divided by the total number of genotypes. The estimates were calculated using the *–hardy* flag in PLINK using the default settings.

4.3.4 Consistency of gametic phase

The consistency of gametic phase was measured as the Pearson correlation of the signed *r* values (gametic phase) between all pairs of breeds ($n = 5$). For each marker pair with a measure of r^2, the signed *r* value was determined by taking the square root of the r^2 value and assigning the appropriate sign based on the calculated disequilibrium (LD) value. The data were grouped into bins based on the pair-wise marker distances of 0–1, 1–10, 10–50, 50–100, 100–500, and 500–1,000 kb to assess the consistency of the gametic phase at the smallest distances possible, given the number of genotyped SNPs. For each distance bin, the signed *r* values were then correlated between all pairs of breeds using the COR function procedure in the R software (R CORE TEAM, 2019).

4.3.5 Admixture

To examine the most likely number of founder breeds used to create the Montana Tropical® composite, we used the Maximum Likelihood approach implemented in ADMIXTURE software (ALEXANDER; NOVEMBRE; LANGE, 2009). In this context, a cross-validation procedure was used to choose the optimum number of founder populations, by changing the K-value from 5 to

30 using the flag --cv. The level of ancestry divergence among populations (Fst) was also obtained using the ADMIXTURE software in pairwise analyses. Additionally, a partial supervised approach was carried out assuming the known ancestry K for samples corresponding to well-known established pure breeds. A total of 51,709 markers (after quality control) were selected from these analyses from all beef cattle breeders included in this study.

4.3.6 Principal Component Analysis

Principal component analysis (PCA) was used to graphically visualize the genetic diversity in this dataset. Therefore, the *–pca* flag available in PLINK (PURCELL et al., 2007) was used. PCA was performed on the genotypic data by applying the eigen value decomposition on the covariance matrix of the standardized genotype data (i.e., each SNP column was mean centered and divided by the standard deviation).

4.4 RESULTS AND DISCUSSION

Pedigree structure. The characterization of the pedigree structure is presented in Table 4.4. The main factor to discuss is the lack of pedigree information (16.7% of total animals had both parents unknown). Additionally, the proportion of animals with only one parent information reached 30%. In this case, the use of genomic information may be important to help improve the quality of pedigree information.

Table 4.5 shows the distribution of animals according to their inbreeding coefficient. The average inbreeding coefficient for the genotyped animals was low in the population studied (0.001). Only a relatively small portion of the role of pedigree data (397 animals) presented an inbreeding coefficient higher than 0.20 (Table 4.5). These results were as expected, being related to a composite population, as inbreeding (i.e., increases in the homozygosity) has the opposite

effect of heterosis (i.e., increases in the heterozygosity). In this particularly complex population, the level of heterozygosity was high in the genetic background of each animal.

Table 4.4 Characterization of pedigree structure of Montana Tropical® composite cattle population.

	Number of Animals	**Percentage (%)**
Sire	4,172	0.8
Dam	195,524	37.2
Individuals with progeny	199,696	38
Individual without progeny	326,110	62
Only with known sire	1,786	0.34
Only with known dam	158,622	30.2
With both known	277,720	52.8
With both unknown	87,678	16.7
Total	**525,806**	**100**

This information is useful for the imputation approach, as it is possible to construct the haplotypes and use the relatives (SARGOLZAEI et al., 2014) to improve the imputation results. This could also provide substantial information regarding the relationship between all animals to further genomic selections analysis. Therefore, Montana genotyped animals included selected candidates with a higher performance from the genetic evaluation of the composite breeding program, including certain key tested animals. The Nellore animals included here were important sires and dams that were present in the Montana background. In constructing the necessary haplotypes to increase the power of imputation, crossbreed genotyped animals from Angus and Simmental belonging to the Purdue University, Department of Animal Science were added in this study. Angus and Senepol animals represent crucial breeds for this composite population.

Table 4.5 The rate of inbreeding using pedigree information for Montana composite animals.

Inbreeding	Number of Animals
0.00 < F <= 0.05	30,254
0.05 < F <= 0.10	1,565
0.10 < F <= 0.15	798
F > 0.20	397

Genetic diversity metrics

The minor allele frequency (MAF), the expected (i.e., the expected heterozygosity level under random mating conditions) and the observed heterozygosity (i.e., the observed occurrence of heterozygous individuals at a locus), and the average linkage disequilibrium of adjacent SNPs estimated for each breed are shown in Table 4.6. Figure 4.3 presents the distribution of SNP by MAF range across breeds. Conversely, the Angus and Senepol breeds presented a more related SNP distribution for the average MAF ranges. Nellore cattle had the highest rate of low MAF (14.03%) which could imply that the major allele for the SNP is conserved. The Montana Tropical® population presented a higher proportion of SNPs with MAF < 0.15 (88.97%) compared to the crossbred population (77.41%).

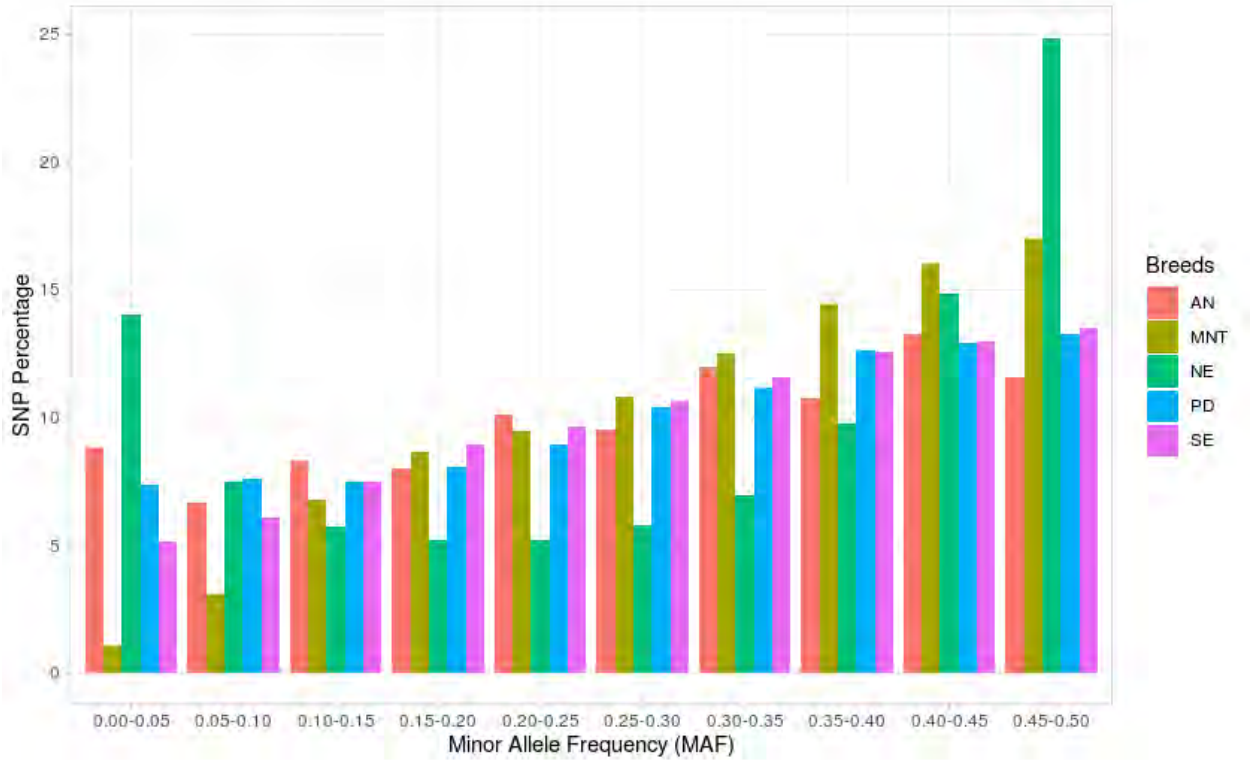


Figure 4.3 Distribution of SNPs by MAF ranges and breed **(**AN = Angus; MNT = Montana; NE = Nellore; PD = Purdue; SE = Senepol).

As expected, composite and crossbred animals had the highest heterozygosity when compared to the purebreds. Similar results were found between Angus and Senepol. The greater levels of heterozygosity indicated that the composite population had greater adaptive genetic variation and fitness in comparison to populations with lower levels of heterozygosity, which may be due to the hybrid vigor. This statement can be confirmed by the fact that the exploration of heterosis is directly proportional to the heterozygosity level obtained by crossing different breeds (FALCONER; MACKAY, 1996). Our estimates were similar to the ones reported by Kelleher et al. (2017; HO = 0.385 and HE=0.387 for Angus, and HO=0.384 and HE=0.386 for Simmental). Similar estimates were also obtained by Zhang et al. (2018) for Angus beef cattle (HO = 0.395 and HE=0.386). For the indicine breeds, here represented by Nellore, the lowest HO value found in

the literature was 0.16 (OROZCO-TER WENGEL et al., 2015). These authors also estimated HO for other indicine breeds, such as Brahman (0.19), and Gyr (0.16).

Table 4.6 Summary of number of genotyped animals, observed (HO) and expected (HE) heterozygosity, standard deviation (SD), average of minor frequency allele (MAF), and average of linkage disequilibrium for adjacent SNPs (LD) in all beef cattle breeds studied.

Breed	**Montana**	**Nellore**	**Angus**	**Senepol**	**Purdue**
Sub-species	(*Bos taurus indicus* x *Bos taurus taurus*)	*Bos taurus indicus*	*Bos taurus taurus*	*Bos taurus taurus*	*Bos taurus taurus*
Abbreviation	MNT	NE	AN	SE	PD
Sample Size	1,900	1,590	65	50	487
HO_O	0.39	0.27	0.39	0.38	0.35
(SD)	(0.11)	(0.17)	(0.15)	(0.14)	(0.15)
HO_I	0.41	0.37	0.35	0.36	0.38
(SD)	(0.10)	(0.15)	(0.15)	(0.14)	(0.14)
HE_O	0.38	0.28	0.39	0.37	0.34
(SD)	(0.11)	(0.16)	(0.11)	(0.13)	(0.14)
HE_I	0.40	0.37	0.36	0.37	0.37
(SD)	(0.10)	(0.15)	(0.13)	(0.12)	(0.13)
MAF_O	0.30	0.26	0.29	0.25	0.23
MAF_I	0.32	0.30	0.28	0.29	0.28
LD_O	0.19	0.42	0.32	0.26	0.19
LD_I	0.20	0.20	0.24	0.23	0.21

O = original panel; I = imputed panel.

The total autosomal genome length containing 29 autosomal pairs was 4,256.7 Mb, with the shortest *Bos taurus* autosome (BTA) being 110.4 Mb (BTA18) and the longest BTA being 184.9 Mb (BTA13). A refined descriptive summary of the chromosome and SNP parameters for Montana composite genotyped animals is shown in the Supplementary Material: Table S4.1.

Extent of linkage disequilibrium. The average LD between adjacent SNPs estimated for Montana animals and the average distance between adjacent SNPs (Mb) are presented in the

Supplementary File – Table S4.1. The lowest average LD overall distances were estimated for the Montana Composite (MNT; 0.19) and the crossbred (PD; 0.19) populations over genomic distances (Table 4.6). These findings suggested the maintainance of low levels of LD through genetic drift balanced by the mutation and recombination rates. In agreement with other studies, the LD provided information regarding the population history, the breeding system, natural and artificial selection, and other forces that cause gene-frequency evolution. The overall LD across adjacent SNP pairs were higher for Nellore and Angus (0.42 and 0.32) in the original panel compared to the other breeds.

For the Montana animals, the chromosomes BTA5 and BTA21 presented higher levels of LD (0.27 and 0.26), and BTA25 and BTA28 were observed with the lowest LD levels (0.20 and 0.21). Based on the number of SNPs the higher levels corresponded to the chromosomes with greater numbers of SNPs.

The LD decay values in certain distance intervals are shown in Figure 4.4 for composite, purebreds and crossbred animals and in Supplementary File – Fig. S4.2 only for composite animals. As expected, high LD values were observed only at small distances between pairs of SNPs because the LD declines as the rate of recombination and the physical distance between the markers increase. Trends across distances were very similar (Fig. 4.4) for AN, SE, PD and NE and the LD level decayed at a very similar rate. The Angus and Senepol breeds showed similar patterns of LD, which could be explained by their common ancestral origin. However, for the composite population, LD decayed rapidly compared to other breeds. The average r^2 estimates for the Montana population were the lowest values across all distances.

The extent of LD decreased substantially from the first (up to 0.01 Mb; r^2=0.60) to the second range of distances (between 0.01 and 0.02 Mb; r^2=0.49) for the Montana Tropical

composite. Between the range of distance from 0.07 to 0.12 Mb, the extent of LD decay appeared to follow a "plateau" of r^2 values. Therefore, after 1.2 Mb, the mean r^2 value decreased more slowly and seemed to consist of the decay curve.

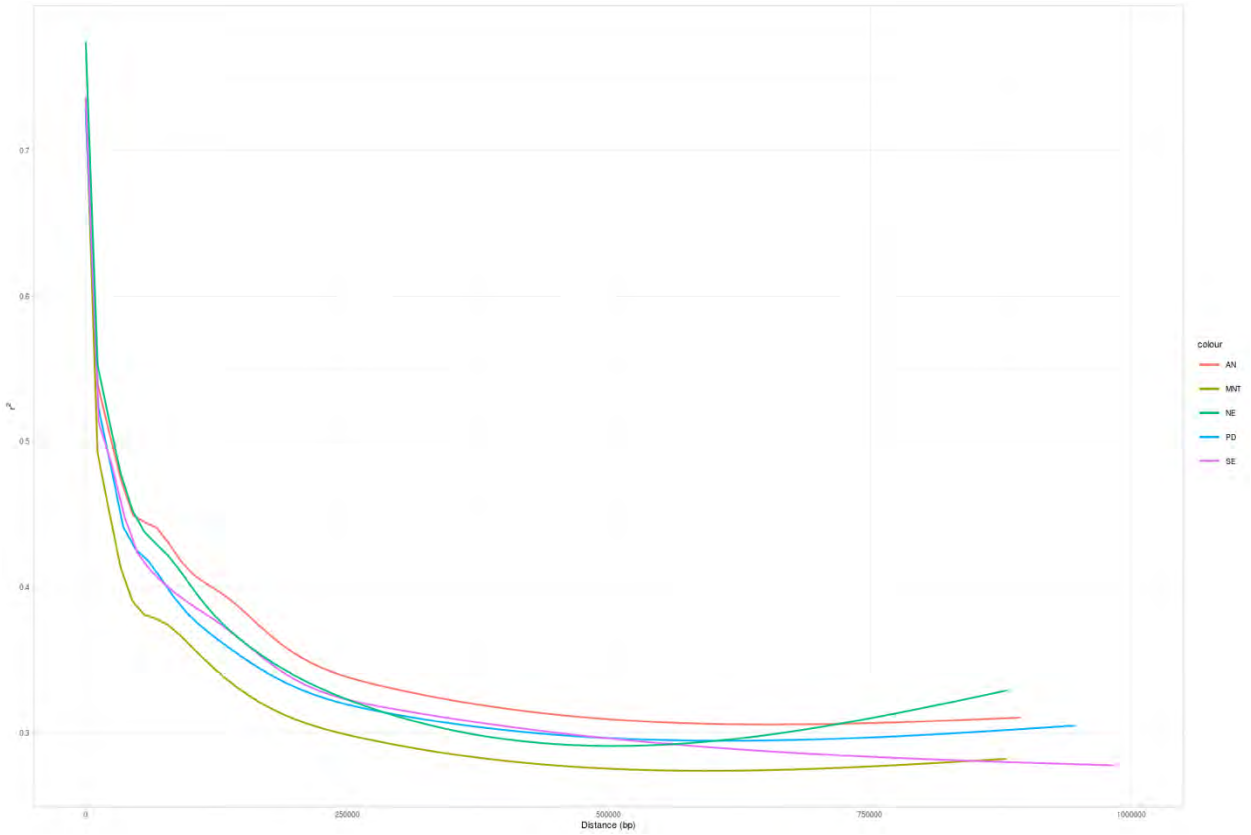


Figure 4.4 Linkage disequilibrium decay for the beef cattle breeds (AN=Angus; MNT=Montana Composite; NE=Nellore; PD=Crossbred AngusxSimmental; SE=Senepol).

Consistency of the gametic phase. The consistency of the gametic phase (Pearson correlation between signed r values) between Montana and the other breeds are shown in Figure 4.5 and Table 4.7. This metric has relevant implications for the design of further successful genomic applications between breeds as related by genetic linkages, specifically, in further genomic prediction analyses, with influences on breed combinations to estimate the SNP effects.

The highest consistency of phase was found between the Montana Tropical® composite and the crossbreed from Purdue (Angus x Simmental), suggesting a greater level of relatedness between these breeds. In general, the estimated correlations were high for all breeds, especially for short distances as expected. Likewise, Mokry et al (2014) observed a consistent phase over 0.80 between a composite beef cattle called Canchim (5/8 parts Charolais and 3/8 parts zebu) at a 100 Kb distance. In agreement with those results, Brito et al. (2015) analyzed different goat breeds and reported higher levels of consistency of phase at short distances compared to this study.

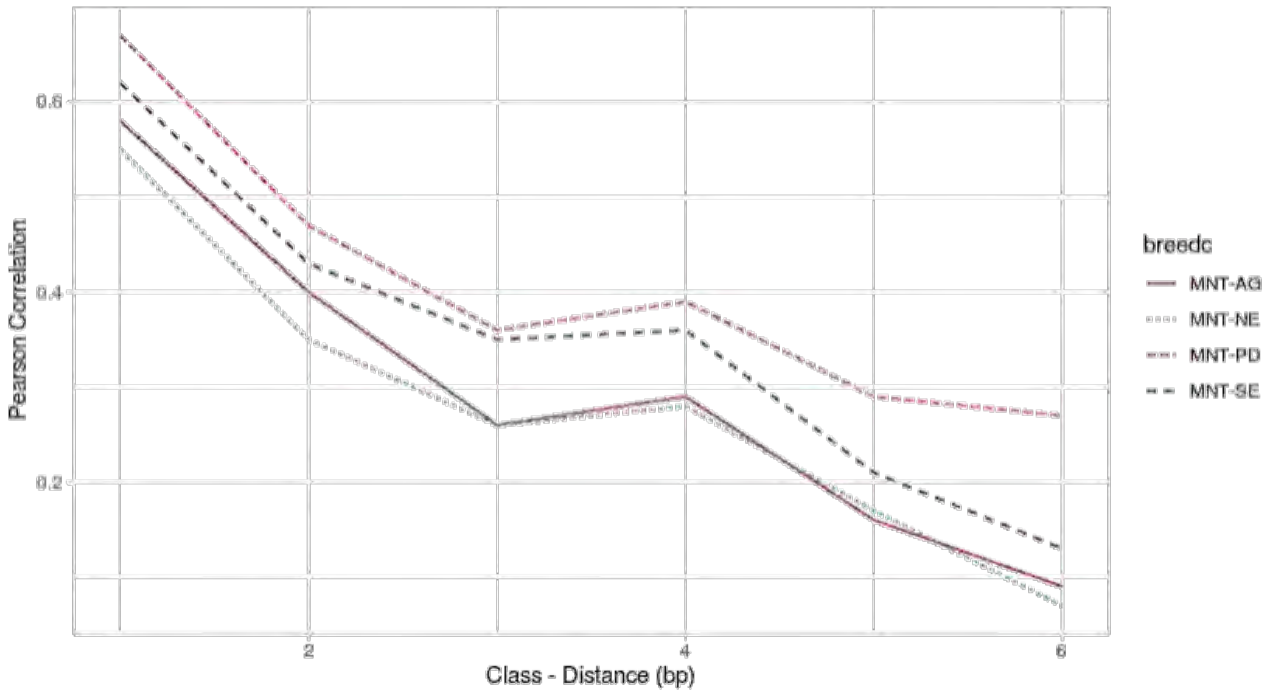


Figure 4.5 Consistency of gametic phase (Pearson correlations of signed r values) at given distances between 5 breeds pairs. MNT: Montana, AN: Angus, SE: Senepol, PD: Angus x Simmental, and NE: Nellore. The x-axis corresponds to a class of distance in bp: 1: 0-1,000; 2:1,000-10,000; 3:10,000-50,000; 4:50,000-100,000; 5:100,000-500,000; 6:500,000-1,000,000.

Table 4.7 Consistency of gametic phase for Montana and the other breeds, at different distances (kb).

Breeds[1]	Distance (kb)					
	0-1	1-10	10-50	50-100	100-500	500-1000
MNT-AG	0.58	0.40	0.26	0.29	0.16	0.09
MNT-SE	0.62	0.43	0.35	0.36	0.21	0.13
MNT-PD	0.67	0.47	0.36	0.39	0.29	0.27
MNT-NE	0.55	0.35	0.26	0.28	0.17	0.07

[1]MNT (Montana tropical composite), ANG (Angus), SEN (Senepol), PD (crossbreed animals from Purdue University), NEL (Nellore).

Admixture and principal component analysis. To characterize the population structure resulting from heterogeneous breed ancestry in this present study, we performed PCA and an ADMIXTURE analysis. PCA has previously been used to visualize the genomic relationships among numerous cattle breeds (LEWIS et al., 2011; CAÑAS-ÁLVAREZ et al., 2015; KELLEHER et al., 2017; UPADHYAY et al., 2017), as has ADMIXTURE (BUZANSKAS et al., 2017; UPADHYAY et al., 2017; ZHANG et al., 2018; AKANNO et al., 2018; AHMAD et al., 2019; CHHOTARAY et al., 2020).

Figure 4.6 shows the plot of the first and second principal components (PC1 and PC2), which explained 7.3% and 6.9% of the variation in the entire genetic data considering only the Montana genotypes, respectively. This indicates that there is not clear population stratification inside the Montana composite. The PC1, PC2, and PC3 were estimated considering all available genotypes and are shown in Figure 4.6. The proportions of the variation explained by these were 28.04%, 8.80%, and 7.24%, respectively. A clear distinction between the *Bos taurus taurus* and *Bos taurus indicus* genotypes was observed in Figure 4.8, represented by Angus, Senepol, crossbred Angus x Simmental, and Nellore. Animals with a higher taurine proportion had more

variation in PC2. This variation could be attributed to the distance of Nellore animals from the initial crossbreeding forming a composite animal. The Montana composite cluster was widely distributed along the principal components (PC1, PC2, and PC3; Figure 4.8), and overlapped mainly with Senepol and Angus characterizing the relatedness between these breeds. The genetic similarities between these two breeds agreed with previous studies (BRENNEMAN et al., 2007; HUSON et al., 2014). The crossbred (Angus x Simmental) cluster resided in close proximity to the composite samples on the PCA plot (Figures 4.8).

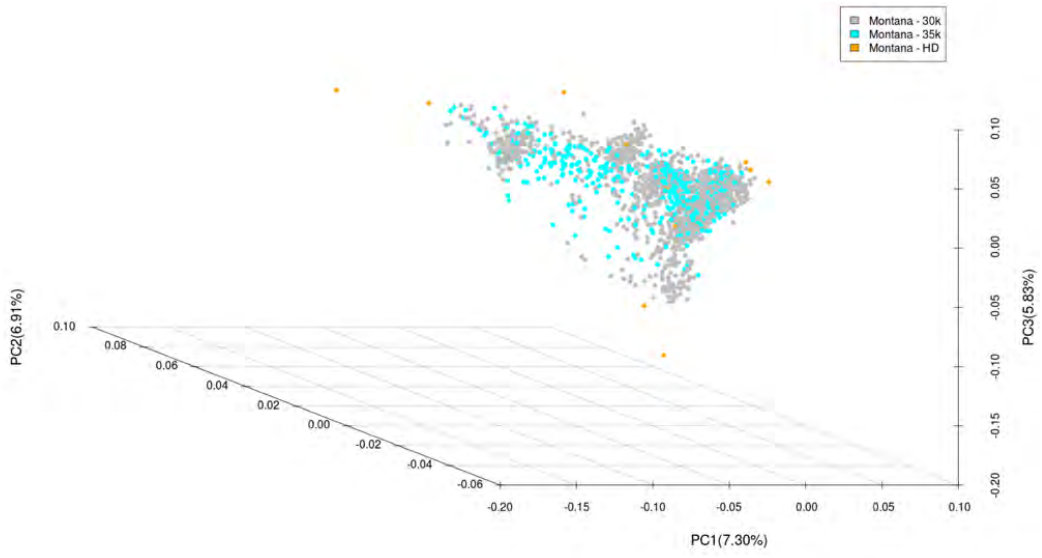


Figure 4.6 Plot of principal components (PC1; 7.30%), PC2 (6.91%), and PC3 (5.83%) of the genomic relationship matrix. The dots represented the Montana animals genotyped included in the imputation as scenario A.

We highlight in Figure 4.8 the position of composite animals in the middle of both subspecies (i.e., *Bos taurus taurus* and *Bos taurus indicus*), confirming the complex genetic background (i.e., the genetic composition) of Montana, using the four different biological types. The distance relationship between the Nellore samples suggested an apparent differentiation between those genotypes likely as a consequence of the genetic background.

Figure 4.7 shows the rates at which the breeds were used since the beginning of the Montana' development. The Nellore breed represents the single largest breed of beef cattle raised in Brazil originally imported from India (i.e., Ongole cattle – *Bos taurus indicus*). This breed was used more frequently, predominantly crossing taurine bulls with zebuine cows. This phenomenon is likely explained through the fact that those animals demonstrated higher heat tolerance, in other words, those animals could better tolerate the tropical conditions. In order to achieve progress, 2002 shows a clear change between the biological types N and A, promoting a higher introduction of the taurine genotype pool in the Montana population, as the breeding goal of improving beef quality and production became mandatory.

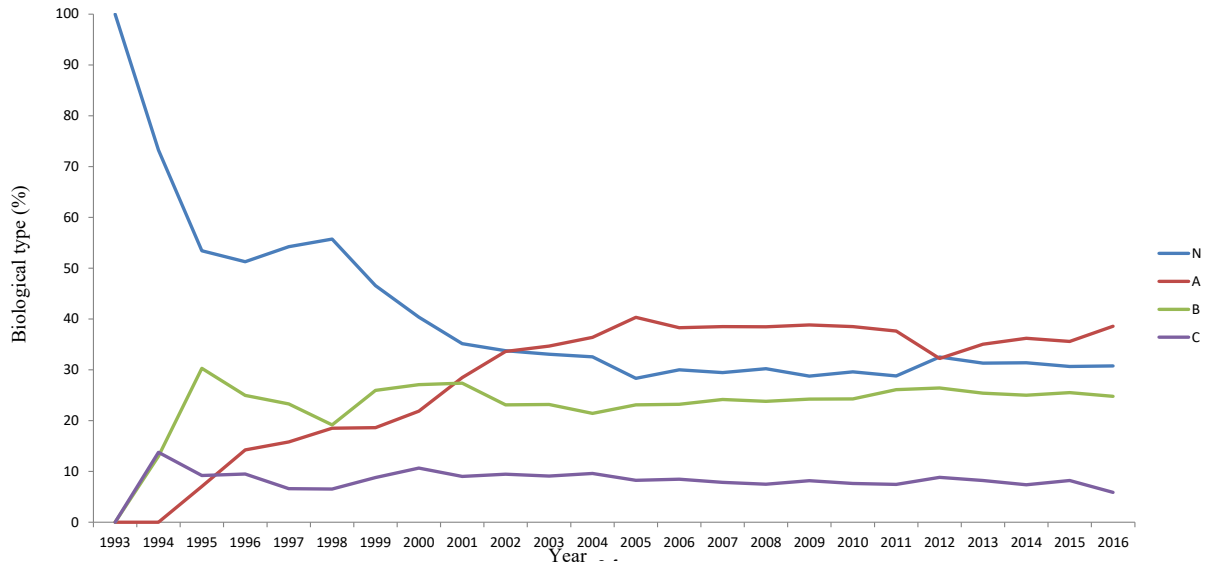


Figure 4.7 The rate in which each biological type was used since the development of the Montana composite population.

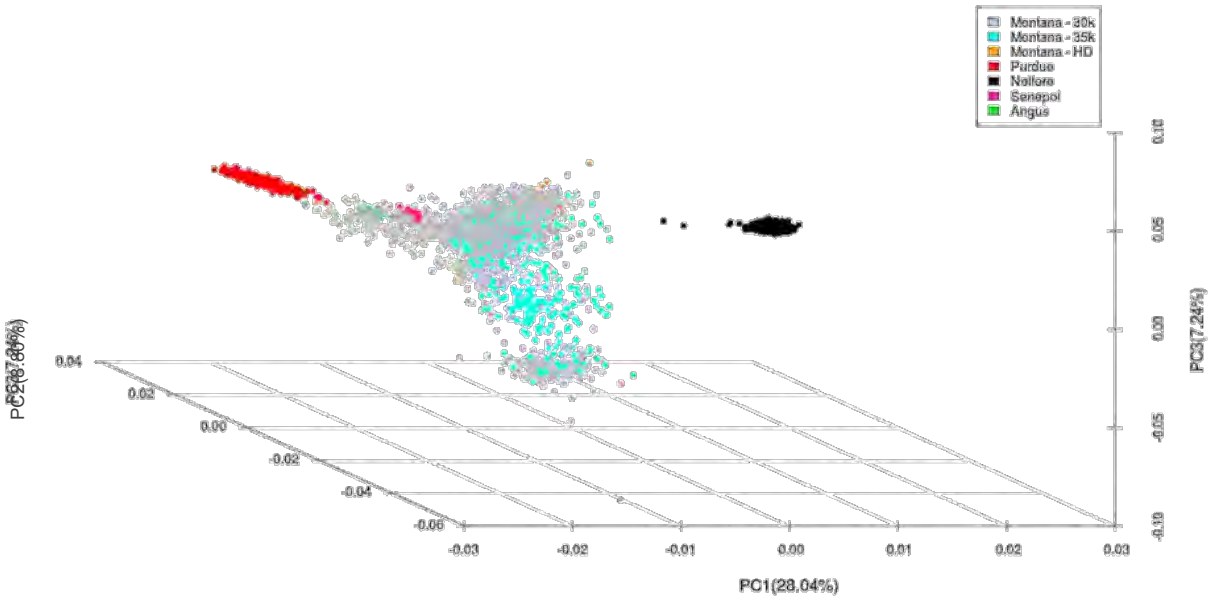


Figure 4.8 Plot of principal components (PC1; 28.04%), PC2 (8.80%), and PC3 (7.24%) of the genomic relationship matrix. The dots animals genotyped included the imputation in scenario B.

The ADMIXTURE results showed that a great number of animals had a significant portion of their genotype coming from another cluster (Figure 4.9). The lowest cross validation error (0.53) was detected at K = 26 (Figure 4.9). As expected, the Montana animals appeared to demonstrate a higher level of genome proportion of admixture compared to the other breeds, indicating a more diverse genetic composition. For K = 6, all animals from the composite and other breeds were clustered into taurine and zebu (or indicine) lineages. The overall admixture patterns were in good agreement with those revealed by PCA (Figure 4.6 and 4.8).

In this study, the PC1 and admixture analyses clearly separated the composite cattle from the other five breeds. This is explained by the fact that the Montana was developed using the NABC system. These facts indicated that the taurine background within this breed may be corresponding to ancient admixture events from the putative number of breeds involved in the Montana Composite development. At K = 6, the Angus (taurine) background is shared in large

proportions with the Purdue (Angus x Simmental) and Senepol animals. At K = 26, the optimal K, the ancestry with taurine largely remained. The admixture analysis indicates, as expected, a zebu background in the composite samples from the Nellore breed, since the criteria to be considered a Montana animal required at least a 25% indicine genome contribution. In agreement, a genetic study in Canchim (62.5% Charolais to 37.5% Zebu) a composite beef breed raised in Brazil, also revealed the proportion of the taurine × zebu admixture (BUZANSKAS et al., 2017).

On average, the Angus samples had an ancestral lineage in common with the Senepol breed. For those breeds, the introgression of the indicine genome indicated the evolutionary history for raising these animals in tropical and subtropical conditions, which is the case in Brazil. PCA (Figure 4.8) also confirmed this relationship, where animals from Angus, crossbred, and Senepol were closely clustered. This sharing of the gene pool may be due to mixing of the breeds as discussed before, especially for Montana population. Similarly, Porto-Neto et al. (2014) also revealed in a Australian tropical composite population diverging between the taurine and indicine genome; however, they detected genome regions in a positive balance within taurine and zebu cattle, reinforcing sub-division of domestic cattle.

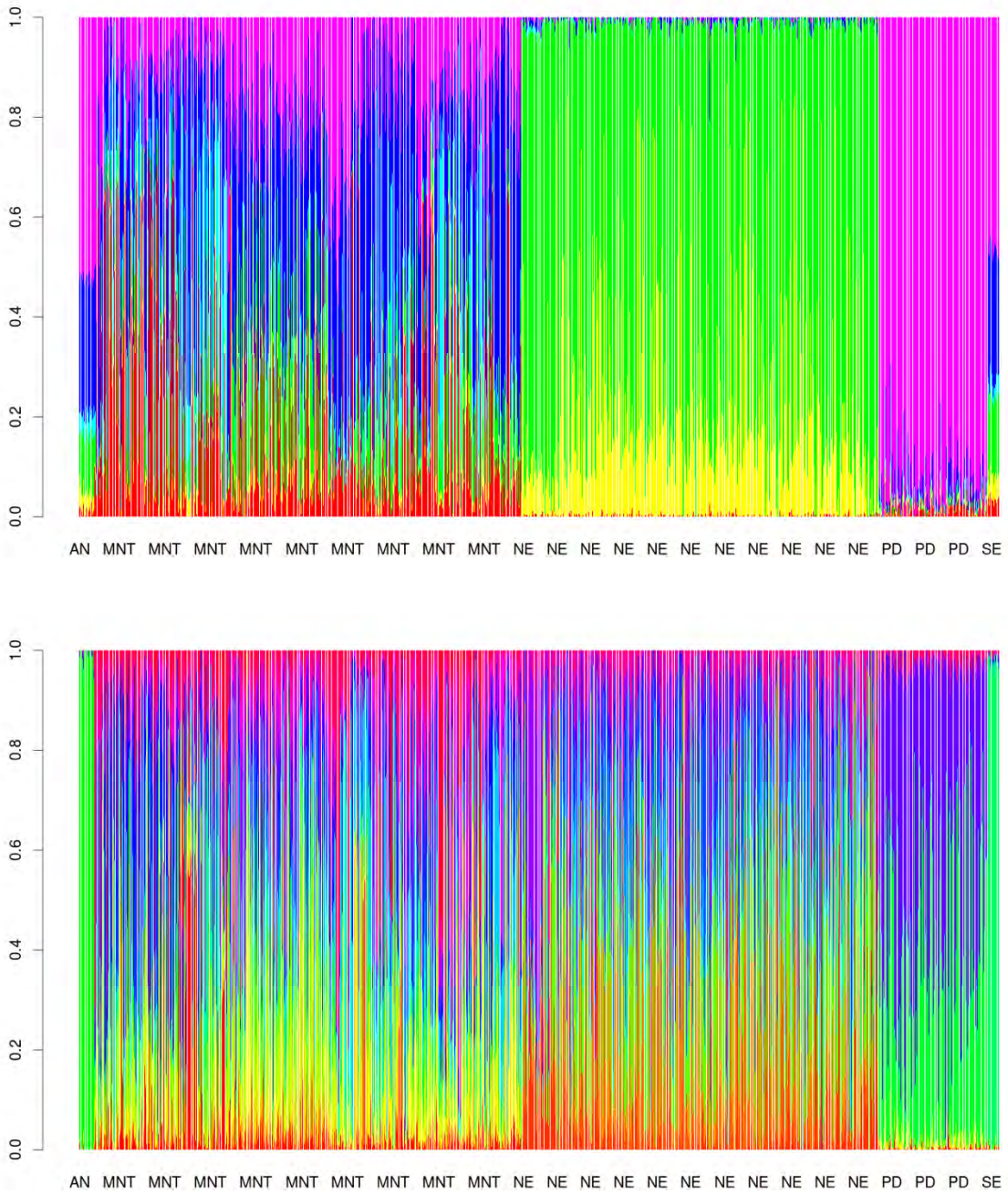


Figure 4.9 Hierarchical clustering of individual animals using 51,709 genome-wide SNPs. Results are shown for a range of assumed values (K = 6; top and K=26; bottom) for the number of ancestral populations. (AN = Angus; MNT = Montana; NE = Nellore; PD = Purdue; SE = Senepol).

In this study, we reported that the genotype imputation method could be suitable for inferring and discriminating animals to assign the metrics evaluated and animals to the correct clusters based on the level of breed introgression. As expected, the levels of inbreeding in this population were low, mainly due to the lack of information regarding pedigree and also due to the high levels of crossbreeding in the population. The levels of LD were comparable to other composite populations (GRIGOLETTO et al., 2019). Montana Tropical® composite beef cattle were developed to achieve high performance through heterosis and complementarity for the adaptability to tropical conditions from the indicine with the precocity and meat quality from the taurine cattle. Therefore, from a practical point of view, the study of introgression in this breed can assist the implementation of further genomic analyses. This is the first work using genomic information to characterize the levels of introgression of beef composite populations. Our study illustrated the diverse genetic composition of the population of zebu and taurine using PCA and genetic admixture techniques.

4.5 CONCLUSIONS

The level of linkage disequilibrium (r^2) decayed rapidly with greater distances (> 0.13 Mb) in composite animals. In this current SNP panel, all breeds (Montana Tropical, Angus, Nellore, Senepol, and Angus x Simmental) exceeded the level of linkage disequilibrium that was useful (r^2 > 0.2) at 0.15 Mb for the average distance between adjacent SNPs. The highest consistency of phase was found for the Montana Tropical® composite and crossbreed from Purdue (Angus x Simmental), suggesting a greater level of relatedness between these breeds, and inferring a possibility for use in further genomic analyses for genotype pooling. In this sense, this crossbred animals' cluster resided in close proximity to the composite samples on the PCA plot. These facts indicated that the taurine background within this breed may be corresponding to ancient admixture

events from the putative number of breeds involved in the Montana Composite development. Our study provided a great deal of information regarding the population structure and genetic diversity of pedigrees as well as genomic information of tropical composite cattle breeds, and the results may be used for further implementation of genomic selection of productive and adaptive traits in this population.

4.6 REFERENCES

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

AKANNO, E. C. et al. Genome-wide association scan for heterotic quantitative trait loci in multi-breed and crossbred beef cattle. **Genetics Selection Evolution**, v. 50, n. 1, p. 48, 2018.

AHMAD, S. F. et al. Population structure and admixture analysis in Frieswal crossbred cattle of India–a pilot study. **Animal biotechnology**, v. 31, n. 1, p. 86-92, 2020.

BRENNEMAN, R. A. et al. Genetic diversity among Angus, American Brahman, Senepol and Romosinuano cattle breeds. **Animal genetics**, v. 38, n. 1, p. 50-53, 2007.

BRITO, L. F. et al. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. **BMC genetics**, v. 16, n. 1, p. 67, 2015.

BUENO, R. S.; TORRES, R. D. A.; FERRAZ, J. B. S. et al. Métodos de estimação de efeitos genéticos não-aditivos para características de peso e perímetro escrotal em bovinos de corte mestiços. **Revista Brasileira de Zootecnia**, v. 41, n. 5, p. 1140–1145, 2012.

BUZANSKAS, M. E. et al. Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. **PLoS ONE**, v. 12, n. 2, 2017.

CAÑAS-ÁLVAREZ, J. J. et al. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip. **Journal of animal science**, v. 93, n. 11, p. 5164-5174, 2015.

CHHOTARAY, S. et al. Ancestry informative markers derived from discriminant analysis of principal components provide important insights into the composition of crossbred cattle. **Genomics**, v. 112, n. 2, p. 1726-1733, 2020.

CHUD, T. C. S et al. Strategies for genotype imputation in composite beef cattle. **BMC genetics**, v. 16, n. 1, p. 99, 2015.

COLE, J. B.; SILVA, M. V. G. B. Genomic selection in multi-breed dairy cattle populations. **Revista Brasileira de Zootecnia**, v. 45, n. 4, p. 195-202, 2016.

CUNDIFF, L. V.; GREGORY, K. E. What is systematic crossbreeding. In: **Cattlemen's College, 1999 Cattle Industry Annual Meeting and Trade Show, National Cattlemen's Beef Association. Charlotte, NC**. 1999.

DIAS, R. A. P.; PETRINI, J.; FERRAZ, J. B. S.; ELER, J. P.; BUENO, R. S.; DA COSTA, A. L. L.; MOURÃO, G. B. Multicollinearity in genetic effects for weaning weight in a beef cattle composite population. **Livestock Science**, v. 142, n. 1-3, p. 188-194, 2011.

MOKRY, F. B. et al. Linkage disequilibrium and haplotype block structure in a composite beef cattle breed. **BMC genomics**, v. 15, n. S7, p. S6, 2014.

FALCONER, D. S.; MACKAY, T. F. C. Introduction to quantitative genetics. 1996. **Harlow, Essex, UK: Longmans Green**, v. 3, 1996.

FERRAZ, J. B. S.; ELER, J. P.; GOLDEN, B. L. A formação do composto Montana Tropical. **Revista Brasileira de Reprodução Animal**, v. 23, p. 115-117, 1999.

GARCÍA-RUIZ, A. et al. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. **Proceedings of the National Academy of Sciences**, v. 113, n. 28, p. E3995-E4004, 2016.

GRIGOLETTO, L.; BRITO, L. F.; MATTOS, E. C.; ELER, J. P.; BUSSIMAN, F. O.; SILVA, B. C. A.; SILVA, R. P.; CARVALHO, F. E.; BERTON, M.P.; BALDI, F.; FERRAZ, J. B. S. Genome-wide associations and detection of candidate genes for direct and maternal genetic

effects influencing growth traits in the Montana Tropical® Composite population. **Livestock Science**, v. 229, p. 64-76, 2019.

GRIGOLETTO, L. FERRAZ, J. B. S.; OLIVEIRA, H. R.; ELER, J. P.; BUSSIMAN, F. O.; ABREU SILVA, B. C.; BALDI, F.; BRITO, L. F. Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical® Composite Beef Cattle. **Frontiers in genetics**, v. 11, p. 123, 2020.

HAYES, B. J.; VISSCHER, P. M.; GODDARD, M. E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics research**, v. 91, n. 1, p. 47-60, 2009.

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

HUSON, H. J. et al. Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle. **Frontiers in genetics**, v. 5, p. 101, 2014.

KELLEHER, M. M. et al. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. **animal**, v. 11, n. 1, p. 15-23, 2017.

LEWIS, J. et al. Tracing cattle breeds with principal components analysis ancestry informative SNPs. **PloS one**, v. 6, n. 4, 2011.

MOURÃO, G. B.; FERRAZ, J. B. S.; ELER, J. P.; BALIEIRO, J. C. D. C.; BUENO, R. S.; MATTOS, E. C.; FIGUEIREDO, L. G. G. Genetic parameters for growth traits of a Brazilian Bos taurus x Bos indicus beef composite. **Genetics and Molecular Research**, v. 6, n. 4, p. 1190-1200, 2007.

OLSON, K. M.; VANRADEN, P. M.; TOOKER, M. E. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. **Journal of Dairy Science**, v. 95, n. 9, p. 5378-5383, 2012.

OROZCO-TERWENGEL, P. et al. Revisiting demographic processes in cattle with genome-wide population genetic analysis. **Frontiers in genetics**, v. 6, p. 191, 2015.

PERIPOLLI, E. et al. Genome-wide scan for runs of homozygosity in the composite Montana Tropical® beef cattle. **Journal of Animal Breeding and Genetics**, v. 137, n. 2, p. 155-165, 2020.

PETRINI, J.; DIAS, R. A. P.; PERTILE, S. F. N.; ELER, J. P.; FERRAZ, J. B. S.; MOURÃO, G. B. Degree of multicollinearity and variables involved in linear dependence in additive-dominant models. **Pesquisa Agropecuária Brasileira**, v. 47, n. 12, p. 1743-1750, 2012.

PORTO-NETO, L. R.; KIJAS, J. W.; REVERTER, A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. **Genetics Selection Evolution**, v. 46, n. 1, p. 22, 2014.

PURCELL, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

R CORE TEAM (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. < https://www.R-project.org>.

SANTANA, M. L.; ELER, J. P.; CARDOSO, F. F.; ALBUQUERQUE, L. G. D.; BIGNARDI, A. B.; FERRAZ, J. B. S. Genotype by environment interaction for birth and weaning weights of composite beef cattle in different regions of Brazil. **Livestock science**, v. 149, n. 3, p. 242-249, 2012.

SANTANA, M. L. et al. Genotype by production environment interaction for birth and weaning weights in a population of composite beef cattle. **animal**, v. 8, n. 3, p. 379-387, 2014.

SANTANA, M. L. et al. Genotype by environment interaction for post-weaning weight gain, scrotal circumference, and muscling score of composite beef cattle in different regions of Brazil. **Genetics and Molecular Research**, p. 3048-3059, 2013.

SARGOLZAEI, M. et al. CFC: a tool for monitoring genetic diversity. **Proc. 8th World Congr. Genet. Appl. Livest. Prod., CD-ROM Communication**, n. 27-28, p. 13-18, 2006.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC genomics**, v. 15, n. 1, p. 478, 2014.

VENTURA, R. V. et al. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. **Journal of animal science**, v. 92, n. 4, p. 1433-1444, 2014.

UPADHYAY, M. R. et al. Genetic origin, admixture and population history of aurochs (Bos primigenius) and primitive European cattle. **Heredity**, v. 118, n. 2, p. 169, 2017.

WIGGANS, G. R.; VANRADEN, P. M.; COOPER, T. A. The genomic evaluation system in the United States: Past, present, future. **Journal of dairy science**, v. 94, n. 6, p. 3202-3211, 2011.

ZHANG, W. et al. Genome-wide assessment of genetic diversity and population structure insights into admixture and introgression in Chinese indigenous cattle. **BMC genetics**, v. 19, n. 1, p. 114, 2018.

CHAPTER 5. GENOMIC PREDICTION OF BREEDING VALUES FOR GROWTH REPRODUCTIVE AND CARCASS TRAITS IN MONTANA TROPICAL COMPOSITE CATTLE

Laís Grigoletto[a,b], Hinayah R. Oliveira[b], Joanir Pereira Eler[a], José Bento Sterman Ferraz[a] and Luiz F. Brito[b]

[a]Department of Animal Sciences, College of Animal Sciences and Food Engineering, University of São Paulo, Pirassununga, Sao Paulo, Brazil

[b]Department of Animal Sciences, Purdue University, West Lafayette, Indiana, USA

ABSTRACT

The incorporation of genomic information in genetic evaluations has substantially contributed to increase the accuracy of breeding values. However, genomic evaluations in Composite populations is still challenging. For instance, the Montana Tropical® Composite beef cattle population was developed in Brazil based on the crossing among four different biological types derived from various *Bos taurus taurus* and *Bos taurus indicus* breeds. In this study, the forward validation procedure was used to evaluate the predictive ability in the economically important traits, such as: growth, reproductive, carcass and meat quality, incorporating genomic information. Through the forward validation approach, these two populations were divided based on birth year. Animals were divided based on older animals composing the training (n=1,452) population and animals (without progeny) the validation (n=448) population. The traditional BLUP (based on pedigree-based relationship matrix), Genomic BLUP (GBLUP), single-step genomic BLUP (ssGBLUP), and BayesR models were compared to estimate genomic breeding values. We used around 300,000 and 9,000 growth and carcass/ meat quality phenotypic records, respectively. The pedigree file included 525,806 animals, from which 1,900 were genotyped using a 30K, 35K, 50K, or HD SNP panel and imputed to a common moderate SNP panel containing 51,709 SNPs. The predictive performance of each model was evaluated based on the accuracy (Pearson correlation coefficient) and the linear regression coefficient estimated between EBVs and GEBVs. The average observed accuracies (± SD) in the growth traits for ssGBLUP, GBLUP, and BayesR were 0.58 ± 0.04, 0.55 ± 0.04, and 0.61 ± 0.03, respectively. The highest increase in accuracy was observed for LMA (0.14) indicated a 33% increase in accuracy through the BayesR. Bayesian approach outperformed compared to ssGBLUP and GBLUP for the majority of traits analyzed with few exceptions. These initial findings can guarantee the success of genomic selection implementation and the gain of genomic information including for commercial applications on this composite population.

Keywords: accuracy, bayesian, composite, genomic breeding values, prediction

5.1 INTRODUCTION

The Montana Tropical® composite population was developed following the crossing between two cattle sub-species, to combine rusticity and climatic adaptability (e.g., indicine breeds) and productive performance (e.g., taurine breeds) (FERRAZ; ELER; GOLDEN, 1999; GREGORY et al., 1993;1994). Genomic selection is widely used in livestock breeding programs. However, the majority is performed on purebred animals. In this sense, there still a lack of genomic selection on crossbred and composite animals due to the complexity of the genetic background and the lack of sufficiently large reference populations (COLE; SILVA, 2016). For beef cattle, the incorporation of genomic information is attractive to take advantage of estimated the genetic merit of many traits that affect the profitability of beef production, such as those that are expensive and difficult to select (e.g., carcass and meat quality traits), and with low heritability (BOICHARD et al., 2016). Also, it leads to greater genetic gain reducing the generation interval and increasing the accuracy of breeding values for these traits (XU et al., 2020).

Nowadays, several statistical methods were used to develop genomic selection approach for numerous beef cattle breeds (SAATCHI et al., 2011; NEVES et al., 2014; CHEN et al., 2015; MEHRBAN et al., 2017; FARAH et al., 2016; WANG et al., 2019; MRODE et al., 2019). However, to date, the predictive performance of genomic selection in Montana composite cattle has not been investigated.

Compared with traditional Best Linear Unbiased Predictor (BLUP), the Genomic BLUP method uses the genomic-based relationship providing a more accurate relationship coefficient among individuals, contributing to the increased of estimated accuracy breeding values (VANRADEN, 2008). Legarra et al. (2009) and Aguilar et al. (2010) proposed the single-step GBLUP (ssGBLUP) method, allowing the use of pedigree-based relationship matrix to include

non-genotyped animals, especially, for situations of missing pedigree, where is an appropriate alternative to obtain more accurate and less biased breeding values for young and genotyped animals with missing pedigree. Besides, the influences of Bayesian approaches on prediction methods could improve the identification of every single nucleotide polymorphism (SNP) with a large effect and produce higher accuracies for multi-breed or composite populations when traits are influenced by genes with large effects (VAN DEN BERG et al., 2017; HAYES et al., 2010; DAETWYLER et al., 2010; ZHANG et al., 2010; CLARK et al., 2011; ROLF et al., 2015; VAN DER BERG et al., 2017).

Hereof, a major issue with the implementation of genomic selection is the number of genotyped animals, and the complexity of the composite population structure constitutes one of the obstacles. Several studies were conducted to examine the use of different density panels (low-to-high) and imputed data on the accuracy of genomic prediction (BERRY et al., 2014; ALILOO et al., 2018; BERNARDES et al., 2019). According to Boison et al. (2017), the use of low density (LD) SNP panels have similar predictive performance compared with high-density panels (HD) in indicine breed. Investigating the efficacy of imputation on dairy crossbred cattle, Aliloo et al. (2018) reported increasing accuracy using a reference population consisting of a mixture of crossbred and ancestral purebred animals.

The main objective of this study was to evaluate the predictive performance through the accuracy and bias (regression coefficient) of genomic breeding values (GEBVs) predicted using the BLUP, GBLUP, single-step GBLUP (ssGBLUP) and BayesR methods in Montana Tropical® Composite beef cattle.

5.2 MATERIAL AND METHODS

All the analyses were performed using pre-existing databases. Therefore, Animal Care Committee approval was not required for this study.

5.2.1 Animals and phenotypes

The animals were classified within each biological group (NABC) as described by Grigoletto et al. (2020a): 1) N: Zebu breeds, mainly represented by Nellore; 2) A: Taurine breeds adapted to tropical conditions (Senepol, Belmont Red, Bonsmara, and Caracu); 3) B: Taurine breeds of British origin (mainly Angus, Devon, and Hereford); and, 4) C: Continental European breeds (mainly Charolais, Limousin, and Simmental). The main contributing breeds to the development of this composite population were Aberdeen Angus, Red Angus, Nellore, Senepol, Limousin, Simmental, Hereford, and Bonsmara. Breed was recorded by the producers/technicians or calculated based on pedigree relationship between the animals. The pedigree file included 525,806 animals and their breed composition. In relation on this, there is a lack in pedigree information (16.7% of total animals have both parents unknown). In this case, the use of genomic information may be important to help improve the quality of pedigree information (GRIGOLETTO et al., 2020b; unpublished).

The descriptive statistics of the phenotypic data is shown in Table 5.2. Five growth traits (BW, WW, YW, WG, and MUSC), two reproductive traits (SC365 and SC420), and four ultrasound-based carcass and meat quality traits (LMA, BFT, RFT, and MARB), recorded on animals born between 1994 and 2017, were included in this study. Weight recording was obtained at birth and weaning. Further records of yearling weight, scrotal circumference, and other

productive traits were collected between 14 and 18 months. More details for are presented in Grigoletto et al. (2019; 2020a).

Phenotypic quality control removed records deviating 3.5 SD from the overall mean within contemporary group (CG). The CGs were defined based on farms, years, and seasons of birth, sexes, and management groups. The CGs with less than five records were excluded from subsequent analyses.

5.2.2 Genotype and quality control

Genotype imputed data from a total of 1,900 animals were performed using the GeneSeek Genomic 325 ProfilerTM LDv4-GGP Bovine 30,105K (n = 1,436), GGP *Bos indicus:* 35,139K (n=448) and HD 777 K (n = 16) Beadchips (Illumina Inc., San Diego, CA, USA) as described by Grigoletto et al. (2020b; unpublished). Common SNPs between the moderate to high panels were selected which resulted in 55,614 SNPs. SNPs were excluded from further analyses if their minor allele frequency (MAF) lower than 0.01, call rate lower than 95%, SNPs located on the sexual chromosome, and that did not follow the Hardy Weinberg Equilibrium (HWE; considering as threshold 1e-15) were excluded from the analysis. After the genotype quality control, 51,709 SNPs remained in this study.

5.2.3 Statistical model and genetic analyses

Variance components and breeding values. Single- trait linear animal models and the average-information restricted maximum likelihood (AI-REML) procedure were used to estimate heritability and variance components, using the AIREMLF90 package from the BLUPF90 family programs (MISZTAL et al., 2002; 2014). The single-trait animal models used in this study

included the direct additive genetic and residual as random effects. CG, direct (individual) heterozygosity (described below), and age of the animal at the measurement were included as fixed effects in the model. Thus, the general statistical model used in this study can be described as:

$$\mathbf{y}_{ijkl} = \mathbf{CG}_i + \mathbf{b_1}(\mathbf{Age}_j - \overline{\mathbf{Age}}) + \mathbf{b_2}\left(\mathbf{H_D}_k - \overline{\mathbf{H_D}}\right) + \boldsymbol{\alpha}_l + \boldsymbol{\varepsilon}_{ijkl},$$

where y_{ijkl} is the phenotypic record for each trait recorded on the animal *l*, belonging to the CG *i*, at age *j*, and heterozygosity (H_D) k. b_1 and b_2 are the linear regression coefficients related to the Age and H_D effects, respectively, which were considered as deviations from the mean ($\overline{Age}$ and $\overline{H_D}$). The α_l is the direct additive genetic random effect for the animal *l*, and ε_{ijkl} is the residual random effect associated with the animal *l*, heterozygosity *k*, age *j* and CG *i*. The statistical model (effects) used for each trait analyzed is presented in Table 5.1. Assuming a matrix notation, the previous model can be written as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z\alpha} + \boldsymbol{\varepsilon},$$

where, $\mathbf{y}$ is the vector of phenotypic observations for each trait; $\boldsymbol{\beta}$ is the vector of solutions for fixed effects; $\boldsymbol{\alpha}$ is the vector of predictions for random additive genetic animal effect; $\boldsymbol{\varepsilon}$ is the vector of random residual terms; $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices of fixed and random effects, respectively. It was assumed that: $\alpha \sim N(0, \mathbf{A}\sigma^2_\alpha)$ and $\varepsilon \sim N(0, \mathbf{I}\sigma^2_\varepsilon)$; where σ^2_α is the additive genetic variance; σ^2_ε is the residual variance; and $\mathbf{I}$ is an identity matrix.

5.2.4 Genomic Prediction

The prediction of genomic breeding values (GEBVs) was performed through alternative methods: genomic BLUP (GBLUP, VANRADEN, 2008), single-step GBLUP (ssGBLUP;

LEGARRA et al., 2009, AGUILAR et al., 2010; CHRISTENSEN; LUND, 2010), and BayesR (ERBE et al., 2012).

5.2.4.1 Genomic Best Linear Unbiased Prediction (GBLUP)

GBLUP was applied using the BLUPF90 software (MISZTAL et al., 2002; 2014) as follows:

$$\boldsymbol{y} = \mathbf{1}\boldsymbol{\mu} + \boldsymbol{Z_g} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is a vector of the trait of interest, which was adjusted for fixed effects (contemporary group, and age as a covariate) based on the full dataset (please see Table 5.1); $\mathbf{1}$ is a vector of 1 s; $\boldsymbol{\mu}$ is the overall mean; $\boldsymbol{Z}$ is the incidence matrix of direct genomic breeding values (DGV) and $\boldsymbol{g}$ is the vector of DGV and is assumed to follow a normal distribution N(0, $\mathbf{G}\sigma_g^2$); , where $\mathbf{G}$ is the marker-based genomic relationship matrix as a genomic relationship matrix and $\boldsymbol{\sigma_g^2}$ the genetic variance captured by the markers; $\boldsymbol{e}$ is a vector of random residual effects and is assumed to follow a normal distribution N(0, $\mathbf{I}\sigma_\varepsilon^2$), where is an identity matrix; and $\boldsymbol{\sigma_e^2}$ is the residual variance.

Table 5.1 Final fixed effects, covariables, and random effects used for mixed models for growth, reproductive, carcass and meat quality traits for tropical composite beef cattle.

TRAIT[1]	FIXED EFFECTS	COVARIABLES[2]	RANDOM EFFECTS
BW	Season, year of birth, sex, birth farm = CG	agedam[3], A, B, C, A_m, C_m, HetIndNABC, HetMatNABC	Animal, Maternal, Maternal perm. Envir.
WW	Season, year of birth, sex, weaning farm = CG	agedam, A, B, C, A_m, C_m, HetIndNABC, HetMatNABC	Animal, Maternal, Maternal perm. Envir.
YW	Season, year of birth, sex, farm = CG	ageanimal, agedam, A, B, C, A_m, C_m, HetIndNABC, HetMatNABC	Animal, Maternal, Maternal perm. Envir.
WG	Year of birth, sex, birth farm = CG	ageanimal, agedam[4], A, B, C, B_m, C_m, HetIndNABC, HetMatNABC	Animal, Maternal, Maternal perm. Envir.
MUSC	Season, year of birth, sex, farm = CG	ageanimal, agedam[4], A, B, C, A_m, B_m, C_m, HetIndNABC, HetMatNABC	Animal, Maternal
SC365	Season, year of birth, farm = CG	ageanimal, A, B, C, C_m, HetIndNABC, HetMatNABC	Animal, Maternal, Maternal perm. Envir.
SC420	Season, year of birth, farm = CG	A, B, C, A_m, C_m, HetIndNABC	Animal
LMA	Season, year of birth, sex, farm = CG	ageanimal, N, A, B, HetIndNABC	Animal
RFT	Season, year of birth, sex, farm = CG	ageanimal, N, A, B, HetIndNABC	Animal
BFT	Season, year of birth, sex, farm = CG	ageanimal, N, A, B, HetIndNABC	Animal
MARB	Season, year of birth, sex, farm = CG	ageanimal, N, A, B, HetIndNABC	Animal

[1] BW = Birth weight; WW = Weaning weight; YW = Yearling weight; WG = Weight gain; MUSC = Muscularity; SC365 = Scrotal circumference at 365 days; SC420 = Scrotal circumference at 420 days; LMA = *Longissimus* muscle area; RFT = Rump fat thickness; BFT = Backfat thickness; MARB = Marbling. N = Zebu breeds; A = Adapted breeds, B = British breeds, C = Continental breeds. [2]HetIndNABC = heterozygosity for individual: NXA, NXB, NXC, AXB, AXC, BXC. HetMatNABC = heterozygosity for maternal effect. [3] agedam = age of dam at the first calf considering in this study as linear and quadratic covariable.

5.2.4.2 Single-step GBLUP (ssGBLUP)

Genomic breeding values for all traits were directly predicted using the ssGBLUP procedure (MISZTAL et al., 2009; AGUILAR et al., 2010; CHRISTENSEN; LUND, 2010). The ssGBLUP is a modified version of the traditional BLUP, in which the inverse of the pedigree-based relationship matrix ($\mathbf{A}^{-1}$) is replaced by the $\mathbf{H}^{-1}$ matrix. The $\mathbf{H}^{-1}$ is defined as follow (LEGARRA et al., 2009; AGUILAR et al., 2010):

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\tau}\mathbf{G}^{-1} - \boldsymbol{\omega}\mathbf{A}_{\mathbf{22}}^{-1} \end{bmatrix},$$

where $\mathbf{A}^{-1}$ was previously defined, τ and ω are the scaling factors used to combine **G** and $\mathbf{A_{22}}$ (assumed as τ=1.0 and ω=0.7 in order to reduce bias; MISZTAL et al., 2010; TSURUTA et al., 2011; GRIGOLETTO et al., 2020a), $\mathbf{A}_{\mathbf{22}}^{-1}$ is the inverse of pedigree-based relationship matrix for the genotyped animals, and $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix (**G**), which was calculated as (VANRADEN, 2008):

$$\mathbf{G} = \mathbf{ZZ'/k},$$

where **Z** is the matrix containing the centered genotypes (-1, 0, 1) accounting for the observed allelic frequencies; and k is a scaling parameter, defined as $2\sum p(1-p)$, in which p is the observed allele frequency of each marker. The weighting factor can be derived either based on SNP frequencies (VANRADEN, 2008), or by ensuring that the average diagonal of **G** is close to one as in $\mathbf{A_{22}}$ (VITEZICA et al., 2011). In order to minimize issues with **G** inversion, 0.05 of **A** was added to 0.95 of the **G** matrices.

5.2.4.3 BayesR

BayesR is an extension of BayesCπ, where SNP effects are assumed to be sampled from a mixture of normal distributions (ERBE et al., 2012; BOLORMAA et al., 2013). The statistical model of BayesR could be written as:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zg} + \boldsymbol{e},$$

Where $\boldsymbol{y}$ is a vector of the deregressed EBV (dEBV) as the response variable obtained following the VanRaden (2009) methodology; **X** is the incidence matrices of fixed (the CGs were used as a sequence of ones (1), since in this method we used dEBVs); $\boldsymbol{Z}$ is an n by m matrix of genotypes encoded as 0, 1, 2 copies of the reference allele; $\boldsymbol{g}$ is the sum of m-dimensional vector of SNP effects that derived from four independent normal distributions with mean of zero, and e as residual effect. The relative variance for each distribution is fixed as:

$$\boldsymbol{\rho}\left(\boldsymbol{g} \mid \boldsymbol{\pi}, \boldsymbol{\sigma_g^2}\right) = \boldsymbol{\pi_1}\, \boldsymbol{x}\, \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{0}\, \boldsymbol{x}\, \boldsymbol{\sigma_g^2}\right) + \boldsymbol{\pi_2}\, \boldsymbol{x}\, \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{10^{-4}} \boldsymbol{x}\, \boldsymbol{\sigma_g^2}\right) + \boldsymbol{\pi_3}\, \boldsymbol{x}\, \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{10^{-3}}\, \boldsymbol{x}\, \boldsymbol{\sigma_g^2}\right) + \boldsymbol{\pi_4}\, \boldsymbol{x}\, \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{10^{-4}}\, \boldsymbol{x}\, \boldsymbol{\sigma_g^2})$$

Where σ_g^2 is the additive genetic variances explained by SNPs, and $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$. The unknown parameters (b, π, g, σ_g^2, σ_e^2) are obtained from a Gibbs scheme based MCMC iterations. A Gibbs sampler was applied and a total of 500,000 iterations of sampling was performed, with a burn-in of 25,000 iterations.

5.3.5 Expected accuracy of genomic prediction

The expected accuracies (AccE) were estimated as the correlation between true and estimated genomic values, i.e. $\sqrt{\frac{N_p\, h_2}{N_p\, h_2 + M_e}}$ (DAETWYLER et al., 2010), where N_p is the number

of individuals in the training population (genotyped and measured for each trait), h_2 is the heritability of each trait (Table 5.3), M_e is the effective number of loci, which can be calculated as $2N_e$L/log($4N_e$L) (GODDARD, 2008), where assumed genome length (L) was 30.58 Morgans (ARIAS et al., 2009) and N_e was a value of 128 animals (PERIPOLLI et al., 2020).

5.3.5 Validation of the models

The dataset was split into training and validation populations to estimate the accuracy of the prediction equation in the validation population following the forward validation approach divided the animals based on birth year. Animals from 2001 to 2014, considering the older animals, composing the training (n=1,452) population and animals (without progeny from 2016) the validation (n=448) population.

5.3.6 Predictive performance

The observed accuracy (AccO) of genomic breeding values (GEBVs) of each method (GBLUP, ssGBLUP, and BayesR) trait were derived:

i) as the Pearson's correlation coefficients calculated between GEBVs and Estimated Breeding Values (EBVs; based on the pedigree-based relationship matrix estimated by BLUP) of validation population.

ii) In addition, the regression coefficient (an indicator of inflation or deflation of the EBVs on GEBVs) was assessed using a linear regression model of EBVs on GEBVs, for the validation animals using R software (R CORE TEAM, 2019).

5.4 RESULTS AND DISCUSSION

5.4.1 Descriptive statistics and genetic parameters

Table 5.2 summarizes all phenotypic traits based on the following parameters: number of observations (N), mean, standard deviation (SD), and phenotypic range for growth, reproductive, ultrasound carcass, and meat quality traits. The size of training (N=1,452) and validation (N=448) populations comprises all animals with genotypic and phenotypic information corresponding in 75% (training) and 25% (validation) of total genotyped animals.

Table 5.2 Descriptive statistics for growth, reproductive and ultrasound carcass and meat quality traits.

Trait, measurement unit	Abbreviation	N	Mean ± SD	Range
Birth weight, kg	BW	352,732	33.07±4.43	21-44
Weaning weight, kg	WW	381,269	194.37±37.42	101-298
Yearling weight, kg	YW	146,795	410.58±36.96	271-563
Muscularity, score	MUSC	134,527	4.28±1.15	1-6
Scrotal circumference at 365 days, cm	SC365	52,863	28.42±3.78	19.1-36.9
Scrotal circumference at 420 days, cm	SC420	52,863	27.42±3.70	16.3-36.57
***Longissimus* muscle area, cm**	LMA	8,891	57.89±12.6	23.1-89.9
Backfat thickness, cm	BFT	7,749	2.53±1.27	0.00-4.95
Rump fat thickness, cm	RFT	7,999	2.07±1.00	0.00-3.95
Marbling, percentage	MARB	1,801	3.21±1.09	0.56-5.98

N: number of observations; SD: standard deviation

The variance components and genetic parameter estimates for all traits are presented in Table 5.3. Heritability estimates ranged from 0.17 to 0.30 (average: 0.24 ± 0.03) for growth traits (BW, WW, YW, and MUSC), 0.21 to 0.24 (average: 0.22 ± 0.03) for reproductive (SC365 and

SC420), and 0.15 to 0.26 (average: 0.21 ± 0.02) for ultrasound carcass and meat quality traits (LMA, BFT, RFT, and MARB). These estimates are in agreement to the ones reported by Grigoletto et al. (2019; 2020a) for the same population and database using genomic information through the genome-wide association studies (GWAS), indicating that those moderate to high heritability will enable faster genetic progress for those important economically traits in the Montana Tropical® Composite population through selective breeding.

Table 5.3. Variance components and genetic parameter estimate for growth, reproductive, ultrasound carcass and meat quality traits in the Montana Tropical® Composite cattle population.

Trait	σ^2_a± SE	σ^2_m± SE	r_{am}± SE	σ^2_c± SE	σ^2_e± SE	h^2_a± SE	h^2_m± SE
BW	4.64± 0.16	1.10± 0.07	-0.42± 0.03	0.55± 0.04	10.01± 0.10	0.30± 0.05	0.07± 0.01
WW	159.08± 5.97	73.21± 3.46	-0.47± 0.04	88.87± 2.11	391.45± 3.72	0.24± 0.03	0.11± 0.01
YW	265.40± 13.6	80.05± 7.32	-0.30± 0.02	9.61± 4.34	603.44± 8.36	0.29± 0.04	0.08± 0.01
MUSC	0.18± 0.11	0.03± 0.004	-0.25± 0.02	-	0.58± 0.006	0.18± 0.02	0.03± 0.01
SC365	1.93± 0.17	0.77± 0.12	-0.57± 0.04	0.32± 0.09	6.70± 0.12	0.21± 0.03	0.08± 0.01
SC420	2.16± 0.15	-	-	-	6.80± 0.12	0.24± 0.03	-
LMA	12.42± 1.68	-	-	-	34.36± 1.45	0.26± 0.03	-
BFT	0.20± 0.05	-	-	-	0.83± 0.04	0.19± 0.02	-
RFT	0.09± 0.01	-	-	-	0.49± 0.01	0.15± 0.01	-
MARB	0.14± 0.04	-	-	-	0.39± 0.03	0.26± 0.04	-

σ^2_a: Additive genetic variance; σ^2_m: Maternal genetic variance; r_{am}: correlation between additive and maternal genetic effects; σ^2_c: Permanent environmental variance; σ^2_e: Residual variance; h^2: Direct heritability; h^2_m: Maternal heritability; c^2: Permanent environmental effect; SE: standard error.

5.4.2 Accuracy and regression coefficient of genomic predictions

The accuracies of genomic predictions are presented in Table 5.4 and 5.5 for all traits and methods. The expected average accuracies (AccE) for traits measured in the growth (BW, WW, YW, and MUSC), reproductive (SC365 and SC420), ultrasound carcass traits and meat quality (LMA, BFT, RFT, and MARB) traits were 0.46 ± 0.04, 0.45 ± 0.02, and 0.43 ± 0.05, respectively (Table 5.4 and 5.5).

Table 5.4 Expected (AccE) and observed (AccO) accuracy and regression coefficient for growth and reproductive traits using alternative genomic approaches.

Trait[1]	Methods	AccE	AccO	Regression Coefficient
BW	BLUP	0.51	0.55	0.89
	ssGBLUP		0.61	0.68
	GBLUP		0.56	0.89
	Bayes R		0.62	1.46
WW	BLUP	0.47	0.52	0.70
	ssGBLUP		0.62	0.92
	GBLUP		0.60	0.89
	Bayes R		0.59	1.01
YW	BLUP	0.50	0.52	0.70
	ssGBLUP		0.56	0.76
	GBLUP		0.56	0.70
	Bayes R		0.58	0.82
MUSC	BLUP	0.42	0.43	0.77
	ssGBLUP		0.54	0.71
	GBLUP		0.49	0.62
	Bayes R		0.64	0.90
SC365	BLUP	0.44	0.48	0.83
	ssGBLUP		0.54	0.74
	GBLUP		0.58	1.55
	Bayes R		0.60	1.54
SC420	BLUP	0.47	0.53	0.79
	ssGBLUP		0.53	0.63
	GBLUP		0.55	0.73
	Bayes R		0.53	1.27

[1]BW: Birth weight; WW: Weaning weight; YW: Yearling weight; MUSC: muscularity; SC365: scrotal circumference at 365 days; SC420: scrotal circumference at 420 days.

The average observed accuracies (± SD) for traits measured in the growth traits for ssGBLUP, GBLUP, and BayesR were 0.58 ± 0.04, 0.55 ± 0.04, and 0.61 ± 0.03, respectively. Therefore, an increase of 15.8% in accuracy was observed on average between the three genomic methods used in this composite population. On average across growth traits, the slope of the regression of EBV on GEBV was equal to 0.76 and 0.77 for ssGBLUP and GBLUP, respectively, revealing an inflated slope on the estimations. And a value of regression coefficient close to 1 (1.04) to BayesR. In this sense, BayesR method outperformed compared to ssGBLUP and GBLUP, which obtained inflated in terms of scale. Higher observed accuracies values (AccO) for reproductive traits was 0.60 for SC365 using BayesR method, and 0.55 for SC420 through GBLUP. Moreover, for ultrasound carcass and meat quality traits, the highest increase in accuracy was observed for LMA (0.14) indicated a 33% increase in accuracy of breeding values for ultrasound-based carcass traits by incorporating genomic information through the Bayesian approach. In addition, the estimated GEBVs were less biased (i.e., regression coefficient was closer to 1.0) for BayesR (average = 0.10 between ultrasound carcass and meat quality traits) compared to GBLUP (average = 0.12 between ultrasound carcass and meat quality traits). Therefore, in general, Bayesian approach outperformed GBLUP and ssGBLUP in terms of the regression coefficient; for most traits, predictions of GEBVs obtained with BayesR were deflated, closer to 1 (Table 5.5). Similar was suggested by Neves et al. (2014) for Nellore cattle using two Bayesian (BayesC and BLASSO) methods compared with GBLUP approach. For a beef cattle composite population, Hay and Roberts (2017) obtained more accurate GEBV predictions using Bayesian models (BayesA and BayesB) than through GBLUP for female longevity. One feasible explanation for that could be related to the capacity of Bayesian methods modeling SNP effects affected by genes of moderate-to-large effects (FERNANDO et al., 2014).

Table 5.5 Expected (AccE) and observed (AccO) accuracy and regression coefficient for ultrasound carcass and meat quality traits using alternative genomic approaches.

Trait[1]	Methods	AccE	AccO	Regression Coefficient
LMA	BLUP	0.48	0.44	0.81
	ssGBLUP		0.51	0.69
	GBLUP		0.50	0.81
	Bayes R		0.58	0.91
BFT	BLUP	0.40	0.49	0.70
	ssGBLUP		0.53	1.10
	GBLUP		0.49	0.90
	Bayes R		0.49	1.19
RFT	BLUP	0.38	0.51	0.70
	ssGBLUP		0.56	0.65
	GBLUP		0.57	0.78
	Bayes R		0.57	1.14
MARB	BLUP	0.48	0.56	0.97
	ssGBLUP		0.59	0.86
	GBLUP		0.57	1.01
	Bayes R		0.56	1.27

[1]LMA: Longissimus muscle area; BFT: Backfat thickness; RFT: rump fat thickness; MARB: marbling.

Most studies of genomic predictions in beef cattle report the slope of EBV regressed on the GEBVs as a measure of genomic performance following a deflated slope of regression (HAY; ROBERTS, 2017; NEVES et al., 2014) similar obtained in this current study for some traits and methods with few exceptions. These results indicated that the GEBVs are less spread (i.e. regression coefficient higher than 1). Similar scale (on average) was reported for a sheep composite population by Brito et al. (2017) using GBLUP method.

In relation to the single-step approach, it is still unclear how factors could affect GEBV accuracy in terms of population structure and the relationship between training and validation populations due to the genetic relationships among individuals captured by the markers (HABIER et al., 2007; ZHOU et al., 2018). The linkage disequilibrium is an important factor to successfully

predict GEBVs and their accuracy (GODDARD; HAYES; MEUWISSEN, 2011). Thus, Zhou et al. (2018) showed that ssGBLUP mainly utilized the genomic relationship among the training and validation populations (Figure 5.1), instead of capturing the LD between markers and QTL. In this composite population, the level of linkage disequilibrium exceeded the level of linkage disequilibrium threshold ($r^2 > 0.2$) at short distance which could be enough for genomic predictions.

The accuracies observed for most traits indicate that genomic selection is a very important tool to increase the rate of genetic gains in the Montana composite beef cattle population. The forward prediction validation design was the only scheme used in this study because it can mimic the current genomic selection in this population. However, it is important to increase the number of animals in the training population and control the relatedness between reference and validation of animals in order to avoid inflated accuracies (BOISON et al., 2017). The traditional EBV was used as a reference to compare the gain in accuracy attribute to the inclusion of genomic information because the selected candidates in this population have their own performance data recorded. In addition, GEBV accuracies are expected to increase by adding more genotyped animals.

Figure 5.1 shows the principal components analysis (PCA) of the training and validation population by the genetic relationship matrix among those animals. The first principal component, represented on the x-axis, captures 7% of the total variation, clearly there are no evidence of population stratification (i.e. existence of subgroups among individuals). Further details are described by Grigoletto et al. (2020b; unpublished)

Another topic is the moderate SNP panel used (~52,000 markers) which showed be an optimal density and feasible to genomic prediction in the population analyzed. Previous studies that compared genomic predictions obtained with low to high-density (HD) in beef and dairy cattle populations reported a marginal increase in accuracies using HD panels (ERBE et al., 2012; SU et al., 2012; BRESOLIN et al., 2017; OGAWA et al., 2017; ZHANG et al., 2019).

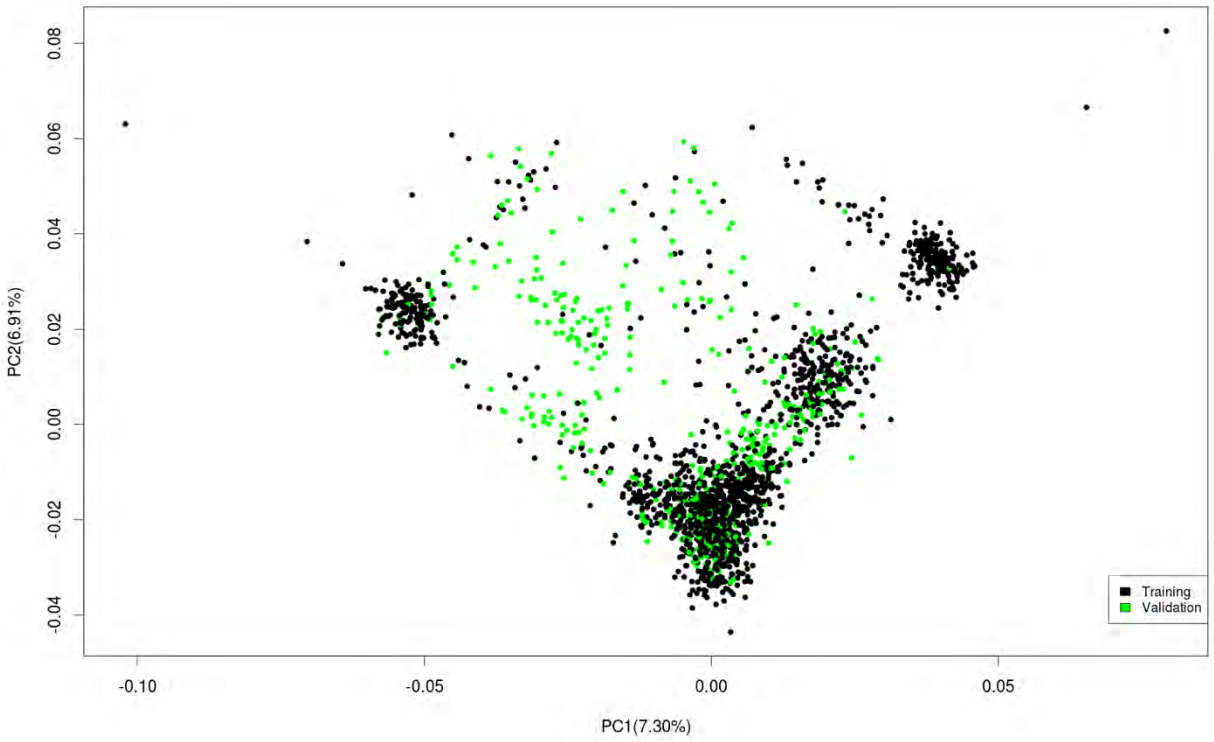


Figure 5.1 Plot of principal components (PC1; 7.30%) and PC2 (6.91%) of the genomic relationship matrix. The dots animals genotyped included the training (black) and validation (green) populations.

This study demonstrates that genomic selection for this complexity composite beef cattle population is possible across several economic traits from a comprehensive analysis of genomic prediction using imputed moderate SNP chip. We reported that there is enough phenotypic variability to implement genetic and genomic selection for growth traits in the Montana Tropical® Composite. This will make it possible to select beef cattle animals at an earlier age for breeding,

thus reducing generation interval. It is also important to highlight the need to maintain the performance recording to continuously update and increase the number of training populations over time since the predictive performance is directly influenced by the number of animals in the reference population.

5.5 CONCLUSIONS

The accuracies reported in this study support the feasibility of applying genomic selection for growth, reproductive, and ultrasound carcass and meat quality traits in Montana composite beef cattle population using the moderate SNP chip. Our findings indicate that relatively accurate GEBVs can be estimated for various productive traits and different statistical methodologies (GBLUP, single-step GBLUP, and BayesR). Bayesian approach outperformed compared to ssGBLUP and GBLUP for the majority of traits analyzed with few exceptions. Further work is needed on the design of genomic selection for this particular population, especially the inclusion of a closer genetic population (e.g. crossbred Angus x Simmental) to increase the number of animals on the reference population. In summary, these initial findings can guarantee the success of genomic selection implementation and the gain of genomic information including for commercial applications on this composite population.

5.6 REFERENCES

AGUILAR, I.; MISZTAL, I.; JOHNSON, D. L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T. J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **Journal of Dairy Science,** v. 93, p.743-752, 2010. Disponível em: <doi: 10.3168/jds.2009-2730>.

ALILOO, H. et al. The feasibility of using low-density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa. **Journal of dairy science**, v. 101, n. 10, p. 9108-9127, 2018.

ARIAS, J. A. et al. A high density linkage map of the bovine genome. **BMC genetics**, v. 10, n. 1, p. 18, 2009.

BERNARDES, P. A. et al. Evaluation of imputation accuracy using the combination of two high-density panels in Nelore beef cattle. **Scientific Reports**, v. 9, n. 1, p. 1-10, 2019.

BERRY, D. P.; MCCLURE, M. C.; MULLEN, M. P. Within-and across-breed imputation of high-density genotypes in dairy and beef cattle from medium-and low-density genotypes. **Journal of Animal Breeding and Genetics**, v. 131, n. 3, p. 165-172, 2014.

BOICHARD, D. et al. Genomic selection in domestic animals: principles, applications and perspectives. **Comptes rendus biologies**, v. 339, n. 7-8, p. 274-277, 2016.

BOISON, S. A. et al. Accuracy of genomic predictions in Gyr (Bos indicus) dairy cattle. **Journal of dairy science**, v. 100, n. 7, p. 5479-5490, 2017.

BOLORMAA, S. et al. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus, and composite beef cattle. **Journal of animal science**, v. 91, n. 7, p. 3088-3104, 2013.

BRESOLIN, T. et al. Effect of quality control, density and allele frequency of markers on the accuracy of genomic prediction for complex traits in Nellore cattle. **Animal production science**, v. 59, n. 1, p. 48-54, 2019.

BRITO, L. F. et al. Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. **BMC genetics**, v. 18, n. 1, p. 7, 2017.

CHEN, C.; TEMPELMAN, R. J. An integrated approach to empirical Bayesian whole genome prediction modeling. **Journal of agricultural, biological, and environmental statistics**, v. 20, n. 4, p. 491-511, 2015.

CHRISTENSEN, O. F.; LUND, M. S. Genomic prediction when some animals are not genotyped. **Genetics Selection Evolution**, v. 42, n. 1, p. 2, 2010.

CLARK, S. A.; LEE, S. H.; VAN DER WERF, J. H. J. Using Genomic Information to Estimate Genotype by Environment Interactions. **Proc. Assoc. Advmt. Breed. Genet**, v. 21, p. 137-140, 2015.

COLE, J. B.; SILVA, M. V. G. B. Genomic selection in multi-breed dairy cattle populations. **Revista Brasileira de Zootecnia**, v. 45, n. 4, p. 195-202, 2016.

DAETWYLER, H. D. et al. The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, n. 3, p. 1021-1031, 2010.

ERBE, M. et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. **Journal of dairy science**, v. 95, n. 7, p. 4114-4129, 2012.

FARAH, M. M. et al. Accuracy of genomic selection for age at puberty in a multi-breed population of tropically adapted beef cattle. **Animal genetics**, v. 47, n. 1, p. 3-11, 2016.

FERNANDO, R. L.; DEKKERS, J. C. M; GARRICK, D. J. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. **Genetics Selection Evolution**, v. 46, n. 1, p. 50, 2014.

FERRAZ, J. B. S.; ELER, J. P.; GOLDEN, B. L. A formação do composto Montana Tropical. **Revista Brasileira de Reprodução Animal**, v. 23, n. 2, p. 115-117, 1999.

GODDARD, M. E.; HAYES, B. J.; MEUWISSEN, T. H. E. Using the genomic relationship matrix to predict the accuracy of genomic selection. **Journal of animal breeding and genetics**, v. 128, n. 6, p. 409-421, 2011.

GREGORY, K. E.; CUNDIFF, L. V.; KOCH, R. M.; LUNSTRA, D. D. Germplasm utilization in beef cattle. **Beef Res. Prog. Progress Report**, v. 4, p. 7-19, 1993. http://digitalcommons.unl.edu/hruskareports/140

GREGORY, K. E.; CUNDIFF, L. V.; KOCH, R. M.; DIKEMAN, M. E.; KOOHMARAIE, M. Breed effects, retained heterosis, and estimates of genetic and phenotypic parameters for carcass and meat traits of beef cattle. **Journal of animal science**, v. 72, n. 5, p. 1174-1183, 1994.

GRIGOLETTO, L. FERRAZ, J. B. S.; OLIVEIRA, H. R.; ELER, J. P.; BUSSIMAN, F. O.; ABREU SILVA, B. C.; BALDI, F.; BRITO, L. F. Genetic Architecture of Carcass and Meat Quality Traits in Montana Tropical® Composite Beef Cattle. **Frontiers in genetics**, v. 11, p. 123, 2020a.

GRIGOLETTO, L.; BRITO, L. F.; MATTOS, E. C.; ELER, J. P.; BUSSIMAN, F. O.; ABREU SILVA, B. C.; SILVA, R. P.; CARVALHO, F. E.; BERTON, M. P.; BALDI, F.; FERRAZ, J. B. S. Genome-wide associations and detection of candidate genes for direct and maternal

genetic effects influencing growth traits in the Montana Tropical® Composite population. **Livestock Science**, v. 229, p. 64-76, 2019.

GRIGOLETTO, L. FERRAZ, J. B. S.; OLIVEIRA, H. R.; ELER, J. P.; BRITO, L. F. Predictive performance of genomic selection for composite beef cattle through the comparison of different methodologies. **Unpublised,** 2020b.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J.C.M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, n. 4, p. 2389-2397, 2007.

HAY, E. H.; ROBERTS, A. Genomic prediction and genome-wide association analysis of female longevity in a composite beef cattle breed. **Journal of animal science**, v. 95, n. 4, p. 1467-1471, 2017.

HAYES, B. J. et al. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. **PLoS Genetics**, v. 6, n. 9, 2010.

LEGARRA, A.; AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. **Journal of dairy science**, v. 92, n. 9, p. 4656-4663, 2009.

MEHRBAN, H. et al. Predictive performance of genomic selection methods for carcass traits in Hanwoo beef cattle: impacts of the genetic architecture. **Genetics Selection Evolution**, v. 49, n. 1, p. 1, 2017.

MISZTAL, I. et al. BLUPF90 and related programs (BGF90). In: **Proceedings of the 7th world congress on genetics applied to livestock production**. 2002. p. 743-744.

MISZTAL, I.; TSURUTA, S.; LOURENCO, D.; AGUILAR, I.; LEGARRA, A.; VITEZICA, Z. Manual for BLUPF90 family of programs. **Athens: University of Georgia**, 2014. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

MISZTAL, I. et al. Choice of parameters for single-step genomic evaluation for type. **Journal of Dairy Science**, v. 93, n. August, p. 166, 2010.

MRODE, R. et al. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: current status and future prospects. **Frontiers in genetics**, v. 9, p. 694, 2019.

NEVES, H. H. R. et al. Accuracy of genomic predictions in Bos indicus (Nellore) cattle. **Genetics Selection Evolution**, v. 46, n. 1, p. 17, 2014.

OGAWA, S. et al. Genomic prediction for carcass traits in Japanese Black cattle using single nucleotide polymorphism markers of different densities. **Animal Production Science**, v. 57, n. 8, p. 1631-1636, 2017.

PERIPOLLI, Elisa et al. Genome-wide scan for runs of homozygosity in the composite Montana Tropical® beef cattle. **Journal of Animal Breeding and Genetics**, v. 137, n. 2, p. 155-165, 2020.

R CORE TEAM (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. < https://www.R-project.org>.

ROLF, M. M. et al. Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. **Genetics Selection Evolution**, v. 47, n. 1, p. 23, 2015.

SAATCHI, M. et al. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. **Genetics Selection Evolution**, v. 43, n. 1, p. 40, 2011.

SU, G. et al. Comparison of genomic predictions using medium-density (~ 54,000) and high-density (~ 777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. **Journal of dairy science**, v. 95, n. 8, p. 4657-4665, 2012.

TSURUTA, S.; MISZTAL, I.; AGUILAR, I.; LAWLOR, T. J. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. **Journal of Dairy Science**, v. 94, n. 8, p. 4198-4204, 2011.

VAN DEN BERG, I. et al. Multi-breed genomic prediction using Bayes R with sequence data and dropping variants with a small effect. **Genetics Selection Evolution**, v. 49, n. 1, p. 70, 2017.

VANRADEN, P. M. Efficient Methods to Compute Genomic Predictions. **Journal of Dairy Science**, v. 91, n. 11, p. 4414–4423, 2008. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0022030208709901>.

VANRADEN, P. M. et al. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of dairy science**, v. 92, n. 1, p. 16-24, 2009.

VITEZICA, Z. G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. Bias in genomic predictions for populations under selection. **Genetics Research**, v. 93, n. 5, p. 357-366, 2011.

XU, Y. et al. Enhancing genetic gain through genomic selection: from livestock to plants. **Plant Communications**, v. 1, n. 1, p. 100005, 2020.

WANG, H. et al. Precision Lasso: accounting for correlations and linear dependencies in high-dimensional genomic data. **Bioinformatics**, v. 35, n. 7, p. 1181-1187, 2019.

ZHANG, H. et al. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. **Frontiers in genetics**, v. 10, p. 189, 2019.

ZHANG, Z.; DRUET, T. Marker imputation with low-density marker panels in Dutch Holstein cattle. **Journal of Dairy Science**, v. 93, n. 11, p. 5487-5494, 2010.

ZHOU, L. et al. Factors affecting GEBV accuracy with single-step Bayesian models. **Heredity**, v. 120, n. 2, p. 100-109, 2018.

SUPPLEMENTARY FILE

CHAPTER 2. GENOME-WIDE ASSOCIATIONS AND DETECTION OF CANDIDATE GENES FOR DIRECT AND MATERNAL GENETIC EFFECTS INFLUENCING GROWTH TRAITS IN THE MONTANA TROPICAL® COMPOSITE POPULATION.

Table S2.1 Summary of the analyzed SNP markers and mean (r^2) linkage disequilibrium (LD) between the adjacent markers obtained for each autosome (BTA)[a].

BTA	Size (Mbp)	SNP (n)	Average distance (kbp)	Mean r^2±SD
1	158.72	1,552	9.91	0.19±0.008
2	136.68	1,311	10.0	0.16±0.007
3	121.32	1,265	9.28	0.15±0.006
4	119.88	1,196	9.66	0.26±0.014
5	120.98	1,412	7.87	0.22±0.008
6	121.25	1,310	8.62	0.15±0.009
7	112.55	1,134	9.55	0.14±0.012
8	113.17	1,112	9.94	0.13±0.013
9	105.54	1,089	9.34	0.22±0.008
10	104.14	1,054	9.64	0.12±0.013
11	107.18	1,145	9.12	0.13±0.012
12	90.99	910	9.64	0.14±0.014
13	83.95	938	8.66	0.12±0.011
14	83.47	988	8.12	0.13±0.011
15	85.15	914	8.96	0.11±0.005
16	81.53	841	9.37	0.19±0.008
17	74.90	803	9.09	0.12±0.013
18	65.85	856	7.30	0.10±0.005
19	63.91	859	7.18	0.14±0.006
20	71.86	856	7.96	0.22±0.007
21	71.20	808	8.51	0.20±0.008
22	61.23	670	8.93	0.11±0.014
23	52.38	678	7.55	0.13±0.007
24	62.20	686	8.69	0.13±0.013
25	42.82	569	7.28	0.12±0.013
26	51.58	585	8.59	0.12±0.015
27	42.33	502	8.87	0.10±0.017
28	46.20	536	8.36	0.12±0.016
29	51.19	620	7.92	0.11±0.012

[a] *Bos taurus* autosome

Table S2.2 List of genomic window locations, proportion of genetic variance explained by each window (Var%) and candidate genes for Birth weight (Direct and Maternal effects) in the Montana Tropical® Composite population.

Effect	BTA:Start-End(bp)*	Var%	Candidate genes
Direct	3:14014602-14859408	4.90	*ETV3L, ARHGEF11, LRRC71, PEAR1, NTRK1, INSRR, SH2D2A, PRCC, HDGF, MRPL24, RRNAD1, ISG20L2, CRABP2, NES, BCAN, HAPLN2, GPATCH4, NAXE, TTC24, IQGAP3, MEF2D, RHBG, TSACC, CCT3, GLMP, TMEM79, SMG5, PAQR6, BGLAP, PMF1, SLC25A44, SEMA4A, LMNA, MEX3A, RAB25, LAMTOR2, UBQLN4, SSR2, ARHGEF2, KHDC4, RIT1, SYT11, GON4L, MSTO1, DAP3*
	3:88950716-89911846	1.21	*DAB1 C8B C8A FYB2 PRKAA2 PLPP3*
	4:5262797-6180362	2.33	*GRB10 DDC FIGNL1 IKZF1 SPATA48 ZPBP VWC2*
	4:6298054-7232160	4.42	*ABCA13*
	4:75002564-75951231	1.34	*TNS3 IGFBP3 IGFBP1 ADCY1*
	9:8668733-9748766	1.26	*LMBRD1 COL9A1 FAM135A SDHAF4 SMAP1 B3GAT2*
	10:29475029-30367930	1.17	*TMCO5B FMN1 GREM1 SCG5 ARHGAP11A GJD2 ACTC1 AQR ZNF770*
	12:1648104-2209666	1.56	
	12:2249044-2618039	1.55	*TDRD3 DIAPH3*
	12:2682385-3279272	1.25	
	12:84466783-85329377	1.40	
	13:62850271-63394770	1.06	*BPIFA2B BPIFA3 BPIFA1 BPIFB1 BPIFB5 CDK5RAP1 SNTA1 NECAB3 C13H20orf144 E2F1 PXMP4 ZNF341 CHMP4B RALY EIF2S2*
	14:25019900-25541189	4.74	*NSMAF TOX*
	17:68678716-69312539	1.77	*KREMEN1 EMID1 RHBDD3 EWSR1 GAS2L1 RASL10A AP1B1 NEFH THOC5 NIPSNAP1 NF2 CABP7 ZMAT5 UQCR10 ASCC2 MTMR3 HORMAD2 LIF OSM CASTOR1 TBC1D10A SF3A1 CCDC157 RNF215 SEC14L2 MTFP1 SEC14L3 SEC14L4 SEC14L6 GAL3ST1 PES1 RASIP1 IZUMO1 FUT1 FGF21 BCAT2 HSD17B14 PLEKHA4 PPP1R15A TULP2*
	18:55621823-56592023	1.34	*NUCB1 DHDH BAX FTL GYS1 RUVBL2 LHB NTF4 KCNA7 SNRNP70 LIN7B PPFIA3 HRC TRPM4 SLC6A16 TEAD2 CD37 DKKL1*

	19:19587050-20639296	1.75	*CCDC155 PTH2 GFY SLC17A7 PIH1D1 ALDH16A1 FLT3LG RPL13A RPS11 FCGRT RCN3 NOSIP PRRG2 PRR12 RRAS SCAF1 IRF3 BCL2L12 PRMT1 CPT1C TSKS AP2A1 FUZ MED25 PTOV1 PNKP AKT1S1 TBC1D17 IL4I1 NUP62 ATF5 VRK3 ZNF473 IZUMO2 MYH14 KCNC3 NAPSA NR1H2 POLD1 SPIB MYBPC2 FAM71E1 EMC10 JOSD2 LRRC4B SYT3 C18H19orf81 SHANK1 CLEC11A*
	23:3626219-4071392	3.88	*ABHD15 TP53I13 GIT1 ANKRD13B CORO6 SSH2*
	28:1934102-2814450	1.66	*NOT FOUND*
	28:40993536-41746754	2.04	*GALNT2 PGBD5 FMN2 GRID1 WAPL OPN4 LDB3 BMPR1A MMRN2 SNCG ADIRF*
	28:41961649-42445200	7.61	*GLUD1 SHLD2 SYT15 GPRIN2 PPYR1 ANXA8L1 RBP3 GDF2 GDF10 FRMPD2*
	29:49838844-50059018	1.10	*TNNT3 LSP1 TNNI2 SYT8 CTSD IFITM10 DUSP8 MOB2 TOLLIP*
Maternal	4:4645352-5182281	1.62	*COBL RF00020 GRB10 DDC FIGNL1 IKZF1*
	4:5318182-6298054	3.24	*SPATA48 ZPBP VWC2*
	5:120581133-120886093	1.12	*NOT FOUND*
	7:40136380-41114885	1.01	*TRIM7 TRIM41 RACK1 TRIM52 IFI47 ZNF496 NLRP3 RF00377 OR2B11 GCSAML OR2G2 OR2G3 OR6F1*
	10:821553-1346901	1.25	*CASC4 CTDSPL2 EIF3J SPG11 PATL2 B2M TRIM69 TERB2*
	10:25598535-26443555	1.18	*OR10G3 RAG1 SALL2 METTL3 TOX4 RAB2B CHD8 SUPT16H RPGRIP1 HNRNPC OR5AU1 TMEM253 ZNF219 ARHGEF40 RNASE13 TPPP2 NDRG2 SLC39A2*
	10:102758856-103382526	1.23	*METTL17 RF00001 RNASE2 RAG2 RNASE1 BRB RNASE6 RNASE4 ANG2 MCC DCP2 REEP5 SRP19 APC*
	11:1288880-1901787	1.04	*ACOXL BUB1 TPC3 NPHP1 MALL MAL DIAPH3*
	12:2682385-3279272	1.58	*IRS2*
	12:84559270-85343140	1.50	*COL4A1 COL4A2 RAB20 NAXD CARS2 ING1 ANKRD10*
	13:62224699-63067797	1.59	*EFCAB8 SUN5 BPIFB2 BPIFB6 BPIFB3 BPIFB4 BPIFA2A BPIFA2C BPIFA2B BPIFA3 BPIFA1 BPIFB1 BPIFB5 CDK5RAP1 SNTA1*
	15:75314110-76192434	1.25	*MAPK8IP1 C15H11orf94 PEX16 LARGE2 PHF21A SYT13 CHST1 SLC35C1 CRY2*
	21:8128853-8955497	2.45	*ARRDC4*

22:21203326-22043400	1.13	*EDEM1 ARL8B BHLHE40 ITPR1 SUMF1 SETMAR*
23:11260313-12582837	1.13	*RNF8 CMTR1 CCDC167 MDGA1 ZFAND3 BTBD9 GLO1 RF02271 DNAH8*
25:38198723-38687000	1.27	*CYTH3 RAC1 DAGLB KDELR2 GRID2IP ZDHHC4 C25H7orf26 ZNF853 ZNF316 ZNF12 RBAK RNF216*
26:23270556-24173007	1.40	*TRIM8 ARL3 SFXN2 WBP1L CYP17A1 AS3MT CNNM2 NT5C2 INA PCGF6 TAF5 RF00055 ATP5MD PDCD11 CALHM2*
29:49838844-50059018	2.72	*CALHM1 CALHM3 NEURL1 SH3PXD2A DUSP8 MOB2*

*bp=pairs of bases

Table S2.3 List of genomic window locations, proportion of genetic variance explained by each window (Var%) and candidate genes for Weaning weight (Direct and Maternal effects) in the Montana Tropical® Composite population.

Effect	BTA:Start-End(bp)*	Var%	Candidate genes
Direct	1:18170722-19450329	2.08	*TMPRSS15 CHODL C1H21orf91 RF00026 BTG3 CXADR*
	3:88652002-89697477	1.09	*DAB1 C8B C8A FYB2 PRKAA2 PLPP3 MUC1 EFNA4 UBE2Q1 TDRD10 SHE IL6R ATP8B2 AQP10 HAX1 UBAP2L C3H1orf43 C3H1orf189 TPM3 NUP210L RPS27 RAB13 JTB CREB3L4 SLC39A1 CRTC2 DENND4B GATAD2B SLC27A3 INTS3 NPR1 ILF2 SNAPIN CHTOP*
	3:15188884-16583646	3.68	*SYT11 GON4L MSTO1 PKLR HCN3 CLK2*
	10:97174992-98223275	1.26	*FLRT2*
	10:100825970-101452362	2.95	*FOXN3 EFCAB11 TDP1*
	12:1314466-1966027	2.44	*TDRD3 DIAPH3*
	12:59935901-60881012	2.20	*SLITRK6*
	12:89006700-89424324	4.09	*NOT FOUND*
	13:41678593-42154630	2.56	*FOXA2 THBD CD93 NXT1 GZF1 NAPB MGC133636 CST8 CST3*
	13:70485807-71262962	1.66	*PTPRT*
	14:63404391-64023519	1.42	*ZNF706 YWHAZ PABPC1 SNX31 ANKRD46 RNF19A SPAG1 POLR2K FBXO43*
	15:80404264-81348627	1.58	*LRRC55 APLNR TNKS1BP1 RF00432 SSRP1 RF00026 P2RX3 PRG3 MGC140461 TMX2 CTNND1 OR6Q1 OR9I1 OR5B17*
	16:6809499-8648376	1.84	*KCNT2*
	16:8758408-9815328	3.66	*NOT FOUND*
	18:5857571-6756194	1.66	*WWOX MAF*
	18:30816769-32393402	1.03	*CDH11 CDH5 BEAN1 CKLF CMTM2*
	18:33062674-34292144	1.18	*CMTM3 CMTM4 DYNC1LI2 TERB1 NAE1 CA7*
	19:60198877-60746300	2.60	*KCNJ2 KCNJ16 MAP2K6*
	19:60757465-61136715	4.92	*ABCA5*
	22:1342293-1930311	6.47	*EGFR SEC61G NEK10 SLC4A7 EOMES*
Maternal	1:72950588-73821868	1.40	*FAM43A LSG1 TMEM44 ATP13A3 GP5 LRRC15 CPN2 HES1 OPA1 ATP13A4 ATP13A5*
	2:105488933-106428272	1.01	*TNS1 RUFY4 CXCR2 CXCR1 ARPC2 GPBAR1 AAMP PNKD TMBIM1 PNKD CATIP SLC11A1 CTDSP1 VIL1 USP37 CNOT9 RF00568 PLCD4 ZNF142 BCS1L RNF25 STK36 TTLL4*
	7:17754418-18238845	5.69	*ADGRE1 VAV1 SH2D3A TRIP10 GPR108 C3 TNFSF14 CD70 TNFSF9*

		TUBB4A DENND1C CRB3 SLC25A23 SLC25A41 KHSRP GTF2F1 PSPN ALKBH7CLPP ACER1 MLLT1 SPTBN5 ACSBG2 RFX2 RANBP3 CAPS VMAC NDUFA11
10:612683-126666	9.15	*MCC DCP2*
10:1206871-1553984	1.16	*REEP5 SRP19 APC EPB41L4A CCDC9B PHGR1 DISP2 KNSTRN IVD BAHD1 CHST14 CCDC32 RPUSD2 KNL1 RAD51 RMDN3 GCHFR DNAJC17 C10H15orf62 ZFYVE19 PPP1R14D SPINT1*
10:36204749-37393241	1.20	*RHOV VPS18 DLL4 CHAC1 INO80 EXD1 CHP1 OIP5 NUSAP1 MGA MAPKBP1 JMJD7 PLA2G4B EHD4 PLA2G4E PLA2G4D PLA2G4FNDUFAF1 RTF1 ITPKA RPAP1 MIR26B TYRO3 VPS39 TMEM87A GANC*
16:80127630-80717601	1.05	*ADIPOR1 KLHL12 RABIF PTPN7 ARL8A GPR37L1 ELF3 RNPEP TIMM17A LMOD1 SHISA4 IPO9 NAV1*
21:13173337-13937115	2.26	*RGMA CHD2*
22:51967543-52912371	2.54	*TMIE ALS2CL LRRC2 TDGF1 RTP3 LTF CCRL2 CCR5 CCR2 CCR3 PRSS45 SETD2 NRADD CCDC12 PTH1R MYL3 PRSS42P KLHL18 KIF9 NGP SCAP MAP4 DHX30 SMARCC1 CSPG5 ELP6*
23:1914507-2492058	1.56	*PRIM2*
27:3938392-4498454	1.66	*NOT FOUND*

*bp=pairs of bases

Table S2.4 List of genomic window locations, proportion of genetic variance explained by each window (Var%) and candidate genes for Yearling weight (Direct effects) in the Montana Tropical® Composite population.

Effect	BTA:Start-End(bp)*	Var%	Candidate genes
Direct	1:69097650-70017168	1.33	*UMPS ITGB5 MUC13 HEG1 SLC12A8 ZNF148 SNX4 OSBPL11*
	3:28853229-29707933	2.31	*TRIM33 SYT6 OLFML3 HIPK1 DCLRE1B*
	3:29789003-30601685	1.04	*AP4B1 BCL2L15 PTPN22 RSBN1 PHTF1*
	3:44959256-46237605	2.18	*MAGI3 LRIG2 SLC16A1 DPYD*
	6:40983913-41414050	1.89	
	6:44023813-44926243	3.70	
	6:45056550-45733613	1.89	*KCNIP4 DHX15 SOD3 CCDC149 LGI2 SEPSECS PI4K2B ZCCHC4 ANAPC4 SLC34A2 SEL1L3 SMIM20*
	7:3714956-4358132	2.66	*PBX4 CILP2 NDUFA13 TSSK6 GATAD2A MAU2 SUGP1 TM6SF2 HAPLN4 NCAN NR2C2AP RFXANK BORCS8 MEF2B TMEM161A SLC25A42 ARMC6 SUGP2*
	7:66893384-68156093	1.62	*HOMER3 DDX49 COPE CERS1 GDF1 UPF1 SGCD*
	9:9870836-10629506	2.00	*SMAP1 B3GAT2 OGFRL1*
	10:27589941-28647003	2.69	*OR4F15 LPCAT4 NUTM1 NOP10 SLC12A6 EMC4 KATNBL1 EMC7 CHRM5 AVEN*
	10:41687709-42814274	3.35	*RPS29 LRR1 RPL36AL MGAT2 DNAAF2 POLE2 KLHDC1 KLHDC2 NEMF*
	10:102794657-103478776	1.57	*CASC4 CTDSPL2 EIF3J SPG11 PATL2 B2M TRIM69 TERB2*
	11:73976212-74950301	3.41	*DNMT3A POMC EFR3B DNAJC27 ADCY3 CENPO PTRHD1 NCOA1 ITSN2 FAM228A FAM228B*
	16:64560032-65415748	2.49	*SMG7 NCF2 ARPC5 APOBEC4 RGL1 COLGALT2 TSEN15*
	19:26264145-27099600	1.74	*SCIMP ZFP3 KIF1C INCA1 CAMTA2 SPAG7 ENO3 PFN1 RNF167 SLC25A11 GP1BA CHRNE C19H17orf107 MINK1 PLD2 PSMB6 GLTPD2 VMO1 TM4SF5 ZMYND15 CXCL16 MED11 PELP1 ARRB2 ALOX15 ALOX12E ALOX12 RNASEK C19H17orf49 BCL6B SLC16A13 SLC16A11 CLEC10A ASGR2 ASGR1 DLG4 ACADVL DVL2 PHF23 GABARAP CTDNEP1 ELP5 CLDN7 SLC2A4 YBX2 EIF5A GPS2 NEURL4 KCTD11 TMEM95 TNK1 PLSCR3 TMEM256 NLGN2 SPEM1*
	19:61016756-61467107	1.88	*MAP2K6 ABCA5 ABCA10 ABCA6*
	21:13173337-13937115	1.40	
	21:48911422-49856776	2.62	*SEC23A GEMIN2 TRAPPC6B PNN FBXO33*
	22:6064248-7313043	1.33	*STT3B OSBPL10 GPD1L CMTM8 CMTM7 CMTM6 DYNC1LI1 CNOT10 TRIM71 CCR4*

22:60186301-60543119	1.88	*PLXNA1 CHCHD6 TXNRD3 C22H3orf22 CHST13 UROC1 ZXDC SLC41A3*
23:36759291-37931409	2.57	*CDKAL1 E2F3 MBOAT1*
27:15166460-16300721	3.45	*ACSL1 HELT SLC25A4 CFAP97 SNX25 LRP2BP ANKRD37 UFSP2 C27H4orf47 CCDC110 PDLIM3 SORBS2 TLR3 FAM149A CYP4V2 KLKB1 F11*
29:49922603-50116711	1.47	*DUSP8 MOB2*

*bp=pairs of bases

Table S2.5 List of genomic window locations, proportion of genetic variance explained by each window (Var%) and candidate genes for Post weaning weight gain (Direct effects) in the Montana Tropical® Composite population.

Effect	BTA:Start-End(bp)*	Var%	Candidate genes
	1:8437530-9384280	1.62	*H4 ADAMTS5 ADAMTS1*
	1:15935780-17330771	1.86	*NOT FOUND*
	5:42601159-43565856	1.28	*CPNE8 PTPRR PTPRB KCNMB4 CNOT2 MYRFL RAB3IP BEST3 HMGA2 MSRB3*
	5:48208853-48600539	1.98	*LEMD3 WIF1*
	5:18109413-18584348	4.99	*CEP290 TMTC3 KITLG DUSP6 POC1B ATP2B1*
	5:18644422-19001585	1.34	
	5:19626879-20231527	2.03	
	6:18105361-18593625	2.50	*DKK2 GIMD1 AIMP1 TENM2 WWC1 RARS FBLL1 ZFYVE16 FAM151B ANKRD34B DHFR MSH3*
	7:79753818-80794531	1.93	*RASGRF2*
	10:77254858-78265085	2.16	*SPTB CHURC1 RAB15 FNTB MAX FUT8*
	12:46104077-47136309	1.22	*DACH1*
	16:16270589-17789937	4.85	*KCTD3 USH2A*
Direct	16:17931777-18989429	1.18	
	17:72554429-72961394	1.83	*CRKL AIFM3 LZTR1 THAP7 TUBA3E LRRC74B P2RX6 SLC7A4 MZT2B SMPD4 MED15 KLHL22 SCARF2 ZNF74 DGCR2 TSSK1B TSSK2 ESS2 GSC2 SLC25A1 HIRA MRPL40 UFD1 CDC45 CLDN5 SEPT5 GP1BB TBX1 GNB1L TXNRD2 COMT ARVCF TANGO2 DGCR8 TRMT2A RANBP1 ZDHHC8 CCDC188 RTN4R PRODH DGCR6L*
	19:2378815-3168492	2.34	*NOT FOUND*
	22:5111080-5952544	5.30	*TGFBR2 GADL1 STT3B OSBPL10 ATP2B2 SLC6A11 SLC6A1 ATG7 VGLL4*
	22:54827086-55618361	10.80	*TAMM41 C25H16orf82 KDM8 NSMCE1 IL4R IL21R*
	25:23888782-24765594	1.87	*GTF3C1*

*bp=pairs of bases

CHAPTER 3. GENETIC ARCHITECTURE OF CARCASS AND MEAT QUALITY TRAITS IN MONTANA TROPICAL® COMPOSITE BEEF CATTLE.

Table S3.1 Genome-wide regions and candidate genes that explain more than 1% of the total genetic variance for *Longissimus* muscle area (LMA) in the Montana Tropical Composite beef cattle population.

BTA	Position (bp)		Gene name
	Start	End	
1	51,940,576	51,941,806	*CCDC54*
2	64,157,472	64,872,456	*NCKAP5*
2	64,874,976	64,937,074	*LYPD1*
2	64,935,007	65,196,499	*GPR39*
4	12,191,129	12,534,305	*PPP1R9A*
4	12,542,354	12,576,328	*PON1*
4	12,593,909	12,631,421	*PON3*
4	12,645,430	12,673,193	*PON2*
4	12,750,329	12,843,879	*ASB4*
4	12,881,889	12,895,362	*PDK4*
4	13,061,840	13,440,805	*DYNC1I1*
6	102,192,504	102,212,283	*HSD17B13*
6	102,227,496	102,276,940	*HSD17B11*
6	102,320,394	102,357,766	*NUDT9*
6	102,370,630	102,423,199	*SPARCL1*
6	102,494,936	102,500,867	*DSPP*
6	102,526,308	102,543,087	*DMP1*
6	102,716,789	102,881,084	*PPP2R2C*
6	102,896,904	102,928,920	*WFS1*
6	103,013,174	103,169,423	*JAKMIP1*
7	496,572	537,888	*FLT4*
10	5,450,044	5,474,341	*HRH2*
10	5,613,866	5,674,977	*SFXN1*
10	5,715,882	5,718,113	*DRD1*
14	22,640,221	22,953,771	*XKR4*
14	23,034,280	23,070,124	*TMEM68*
14	23,070,145	23,095,949	*TGS1*
14	23,134,995	23,244,752	*LYN*
14	23,278,316	23,279,689	*RPS20*
14	23,299,177	23,300,199	*MOS*
14	23,330,541	23,375,751	*PLAG1*
18	25,722,016	25,756,803	*KIFC3*
18	25,813,885	25,882,542	*CNGB1*
18	25,888,368	25,895,551	*TEPP*
18	25,902,577	25,904,786	*ZNF319*
18	25,907,856	25,923,157	*USB1*

18	25,939,947	25,960,720	*MMP15*
18	25,994,990	26,008,531	*CFAP20*
18	26,016,093	26,051,043	*CSNK2A2*
18	26,085,843	26,119,863	*CCDC113*
18	26,119,132	26,132,035	*PRSS54*
18	26,198,116	26,204,583	*GINS3*
18	26,255,289	26,298,424	*NDRG4*
18	26,299,649	26,305,367	*SETD6*
18	26,304,976	26,388,347	*CNOT1*
18	26,429,234	26,440,641	*SLC38A7*
18	26,447,740	26,471,387	*GOT2*
18	5,240,349	6,170,333	*WWOX*
24	47,693,446	47,994,499	*ZBTB7C*
24	48,220,180	48,542,521	*CTIF*
29	49,970,526	49,980,774	*DUSP8*
29	49,990,564	50,043,414	*MOB2*

BTA: *Bos taurus* Autossome.

Table S3.2 Genome-wide regions and candidate genes that explain more than 1% of the total genetic variance for backfat thickness (BFT) in the Montana Tropical Composite beef cattle population.

BTA	Position (bp)		Gene name
	Start	End	
1	114,908,447	114,910,097	*P2RY1*
1	115,303,718	115,524,968	*MBNL1*
4	10,600,800	10,732,873	*VPS50*
4	10,781,954	10,893,236	*CALCR*
4	117,125,786	117,138,560	*INSIG1*
4	117,222,455	117,226,427	*EN2*
4	117,261,920	117,301,793	*CNPY1*
4	117,366,283	117,462,802	*RBM33*
5	109,423,829	109,432,050	*CDC42EP1*
5	109,432,450	109,438,746	*LGALS2*
5	109,458,483	109,477,483	*GGA1*
5	109,484,519	109,497,008	*SH3BP1*
5	109,498,401	109,504,537	*PDXP*
5	109,512,489	109,515,793	*LGALS1*
5	109,519,065	109,526,829	*NOL12*
5	109,532,805	109,591,861	*TRIOBP*
5	109,621,889	109,634,246	*GCAT*
5	109,630,804	109,639,001	*GALR3*
5	109,643,325	109,653,939	*ANKRD54*
5	109,662,810	109,683,993	*EIF3L*
5	109,709,317	109,737,117	*MICALL1*
5	109,738,012	109,745,876	*C5H22orf23*
5	109,745,905	109,755,102	*POLR2F*
5	109,757,715	109,768,623	*SOX10*
5	109,834,675	109,851,541	*PICK1*
5	109,853,974	109,858,626	*SLC16A8*
5	109,861,367	109,891,295	*BAIAP2L2*
5	109,891,806	109,938,952	*PLA2G6*
9	10,397,908	10,413,050	*OGFRL1*
10	101,466,634	101,634,822	*EFCAB11*
10	101,640,018	101,708,726	*TDP1*
10	101,723,038	101,828,432	*KCNK13*
11	105,021,683	105,114,871	*RXRA*
11	105,268,089	105,419,303	*COL5A1*
11	105,440,661	105,450,692	*FCN1*
13	61,464,644	61,471,874	*PDRG1*
13	61,487,256	61,512,327	*XKR7*
13	61,525,479	61,545,528	*CCM2L*
13	61,563,070	61,608,503	*HCK*
13	61,622,078	61,675,062	*TM9SF4*
13	61,683,029	61,694,301	*PLAGL2*

13	61,694,343	61,745,905	*POFUT1*
13	61,747,316	61,787,770	*KIF3B*
13	61,807,148	61,871,197	*ASXL1*
13	61,874,277	61,955,381	*NOL4L*
13	62,102,412	62,131,186	*COMMD7*
13	62,142,537	62,176,363	*DNMT3B*
13	62,182,219	62,210,102	*MAPRE1*
13	62,220,966	62,268,229	*EFCAB8*
13	62,276,784	62,297,291	*SUN5*
13	62,304,795	62,323,150	*BPIFB2*
14	58,819,937	59,239,114	*ZFPM2*
15	75,918,764	75,938,190	*MAPK8IP1*
15	75,938,260	75,938,905	*C15H11orf94*
15	75,941,284	75,946,922	*PEX16*
15	75,950,774	75,956,775	*LARGE2*
15	75,955,555	76,148,892	*PHF21A*
15	76,282,608	76,319,986	*CREB3L1*
15	75,702,165	75,705,849	*CHST1*
15	75,852,557	75,860,286	*SLC35C1*
15	75,879,778	75,915,413	*CRY2*
17	55,569,766	55,740,689	*CIT*
17	55,750,475	55,759,908	*PRKAB1*
17	55,795,742	55,830,506	*TMEM233*
17	55,876,002	56,046,137	*CCDC60*
17	56,156,718	56,169,968	*HSPB8*
17	56,189,388	56,355,937	*SRRM4*
22	12,106,802	12,195,313	*SCN10A*
22	12,236,370	12,317,469	*SCN11A*
22	12,377,233	12,429,953	*WDR48*
22	12,431,752	12,443,560	*GORASP1*
22	12,455,223	12,477,183	*TTC21A*
22	12,478,903	12,492,412	*CSRNP1*
22	12,509,227	12,518,446	*XIRP1*
22	12,569,452	12,583,576	*CX3CR1*
22	12,634,614	12,635,672	*CCR8*
22	12,645,751	12,660,388	*SLC25A38*
22	12,684,159	12,710,177	*RPSA*
22	12,723,390	12,756,307	*MOBP*
22	12,922,611	13,149,096	*MYRIP*
22	13,170,136	13,172,696	*EIF1B*
22	13,227,438	13,261,914	*ENTPD3*
22	13,292,655	13,296,110	*RPL14*
22	13,302,897	13,310,529	*ZNF619*
22	13,325,988	13,336,394	*ZNF621*
22	13,635,727	13,637,448	*HSPD1*
22	52,089,753	52,102,048	*CSPG5*

22	52,133,185	52,149,780	*ELP6*
22	52,163,455	52,264,116	*SCAP*
22	52,263,501	52,284,441	*PTPN23*
22	52,297,736	52,301,963	*NGP*
22	52,310,680	52,361,884	*KLHL18*
22	52,361,996	52,398,789	*KIF9*
22	52,450,147	52,509,045	*SETD2*
22	52,511,365	52,515,665	*NRADD*
22	52,548,484	52,585,071	*CCDC12*
22	52,597,375	52,624,332	*PTH1R*
22	52,638,016	52,644,842	*MYL3*
22	52,671,297	52,675,596	*PRSS42P*
22	52,731,581	52,739,251	*PRSS45*
24	2,145,784	2,156,850	*GALR1*
24	2,224,463	2,328,579	*MBP*
24	2,336,527	2,399,019	*ZNF236*
26	17,925,643	18,029,878	*PIK3AP1*
26	18,219,077	18,256,221	*LCOR*
26	18,272,204	18,444,938	*SLIT1*
27	25,180,847	25,190,311	*PPP1R3B*
27	25,497,637	25,654,276	*TNKS*

BTA: *Bos taurus* Autossome.

Table S3.3 Genome-wide regions and candidate genes that explain more than 1% of the total genetic variance for rump fat thickness (RFT) in the Montana Tropical Composite beef cattle population.

BTA	Position (bp)		Gene name
	Start	End	
2	29,840,643	29,898,509	*SCN7A*
2	29,996,676	30,088,178	*SCN9A*
2	30,224,956	30,320,883	*SCN1A*
2	30,336,543	30,434,725	*TTC21B*
11	24,464,835	24,474,936	*PKDCC*
11	24,585,989	24,739,169	*EML4*
11	24,756,935	24,769,582	*COX7A2L*
11	24,832,853	24,885,462	*KCNG3*
11	24,886,922	25,096,963	*MTA3*
13	47,916,971	48,028,742	*SHLD1*
13	48,069,187	48,084,157	*CHGB*
13	48,110,221	48,120,755	*TRMT6*
13	48,123,958	48,166,976	*MCM8*
13	48,186,875	48,209,791	*CRLS1*
13	48,213,047	48,233,309	*LRRN4*
13	48,251,375	48,312,017	*FERMT1*
13	61,622,078	61,675,062	*TM9SF4*
13	61,683,029	61,694,301	*PLAGL2*
13	61,694,343	61,745,905	*POFUT1*
13	61,747,316	61,787,770	*KIF3B*
13	61,807,148	61,871,197	*ASXL1*
13	61,874,277	61,955,381	*NOL4L*
13	62,102,412	62,131,186	*COMMD7*
13	62,142,537	62,176,363	*DNMT3B*
13	62,182,219	62,210,102	*MAPRE1*
13	62,220,966	62,268,229	*EFCAB8*
13	62,276,784	62,297,291	*SUN5*
13	62,304,795	62,323,150	*BPIFB2*
13	62,331,559	62,347,090	*BPIFB6*
13	62,356,043	62,372,303	*BPIFB3*
13	62,381,695	62,408,299	*BPIFB4*
14	7,865,552	7,951,420	*ST3GAL1*
14	8,073,208	8,129,604	*NDRG1*
14	8,137,189	8,169,248	*CCN4*
18	60,741,854	60,749,312	*ZNF331*
18	60,856,348	60,866,029	*MGC139164*
18	60,905,906	60,925,131	*NLRP12*
18	61,009,582	61,018,636	*MGC157082*
18	65,375,097	65,405,941	*ZNF814*
21	18,837,293	19,260,319	*NTRK3*
21	19,474,861	19,484,599	*MRPL46*

21	19,484,255	19,495,138	*MRPS11*
21	19,522,497	19,547,431	*DET1*
21	19,591,620	19,601,252	*AEN*
21	19,607,426	19,625,071	*ISG20*
23	3,051,079	3,142,445	*ZNF451*
23	3,189,366	3,252,666	*BEND6*

BTA: *Bos taurus* Autossome.

Table S3.4 Genome-wide regions and candidate genes that explain more than 1% of the total genetic variance for marbling score (MARB) in the Montana Tropical Composite beef cattle population.

BTA	Position (bp)		Gene name
	Start	End	
2	96,101,252	96,318,495	*PLEKHM3*
2	96,395,739	96,397,590	*CRYGD*
2	96,405,966	96,407,968	*CRYGC*
2	96,417,867	96,420,083	*CRYGB*
2	96,432,628	96,435,050	*CRYGA*
2	96,437,069	96,460,618	*C2H2orf80*
2	96,510,261	96,532,002	*IDH1*
2	96,557,844	96,625,577	*PIKFYVE*
2	96,667,717	96,752,328	*PTH2R*
5	94,669,118	94,824,113	*RERG*
6	4,138,065	4,200,076	*QRFPR*
10	7,084,931	7,115,694	*POC5*
10	7,361,671	7,574,347	*SV2C*
12	22,798,940	23,004,081	*LHFPL6*
12	23,178,045	23,219,416	*NHLRC3*
12	23,219,451	23,251,116	*PROSER1*
12	23,258,105	23,284,263	*STOML3*
12	23,297,835	23,449,177	*FREM2*
13	47,720,768	47,778,950	*GPCPD1*
13	47,916,971	48,028,742	*SHLD1*
13	48,069,187	48,084,157	*CHGB*
13	48,110,221	48,120,755	*TRMT6*
13	48,123,958	48,166,976	*MCM8*
13	48,186,875	48,209,791	*CRLS1*
13	48,213,047	48,233,309	*LRRN4*
13	48,251,375	48,312,017	*FERMT1*
14	48,629,593	48,904,459	*TRPS1*
15	24,284,251	24,319,569	*ZW10*
21	14,842,445	15,219,453	*SLCO3A1*
27	33,387,283	33,416,172	*DDHD2*
27	33,416,459	33,421,456	*PLPP5*
27	33,427,742	33,524,431	*NSD3*
27	33,528,788	33,544,151	*LETM2*
27	33,549,267	33,599,083	*FGFR1*

BTA: *Bos taurus* Autossome.

CHAPTER 4. GENOMIC IMPUTATION AND CHARACTERIZATION OF POPULATION STRUCTURE, LINKAGE DISEQUILIBRIUM, CONSISTENCY OF GAMETIC PHASE AND ADMIXTURE IN TROPICAL COMPOSITE BEEF CATTLE

Table S4.1 Cross-validation (CV) error estimated using different numbers of ancestral populations (k). The lowest CV error was obtained at k=26.

K	Value
1	0.589
2	0.586
3	0.582
4	0.577
5	0.572
6	0.568
7	0.565
8	0.562
9	0.558
10	0.556
11	0.554
12	0.551
13	0.550
14	0.548
15	0.547
16	0.544
17	0.543
18	0.541
19	0.540
20	0.538
21	0.538
22	0.537
23	0.536
24	0.535
25	0.534
26	0.533
27	0.538
28	0.542
29	0.547
30	0.551

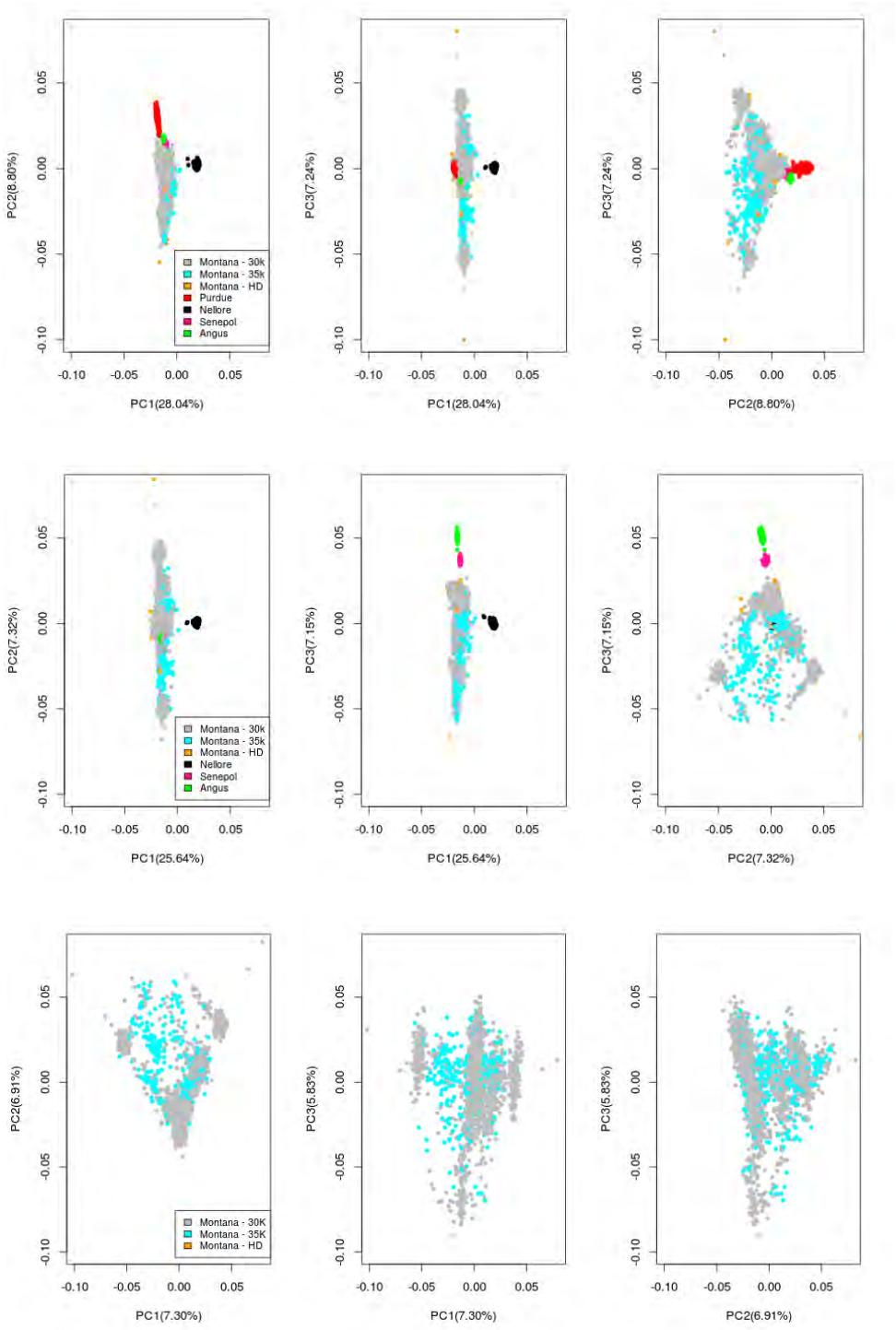


Figure S4.1 Principal components analysis plot for the first two PC of the genomic relationship matrix. The dots animals genotyped included the imputation in different scenario.

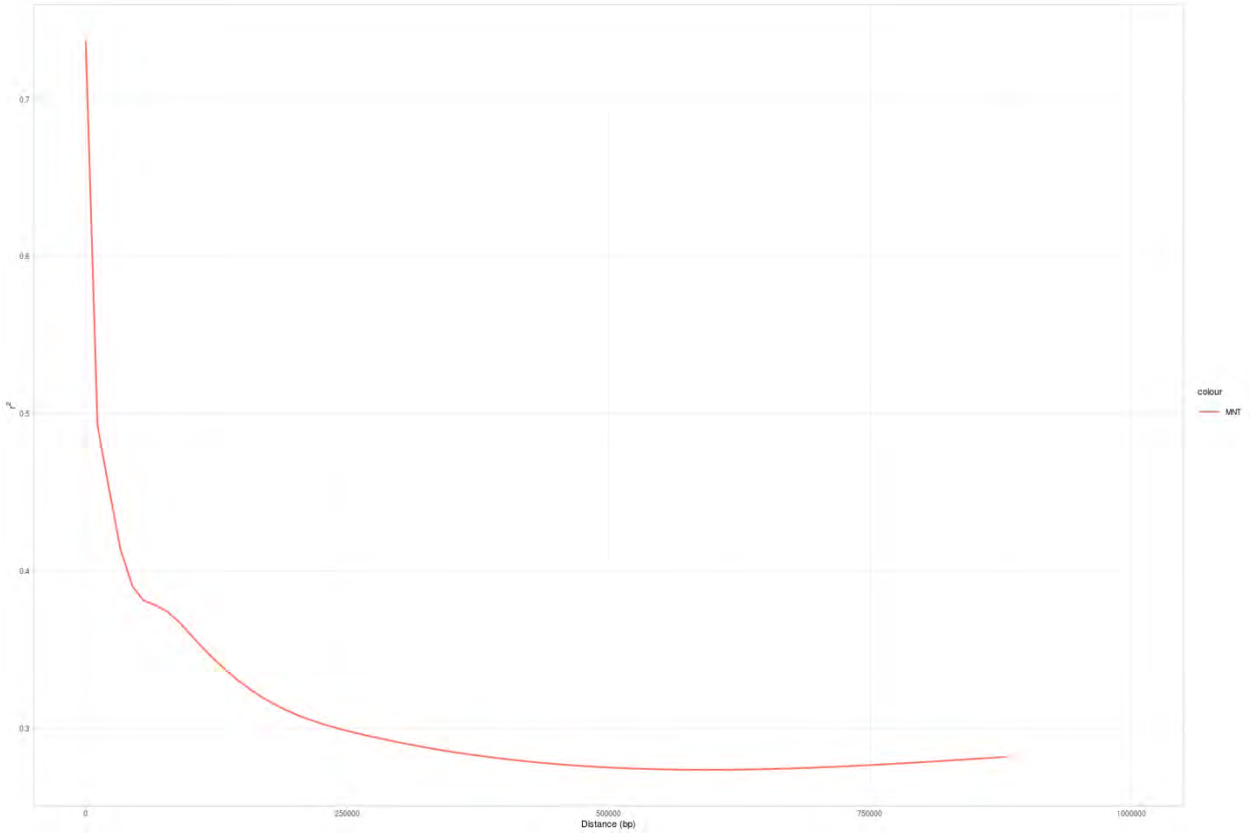


Figure S4.2 Linkage disequilibrium decay for the Montana Tropical Composite.