University of São Paulo
"Luiz de Queiroz" College of Agriculture

Genomic confirmatory factor analysis on milk fatty acid profile in dairy cattle reared in tropical conditions

Brayan Dias D'auria

Thesis presented to obtain the degree of Doctor in Science.
Area: Animal Science and Pastures

Piracicaba
2021

Brayan Dias D'auria
Animal Scientist

Genomic confirmatory factor analysis on milk fatty acid profile in dairy cattle reared in tropical conditions

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **GERSON BARRETO MOURÃO**

Thesis presented to obtain the degree of Doctor in Science.
Area: Animal Science and Pastures

Piracicaba
2021

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

D'auria, Brayan Dias

Genomic confirmatory factor analysis on milk fatty acid profile in dairy cattle reared in tropical conditions / Brayan Dias D'auria. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2021.

68 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Raça holandesa 2. Variável latente 3. Associção genômica ampla 4. Estresse térmico I. Título

DEDICATION

To my mom,
Lucimara de Oliveira Dias.

To my father,
Charles D`auria.

To them,
I *would like dedicate this work for all the support along my journey.*

ACKNOWLEDGMENTS

To my parents Charles D'auria and Lucimara de Oliveira Dias, and my brothers (Kinsya Muriel Dias D'auria and Dener Dias) for their love and support during all my life.

To all my family and friends for their support, love, patient, and motivation. In especial, I would like to thank my girlfriend (Paola Bóscollo) for her daily support, advice, and love.

To my advisor, professor Dr. Gerson Barreto Mourão, for his supervision, motivation and confidence during my doctorate, and also Barreto Mourão family for their friendship and support.

To my co-advisor, professor Dr. Francisco Peñagaricano (University of Florida – Gainesville), for receiving me on my intership abroad and for being so helpful, patient and supportive during my stay in Florida.

To members of my comitte, professor Dr. Luiz Lehmann Coutinho (USP/ESALQ) and professor Dr. Paulo Fernando Machado (USP/ESALQ), for their advice, support and suggestions. Additionaly, To thank professor Dr. Paulo Fernando Machado and staff from Clínica do leite – USP/ESALQ to provide the test-day milk yield from farms. To thank professor Dr. Luiz Lehmann Coutinho and members from Laboratório de Biotecnologia Animal – USP/ESALQ to provide personal access to the server for genetic analysis and support in genotyping.

To members of my qualification exam, professor Dr. Roberto Sartori Filho (USP/ESALQ), professor Dr. Júlio Cesar de Carvalho Balieiro (USP/FZEA) and professor Daniela Lino Lourenco (University of Georgia – UGA), for their suggestions, constructive criticism, and collaboration.

To all past and current especial collegues and friends from GEMA (Grupo de Estudos em Estatística e Melhoramento Animal USP/ESALQ), Daysi Orzuza, Eula Regina Carrara, Fabricio Pilonetto, Fátima Bogdanski, Giovanni Ladeira, Gregorí Rovadoscki, Izally Gervásio, Juliana Petrini, Luiz Losada, Mayara Salvian, Paola Bóscollo, Sofia Andrade, Vamilton Franzo for friendship, patient and collaboration.

To my collegues and friends from Quantitative Genetics and Genomics at University of Florida, Anil Sigdel, Hendyel Pacheco, Laila Talarico, Lihe Liu, Maria André, Rodrigo Teixeira and Rostam Abdolahii-Arpanahi for friendship and the good times.

To all professors and staff from departament of zootecnia USP/ESALQ, in special, professor Dr. Carla Maris Machado Bittar and professor Dr. Roberto Saroti Filho for their support and advices.

To the Government Funding Agency CNPq (process: 482327/2010-3, 484560/2012-3, and 433056/2016-9) and the Government Funding Agency CAPES for the fellowship financial support.

“The saddest aspect of life right now is that science gathers knowledge faster than society gathers wisdom.”

Isaac Asimov

SUMMARY

RESUMO

Análise fatorial confirmatória e genômica do perfil de ácidos graxos do leite em bovinos leiteiros criados em condições tropicais

O perfil de ácidos graxos do leite (AGL) bovino é um dos mais complexos e únicos entre mamíferos terrestres. A sua composição é extremamente variável e dependente de fatores como fonte da dieta, estado fisiológico do animal e uma fração atribuída ao componente genético. AGL são principalmente sintetizados pela *de novo* síntese que ocorre na glândula mamária ou derivados do processo de biohidrogrenação ruminal. Recentemente, esforços têm sido realizados na tentativa de elucidar os principais mecanismos associados às suas vias metabólicas. Além disso, o perfil de AGL tem sido investigado como biomarcador para o estresse térmico devido a grande influência da dieta na sua composição. Esses estudos são realizados na sua maioria pela análise de estimação de parâmetros genéticos, associação genômica ampla (GWAS) e de forma complementar com análise de enriquecimento. No entanto, pouco ainda tem sido explorado sobre abordagens que possam avaliar características quantitativas de forma multivariada. Neste contexto, foram conduzidos dois estudos utilizando estratégias e objetivos diferentes. Objetivou-se, no 1 estudo, estimar os parâmetros genéticos para AGL sob condições de estresse térmico (declínio) e sem estresse térmico (intercepto), posteriormente, comparar valores genômicos entre rankings (intercepto e declinio). Dados de produção de 7 AGL (AGL saturado - SAT, ALG insaturado - INSAT, ácido graxo monoinsaturado - MONO, AGL poliinsaturado - POLI, AGL palmítico - C16:0, AGL oleico - C18:0, e AGL esteárico - C18:1) foram utilizados nas análises genéticas. Os componentes de variância foram obtidos por meio de um modelo de repetibilidade com regressão aleatória de uma função de THI (índice de temperatura e umidade). AGL saturados (saturado e C16:0) apresentaram menores estimativas de herdabilidade sobre condições de estresse térmico. C18:1 apresentou maior sensibilidade ao calor em condições de estresse térmico. Correlacões entre os rankings de valores genéticos genomicos variaram entre -0.27 a 0.99. Nossos resultados demonstraram uma oportunidade para investigar novos biomarcadores e melhorar os processos seletivos para termotolerância. No experimento 2, objetivou-se ajustar variáveis latentes (variável não observável) e predizer *fator scores* para utilizá-las como fenótipo na estimação de parâmetros genéticos, GWAS e análise de enriquecimento. Os ajustes foram obtidos por meio de uma análise fatorial confirmatória (método multivariado) que tem como principal objetivo reduzir a dimensão dos dados. O conjunto de variáveis observáveis que obtiveram melhor ajuste no modelo foram SAT, POLI, C18:0 e C18:1. Essas variáveis foram combinadas em um fator que reprentou 3 estágios de lactação (início: 40-60; meio: 160-180; final: 250-270), baseados em intervalos fixos em dias em lactação (DEL) e posteriormente, foram combinadas para representar ordens de parto (1-3) utilizadas nas análises genéticas. As estimativas de herdabilidades para variáveis latentes foram de baixa magnitude (0.07 to 0.11). Dos resultados de GWAS, 11 genes candidatos (*PLD1*, *TM6SF2*, *NUDT7*, *LIPT1*, *AKPA1*, *APOH*, *RPGRIP1L*, *FTO*, *GMDS*, *ALDH3B1*, and *PC)* foram localizados em 9 cromossomos, incluindo na sua maioria genes que ainda não foram discutidos na literatura. Na análise de enriquecimento, foram revelados termos funcionais incluindo síntese de ácidos graxos, síntese do triglicerol e metabolismo de lipídeos e lipoproteínas. No geral, nosso estudo contribuiu como molde para novos estudos e para melhorar a base de conhecimento sobre os mecanismos genéticos subjacente à composição de AGL.

Palavras-chave: Bovinos leiteiros, Variável latente, Estresse térmico, Ácidos graxos

ABSTRACT

Genomic confirmatory factor analysis on milk fatty acid profile in dairy cattle reared in tropical conditions

The bovine milk fatty acid (FA) profile is one of the most complex and unique among terrestrial mammals. Its composition is extremely variable and depends on factors such as the source of the diet, the physiological state of the animal, and a fraction attributed to the genetic component. FA is mainly synthesized by the new synthesis that occurs in the mammary gland or derived from the ruminal biohydrogenation process. Recently, efforts have been made in an attempt to elucidate the main mechanisms associated with their metabolic pathways. In addition, the FA profile has been investigated as a biomarker for heat stress due to the great influence of diet on its composition. Most of these studies are carried out by analyzing genetic parameter estimation, genome-wide association study (GWAS), and in a complementary tool with enrichment analysis. However, little has been explored on approaches that can assess quantitative characteristics in a multivariate method. In this context, two studies were conducted using different strategies and objectives. The objective of this study was to estimate the genetic parameters for FA under conditions of thermal stress (decline) and without thermal stress (intercept), subsequently to compare genomic values between rankings (intercept and decline). Records of test-day milk of 7 FA (saturated FA - SFA, unsaturated FA - UFA, monounsaturated fatty acid - MUFA, polyunsaturated FA - PUFA, palmitic FA - C16:0, oleic FA - C18:0, and stearic FA - C18:1) were used in genetic analysis. The components of variance were obtained using a repeatability model with random regression of a THI function (temperature and humidity index). Saturated FA (saturated and C16:0) showed lower estimates of heritability under thermal stress conditions. C18:1 showed greater sensitivity to heat under conditions of thermal stress. Correlations between the rankings of genomic genetic values ranged from -0.27 to 0.99. Our results demonstrated an opportunity to investigate new biomarkers and improve selection processes for thermotolerance. In experiment 2, the objective was to adjust latent variables (unobservable variable) and predict factor scores to use them as a phenotype in the estimation of genetic parameters, GWAS, and enrichment analysis. The adjustments were obtained through confirmatory factor analysis (multivariate method) whose main objective is to reduce the size of the data. The set of observable variables that obtained the best fit in the model were SAT, POLI, C18: 0, and C18:1. These variables were combined into a factor that represented 3 stages of lactation (40-60; 160-180; 250-270), based on fixed intervals in days in milk (DIM) and later, were merged to represent lactation order (1 to 3) used in genetic analysis. The heritability estimates for latent variables were low (0.07 to 0.11). From the GWAS results, 11 candidate genes *(PLD1, TM6SF2, NUDT7, LIPT1, AKPA1, APOH, RPGRIP1L, FTO, GMDS, ALDH3B1*, and *PC*) were located on 9 chromosomes, mostly including genes that have not yet been discussed in the literature. In the enrichment analysis, functional terms were revealed including fatty acid synthesis, triacylglycerol synthesis, and lipid and lipoprotein metabolism. Overall, our study contributed as a design for further studies and to improve the knowledge base on the genetic mechanisms underlying the composition of FA.

Keywords: Dairy cattle, Latent variable, Heat stress, Fatty acids

1. INTRODUCTION

In dairy cows, recent efforts have been directed through elucidate genetic determinants related to physiological regulation, lipid metabolism and analyze the effects of milk fatty acids (FA) on human health (Palmquist, 2006). Milk FA derive from de novo synthesis in the mammary gland and from diet, biohydrogenation ruminal, or lipid mobilization (Palmquist, 2006). FA composition can be manipulated by diet and management (Rennó et al., 2013) and genetic selection (Bastin et al., 2012). However, it is necessary to emphasize that changes in composition could affect the physical and sensory properties of dairy products (Chiliard et al., 2000). Moreover, breeding goals can be challenging because of pattern phenotypic and genetic correlations among FA traits (Petrini et al., 2016). Bastin et al. (2012) reinforced that the directions of change in FA composition remain unclear and should be defined before including these traits in breeding programs.

Statistical method based on data reduction, such as confirmatory factor analysis (CFA), may be adopted to investigated the covariance structure of complex patterns among FA traits. CFA attempts to determine which sets of observed variables forward common variance and covariance traits that defined theoretical constructs or factors, namely latent variables. (Schumaker & Lomar, 2004). In dairy cattle, CFA has been used to study milk FA profile and milk composition (Conte et al., 2016; Mele et al., 2016). Moreover, genetic parameters of latent variables have been estimated and subsequently used in GWAS and pathways-based approach (Dadousis et al., 2017; Cecchinato et al., 2019; Palombo et al., 2020a, Palombo et al., 2020b). These studies indicate a relevant potential of using this tool in dairy cattle breeding.

There is growing evidence that there is a substantial increase of knowledge of the biological functions of milk fatty acids, a new appreciation for studies of genetic effects and nutritional strategies on milk fat composition (Bastin et al., 2012; Rennó et al., 2013). In addition, FA composition have used as biomarker for several metabolic patterns in animals (e.g., enteric methane emission, ketosis, acidosis, feed efficiency, lipid mobilization) (Cecchinato et al., 2019). They also have been suggested as one potential biomarker for heat stress in dairy cattle (Hammami et al., 2015; Nguyen et al., 2016).

In the context of biomarker function and modeling latent variables, this thesis has two main objectives: I) To determine if milk FA is a relevant biomarker to capture the effects of heat stress. II) To investigate the possibility to use latent variables to estimate genetic parameters and identify genomic regions and functional terms associated with milk FA traits.

In chapter 2, we measured genetic parameters under two conditions (heat stress and thermo-neutral). Repeatability test day models with random regressions on a function of temperature-humidity index (THI) values were used for genetic analyses. In addition, in this study we performed a comparation between rankings of GPTAs (genomic predicted transmitting abilitty) evalutead under two conditions (heat stress and thermo-neutral). We report remarkable differences between estimates of heritabilities. In general, saturated groups (SFA, C16:0 and C18:0) had lower estimates of heritability under heat stress conditions. In contrast, unsaturated groups had higher estimates of heritability under heat stress conditions (PUFA, UFA ,MUFA and C18:1). We suggested that this difference was due to the origin of milk FA, in which short-chain and medium-chain (saturated group) are less sensitive to environmental changes and derived mainly from de novo synthesis in the mammary gland. Briefly, saturated variables presented an antagonistic effect (negative correlation) between rankings correlation.

In chapter 3, we performed a study with latent variables related to stages and orders of lactations. Milk FA were fitted using CFA to construct latent variables and subsequently extracted factor scores. Genetic parameters were estimated using factor scores as phenotype. We carried out a GWAS and enrichment analysis for three lactations representing by latent variables. We report low estimates of heritability for latent variables. Our genomic study revealed some novel genomic regions that explained a small fraction of additive variance. In total, 11 putative genes were associated with a biological function in FA synthesis. We found significant functional terms of enrichment analysis related to lipid metabolism, FA synthesis, and triacylglycerol synthesis. Our results showed a relevant opportunity to design further studies and validated these putative genes. In chapter 4, we present the general conclusion of this thesis.

References

Bastin C., Soyeurt H., Gengler N. (2012) Genetic parameters of milk production traits and fatty acid contents in milk for Holstein cows in parity 1–3. J. ***Anim. Breed. Genet***. 130, 118–127.

Conte, G., A. Serra, P. Cremonesi, S. Chessa, B. Castiglioni, A. Cappucci, E. Bulleri, and M. Mele. (2016). Investigating mutual relationship among milk fatty acids by multivariate factor analysis in dairy cows. ***Livest. Sci.*** 188:124–132.

Cecchinato A, Macciotta NPP, Mele M et al (2019) Genetic and genomic analyses of latent variables related to the milk fatty acid profile, milk composition, and udder health in dairy cattle. ***J. Dairy. Sci.*** 102:5254– 5265. https://doi.org/10.3168/jds.2018-15867.

Chilliard Y., Ferlay A., Mansbridge R.M., Doreau M. (2000) Ruminant milk fat plasticity: nutritional control of saturated, polyunsaturated, trans and conjugated fatty acids**. Ann. Zootech.,** 49, 181–205.

Dadousis, C., C. Cipolat-Gotet, G. Bittante, and A. Cecchinato. (2017). Inferring genetic parameters on latent variables underlying milk yield and quality, protein composition, curd firmness and cheese-making traits in dairy cattle. ***Animal***. https://doi.org/10 .1017/S1751731117001616.

Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bas- tin, and N. Gengler. (2015). Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. ***J. Dairy Sci***. 98:4956–4968.

Mele, M., N. P. P. Macciotta, A. Cecchinato, G. Conte, S. Schiavon, and G. Bittante. (2016). Multivariate factor analysis of detailed milk fatty acid profile: Effects of dairy system, feeding, herd, parity, and stage of lactation. ***J. Dairy Sci***. 99:9820–9833. https://doi.org/10 .3168/jds.2016-11451.

Nguyen, T. T. T., Bowman, P. J., Haile-Mariam, M., Pryce, J. E., Hayes, B. J. (2016). Genomic selection for tolerance to heat stress in Australian dairy cattle. ***J. Dairy Sci***. 99, 2849–2862. doi: 10.3168/jds.2015-9685

Palmquist DL. (2006). Milk fat: Origin of fatty acids and influence of nutritional factors thereon. ***Advanced Dairy Chemistry: Lipids***. Edited by: Fox PF, McSweeney PLH, New York, NY: Springer, 2: 43-92.

Palombo, V., Conte, G., Mele, M., Macciotta, N. P. P., Stefanon, B., Marsan, P. A., & D'Andrea, M. (2020). Use of multivariate factor analysis of detailed milk fatty acid profile to perform a genome-wide association study in Italian Simmental and Italian Holstein. ***J. App. Genet***. 61(3), 451–463. https://doi.org/10.1007/s1335 3-020-00568-2.

Palombo, V., Milanesi, M., Sgorlon, S., Capomaccio, S., Mele, M., Nicolazzi, E., Ajmone-Marsan, P., Pilla, F., Stefanon, B., & D'Andrea, M. (2018). Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. ***J. Dairy Sci***. 101(12), 11004–11019. https://doi.org/10.3168/ jds.2018-14413.

Petrini J, Iung LHS, Rodriguez MAP, Salvian M, Pértille F, Rovadoscki GA, Cassoli LD, Coutinho LL, Machado PF, Wiggans GR and Mourão GB. (2016). Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. ***J. Anim. Breed. Genet***. 133, 384–395. doi:10.1111/jbg.12205.

Rennó, F. P, Freitas J, José E, Gandra, J. R, Verdurico, L. C, Santos, M. V. B, R V, Venturelli, B. C, & Vilela, F. G. (2013). Fatty acid profile and composition of milk protein fraction in dairy cows fed long-chain unsaturated fatty acids during the transition period. **Revista Brasileira de Zootecnia,** 42(11), 813-823. https://doi.org/10.1590/S1516-35982013001100008

Schumacher, R. E., & Lomax, R. G. (2004). *A Beginner's Guide to Structural Equation Modeling. Mahwah, New Jersey: Lawrence Erlbaum Associate.*

2. ASSESSING GENETIC COMPONENTS OF MILK FATTY ACIDS TRAITS UNDER HEAT STRESS IN A BRAZILIAN HOLSTEIN POPULATION

Brayan Dias Dauria[1], Anil Sigdel[2], Francisco Peñagaricano[2], Fabricio Pilonetto[1], Juliana Petrini[1], Mayara Salvian[1], Paola Perez Bóscollo[1], Paulo Fernando Machado[1], Luiz Lehman Coutinho[1], & Gerson Barreto Mourão[1*]

1 Department of Animal Science, University of São Paulo (USP) / Luiz de Queiroz College of Agriculture (ESALQ), Piracicaba, SP, Brazil
2 Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

Abstract

The present study aimed to estimate covariance components of milk fatty acids (FA) and to compare the genomic-predicted genetic values (GPTA) under general and heat stress effects. Data consisted of 38,762 test-day records from 6,344 Holstein cows obtained from May 2012 through January 2018 on four dairy herds from Brazil. Single-trait repeatability test day models with random regressions as a function of temperature-humidity index (THI) values were used for genetic analyses. The models included contemporary groups, parity order (1-6), and DIM classes as fixed effects, permanent environmental and general additive and heat stress additive genetic as random effects. Notably, heritability estimates increased (0.03 to 0.06) for unsaturated FA traits (UFA = unsaturated; MUFA = monounsaturated; PUFA = polyunsaturated; and C18:1 = oleic acid) at heat stress levels. In contrast, heritability estimated for saturated FA traits (SFA = saturated; C16:0 = palmitic acid; and C18:0 = stearic acid) decreased (-0.01 to -0.04) under heat stress conditions. In addition, our study revealed negative genetic correlation between general and heat stress additive genetic effects (antagonistic effect) for SFA, C16:0, C18:0, and C18:1 ranged from -0.007 to -0.32. Estimation of genetic correlation within traits was generally low to moderated (<0.80), which suggests animals re-ranking at heat stress conditions. The lowest genetic correlation between additive effects was observed for C18:1 (-0.07), suggesting being more sensitivity to heat stress. Spearman's ranking correlation between GPTAs ranged from -0.27 to 0.99. Results indicated that most FA traits are affecting the selection response if practiced in a limiting environment. Our findings point out novel opportunities to explore the use of FA milk profile as a potential biomarker of heat stress in dairy cattle.

Key words: temperature-humidity index, heat tolerance, variance component, milk fatty acid

2.1 Introduction

In tropical environment, climatic factors could be challenging for dairy farming, where climate is characterized by high temperature, humidity, and rainfall. Under such climatic conditions, high-producing dairy cows experience heat stress and show decline in health, fertility, and production traits. Heat stress is a condition in which animal is not able to dissipate endogenous or exogenous heat from its body (Bernabucci et., 2010). Heat stress is an important economic issue in dairy farming. In Brazil, there are no economic loss estimates due to heat

stress found in literature. However, this effect could be a relevant factor, mainly because most of the high-producing dairy cows (purebred) are not adapted genetically to tropical conditions. In the US dairy industry, the economic losses due to heat stress are estimated to be between $897 million to $1,500 million per year (St-Pierre et al., 2003).

As heat stress is a costly problem for dairy producers, different strategies such as environmental conditioning of the warehouses using fans, sprinklers, shades, and improved nutritional and management practices have been used to alleviate the effects of heat stress (Moran, 2005). However, these practices increase production costs, and, in general, cannot eliminate heat stress completely. Santana Jr et al. (2016) reinforced the hypothesis that heat stress reduces milk production and quality of Brazilian Holstein herds even maintained in modified physical environments. One complementary strategy for reducing the effects of heat stress on dairy cattle performance is the identification and subsequent selection of animals that are genetically more thermotolerant. Selecting animals that are more thermo-tolerant is the most cost-effective approach to raise dairy purebred under tropical conditions as the gains made through genetic selection are cumulative and permanent (Wall et al., 2010).

The main challenges in genetic analyses of heat stress are accurate selection of phenotypes and the choice of models for variance components and parameter estimation to quantify the level of heat stress. Several studies have used physiological traits such as rectal temperature, respiration rate and intra-vaginal temperature as indicator traits of heat stress (Kendal et al., 2007; Dikmen and Hansen, 2009; Kaufman et al., 2018). However, collection of such data at national level is both logistically challenging and time consuming. Regarding the choice of heat stress variable indicator, Ravagnolo et al. (2000) examined several functions for quantifying heat stress and suggested that temperature-humidity index (THI) was a good environmental indicator. Moreover, THI has been used for decades to measures heat stress effects into genetic models by several studies related to dairy cattle (Ravagnolo & Misztal, 2000; Aguilar et al., 2009; Hammami et al., 2015; Nguyen et al., 2016; Sigdel et al., 2019).

Most of the genetic studies in dairy cattle are interested in estimated a fixed threshold level of THI for production and quality traits (Aguilar et al., 2009; Bernabucci et al., 2014; Hammami et al., 2015). These models usually assumed that production declines linearly with increasing of THI, and the declines can be attributed to genetic effects. In addition, Sánchez et al. (2009) proposed a more complex model that assumed each cow had different threshold and slope based on a hierarchical Bayes model. However, this model could be computationally challenging as it has a large number of parameters. Therefore, a repeatability test-day model

with constant threshold and random slope for heat load function could result in reasonable estimates of genetic parameters under heat stress (Aguilar et al., 2009; Nguyen et al, 2016).

Recently, there has been a considerable interest in milk fatty acid profile because of its effects on human health, technological, sensorial, and nutritional properties of milk and dairy products (Hanus et al., 2018). Hammami et al. (2015) have assessed the potential of milk fatty acid profile as biomarkers of thermotolerance in Holstein population from Belgium. Also, Nguyen et al. (2016) suggested that fatty acid profile could be a potential biomarker for heat tolerance as it can be measured on a large scale through mid-infrared spectroscopy of milk samples and hence can be applied to datasets as large as those utilized for the national evaluation. Thus, the first objective of this study was to estimate covariance components of FA as a function of THI values through single-trait repeatability test-day model. The second objective was to estimate GPTA under general and heat stress conditions for milk FA and subsequently to compare the Spearman ranking correlations between sires.

2.2 Material and Methods

2.2.1. Phenotypic and genotypic data

Data consisted of 38,762 FA records from 6,344 Holstein cows obtained from 2012 through 2018 on four dairy herds in Brazil (Table 1). Cows were from first through sixth parity and were daughters of 535 sires, with days in milk (DIM) between 5 and 305. Test-day records ranged from 1 to 29, with an average milk frequency of three times a day. Contemporary group (CG) was formed by the combination of herd, calving year, and month of test-day record. Only records within the acceptable range of three standard deviations from the respective mean and from CG with a minimum of five animals were used in the genetic analyses. Pedigree was created by tracing the pedigrees of cows back to five generations. The pedigree file included 9,759 animals.

In this study, we considered seven (7) traits related to FA milk profile: saturated **(SFA)**, unsaturated **(UFA)**, monounsaturated **(MUFA)**, and polyunsaturated **(PUFA)** fatty acids groups; and the individual fatty acids: palmitic acid **(C16:0),** stearic acid **(C18:0),** and oleic acid **(C18:1)**. FA were determined as grams per 100 g of milk and measured by mid-infrared spectroscopy (*Delta Instruments CombiScope Filter, Advanceed Instruments Inc., Norwood*). The key ingredients in cows' diet include corn silage, grass hay, cottonseed, soybean meal, soybean husk, cornmeal, citrus pulp, minerals, and vitamins.

Genotype data for 72,444 SNP markers across the bovine genome were available for 1,152 cows with records, daughters from 165 sires in the pedigree. SNP information was updated to the new bovine reference genome ARS-UCD 1.2. A detailed description of the procedure of imputation analysis for cow's reference population have been reported by Petrini et al. (2016) and Iung et al. (2019). The SNP that mapped to sex chromosomes, were monomorphic, or had minor allele frequency below 1% were excluded from the genotype data. After quality control, a total of 928 cows and 67,471 SNP markers were retained for subsequent genomic analysis.

2.2.2. *Climate Data Information*

Weather data were obtained from NASA Prediction of Worldwide Energy Resource (POWER, 2017), using the farm's location's coordinates: latitude and longitude. Hourly THI (temperature-humidity index) values were calculated as proposed by Ravagnolo et al. (2000):

$$\mathbf{THI = (1.8 \times T + 32) - (0.55 - (0.0055 \times RH) \times (1.8 \times T - 26)}$$

where **T** is the average of temperature in degree ℃ and **RH** is the average of relative humidity, expressed as a percentage. After that, mean daily THI corresponding three days prior each test day was calculated as suggested by Bohmanova et al. (2007). A heat load function, denoted as $\mathbf{f(THI)}$, was calculated to estimate the decrease (slope) in the content of FA in milk under heat stress, as follows:

$$\mathbf{f(THI)} = \begin{cases} 0 & \textbf{if THI} \leq \mathbf{THI_{thr}} \\ \mathbf{THI} - \mathbf{THI_{thr}} & \textbf{if THI} > \mathbf{THI_{thr}} \end{cases}$$

where $\mathbf{THI_{thr}}$ (THI - threshold) was set to 68, and thus $\mathbf{f(THI)}$ was equal to max $(0, \mathbf{THI} - \mathbf{THI_{thr}})$ (Sigdel et al., 2019).

2.2.3. *Genetic analyses*

Single-trait repeatability test-day model proposed by Ravagnolo and Misztal (2000) were fitted to estimate variance components for FA in milk under general and heat stress conditions. Additionally, genomic predicted transmitting ability (GPTA) were

estimated for two environments: general (heat load THI function = 0) and heat stress (heat load THI function The variances are:

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{t} \\ \mathbf{pe} \\ \mathbf{q} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma}^2_{\mathbf{a}} & \boldsymbol{\sigma}_{\mathbf{at}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\sigma}_{\mathbf{at}} & \boldsymbol{\sigma}^2_{\mathbf{t}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma}^2_{\mathbf{pe}} & \boldsymbol{\sigma}_{\mathbf{peq}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma}_{\mathbf{peq}} & \boldsymbol{\sigma}^2_{\mathbf{q}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma}^2_{\mathbf{e}} \end{bmatrix}$$

Where was assumed $\mathbf{a} = [\mathbf{a}'_{\mathbf{n}}\mathbf{t}'_{\mathbf{n}}]$ be a vector of random additive genetic effects and $\mathbf{pe} = [\mathbf{pe}'_{\mathbf{n}}\mathbf{q}'_{\mathbf{n}}]$ be a vector of random permanent effects.

= 10 degrees above to threshold), afterward, sire rankings were compared through Pearson's correlations.

$$\mathbf{y_{ijklm}} = \mathbf{CG_i} + \mathbf{PAR_j} + \mathbf{DIM_k} + \mathbf{a_l} + \mathbf{pe_l} + \mathbf{t_l}[\mathbf{f(THI)}] + \mathbf{q_l}[\mathbf{f(THI)}] + \mathbf{e_{ijklm}}$$

where $\mathbf{y}_{\text{klmn}}$ is the record for the fatty acid milk traits, $\mathbf{CG}_i$ is i$^{\text{th}}$ contemporary group (herd, calving year, and month of test-day record) (i = 1 to 142), $\mathbf{PAR}_{\text{j}}$ is j$^{\text{th}}$ parities (j = 1-6), $\mathbf{DIM}_{\text{k}}$ is the k$^{\text{th}}$ DIM class with classes defined every 20 days (k = 16), $\mathbf{a}_{\text{l}}$ is the general random additive genetic effect (intercept) of animal l, $\mathbf{pe}_{\text{l}}$ is the general random permanent environmental effect (intercept) of animal l, $[\mathbf{f(THI)}]$ is a function of THI, $\mathbf{t}_{\text{l}}$ is the random additive genetic effect (slope) of heat stress of the animal l, $\mathbf{q}_{\text{l}}$ is the random permanent environmental effect (slope) of heat stress of animal l, $\mathbf{e_{ijklm}}$ is the random residual effect.

For all analysis it was assumed a genomic polygenic model a~N $(0, \mathbf{H}\sigma_a^2)$, where σ_a^2 is the additive genetic variance, $\mathbf{H}$ is the combined relationship matrix (pedigree and genomic information) (Aguilar et al., 2010; Christensen & Lund, 2010). This method is known as single-step genomic best linear unbiased prediction (ssGBLUP). The inverse of $\mathbf{H}$ was obtained as follows,

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G_w^{-1}} - \mathbf{A_{22}^{-1}} \end{bmatrix}$$

Where $\mathbf{A^{-1}}$ is the inverse of the pedigree relationship matrix, $\mathbf{G_w^{-1}}$ is the inverse of the genomic relationship matrix, and A_{22}^{-1} is the inverse of the pedigree relationship matrix of the animals

genotyped. $\mathbf{G_w} = \alpha\mathbf{G} + \beta\mathbf{A_{22}}$, where **G** was equal to the genomic relationship according to VanRaden (2008). w is the amount the proportion of the total additive genetic variance by genetic marker. This step was performed using the default parameterization in the preGSf90 ($\alpha = 0.95$ and $\beta = 0.05$).

Variance components for FA milk traits were estimated in a Frequentist framework using the restricted maximum likelihood method under genomic polygenic models in AIREMLF90 software (Misztal et al., 2015). The genomic heritability coefficients $\mathbf{(h^2)}$ were obtained at heat stress level $\mathbf{f(THI)} = 10$ and general effects $\mathbf{f(THI)} = 0$ (Ravagnolo & Misztal, 2000):

$$\mathbf{h^2} = \frac{\boldsymbol{\sigma_a^2} + \mathbf{f(THI)^2}\boldsymbol{\sigma_t^2} + \mathbf{2\ f(THI)}\ \boldsymbol{\sigma_{at}}}{\boldsymbol{\sigma_a^2} + \mathbf{f(THI)^2}\boldsymbol{\sigma_t^2} + \mathbf{2\ f(THI)}\ \boldsymbol{\sigma_{at}} + \boldsymbol{\sigma_{pe}^2} + \mathbf{f(THI)}\ ^2\boldsymbol{\sigma_q^2} + \mathbf{2f(THI)}\ ^2\boldsymbol{\sigma_{peq}} + \boldsymbol{\sigma_e^2}}$$

where $\boldsymbol{\sigma_a^2}$ the variance of general additive genetic effects; $\boldsymbol{\sigma_t^2}$ is the variance of thermotolerance additive genetic effects; $\boldsymbol{\sigma_{at}}$ is the additive genetic covariance among general and thermotolerance genetic effects; $\boldsymbol{\sigma_{pe}^2}$ is the variance of general enviromental permanent effects; $\boldsymbol{\sigma_q^2}$ is the variance of thermotolerance environmental permanent effects; $\boldsymbol{\sigma_{peq}}$ is environmental permanent covariance among TNZ and HS effects; $\mathbf{f(THI)}$ is a function of THI, and $\boldsymbol{\sigma_e^2}$ is the residual variance.

The genetic correlation within trait among general and heat stress additive genetic variances was estimated as:

$$\mathbf{corr}\ [\mathbf{a}, \mathbf{f(THI)t}] = \frac{\mathbf{f(THI)}\boldsymbol{\sigma_{at}}}{\sqrt{\boldsymbol{\sigma_a^2}}\mathbf{f(THI)^2}\boldsymbol{\sigma_t^2}}$$

The GPTA was predicted using BLUPF90 program (Misztal et al., 2015) with a convergence criterion of 10^{-12}. Spearman rank correlation was performed to compare the classification of animals between general and heat stress conditions using the software Rstudio (R Development Core Team 2019). For sires ranking, was selected only sires with a minimum

of twenty daughters reared at least in more than one CG, and split into three subsets (TOP25%, TOP60%, and FULL_RANK).

2.3. Results and Discussion

2.3.1. *Genetic parameter estimation*

Variance components for FA milk traits under general and heat stress conditions were estimated using single-trait repeatability test day models (Table 2). The genetic parameters included genomic heritability estimates and genetic correlation calculated within trait at general (intercept) and at heat stress level equal to $\mathbf{f(THI)}$ (slope). The heat stress level was fixed to 10 THI degrees above the THI threshold (THI = 68 degrees) to obtain larger variances (100 times) (Ravagnolo & Misztal, 2000; Aguilar et al., 2010; Sigdel et al., 2019).

Our studies reveal that there is a remarkable difference in genomic heritability estimates for all FA traits under general and heat stress conditions. Interestingly, unsaturated FA traits (PUFA, MUFA, UFA, and C18:1) showed higher heritabilities estimates under heat stress (0.16, 0.09, 0.08, 0.08, respectively) compared with general conditions (0.10, 0.05, 0.05, and 0.05, respectively). Our findings are similar to those reported by Hammami et al. (2015), who indicated that PUFA and C18:1 also had higher heritabilities under high THI values in Belgium Holstein cows using the linear reaction norm model. Moreover, results reinforced that mostly unsaturated FA traits were lowest heritable than saturated FA traits in both conditions. Other studies (Bastin et al., 2012; Penasa et al., 2014; Petrini et al., 2016) showed that PUFA, MUFA, UFA, and C18:1 had lower heritabilities estimated compared than saturated FA traits. Differences in heritability estimated ranged from 0.03 to 0.06 in the proportion of estimation in unsaturated FA traits under heat stress.

In contrast, saturated FA traits (SFA, C16:0, and C18:0) had lower heritabilities estimates under heat stress (0.23, 0.23, and 0.13, respectively) compared to general conditions

(0.26, 0.27, and 0.14, respectively). In agreement, Hammami et al. (2015) suggested that heritability estimates decreased slightly at high THI for most FA traits with higher heritability. As far as we know, milk FA profiles derived from de novo (short- and medium-chain FA = saturated) are mostly heritable than milk FA profiles obtained from ruminal biohydrogenation on diet and body fat stores (long-chain FA = unsaturated) (Penasa et al., 2014). Despite this, very few studies have been devoted to compare changes in heritability estimation for milk FA profile considering heat stress effects in genetic models. Other reports have focused on the effects of lactation stage, parities, and seasons of year associated with variability in genetic and phenotypic components for FA milk profiles (Bastin et al., 2012; Renna et al., 2010). In general, differences in variances components showed changes in estimation during heat stress for all FA traits that indicated the possibility to investigate the phenotypes with greater sensitivity under heat stress effects.

Genetic correlation between additive genetic effects between general and heat stress conditions were negative and low for SFA, C16:0, C18:0, and C18:1, ranging from -0.007 to -0.32. Negative genetic correlations between general and heat stress additive genetic effects also reported for SFA and C16:0 in primiparous Holstein cows (Hammami et al., 2015). Interestingly, other studies have also reported negative genetic correlation between general and heat stress additive genetic effects for milk yield and composition traits such as protein and fat for high-producing dairy cows (Ravagnolo & Misztal, 2000; Aguilar et al., 2009; Bernabucci et al., 2014; Hammami et al., 2015, Sigdel et al., 2019). These studies suggest that milk yield traits are antagonistic to heat tolerance, and selection for higher yield without considering heat tolerance may results in greater susceptibility to heat stress.

The antagonistic effect is related to physiological change mechanisms and a positive correlation between milk production and metabolic heat production. This effect leads to blood insulin concentrations increase, glucose concentrations decrease, feed intake reduced (negative

energy balance), limitation in lactose synthesis, and hence, milk yield declines (Rhoads et al., 2009; Baumgard et a., 2015). In saturated FA milk traits (SFA, C16:0, and C18:0), it probably occurs because heat stress modifies the metabolic strategies of the uses of body resources such as fat, protein, and energy (Slimen et al., 2016). Herein, C18:1 showed a genetic correlation between additive genetic variances close to 0, which means a genetic relationship almost null compared to two different environmental conditions. These results are in accordance with previously reported findings on the use of C18:1 as a biomarker for heat stress (Hammami et al., 2015).

Positive genetic correlations between general and heat stress additive genetic effects were observed for UFA, MUFA, and PUFA, ranging from low to high genetic relationship (0.12 to 0.72). This is in agreement with the findings of Hammami et al. (2015) in Belgium Holsteins, who reported positive genetic correlations between general and heat stress additive genetic effects for UFA, MUFA, and PUFA (0.03 to 0.38). A high positive correlation found for PUFA (0.72) suggests that the effects attributed to genes between general and heat stress conditions possibly mostly are the same associated with their genetic architecture. In addition, PUFA represents a low percentage (5%) of the bovine milk fat and small variability associated with their composition (Penasa et al., 2015), which indicates be a weak candidate for capture heat stress effects from samples of milk. In general, genetic correlations within traits allowed to identify the genetic behavior of milk FA profile under heat stress conditions. Thus, our findings provide further evidence of the interaction between genotype and environment along with the low and negative genetic correlation between saturated FA and C18:1 unsaturated FA with heat stress environment.

2.3.2. *Genetic evaluation for heat tolerance*

Table 4. shows the ranking correlation between GPTAs of 97 sires under general and heat stress conditions split into three subsets (TOP20%, TOP50%, and FULL_RANK).

Negative ranking correlations between GPTAs ranged from -0.12 to -0.46 for SFA and C16:0; positive ranking correlations between GPTAs ranged from 0.28 to 0.99 for UFA, MUFA, PUFA, and C18:1. For breeding schemes, negative rankings correlation indicates re-ranking of sires and differences in performance for these traits under two conditions of the environment. It was expected for saturated FA because there was an antagonistic relationship between general and heat stress additive genetic variances.

Interestingly, PUFA presented a strong positive ranking correlation between general and heat stress conditions (0.97 to 0.99), suggesting no re-ranking of sires. Moreover, PUFA had the highest positive genetic correlation (0.74) in the parameter estimation within traits observed previously. Rankings results indicate that PUFA would have few changes in GPTAs for both environmental conditions proposed in the present study. It reinforced our hypothesis that PUFA probably is not a good indicator for heat stress.

In general, differences between rankings are reported for low genetic correlations (<0.80), which means that the top sires on a trait in one environment are not necessarily as superior in the other (Hammami et al., 2008). Our results indicate that UFA, MUFA, and C18:1 had ranking sires' correlations lower than 0.55. Therefore, all those traits also are expected differences related to GPTAs with the environment of genetic evaluation. Furthermore, rankings correlations (negative or positive) also suggest possible differences between contents of FA milk profiles. It was observed by Renna et al. (2010), which argues that higher concentrations of saturated FA occur during heat stress conditions in alpine grazing systems. Also, Hammami et al. (2015) reported that heat stress influences lipid synthesis by the mammary gland that alters the content of saturated and unsaturated FA in milk. Diet composition also alters the FA milk profile, particularly for C16:0 and C18:1 when sources of fats are included in the feed (Palmquist, 2006). Therefore, we reinforce that the FA of milk interacts with the genotype-environment.

Our ranking correlation results indicate a re-ranking of sires for mostly FA traits including SFA, UFA, MUFA, C16:0, and C18:1 attributed to low to the moderate genetic correlation between GPTAs under general and heat stress conditions. In genetic aspects, all those traits can affect the selection response if practiced in a limiting environment. Considering full rankings, C18:1 presented the lowest ranking Spearman correlation (0.32), which agrees with previous results from the genetic correlation between additive genetic variances. The present results support the idea that C18:1 is influenced strongly by heat stress, thus showing that environmental conditions may affect their synthesis. Bastin et al. (2011) showed a high range between C18:1 and other FA during the first 100 days in milk for Wallon Holstein cows. The authors reinforced that release of long-chain FA inhibits FA synthesis in the mammary gland while the cow is in negative energy balance. On this basis, Moore et al. (2005) suggest that cows under heat stress conditions became a state of negative energy balance, independently of the lactation stage, which would compromise the milk yield and components. The present study provides further evidence that C18:1 can be a candidate milk biomarker for heat stress in dairy cattle, highlighting the importance of combine records from mid-infrared spectrometry and climate data to be used for heat-stress management or development of new tools for analyzing the samples directly from milk.

2.4. Conclusions

The FA profile in milk changes continuously throughout general and heat stress conditions and these changes can be determined genetically. Interestingly, unsaturated FA have higher heritability estimates under heat stress conditions. The antagonistic relationship between additive genetic variances under general and heat stress conditions indicate that saturated FA could have higher concentrations during heat-stress environment. High ranking correlation at heat stress conditions suggests no genotype by environment interaction for PUFA. However,

for C18:1, larger genetic variation and lower genetic correlation indicate that this trait has the greatest sensitivity to heat stress conditions in a tropical climate. These findings could contribute to a better understanding of interaction of milk FA complex traits with the environment under heat stress conditions in dairy cattle. Also, milk FA can be used as a heat stress biomarker or even for the management of the heat stress conditions open new possibilities for further studies.

ACKNOWLEDGEMENTS

This research received the financial support of National Council for Scientific Technological Development (CNPq) (grants number: 553966/2009-0, 484560/2012-3), Research Foundation (FAPESP) (grants number: 2010/12929-6), and Coordination for the Improvement of Higher Education Personnel (CAPES). The authors thank *Clínica do Leite* and Department of Animal Science, 'Luiz de Queiroz' College of Agriculture, University of São Paulo (ESALQ/USP) for assistance and for providing the database, he Biotechnology Laboratory (ESALQ) for their support in the DNA extraction and genotyping; and the Cooperative Dairy DNA Repository (CDDR) for providing the genotypes used as reference population in imputation analysis.

Table 1. Descriptive analysis of fatty acids of milk (g/100g/milk).

Traits	N	Mean	SD	Min	Max
FA groups					
SFA	38,762	2.30	0.56	0.71	6.00
UFA	38,762	1.03	0.31	0.08	3.08
MUFA	38,762	0.87	0.27	0.08	2.66
PUFA	38,762	0.15	0.05	0.01	0.50
Individual FA					
C16:0	38,762	0.87	0.24	0.17	2.49
C18:0	38,762	0.62	0.16	0.50	1.94
C18:1	38,762	0.66	0.23	0.02	2.00

N = number of observations; SD = standard deviation; Min = minimum; Max = maximum; FA = fatty acids of milk; SFA = saturated, UFA = unsaturated, MUFA = monounsaturated, PUFA = polyunsaturated, C16:0, palmitic, C18:0 = stearic, C18:1 = oleic.

Table 2. Variance component estimates, genomic heritability, and correlations between general and heat stress additive effects for FA traits.

FA traits	σ^2_a	σ^2_t	$\sigma_{a,t}$	σ^2_{pe}	σ^2_q	$\sigma_{pe,q}$	σ^2_e	h^2 (t)	h^2 (a)	r^g (a, t)
FA groups										
SFA	0.07668	0.000066	-0.00063	0.0292	0.00053	-0.0015	0.1825	0.23 (0.026)	0.26 (0.017)	-0.28 (0.39)
UFA	0.00539	0.00421	0.000082	0.00013	0.000016	-0.00040	0.0706	0.08 (0.020)	0.05 (0.008)	0.31 (0.78)
MUFA	0.00372	0.00303	0.000036	0.00011	0.000027	-0.00027	0.0528	0.09 (0.027)	0.05 (0.007)	0.12 (0.67)
PUFA	0.00013	0.00019	0.000006	0.000001	0.0000003	-0.000008	0.0016	0.16 (0.01)	0.10 (0.01)	0.74 (0.04)
Individual FA										
C16:0	0.00470	0.01392	-0.000125	0.000092	0.000010	-0.00020	0.0309	0.23 (0.022)	0.27 (0.018)	-0.32 (0.24)
C18:0	0.00186	0.00310	-0.000016	0.000023	0.000003	-0.000086	0.0170	0.13 (0.021)	0.14 (0.014)	-0.15 (0.27)
C18:1	0.00298	0.00209	-0.000001	0.000065	0.000016	-0.000197	0.0349	0.08 (0.018)	0.05 (0.008)	-0.007 (0.28)

Table 3. Spearman correlation between genomic-predicted genetic values (GPTA) of bulls under general (a) and heat stress (t).

Ranking	Bulls		
	TOP20%	TOP50%	FULL_RANK
FA traits	GPTA (a,t)	GPTA (a,t)	GPTA (a,t)
SFA	-0.40	-0.27	-0.38
UFA	0.52	0.28	0.38
MUFA	0.50	0.36	0.39
PUFA	0.97	0.98	0.99
C16:0	-0.40	-0.45	-0.46
C18:0*	-0.18*	-0.12*	-0.13*
C18:1	0.34	0.31	0.32

*: no significance for p-value ≤ 0.05; SFA = saturated; UFA = unsaturated; MUFA = monounsaturated; PUFA = polyunsaturated; C16:0 = palmitic acid; C18:0 = stearic acid; and C18:1 = oleic acid; TOP20% = 20% of best classification sires; TOP50% = 50% of best classification sires; FULL_RANK = All sires presented at ranking

References

Aguilar, I., Misztal, I., and Tsuruta, S. (2009). Genetic components of heat stress for dairy cattle with multiple lactations. *J. Dairy Sci*. 92, 5702–5711. doi: 10.3168/jds.2008-1928

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci*. 93:743–752. doi: 10.3168/jds.2009-2730

Atrian, P and Shahryar, HA. (2012). Heat stress in dairy cows (a review). *Research in Zoology.* 2, 31–37. doi: 10.5923/j.zoology.20120204.03

Bastin, C., N. Gengler, and H. Soyeurt. (2011). Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. *J. Dairy Sci*. 94:4152–4163.

Baumgard L.H., Keating A., Ross J.W., and Rhoads R.P. (2015). Effects of heat stress on the immune system, metabolism and nutrient partitioning: implications on reproductive success. *Rev. Bras. Reprod. Anim*., 39: 173–183. id: vti-12848

Bernabucci, U., N. Lacetera, L. H. Baumgard, R. P. Rhoads, B. Ron-chi, and A. Nardone. (2010). Metabolic and hormonal acclimation to heat stress in domesticated ruminants. *Animal*. 4:1167–1183. doi: 10.1017/S175173111000090X

Bernabucci, U., Biffani, S., Buggiotti, L., Vitali, A., Lacetera, N., and Nardone, A. (2014). The effects of heat stress in Italian Holstein dairy cattle. *J. Dairy Sci.* 97, 471–486. doi: 10.3168/jds.2013-6611

Bohmanova, J., Misztal, I., and Cole, J. B. (2007). Temperature-humidity indices as indicators of milk production losses due to heat stress. *J. Dairy Sci.* 90, 1947–1956. doi: 10.3168/jds.2006-513

Christensen, O.F., Lund, M.S. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42, 2 (2010). doi: 10.1186/1297-9686-42-2

Dikmen, S. E. R. D. A. L., and Hansen, P. J. (2009). Is the temperature-humidity index the best indicator of heat stress in lactating dairy cows in a subtropical environment?. *J. Dairy Sci.* 92(1), 109-116. doi: 10.3168/jds.2008-1370

Freitas, M. S., I. Misztal, J. Bohmanova, and J. West. (2006). Utility of on- and off-farm weather records for studies in genetics of heat tolerance. *Livest. Sci*. 105:223–228. doi: 10.1016/j.livsci.2006.06.011

Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bastin, and N. Gengler. (2015). Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. *J. Dairy Sci*. 98:4956–4968. doi: 10.3168/jds.2014-9148

Hanuš, O., Samková, E, Křížová, L, Hasoňová, L, and Kala, R. (2018). Role of fatty acids in milk fat and the influence of selected factors on their variability—a review. *Molecules*, 23(7), 1636. doi: 10.3390/molecules23071636

Iung L. H. S., Petrini J, Ramírez-Díaz J, Salvian M, Rovadoscki G. A, Pilonetto F, Dauria B. D, Machado P. F, Coutinho L. L, Wiggans GR, and Mourão G. B. (2019). Genome-wide association study for milk production traits in a Brazilian Holstein population. *J. Dairy Sci*. 1:1–10. doi: 10.3168/jds.2018-14811

Kaufman, J. D., Saxton, A. M., and Ríus, A. G. (2018). Relationships among temperature-humidity index with rectal, udder surface, and vaginal temperatures in lactating dairy cows experiencing heat stress. *J. Dairy Sci.* 101(7), 6424-6429. doi: 10.3168/jds.2017-13799

Kendall, P. E., Verkerk, G. A., Webster, J. R., and Tucker, C. B. (2007). Sprinklers and shade cool cows and reduce insect-avoidance behavior in pasture-based dairy systems. *J. Dairy Sc.* 90(8), 3671-3680. doi: 10.3168/jds.2006-766

Kolmodin, R., and P. Bijma. (2004). Response to mass selection when the genotype by environment interaction is modelled as a linear reaction norm. *Genet. Sel. Evol*. 36:435–454. doi: 10.1186/1297-9686-36-4-435

Moran, J. 2005. Tropical Dairy Farming: Feeding Management for Small Holder Dairy Farmers in the Humid Tropics. Landkinks Press, Clayton, Australia.

Moore, C. E., J. K. Kay, M. J. VanBaale, R. J. Collier, and L. H. Baumgard. (2005). Effect of supplemental conjugated linoleic acids on heat-stressed Brown Swiss and Holstein cows. *J. Dairy Sci*. 88:1732–1740. doi: 10.3168/jds.S0022-0302(05)72846-0

Nguyen, T. T. T., Bowman, P. J., Haile-Mariam, M., Pryce, J. E., and Hayes, B. J. (2016). Genomic selection for tolerance to heat stress in Australian dairy cattle. *J. Dairy Sci*. 99, 2849–2862. doi: 10.3168/jds.2015-9685

Penasa, M., Tiezzi, F, Gottardo, P, Cassandro, M, and De Marchi, M. (2015). Genetics of milk fatty acid groups predicted during routine data recording in Holstein dairy cattle. *Livest. Sci.* 173, 9-13. doi: 10.1016/j.livsci.2014.12.014

Petrini J., Iung L. H. S, Rodriguez MAP, Salvian M, Pértille F, Rovadoscki G. A, Cassoli L. D, Coutinho L. L, Machado P. F, Wiggans G. R, and Mourão GB. (2016). Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. *J. Anim. Breed. Genet*. 133, 384–395. doi:10.1111/jbg.12205

Ravagnolo, O., and Misztal, I. (2000). Genetic component of heat stress in dairy cattle, parameter estimation. *J. Dairy Sci.* 83, 2126–2130. doi: 10.3168/jds.S0022-0302(00)75095-8

Ravagnolo, O., Misztal, I., and Hoogenboom, G. (2000). Genetic component of heat stress in dairy cattle, development of heat index function. *J. Dairy sci*. 83(9), 2120-2125. doi: 10.3168/jds.S0022-0302(00)75094-6

Renna, M., C. Lussiana, V. Malfatto, A. Mimosi, and L. M. Battaglini. (2010). Effect of exposure to heat stress conditions on milk yield and quality of dairy cows grazing on Alpine pasture. Pages 1338–1348 in Proc. 9th Eur. IFSA Symp., Vienna, Austria. I. Darnhofer and M. Grötzer, ed. Universität für Bodenkultur, Vienna, Austria.

Rhoads, M. L., Rhoads, R. P., VanBaale, M. J., Collier, R. J., Sanders, S. R., Weber, W. J., Crooker, B. A, and Baumgard, L. H. (2009). Effects of heat stress and plane of nutrition on lactating Holstein cows: I. Production, metabolism, and aspects of circulating somatotropin. *J. Dairy sci.* 92(5), 1986-1997. doi: 10.3168/jds.2008-1641

Sánchez, J. P., I. Misztal, I. Aguilar, B. Zumbach, and R. Rekaya. (2009). Genetic determination of the onset of heat stress on daily milk production in the US Holstein cattle. *J. Dairy Sci*. 92:4035– 4045. doi: 10.3168/jds.2008-1626

Sigdel A, Abdollahi-Arpanahi R, Aguilar I, and Peñagaricano F (2019) Whole Genome Mapping Reveals Novel Genes and Pathways Involved in Milk Production Under Heat Stress in US Holstein Cows. *Front. Genet.* 10:928. doi: 10.3389/fgene.2019.00928

Santana M. L, Bignardi A. B, Pereira R. J, Menéndez-Buxadera A, and El Faro L. (2016). Random regression models to account for the effect of genotype by environment interaction due to heat stress on the milk yield of Holstein cows under tropical conditions. *Animal*. 57, 119–127. doi:10.1017/S1751731116001725

Slimen B. I, Abdeljelil G, and Manef A. (2016). Heat stress effects on livestock: molecular, cellular and metabolic aspects, a review. *J. Anim. Physiol. Anim. Nutr.* 100: 401–412. doi: 10.1111/jpn.12379

St-Pierre, N. R., Cobanov, B, and Schnitkey, G. (2003). Economic losses from heat stress by US livestock industries. *J. Dairy Sci.* 86, E52–E77. doi: 10.3168/jds.S0022-0302(03)74040-5

Stoop W.M., van Arendonk J.A.M, Heck J.M.L, van Va- lenberg H.J.F, and Bovenhuis H. (2008). Genetic parameters for major fatty acids and milk production traits of Dutch Holstein-Friesians. *J. Dairy Sci*., 91, 385–394. doi: 10.3168/jds.2007-0181

Wall, E., G. Simm, and D. Moran. (2010). Developing breeding schemes to assist mitigation of greenhouse gas emissions. *Animal*. 4:366– 376. doi: 10.1017/S175173110999070X

3. GENETIC AND GENOMIC ANALYSIS OF LATENT VARIABLES RELATED TO MILK FATTY ACIDS IN HOLSTEIN COWS

Brayan Dias Dauria[1], Francisco Peñagaricano[2], Lihe Liu[2], Fabricio Pilonetto[1], Paulo Fernando Machado[1], Luiz Lehman Coutinho[1], & Gerson Barreto Mourão[1*]

1 Department of Animal Science, University of São Paulo (USP)/Luiz de Queiroz College of Agriculture (ESALQ), Piracicaba, SP, Brazil

[2] Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

Abstract

This study aimed to perform genetic, genome-wide association (GWAS), and gene-set analysis with latent variables related to milk fatty acid traits (SFA: saturated, PUFA: polyunsaturated, C18:0: stearic acid, and C18:1: oleic acid) of 4,184 Brazilian Holstein cows collected from May 2012 to January 2018. Latent variables were fitted through confirmatory factor analysis based on prior knowledge and biological interest. In the first step, three latent variables were classified as follows: Early (records between 40 to 60 days in milk), Middle (records between 160 to 180 days in milk) and Late (records between 250 to 270 days in milk). In sequence, these latent variables were combined and obtained 3 latent variables (PAR1, PAR2, and PAR3: parities 1-3) with three repetead records (Early, Middle, and Late). (Co)variance components were estimated for factor scores using a single-trait repeatability model under genomic-based approach. Heritabilities estimated for latent variables ranged from 0.07 for *PAR3* to 0.12 for *PAR1*. Genetic correlations between latent variables varied from 0.81 to 0.97. Whole-genome scans identified at least 11 putative candidate genes (*PLD1*, *TM6SF2*, *NUDT7*, *LIPT1*, *AKPA1*, *APOH*, *RPGRIP1L*, *FTO*, *GMDS*, *ALDH3B1*, and *PC*) located on 9 chromosomes, including novel regions explaining relatively smaller fractions of the genetic additive variance (0.55 to 1.81%). The gene-set enrichement analysis revealed functional terms related to fatty acids synthesis, triacylglycerol synthesis, and lipid and lipoprotein metabolism. Our findings point out novel opportunities to investigate multiple correlated traits and deserves to be further investigated for breeding purposes.

Key words: dairy cattle, fatty acids, confirmatory factor analysis, genome-wide association studies (GWAS), gene-set enrichment

3.1 Introduction

The composition of ruminant milk fat has a particular structure among terrestrial mammals, it is attributed to a great diversity of component fatty acids (FA) (Palmquist, 2006). In dairy cattle, Milk FA derives from the effects of ruminal biohydrogenation on dietary unsaturated FA and the variety of FA synthesized de novo synthesis in the mammary gland (Palmquist, 2006). Moreover, FA in the mammary gland also can be originated from the

mobilization of body fat reserves (Gebreyesus et al., 2019). All of those metabolic pathways involved in FA synthesis can be modified by nutrition and management (Walker et al., 2004). However, only selective breeding schemes would have permanent gain for favorable milk FA profile associated with human health (Hein et al., 2018).

Many studies have been showed the possibility of change of milk FA profile based on results of genetic parameters in dairy cows (Hammami et al., 2015; Petrini et al., 2016; Hein et al., 2018). Nevertheless, the exact mechanisms and impacts by which genetic changes can affect the technological and sensory properties of dairy products are not yet known. In addition, the physiological stage related to variation across lactation also can affect the milk FA composition (Bastin et al., 2012), which indicates needed to investigate the genetics and physiological state for efficient manipulation of milk FA synthesis.

For this purpose, the multivariate statistical approach based on reduction of dimensionality of data, such as confirmatory factor analysis (CFA), may be adopted to elucidate the correlation structure among measured traits and to extracted namely factors or latent variables, that are unobserved variables explained by independent variables (Bollen, 1989). In dairy cattle, the potential use of this method has been investigated in genetic analysis, genome-wide association studies (GWAS), and gene-set analysis for milk FA (Conte et al., 2016; Melee et al., 2016; Cecchinato et al., 2019; Palombo et al., 2020a; Palombo et al., 2020b). CFA analysis could be a useful statistical method for assessing the complex structure of correlation among quantitative traits and construct latent variables with significantly biological meanings to be used as a phenotype in genetic and genomic analysis.

We hypothesize that using latent variables might allow us to elucidate the biological interpretation of genomic analysis, particularly in terms of common effects (pleiotropic effects) related to the original phenotype and provide a powerful approach to identify pathways-associated genes with milk FA synthesis. Therefore, the aims of this study were: (i) to assess the use of latent variables related to parities for estimate genetic parameters for a total of milk FA, (ii) to perform whole-genome scans and subsequent gene-set enrichment analysis in order to identify genes and functional terms responsible for milk FA across lactations.

3.2 Materials and Methods

3.2.1. Phenotypes and genotypes

A total of 6,834 fatty acids (FA) records were collected from May 2012 to January 2018 on four dairy herds from Brazil. Lactation records were obtained from 4,184 Holstein cows from first through third parity, daughters of 249 sires, with days in milk (DIM) between 60 and 270. The number of measures ranged from 1 to 3 records per cow. The pedigree file included 6,163 animals created by tracing the pedigree of cows back to two generations. In this study, we considered 4 phenotypes: saturated (UFA) and polyunsaturated (PUFA) FA groups and stearic acid (C18:0) and oleic acid (C18:1) individuals milk FA. These traits were the candidates to construct latent variables and predict factor scores (Table 1). Milk FA were determined as grams per 100g of milk and measured by mid-infrared spectroscopy (*Delta Instruments CombiScope Filter, Advanceed Instruments Inc., Norwood*). The key ingredients in cows' diet include corn silage, grass hay, cottonseed, soybean meal, soybean husk, cornmeal, citrus pulp, minerals, and vitamins.

Table 1. Number of observations (*N*) mean, standard deviation (*SD),* minimum (*Min)* and maximum (*Max*) obtained from factor scores predicted.

Latent variables	***N***	***Mean***	***SD***	***Min***	***Max***
1st Lactation					
Early1	1,015	0.00039	0.9497	-2.90	5.11
Middle1	957	5.7E-08	0.9555	-3.57	4.33
Late1	1,110	0.00036	0.9428	-4.91	6.23
PAR1	3,082	0.00025	0.9487	-4.91	6.23
2nd Lactation					
Early2	887	-0.00022	0.9741	-3.00	5.14
Middle2	856	-0.00011	0.9629	-4.00	5.25
Late2	734	0.002	0.9684	-2.40	4.58
PAR2	2,468	0.00048	0.9699	-4.03	5.25
3rd Lactation					
Early3	487	0.0002	0.9659	-2.72	3.55
Middle3	469	0.0017	0.9757	-2.46	4.86
Late3	319	-0.0025	0.9704	-2.64	4.40
PAR3	1,275	0.000078	0.9689	-2.72	4.86

Abbreviations: Early, Middle and Late are the latent variables with 1 record per cow represented by days in milk: 40-60, 160-180, 250-270 respectively; PAR1, PAR2 and PAR3 corresponding to factor scores predicted from CFA models with 1 to 3 records per cow (the latent variables were obtained through of combination of Early – Middle – Late for each parity).

Genotype data for 72,445 single nucleotide polymorphism (SNP) markers based on a new reference bovine population (ARS-UCD1.2) were available for 824 cows (689 with records) and 89 sires in pedigree file. A detailed description of the procedure of imputation analysis for cow's reference population have been reported by Iung et al. (2019) and Petrini et al. (2019). Sires genotype was obtained from a total of 143,250 bulls available for a panel of 312,614 SNPs provided by USDA-ARS Animal Improvement Programs Laboratory and University of Florida. Subsequently, this panel was reduced to 72,445 SNPs maintained only markers in common with the cows' SNP panel, the procedure was carried out using snp1101 software (Sargolzaei, 2014). Those SNPs markers that mapped to sex chromosomes, were monomorphic, or had minor allele frequency below 1% were excluded from the genotype data.

3.2.2. Latent variable modeling

Before modeling latent variables, the phenotypes were adjusted for one fix effect, described here as an exogenous variable: contemporary group (CG) was formed by the combination of the herd, calving year, and month of analysis information. The phenotypes were defined based on prior knowledge and biological interest. Two steps were performed to define the latent variables that would be used in genetic analysis. i) We created three latent variables for modeling stages of lactation: 60-80 DIM (Early), 160-180 DIM (Middle), and 250-270 DIM (Late) for each parity (1-3) with a total of 9 datasets. Thereby, each latent variable was explained by four observed phenotypes (UFA, PUFA, C18:0, and C18:1). ii) Finally, latent variables obtained previously (*Early-Middle-Late*) were merged, which generated one dataset related to orders of lactations (1 to 3) (*PAR1*, *PAR2*, and *PAR3*). These latent variables were used to estimate genetic parameters, genome-scan mapping, and pathway enrichment analyses using fatty acid scores (FAS) as phenotype. For instance, one cow could have records of one lactation (*PAR1*) ranged from 1 to 3 (*Early-Middle-Late*) measures, which was represented by one FAS obtained from the prediction of the latent variable.

3.2.3. Confirmatory factor analysis (CFA)

All latent variables were separately evaluated using the following measurement model:

$$x = \Lambda\xi + \delta,$$

Where x is a (4 x 1) vector of adjusted phenotypic variables in each animal, ξ is the (1 x 1) vector of latent variables, elements (λ) of Λ are factor loadings relating latent variables to the observed variables (indicator variables), and δ is the corresponding vector of errors of measurement. In this study, it is assumed that $E(\delta) = 0$, var(δ) = θ, $E = (\xi) = \phi$. Here θ is not a diagonal matrix; when considered appropriate, covariances between error terms were freely estimated.

This model was fitted using maximum likelihood estimation (ML) with robust standard errors and mean- and variance-fitted test statistic and unstandardized loadings were estimated for each latent variable construct (Peñagaricano et al., 2015).

In CFA models a well-fitting implies that the latent variable is able to account for the observed covariances among a set of indicator variables. The good fitness of model was performed by several fit indices as x^2(chi-square), standardized root-mean-square residual (SRMR; Bentler, 1995), root-mean-square residual (RMSEA; Steiger, 1990), Tucker-Lewis index (TLI; Tucker and lewis, 1973) and comparative fit index (CFI; Bentler, 1990). All these analyses were performed using package 'lavaan' (Rosseel, 2012) in R software (R Development Core Team, 2011).

After a latent variable has been fitted, we were interested in predicting factor scores for each individual based on their observed values of the variables indicators from factors. We predicted factor scores through the sum of the individual values from observed variables, with the weights considered by the parameters (standardized factor loadings) obtained in the fitted model. This method was performed using the package 'Psych' of the R software (R Development Core Team 2019).

3.2.4. Genetic analyses

Genetic analyses can be summarized in the following steps: i) we performed a single-trait model (animal) to estimate variance components to each latent variable related to stages of lactations (Early, Middle, and Late) for three parities; ii) repeatability model was estimated using FAS as phenotype for each latent variable classified as parities (*PAR1*, *PAR2*, and *PAR3*). We carried out these steps to evaluate the possibility of reduced dimensionality of data and to investigate the use of factor scores as a phenotype in genomic analysis. Single-trait models were performed to estimate variance components of fatty acid scores to stages of lactation: Early, Middle, and Late, considering as different traits for each parity.

$$y = \mu + Z\mathrm{a} + \varepsilon$$

where y, μ, a, and ε are the vectors of estimated latent variables (fatty acid scores), intercept, additive genetics random effects, and the vector of residual effects, respectively. The join distribution of ε~N (0, Iσ_e^2), where I is an identity matrix and σ_e^2 is the residual variance. In the polygenic model, it was assumed a~N(0, Aσ_a^2) where A represents the matrix of the additive genetic relationship between animals in pedigree. The (co)variance structure was assumed as follows:

$$\boldsymbol{Var}\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{e} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \otimes \Phi & 0 \\ 0 & \boldsymbol{I} \otimes \boldsymbol{R} \end{bmatrix}$$

where $\boldsymbol{A}$ is the numerator relationship matrix, and Φ are (co)variance matrix of random coefficients for additive effect, $\boldsymbol{R}$ is a diagonal matrix of residual variances of the trait, and ⊗ denotes the Kronecker product of matrices.

Repeatability model were used to estimate variance components that was performed using the maximum restricted likelihood method in AIREMLF90 program (Misztal et al., 2002) FAS were considering to the first three orders of lactations as different traits. The following model was fitted:

$$y_{klm} = DIM_{kl} + a_{klm} + pe_{klm} + e_{klm}$$

Where y_{klm} is the record for the FAS, DIM_{kl} days in milk defined for FAS observed in 3 different stages of lactation (60-80 DIM (Early), 160-180 DIM (Middle), and 250-270 DIM (Late), a_{klm} is the random genetic additive effect of animal m in stage of lactation kl, pe_{klm} is the random permanent environment effect of animal m in stage of lactation kl, e_{klm} is the random residual effect. The (co)variance structure was assumed as follows:

$$\begin{bmatrix} a \\ pe \\ e \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \otimes \Phi & 0 & 0 \\ 0 & I \otimes \psi & 0 \\ 0 & 0 & \boldsymbol{I} \otimes \boldsymbol{R} \end{bmatrix}$$

where $\boldsymbol{A}$ is the numerator relationship matrix, Φ and ψ are (co)variances matrices of random genetic additive and permanent environment effects respectively, $\boldsymbol{R}$ is a diagonal matrix of residual variances corresponding to each FAS, and ⊗ denotes Kronecker products of matrices.

For all previous analysis were assumed a genomic polygenic model a~N (0, Hσ_a^2), where σ_a^2 is the additive genetic variance, H is the combined relationship matrix (pedigree and genomic information) (Aguilar et al., 2010; Christensen & Lund, 2010). This method is known as single-step genomic best linear unbiased prediction (ssGBLUP). The inverse of H was obtained as follows,

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

where G^{-1} is the inverse of the genomic relationship matrix and A_{22}^{-1} is the inverse of the pedigree relationship matrix of the animals genotyped. The G^{-1} has the dimension of 582 x 582, 575 x 575, 412 x 412 that includes cows with records and sires in the pedigree for 1, 2 and 3 parity respectively. The A matrix has a dimension of 3,977 x 3,977, 3,196 x 3,196, 1,810 x 1,810 for 1, 2 and 3 parity respectively, which is based on a two-generation pedigree.

3.2.5. Assessment of candidate genomic regions

The genomic regions and candidate genes for latent variables were obtained by the following procedures: i) the ssGBLUP method was performed to obtain the estimated genomic breeding values (GEBV), ii) The effect of the SNP was obtained using the GEBV estimated using the postGSF90 program (Aguilar et al., 2014) following the equation described by Wang et al. (2012), iii) Candidate genomic regions were obtained of the amount of genetic percentage of variance explained by 2.0 Mb windows of adjacent SNPs performed in the postGSF90 software.

ii) Wang et al. (2012).

$$\hat{u} = DZ'[ZDZ']^{-1}\widehat{a_g}$$

where: $\hat{u}$ is the vector of markers effects, D is a diagonal matrix of weights of SNPs, here, $D = I$, I is an identity matrix or weight = 1, Z is a matrix of genotyped of each locus, and $\widehat{a_g}$ is the vector of GEBVs.

iii) Aguilar el al. (2014)

$$\frac{var(u_i)}{\sigma_u^2}\ x\ 100\ =\ \frac{var\ (\sum_{j=1}^{N} Z_j S_j)}{\sigma_u^2}\ x\ 100$$

where u_i is the genetic value of the i^{th} genomic region, N is the total number of adjacent SNPs within the 2.0 Mb genomic region, and Z_j is the genotype information of j^{th} marker, S_j is the marker effect within the i^{th} genomic region.

3.2.6. Gene-set enrichment analysis

Gene-set enrichment or pathways-based analysis is an alternative tool to identify biologically relevant pathways and could support a better understanding genetic of complex traits (Weng et al., 2011). This approach could be defined in three steps: i) the assignment of SNPs to gene (i.e., location of SNPs in gene annotated), ii) the assignment of genes to functional pathways and iii) Verify the association of a given pathway with FAS, which was analyzed using a test of proportions based on the cumulative hypergeometric distribution or also named as Fisher's exact, if there was a statistical difference among latent variables for lactations (Peñagaricano et al., 2013).

The ARS-UCD1.2 bovine genome sequence assembly was used for SNP assignments using biomaRt package in R. Herein, SNPs were assigned to genes if they were located within the genomic sequence of an anottaded gene or within 15 kb either upstream or downstream of the gene. An arbitrary threshold of 5% of the SNP effects distribution (in absolute value) was used to define relevant SNP markers and genes associated with latent variables if that gene contained at least one potential SNP.

For assignment of genes to pathways in each latent variable, we carried out functional enrichment analysis on the list of significant genes using the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Medical Subject Headings (MeSH), InterPro, Reactome, and Molecular Signatures Database (MSigDB) to identify significantly overrepresented pathways to define a functional set of genes. The association between a particular gene-set and FAS of latent variables was assigned using a Fisher's exact test (Peñagaricano et al., 2013).

The P-value of observing k significant genes in pathway-based association analysis was calculated by:

$$\mathbf{p} - \text{value} \;=\; 1 \;-\; \sum_{\mathrm{i}=0}^{\mathrm{k}-1} \frac{\binom{\mathrm{S}}{\mathrm{i}} \binom{N-\mathrm{S}}{\mathrm{m}-\mathrm{i}}}{\binom{N}{\mathrm{m}}}$$

where *S* is the total number of genes that are significantly associated with each parity represented by latent variable, *N* is the total number of genes that were analyzed in the study, and *m* is the number of genes in the pathway.

3.3. Results and Discussion

3.3.1. Latent variable fit model

The main challenge involving latent variables is the goodness of fit of the measurement model, traditionally assessed by several fit indices. Thus, these indices are an essential parameter, which allows using these variables to investigate complex traits. In this study, we presented in Table 2 all of the fit indices extracted from CFA. All p-values were higher than 0.05 suggesting, that observable variables were independents, and alternative indices as CHISQ (ranged from 0.65 to 3.78), RMSEA (ranged from 0.007 to 0.05), SRMR (ranged from 0.003 to 0.007), CFI (close to 1), and TLI (close to 1) indicated that proposed model fitted the data reasonably well. The standardized solutions for the factor loadings (i.e., elements of Λ presented in the CFA model) are shown in supplementary tables S1, S2, and S3.

Table 2. Fit indices from confirmatory analysis

Fit indices	***1st Lactation***			***2nd Lactation***			***3rd Lactation***		
	E	***M***	***L***	***E***	***M***	***L***	***E***	***M***	***L***
DF	1	1	1	1	1	1	1	1	1
CHISQ	3.78	2.93	3.64	0.65	1.21	1.03	0.642	1.59	1.23
P-VALUE	0.054	0.082	0.056	0.419	0.271	0.308	0.423	0.207	0.267
RMSEA	0.05	0.04	0.04	0.01	0.01	0.007	0.001	0.035	0.027
SRMR	0.007	0.005	0.006	0.002	0.003	0.003	0.003	0.004	0.004
CFI	0.99	0.99	0.99	1	1	1	1	1	1
TLI	0.99	0.99	0.99	1	1	1	1	0.99	0.99
N	1,015	957	1,110	887	887	887	487	487	487

Abbreviations: DF: Freedom degree; CHISQ: Chi-square; RMSEA: Root mean square error of approximation; SRMR: Standardized root mean residual; CFI: Comparative fit index; TLI: Tucker Lewis index; N: number of observations.

Penãgaricano et al. (2015) suggested that factor loadings can be interpreted as the correlation between the observed phenotype and the corresponding latent variable. The wald test (Z-value) was obtained by dividing the parameter value by its standard error, and noticeably all Z-values were higher than 2.58, thereby all factors loading had statistically different from 0 (P(>|Z|)) in our population, which means that observed variables used here had a relevant association with the corresponding latent variable. As shown in Figure 1. below, we modeled each latent variable separately related to one specific interval across lactation. In our study, the limitation for working with a simultaneous model (structural equation model) was attributed to latent variables be correlated and did not a satisfactory fit in CFA.

Figure 1.

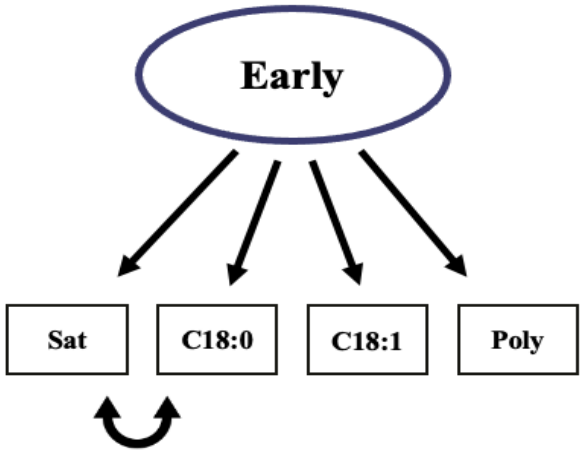


Figure 1. An illustrative example of a measurement model with one latent variable evaluated using confirmatory factor analysis. Curved arrows between boxes represent the presence of (co)variance between error terms. The goodness of fit of the measurement model was evaluated using a $\chi 2$ (scaled) test and alternative fit indices cited previously.

Generally, we found studies considering different latent variables as quality, fertility, or production investigating the causal and no causal structure between latent variables defined by the structural equation model (Peñagaricano et al., 2015; Leal-Gutiérrez et al., 2018; Pegolo et al., 2020). In contrast, in our study, we are interested in assessing different parities been represented by factors or latent variables constructed by a group of milk fatty acids. The potential of this approach (multivariate factor analysis) has been reported by several authors that suggested be a powerful tool for reducing complexity in genetic and genomic studies through separated groups of variables with similar origins and functions, maintaining the biological significance and behavior of the original phenotype (Mele et al., 2016; Dadousis et al., 2017; Olasege et al., 2019; Cecchinato et al., 2019).

3.3.2. Genetic parameters for latent variables

The estimates of heritabilities and genetic correlation for latent variables for the stage of lactation (*Early*, *Middle*, and *Late*) are summarized in Table 3. Overall, genomic heritabilities to lactation stage factors ranged from low to high for three parities. Factors *Early1* (0.18) and *Early3* (0.23) presented hightest heritabilities, whereas *Middle3* (0.02) and *Late3* (0.02) had

lowest heritabilities estimates. The other factors had low to moderate heritabilities ranging from 0.05 (*Late3*) to 0.15 (*Middle2*).

Table 3. Heritabilities estimated through a genomic-based model of factor scores representing each stage of lactation.

	σ^2_a	σ^2_p	σ^2_e	h^2	*n*
1st lactation					
Early1	0.166 (0.086)	0.910	0.743 (0.081)	0.18 (0.089)	1.015
Middle1	0.062 (0.053)	0.914	0.846 (0.062)	0.06 (0.060)	957
Late1	0.063 (0.034)	0.889	0.826 (0.052)	0.07 (0.050)	1.110
2nd lactation					
Early2	0.030 (0.033)	0.948	0.917 (0.048)	0.03 (0.035)	887
Middle2	0.130 (0.069)	0.928	0.798 (0.072)	0.14 (0.086)	856
Late2	0.047(0.059)	0.936	0.888 (0.073)	0.05 (0.063)	734
3rd lactation					
Early3	0.218(0.086)	0.933	0.715 (0.087)	0.23(0.086)	487
Middle3	0.019 (0.019)	0.950	0.931 (0.064)	0.02 (0.019)	469
Late3	0.023 (0.025)	0.933	0.914 (0.077)	0.02 (0.025)	319

Early: 40-60 days in milk; *Middle*: 160-180 days in milk; *Late*: 250-270 days in milk; σ^2_a: genetic variance additive; σ^2_p: phenotypic variance; σ^2_e: residual variance; h^2: heritability; **n**: number of observations.

The heritability estimates found here showed a relevant variability between lactation stages, agreeing with those of the individual traits findings out by Park et al. (2020) that suggested a significant contribution of lactation stage, energy balance, and parity on variation in milk fat composition. Bilal et al. (2014) also reported significant effects of the stage of lactation and parity on milk fatty acid profile for the Canadian Holstein population. Interestingly, the authors argued that the proportions of short- and medium-chain fatty acids (C6:0 to C14:0) were low at the beginning of lactation and increased in the early part of lactation, different than observed for long-chain fatty acids (C18:0 and C18:1) that had an opposed trend compared to others fatty acids.

In our study, we merged the (co)variance structure of two milk fatty acids groups (saturated and polyunsaturated) and two individual milk fatty acids (C18:0 and C18:1) into one variable not observable (latent variable) to evaluate these complex traits together. These phenotypes represented more than 70% of the fatty acid content of milk and involved several pathways about fatty acid synthesis (*de novo*, biohydrogenation, desaturation, and short -and-long-chain FA). The differences observed here in genetic components between stages within lactation probably occurred due to metabolic state and rumen activity, attributed to diet

composition to achieve the nutrient requirements of cows during lactation. In agreement, Mele et al. (2016) also reported a significant effect of the lactation stage on the majority of latent variables related to milk FA, in particular, those associated with fatty acids derived from mammary enzymes (de novo, desaturation, and long-chain FA) presenting an increase across lactation. In agreement, Mele et al. (2016) also reported a significant effect of the lactation stage on the majority of latent variables related to milk FA, in particular, those associated with fatty acids derived from mammary enzymes (de novo, desaturation, and long-chain FA) presenting an increase during lactation. This fact reinforces the importance of carrying out milk quality measurements constant for a better understanding of the genetic basis of fatty acids profile.

Estimates of genetic correlation between lactation stages represented by factors are in Table 4. Intervals across lactation for cows of first parity had positive and moderate (< 0.70) genetic correlation between *Early1* and *Middle1* (0.68) and *Early1* and *Late1* (0.63), whereas a strong genetic correlation, was found among *Middle1* and *Late1* (0.99) latent factors; For the second lactation, the factor *Early2* was strong and positive genetic correlated with *Middle2* (0.99) and *Late2* (0.99), in exception of *Middle2* and *Late2* that presented moderate genetic correlation (0.47); For the third lactation, the factor *Early3* presented a strong and positive genetic correlation with *Middle3* (0.99) and *Late3* (0.99), and conversely, a negative and strong genetic correlation among *Middle3* and *Late3* (-0.99). The estimates of genetic correlations for latent factors disagree with genetic correlations among individual fatty acid contents in milk reported by Bastin et al. (2012) that showed an intermediate magnitude ranging from 0.52 to 0.70. However, a strong association of genetic correlation found here (0.99) associated with high standard errors (0.74 to 1.70) can be due to small variation between factors due to a low phenotypic variation presented for some individual fatty acids, or even the reduced number of records used for this analysis.

Table 4. Genetic correlations among stages of lactations considering latent variables.

1st lactation	***Early***	***Middle***	***Late***
Early1	1	0.68 (1.10)	0.63 (1.70)
Middle1	0.68 (1.10)	1	0.99 (0.55)
Late1	0.63 (1.70)	0.99 (0.55)	1
2nd lactation	***Early***	***Middle***	***Late***
Early2	1	0.99 (1.00)	0.99 (0.74)
Middle2	0.99 (1.00)	1	0.47 (1.67)
Late2	0.99 (0.74)	0.47 (1.67)	1
3rd lactation	***Early***	***Middle***	***Late***
Early3	1	0.99 (1.34)	0.99 (0.70)
Middle3	0.99 (1.34)	1	-0.99 (1.17)
Late3	0.99 (0.70)	-0.99 (1.17)	1

Early: 40-60 days in milk; *Middle*: 160-180 days in milk; *Late*: 250-270 days in milk.

Regarding the genetic parameters estimated for parities as factors, we merged latent factors (*Early*, *Middle*, and *Late*) by ID number cow (identification number of each animal) and parity to perform a repeatability genetic additive model. Estimates of heritability (Table 5.) were low to all factors ranged from 0.07 (*PAR3*) to 0.12 (*PAR1*). Genetic correlations between parities considered by latent variables are presented in Table 6. A high and strong genetic correlation was found between latent factors varying from 0.81 to 0.97.

Table 5. Heritabilities estimated through a genomic-based model of latent variables representing parities.

	σ^2_a	σ^2_p	σ^2_{pe}	σ^2_e	h^2	***n***
PAR1	0.071	1.548	0.118	0.712	0.12 (0.033)	3,082
PAR2	0.073	0.943	0.034	0.836	0.11 (0.036)	2,468
PAR3	0.059	0.943	0.016	0.875	0.07 (0.046)	1,275

PAR1: latent variable referent to first parity; *PAR2*: latent variable referent to second parity; *PAR3*: latent variable referent to third parity. σ^2_a: genetic variance additive; σ^2_p: phenotypic variance, σ^2_e: residual variane, h^2: coefficient of heritability; ***n***: number of records.

Table 6. Genetic correlations among parities of the latent variables genomic-based approach.

Latent variables	*Lac1*	*Lac2*	*Lac3*
PAR1	1	0.96 (0.023)	0.97 (0.147)
PAR2	0.96 (0.023)	1	0.81 (0.547)
PAR3	0.97 (0.147)	0.81 (0.547)	1

PAR1: first parity; *PAR2*: second parity; *PAR3*: third parity.

Bastin et al. (2012) also reported a relevant variation in heritabilities between parities for all individual fatty acids used here (SFA, PUFA, C18:0, and C18:1). Interestingly, the authors suggested that de novo synthesis and saturated FA were more heritable than C18:0 (stearic), C18:1 (oleic), PUFA (polyunsaturated), MUFA (monounsaturated), and LCFA (long-chain fatty acids) that are originated mainly from the diet and the body fat mobilization. Under multivariate factor analysis, Mele et al. (2016) classified several latent factors related to different pathways of fatty acids milk (de novo, biohydrogenation, long-chain FA, desaturation, short-chain FA, odd FA, linolenic, vaccenic, CLA, and milk yield – branched FA) combining variables in common, and investigated the effect of parity of cows on a pattern of milk fatty acid factors. The authors showed parity significantly (P-value < 0.01) affected the scores of almost all the latent factors suggesting that these variables were consistently based on current knowledge of the physiological changes occurring during lactation.

The results obtained through the repeatability additive model suggested that latent variables representing parities had a similar relationship between them due to high genetic correlation estimates (0.81 to 0.97). Thus, considering the latent factor lactations represented by the same group of phenotypes probably would not present relevant differences between estimates of genetic correlation. This limitation could be related to the lack of the other variables to construct latent factors and, if possible, define this factor based on pathways related to fatty acids synthesis (i.e., mammary gland and *de novo*).

3.3.3. Whole-genome mapping for latent variables

ssGBLUP methodology was utilized to identify genomic regions and putative candidate genes related to fatty acids factors. Our interest here was to detect the structure underlying the variables that could interact in a biological role. Figure 1. displays Manhattan plots for fatty acids factors for the three parities in context. The results are presented in terms of the proportions of genetic variance explained by 2.0 Mb SNP windows.

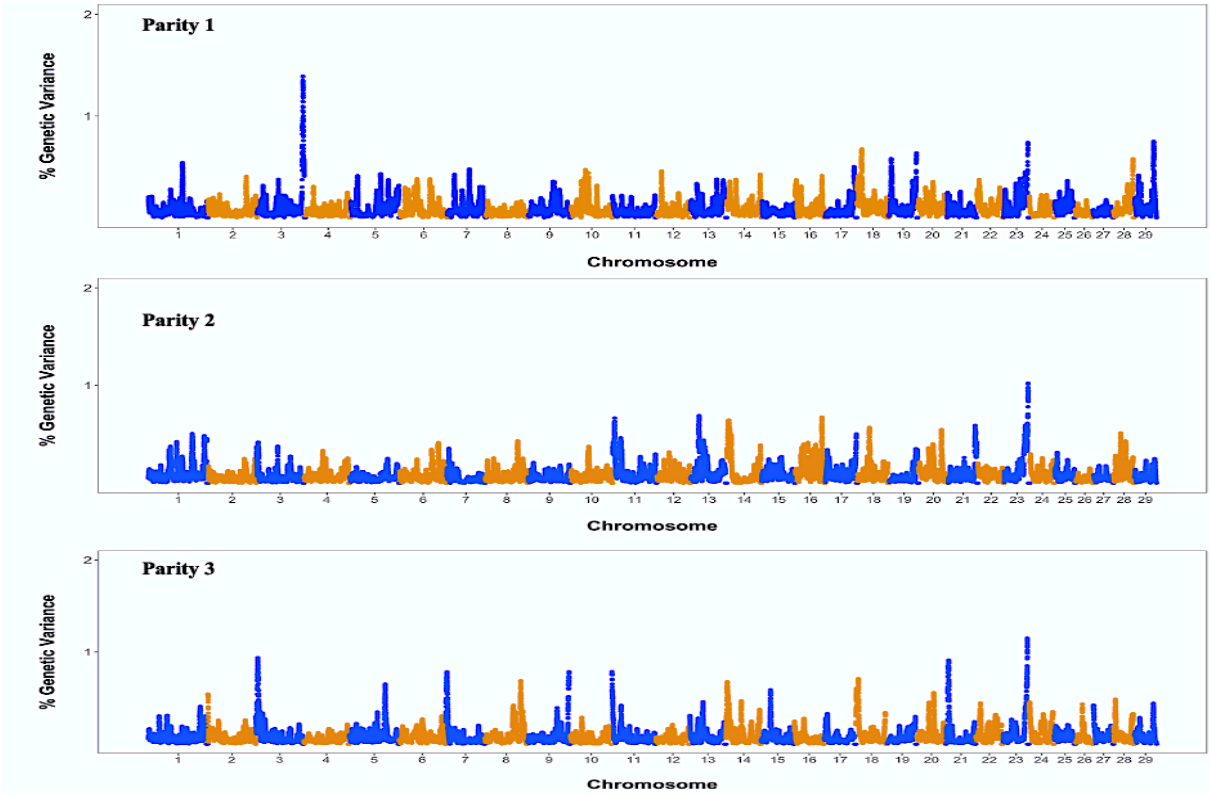


Figure 2. Manhattan plots showing the results of the whole-genome scans for the cows of first to third parity using fatty acids scores.

A summary of the results of GWAS for three latent variables is shown in Table 7. In total, we detected 11 putative candidate genes located on 9 Bos taurus autosomes. We observed one genomic region window in common in BTA23 (50.22–52.22 Mb position) associated with the FA scores across all three lactations. This region on BTA23 that harbors gene *GMDS* explained about 0.66%, 1.02%, and 1.15% of genetic variance for fatty acids scores across the first three parities. Gene *GMDS* acts on the metabolism of amino sugars and nucleotide sugars (Wickramasinghe et al., 2011). In literature, there are reports of association of this gene with the saturated fatty acids profile in intramuscular fat of longissimus thoracis muscle in bovine (Lemos et al., 2016).

Table 7. Putative candidate genes located in 2.0 Mb SNP windows that explain the highest genetic variance for latent variables across the first three parities.

		Genetic Variance (%)			
		Parities			
Chr	***Pos (Mb)***	***PAR1***	***PAR2***	***PAR3***	***Candidate genes***
BTA1	95.42 - 97.42	0.55			*PLD1*
BTA3	117.37 - 119.37	1.81			-
BTA3	119.38 - 121.38	0.61			-
BTA7	2.29 - 4.29			0.78	*TM6SF2, NUDT7*
BTA11	2.85 - 4.85		0.66		*LIPT1*
BTA19	7.78 - 9.78	0.82			*AKPA1*
BTA19	61.94 - 63.94	0.65			*APOH*
BTA21	66.78 - 68.78		0.59		*RPGRIP1L, FTO*
BTA23	50.22 - 52.22	0.66	1.02	1.15	*GMDS*
BTA28	44.00 - 46.00	0.63			*ALDH3B1, PC*
BTA29	47.24 - 49.24	0.73			-

PAR1: first parity; PAR2: second parity; PAR3: third parity.

For PAR1, a total of eight genomic regions window located on chromosomes BTA1, BTA3, BTA19, BTA23, BTA28, and BTA29, explain more than 6.46% of the additive genetic variance for fatty acids scores. In particular, the region on BTA1 (95.42-97.42 Mb) covers the candidate gene *PLD1*, which is involved in lipid metabolism bovine that has a function catalytic activity (Pegolo et al., 2017). In addition, *PLD1* (phospholipase D) has been considered an important enzyme that generates phosphatidic acid (PA) a critical phospholipid constituent in eukaryotic cell membranes, that accounts for 1-4% of the total lipid that can be generated de novo synthesis.

Another 2.0 Mb SNP window on BTA19 (7.78-9.78 Mb) explained about 0.82% of the additive genetic variance for fatty acids scores in PAR1. Notably, this region harbors the gene *AKPA1*, which encodes a member of the *AKAP* family, which has been shown to interact with the *PKA* family (Carlson et al., 2003). Moreover, this gene has been reported as a major adipocyte protein kinase A-binding protein, is the most abundant in adipose tissue and involved in fat metabolism and obesity (Bridges et al., 2006; Marrades et al., 2010). On BTA19, another window from 61.94 to 63.94 Mb was associated with fatty acid scores in PAR1. This region

harbors the gene *APOH* (apolipoprotein H), also known as beta-2-glycoprotein I, which is a component of circulating plasma lipoproteins that binds several kinds of negatively charged substances such as heparin, phospholipids, and dextran sulfate. *APOH* has been involved in several human physiologic pathways, including lipid metabolism, coagulation, and the production of antiphospholipid antibodies (Mehdi et al., 1999).

On BTA28, two genes (*ALDH3B1* and *PC*) were found in genomic region from 44.00 to 46.00 Mb in PAR1. Gene *ALDH3B1* is involved in oxidation of lipid-derived aldehydes generated in the human plasma membrane (Kitamura et al., 2013) and is associated to diatebes in humans (Jeff et al., 2014). Evidence supporting the association between *ALDH3B1* and backfat thickness was previously reported in bovine (Silva et al., 2017). Gene *PC* (pyruvate carboxylase) plays a central role in gluconeogenesis and lipogenis. Previous reports shown a significantly elevated expression of *PC* in dary cattle around the time of calving, during feed restriction, and has been linked to increased fatty acids concentrations and profiles (White et al., 2011). Boesche & Donkin. (2020) suggested that *PC* promoter 1 activity that is mediated by unsaturated fatty acids may determine gene *PC* response during periods of negative energy balance in dairy cows.

For PAR2, a total of three genomic regions window located on chromossomes BTA11, BTA21, and BTA23, explained more than 2.27% of the additive genetic variance for fatty acids scores. On BTA11, a window from 2.85 to 4.85 Mbs harbor the gene *LIPT1* that is associated with hyperlipedemia (accumulation of blood lipids) in human (Soreze et al., 2013), and was strongly associated as marbling trait in beef cattle (Magalhães, 2015).

Another 2.0 Mb SNP window on BTA21 (66.78-68.78 Mb) explained about 0.65% of the additive genetic variance for fatty acids scores in second parity. Interestingly, this region harbors the genes *RPGRIP1L* and *FTO* that were considered associated haplotypes by close genomic regions (Lea et al., 2013). *RPGRIP1*-like (*RPGRIP1L*) gene encodes a protein as a conserved C2-domain, which bind phospholipids, inositol polyphosphates, and intracellular proteins (Tews et al., 2011). The authors isolated primary pre-adipocytes tissue human and concluded that *RPGRIP1L* might be involved in adipogenic differentiation and has a relevant function in the insulin-regulated adipocyte metabolism. *FTO* gene has been reported as the major candidate gene for obesity in human (Dina et al., 2007) and was considering a potential biological locus due to *FTO* protein be conserved with a sequence identity of over 85% among humans, mice, cattle, sheep, dogs, and horses (Fredriksson et al., 2008). Experiments have identified that *FTO* signalizes cellular availability of oxygen, is functionality involved in fatty

acid metabolism and energy homeostasis and has a function in the catalysis of nucleic acid demethylation (Han et al., 2010). Supporting our findings both genes were associated with milk composition variation in Holstein dairy cattle (Zielke et al., 2013), indicating that those genes could be potential markers to investigated milk fat composition.

For PAR3, we found a total of two genomic regions window located on chromosomes BTA7 and BTA23, explained more than 1.93% of the additive genetic variance for fatty acids scores. On BTA7, our study detected a putative candidate gene *TM6SF2* (2.29-4.29 Mb); this gene is a regulator of liver fat metabolism in humans with opposing effects on the secretion of TG-rich lipoproteins and hepatic triglyceride content (Mahdessian et al., 2014). Chen et al. (2017) shown that an SNP E167k in the gene *TM6SF2* may have additive effects on lipid metabolism and the development of NAFLD (nonalcoholic fatty liver disease) by upregulating the expression of *SREBP-1* (sterol receptor-element binding proteins) and *FASN* (fatty acid synthesis). In addition, was found considerable conservation for the predicted protein sequence of *TM6SF2* in the cow, human, dog, guinea, pig, and mouse (Mahdessian et al., 2014). In bovine, fatty acid synthesis TAG-rich chylomicron and very-low-density lipoproteins (VLDL) of plasma are the primary sources of long-chain fatty acids taken up by the mammary glands (Palmquist, 2006). However, Grajales et al. (2020) performed an experimented with differential expression isolated mammary tissue bovine to investigate the association related to lipid metabolism and observed a negatively regulated expression of gene *TM6SF2*.

Gene *NUDT7* (BTA7) encodes for a peroxisomal coenzyme A diphosphatase with the role to remove potentially toxic oxidized CoA disulfide from peroxisomes maintaining the capacity for beta-oxidation of fatty acids (Muñoz et al., 2013). Because this gene is involved in peroxisomal lipid metabolism in mouse and human metabolism (Wanders & Watherham, 2006), we suggest that gene *NUDT7* could be associated with the fatty acid synthesis of milk cows, but it is still unclear in the literature and needs to be more investigated.

Novel regions without potential candidate genes were found on BTA3 (117.37-119.37 and 119.38-121.38 Mbs) for PAR1 and BTA29 (47.24-49.24 Mb) for PAR3 associated with fatty acids scores. The regions on BTA3 explained about 2.42% of the additive genetic variance; On BTA29 the additive genetic variance explained 0.73%. In this study, we found out different genes that still were not well-discussed in the literature, nevertheless, we expected it because our genomic analysis involved a multi-trait model considering a group of milk FA traits with (co)variances structures combined. Overall, we showed that the latent variables approach associated with GWAS could be a powerful tool to investigate genes with a joint biological

function, however, further research on a larger population is necessary to validate the results obtained in this study.

3.3.4. Gene-set analysis for latent variables

The first step in this gene-set analysis of a whole-genome association study was to assign SNPs to genes. Of the 66,460 SNP markers evaluated in the GWAS, representing three lactations, a total of 28,366, 29,370, and 29,919 SNPs, respectively, were located either within annotated genes or at most 10 kb upstream or downstream from annotated genes. This set of SNPs marked a total per parity of 11,635, 11,451, and 11,748 genes annotated in the ARS-UCD1.2 bovine genome sequence assembly. A subset of 513, 454, and 456, respectively, were flagged by at least one relevant SNP (based on the top 5% of the SNP effects distribution) for each parity, hence, these genes were defined as significantly associated with fatty acids scores represented by latent variables.

The functional characterization included six different biological databases: Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, InterPro, Medical Subject Headings (MeSH), and Molecular Signatures Database (MSigDB). Supplementary table 4. reports the full list of significant functional terms, containing term ID, genes, and Fisher's P-value.

Figure 4. shows a set of functional terms significantly enriched with genes affecting fatty acids scores across all lactations. These functional terms were mainly related to fatty acids synthesis, triacylglycerol synthesis, and lipid and lipoprotein metabolism. Noticeably, some of the most relevant terms are directly involved in fatty acid synthesis related to milk fat composition, such as fatty acid biosynthesis [bta00061], long-chain fatty acid uptake [GO:0044539], the chemical reactions and pathways involving in linoleic acid [M15245] and alpha-linoleic acid [M13605], and the chemical reactions and pathways resulting in the formation of a fatty-acyl-CoA [M23818]. These significant fatty acids terms had in common at least three genes, namely, *ACSL1*, *ACSL3*, and *ELOVL2*. All of these genes are involved in process of regulating the channeling of fatty acids in milk fat synthesis (Bionaz & Loor, 2008), lipid biosynthesis, and fatty acids degradation (Popperlreuther et al., 2012), and elongation of PUFA (Castro et al., 2016), respectively.

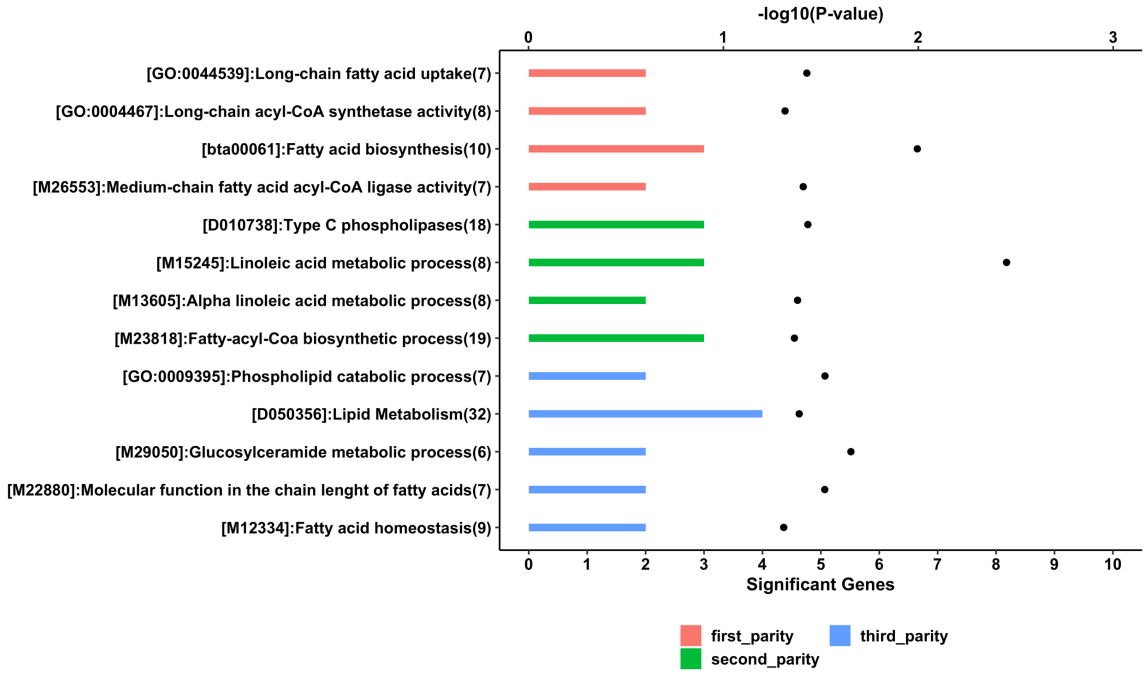


Figure 4. Functional terms and pathways significantly enriched with genes associated with profile of fatty acids in milk across lactations Six gene annotation databases were analyzed: Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Medical Subject Headings (MeSH), InterPro, Reactome and Molecular Signatures Database (MSigDB). The *y*-axis displays the names and the total number of genes of each gene-set. The black dots represent the significance of enrichment ($-\log_{10}$ *P*-value, Fisher's exact test, top *x*-axis) and the bars represent the number of significant genes in each functional term (bottom *x*-axis).

Of special interest, we found gene-sets that showed an overrepresentation of genes that are directly associated with milk fat synthesis, such a *type C phospholipase* [D010738] and *lipid metabolism* [D050356]. These milk fat synthesis terms contain at least three relevant genes, namely *XDH*, *CPT2*, and *CD36*. *XDH* gene is involved in lipid droplet formation with a role in the mechanism encompassing the genes *ADFP* and *BTN1A1* in the mammary gland (Bionaz & Loor, 2008). *CPT2* gene transferred long-chain acyl-CoAs into the mitochondria for *β*-oxidation, which is the key enzyme in lipid oxidation (Eaton, 2002). *CD36* gene act in endothelial long-chain fatty acids transport and appear to be the most important protein related to fatty acids uptake from the blood (Bionaz & Loor, 2008).

The term *fatty acid homeostasis* [M12334] was another functional term significantly enriched with a gene associated with milk fat content. This gene-set harbors gene DGAT2 which was suggested as a strong candidate for studies with dairy cows (Al-shuhabib et al., 2019) and goats (An et al., 2011). Interestingly, Liu et al. (2020) studying functional analysis of *DGAT* family genes, reported that this gene encoding the enzymes interacting with each other,

collectively regulated lipid metabolism, and affected milk secretion and synthesis in mammals. Thus, our findings provide further evidence of the possible involvement of this gene in milk fatty acids profile.

3.4. Conclusions

In summary, we performed confirmatory factor analysis to reducing complexity in genetic and genomic studies of milk fatty acids traits. Genetic parameters for latent variables within (stages of lactation) and between (parities) lactations had strongly genetic correlated and ranged from low to moderate. GWAS for latent variables resulted in the detection of 11 putative candidate genes (*PLD1*, *TM6SF2*, *NUDT7*, *LIPT1*, *AKPA1*, *APOH*, *RPGRIP1L*, *FTO*, *GMDS*, *ALDH3B1*, and *PC*) located on 9 chromosomes, including novel regions explaining relatively smaller fractions of the genetic additive variance. Moreover, the gene-set analysis revealed significant functional terms including fatty acids synthesis, triacylglycerol synthesis, and lipid and lipoprotein metabolism. Overall, this study contributes to designing further studies to improving the knowledge base on the genetics underlying the bovine milk fatty acids composition.

References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. ***J. Dairy Sci***. 93:743–752.

Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A., Wang, H. (2014). "PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs," *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production* (Vancouver, BC).

Al-Shuhaib, M. B., Al-Thuwaini, T. M., Al-Fihan, R. A., & Al-Qutbi, A. A. (2019). Genotyping of diacylglycerol acyltransferase 2 gene in Holstein cattle population**.** ***Agri. Cons. Sci.*** 84(2), 211-218.

An XP, Song SG, Hou JX, Zhu CM, Peng JX, Liu XQ, Liu HY, Xiao WP, Zhao HP, Bai L, Wang JG, Song YX, Cao BY. (2011). Polymorphism identification in goat DGAT2 gene and association analysis with milk yield and fat percentage. ***Small Ruminant Res***.100:107–112.

Bastin, C., N. Gengler, and H. Soyeurt. (2011). Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. ***J. Dairy Sci***. 94:4152–4163.

Bastin C., Soyeurt H., Gengler N. (2012) Genetic parameters of milk production traits and fatty acid contents in milk for Holstein cows in parity 1–3. J. ***Anim. Breed. Genet***. 130, 118–127.

Bentler P. M. (1990). Comparative fit indexes in structural models. ***Psychol***. Bull.107:238–246. doi:10.1037/0033-2909.107.2.238.

Bentler P. M. (1995). EQS structural equations program manual. Multivariate Software. Encino, CA.

Bilal, G., Cue, R. I., Mustafa, A. F., & Hayes, J. F. (2014). Effects of parity, age at calving and stage of lactation on fatty acid composition of milk in Canadian Holsteins. ***Canadian J. Anim. Sci***. 94(3), 401-410.

Bionaz, M., & Loor, J. J. (2008). Gene networks driving bovine milk fat synthesis during the lactation cycle. ***BMC genomics***. 9(1), 366.

Boesche, K. E., & Donkin, S. S. (2020). Bovine pyruvate carboxylase gene proximal promoter activity is regulated by saturated and unsaturated fatty acids in Madin-Darby bovine kidney cells. ***J. Dairy. Sci***.

Bollen K. A. (1989). Structural Equations with Latent Variables Wiley New York, NY.

Bridges D, MacDonald JA, Wadzinski B, Moorhead GB. (2006). Identification and characterization of D-AKAP1 as a major adipocyte PKA and PP1 binding protein. Biochemical and Biophysical Research Communications 346:351–7.

Carlson, C. R., Ruppelt, A., & Taskén, K. (2003). A kinase anchoring protein (AKAP) interaction and dimerization of the RIα and RIβ regulatory subunits of protein kinase A in vivo by the yeast two hybrid system. ***J. Mol. Bio***. *327*(3), 609-618.

Castro, L. F. C., Tocher, D. R., & Monroig, O. (2016). Long-chain polyunsaturated fatty acid biosynthesis in chordates: Insights into the evolution of Fads and Elovl gene repertoire. ***Prog. Lipid Res.*** 62, 25-40.

Cecchinato A, Macciotta NPP, Mele M et al (2019) Genetic and genomic analyses of latent variables related to the milk fatty acid profile, milk composition, and udder health in dairy cattle. ***J. Dairy. Sci.*** 102:5254– 5265. https://doi.org/10.3168/jds.2018-15867.

Chen LZ, Xia HH, Xin YN, Lin ZH, Xuan SY. (2015). TM6SF2 E167K variant, a novel genetic susceptibility variant, contributing to nonalcoholic fatty liver disease. ***J. Clin. Transl. Hepatol***. 3:265–270. doi: 10.14218/JCTH.2015.00023.

Christensen, O.F., Lund, M.S. (2010). Genomic prediction when some animals are not genotyped. ***Genet. Sel. Evol***. 42, 2. https://doi.org/10.1186/1297-9686-42-2.

Conte, G., A. Serra, P. Cremonesi, S. Chessa, B. Castiglioni, A. Cappucci, E. Bulleri, and M. Mele. (2016). Investigating mutual relationship among milk fatty acids by multivariate factor analysis in dairy cows. ***Livest. Sci.*** 188:124–132.

Dadousis, C., C. Cipolat-Gotet, G. Bittante, and A. Cecchinato. (2017). Inferring genetic parameters on latent variables underlying milk yield and quality, protein composition, curd firmness and cheese-making traits in dairy cattle. ***Animal***. https://doi.org/10 .1017/S1751731117001616.

Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P, Carlsson LM, Kiess W, Vatin V, Lecoeur C, Delplanque J, Vaillant E, Pattou F, Ruiz J, Weill J, Levy-Marchal C, Horber F, Potoczna N, Hercberg S, Le Stunff C, Bougneres P, Kovacs P, Marre M, Balkau B, Cauchi S, Chevre JC, Froguel P. (2007). Variation in FTO contributes to childhood obesity and severe adult obesity. ***Nat Genet.*** 39:724–726.

Eaton, S. (2002). Control of mitochondrial β-oxidation flux. Progress in lipid research, 41(3), 197-239.

Fredriksson, R. et al. (2008). The obesity gene, *FTO*, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. ***Endocrinology***. 149, 2062–2071.

Gebreyesus, G., A. J. Buitenhuis, N. A. Poulsen, M. H. P. W. Visker, Q. Zhang, H. J. F. van Valenberg, D. Sun, and H. Bovenhuis. (2019). Multi-population GWAS and enrichment analyses reveal novel genomic regions and promising candidate genes underlying bovine milk fatty acid composition. ***BMC Genomics.*** 20:178. https://doi .org/10.1186/s12864-019-5573-9.

Grajales, S. M. B., Zuluaga, J. J. E., Herrera, A. L., Osorio, N. R., & Vergara, D. M. B. (2020). RNA-seq differential gene expression analysis in mammary tissue from lactating dairy cows supplemented with sunflower oil. ***Ani. Prod. Sci.*** 60(6), 758-771.

Hammami, H., J. Vandenplas, M.-L. Vanrobays, B. Rekik, C. Bas- tin, and N. Gengler. (2015). Genetic analysis of heat stress effects on yield traits, udder health, and fatty acids of Walloon Holstein cows. ***J. Dairy Sci***. 98:4956–4968.

Han, Z. et al. (2010). Crystal structure of the FTO protein reveals basis for its substrate specificity. ***Nature***. 464, 1205–1209.

Hein, L., L. P. Sorensen, M. Kargo, and A. J. Buitenhuis. (2018). Genetic analysis of predicted fatty acid profiles of milk from Danish Holstein and Danish Jersey cattle populations. ***J. Dairy Sci***. 101:2148–2157.

Iung LHS, Petrini J, Ramírez-Díaz J, Salvian M, Rovadoscki GA, Pilonetto F, Dauria BD, Machado PF, Coutinho LL, Wiggans GR, Mourão GB. (2019). Genome-wide association study for milk production traits in a Brazilian Holstein population. ***J. Dairy Sci***. 1:1–10. doi: 10.3168/jds.2018-14811.

Jeff, J.M.; Armstrong, L.L.; Ritchie, M.D.; Denny, J.C.; Kho, A.N.; Basford, M.A.; Wolf, W.A.; Pacheco, J.A.; Li, R.; Chisholm, R.L.; et al. (2014). Admixture mapping and subsequent fine mapping suggests a biologically relevant and novel association on chromosome 11 for type 2 diabetes in African Americans**. PLoS ONE**. 9, e86931.

Kitamura, T., Naganuma, T., Abe, K., Nakahara, K., Ohno, Y., & Kihara, A. (2013). Substrate specificity, plasma membrane localization, and lipid modification of the aldehyde dehydrogenase ALDH3B1. ***Biochim. Biophys. Acta Mol. Cell Biol. Lipids***. 1831(8), 1395-1401.

Leal-Gutiérrez, J. D., Rezende, F. M., Elzo, M. A., Johnson, D., Peñagaricano, F., & Mateescu, R. G. (2018). Structural equation modeling and whole-genome scans uncover chromosome regions and enriched pathways for carcass and meat quality in beef. ***Front. Genet***. 9, 532.

Lemos, M. V., Chiaia, H. L., Berton, M. P., Feitosa, F. L., Aboujaoud, C., Camargo, G. M., Pereira, A. S., Albuquerque, L. G., Ferrinho, A. M., Mueller, L. F., Mazalli, M. R., Furlan, J. J., Carvalheiro, R., Gordo, D. M., Tonussi, R., Espigolan, R., Silva, R. M., de Oliveira, H. N., Duckett, S., Aguilar, I., Baldi, F. (2016). Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. ***BMC Genom***. 17, 213. https://doi.org/10.1186/s12864-016-2511-y.

Liu, J., Wang, Z., Li, J., Li, H., & Yang, L. (2020). Genome-wide identification of Diacylglycerol Acyltransferases (DGAT) family genes influencing Milk production in Buffalo. ***BMC Genet***. 21(1), 1-14.

Magalhães, A. F. B. (2015). Utilização de informações genômicas para o melhoramento genético de características da carne em bovinos da raça Nelore.

Mahdessian, H. Taxiarchis, A., Popov, S., Silveira, A., Franco-Cereceda, A., Hamsten, A., ... & van't Hooft, F. (2014). TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. ***Proc. Natl. Acad. Sci.*** USA 111, 8913–8918.

Marrades MP, Gonzalez-Muniesa P, Martinez JA, Moreno-Aliaga MJ. (2010). A dysregulation in CES1, APOE and other lipid metabolism-related genes is associated to cardiovascular risk factors linked to obesity. ***Obesity Facts***. 3: 312–8.

Medeiros de Oliveira Silva, R., Bonvino Stafuzza, N., de Oliveira Fragomeni, B., Miguel Ferreira De Camargo, G., Matos Ceacero, T., Noely Dos Santos Gonçalves Cyrillo, J., et al. (2017). Genome-wide association study for carcass traits in an experimental nelore cattle population. ***PLoS ONE***. 12:e0169860. doi: 10.1371/journal.pone.0169860.

Mehdi, H., Aston, C. E., Sanghera, D. K., Hamman, R. F., & Kamboh, M. I. (1999). Genetic variation in the apolipoprotein H (β2-glycoprotein I) gene affects plasma apolipoprotein H concentrations. ***Human Genet***. 105(1-2), 63-71.

Mele, M., N. P. P. Macciotta, A. Cecchinato, G. Conte, S. Schiavon, and G. Bittante. (2016). Multivariate factor analysis of detailed milk fatty acid profile: Effects of dairy system, feeding, herd, parity, and stage of lactation. ***J. Dairy Sci***. 99:9820–9833. https://doi.org/10 .3168/jds.2016-11451.

Misztal I., Tsuruta S., Strabel T., Auvray B, Druet T., Lee D. (2002) BLUPF90 and related programs (BGF90). In: Pro- ceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier (France). CD-ROM Communication.

Muñoz, M., Rodríguez, M. C., Alves, E., Folch, J. M., Ibañez-Escriche, N., Silió, L., & Fernández, A. I. (2013). Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. ***BMC Genom***. 14(1), 845.

Olasege, B. S., Zhang, S., Zhao, Q., Liu, D., Sun, H., Wang, Q., ... Pan, Y. (2019). Genetic parameter estimates for body conformation traits using composite index, principal component, and factor analysis. ***J. Dairy Sci.*** 102, 5219–5229. https://doi.org/10.3168/ jds.2018-15561.

Palmquist DL. (2006). Milk fat: Origin of fatty acids and influence of nutritional factors thereon. ***Advanced Dairy Chemistry: Lipids***. Edited by: Fox PF, McSweeney PLH, New York, NY: Springer, 2: 43-92.

Palombo, V., Conte, G., Mele, M., Macciotta, N. P. P., Stefanon, B., Marsan, P. A., & D'Andrea, M. (2020). Use of multivariate factor analysis of detailed milk fatty acid profile to perform a genome-wide association study in Italian Simmental and Italian Holstein. ***J. App. Genet***. 61(3), 451–463. https://doi.org/10.1007/s1335 3-020-00568-2.

Palombo, V., Milanesi, M., Sgorlon, S., Capomaccio, S., Mele, M., Nicolazzi, E., Ajmone-Marsan, P., Pilla, F., Stefanon, B., & D'Andrea, M. (2018). Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays. ***J. Dairy Sci***. 101(12), 11004–11019. https://doi.org/10.3168/ jds.2018-14413.

Park, C. H., Ranaraja, U., Dang, C. G., Kim, J. J., & Do, C. H. (2020). Genetic parameters for milk fatty acid composition of Holstein in Korea. ***Asian. Australas. J. Anim. Sci***. 33(10), 1573–1578. https://doi.org/10.5713/ajas.19.0820.

Pegolo S., Dadousis C., Mach N., Ramayo-Caldas Y., Mele M., Conte G., Schiavon S., Bittante G., Cecchinato A. (2017). SNP co-association and network analyses identify E2F3, KDM5A and BACH2 as key regulators of the bovine milk fatty acid profile. ***Sci. Rep***., 7, p. 17317.

Pegolo, S., Momen, M., Morota, G., Rosa, G. J., Gianola, D., Bittante, G., & Cecchinato, A. (2020). Structural equation modeling for investigating multi-trait genetic architecture of udder health in dairy cattle. ***Sci. Rep***. 10(1), 1-15.

Peñagaricano, F., Weigel, K. A., Rosa, G. J., Khatib, H. (2013). Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. ***Front. Genet.*** 3, 307. doi: 10.3389/fgene.2012.00307.

Peñagaricano, F., Valente, B. D., Steibel, J. P., Bates, R. O., Ernst, C. W., Khatib, H., et al. (2015). Searching for causal networks involving latent variables in complex traits: application to growth, carcass, and meat quality traits in pigs1. ***J. Anim. Sci.*** 93, 4617–4623. doi: 10.2527/jas.2015-9213.

Petrini J, Iung LHS, Rodriguez MAP, Salvian M, Pértille F, Rovadoscki GA, Cassoli LD, Coutinho LL, Machado PF, Wiggans GR and Mourão GB. (2016). Genetic parameters for milk fatty acids, milk yield and quality traits of a Holstein cattle population reared under tropical conditions. ***J. Anim. Breed. Genet***. 133, 384–395. doi:10.1111/jbg.12205.

Petrini, J., Souza Iung, L. H., Petersen Rodriguez, M. A., Salvian, M., Alberto Rovadoscki, G., Colonia, S. R. R., Cassoli, L. D., Lehmann Coutinho, L., Fernando Machado, P., Wiggans, G., & Mourão, G. B. (2019). Assessing the accuracy of prediction for milk fatty acids by using a small reference population of tropical Holstein cows. ***J. Anim. Breed. Genet.*** *136*, 453–463. https://doi.org/10.1111/jbg.12434.

Poppelreuther M, Rudolph B, Du C, Großmann R, Becker M, Thiele C, Ehehalt R, Füllekrug J. (2012). The N-terminal region of acyl-CoA synthetase 3 is essential for both the localization on lipid droplets and the function in fatty acid uptake. ***J. Lipid. Res.*** 53:888–900.

R Development Core Team 2011. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna. Retrieved from: https://www.R-project.org/

Rosseel Y. 2012. lavaan: An R package for structural equation modeling. J. Stat. Softw.48:1–36.

Sargolzei, M. SNP1101 User's guide. **Version**, v. 1, p. 0, 2014.

Soreze Y, Boutron A, Habarou F et al. (2013). Mutations in human lipoyltransferase gene LIPT1 cause a Leigh disease with secondary deficiency for pyruvate and alpha-ketoglutarate dehydrogenase. ***Orphanet J. Rare. Dis.*** 8:192.

Steiger J. 1990. Structural model evaluation and modification: An interval estimation approach. ***Multivariate Behav***. Res.25:173–180. doi:10.1207/s15327906mbr25024.

Tews, D., Fischer-Posovszky, P., & Wabitsch, M. (2011). Regulation of FTO and FTM expression during human preadipocyte differentiation. Hormone and Metabolic Research, 43(01), 17-21.

Tucker L. Lewis C. 1973. A reliability coefficient for maximum likelihood factor analysis. Psychometrika.38:1–10. doi:10.1007/BF02291170.

Walker G. Dunshea F. Doyle P. (2004). Effects of nutrition and management on the production and composition of milk fat and protein: A review. ***Aust. J. Agric***. 55: 1009-1028.

Wanders, R. J., & Waterham, H. R. (2006). Biochemistry of mammalian peroxisomes revisited. ***Annu. Rev. Biochem***. 75, 295-332.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. ***Genet. Res***. (Camb.) 94, 73–83. doi: 10.1017/S0016672312000274.

Wickramasinghe, S., Hua, S., Rincon, G., Islas-Trejo, A., German, J. B., Lebrilla, C. B., & Medrano, J. F. (2011). Transcriptome profiling of bovine milk oligosaccharide metabolism genes using RNA-sequencing. ***PLoS ONE***. 6(4), e18895.

White, H. M., Koser, S. L., & Donkin, S. S. (2011). Differential regulation of bovine pyruvate carboxylase promoters by fatty acids and peroxisome proliferator-activated receptor-α agonist. ***J. Dairy Sci.*** 94(7), 3428-3436.

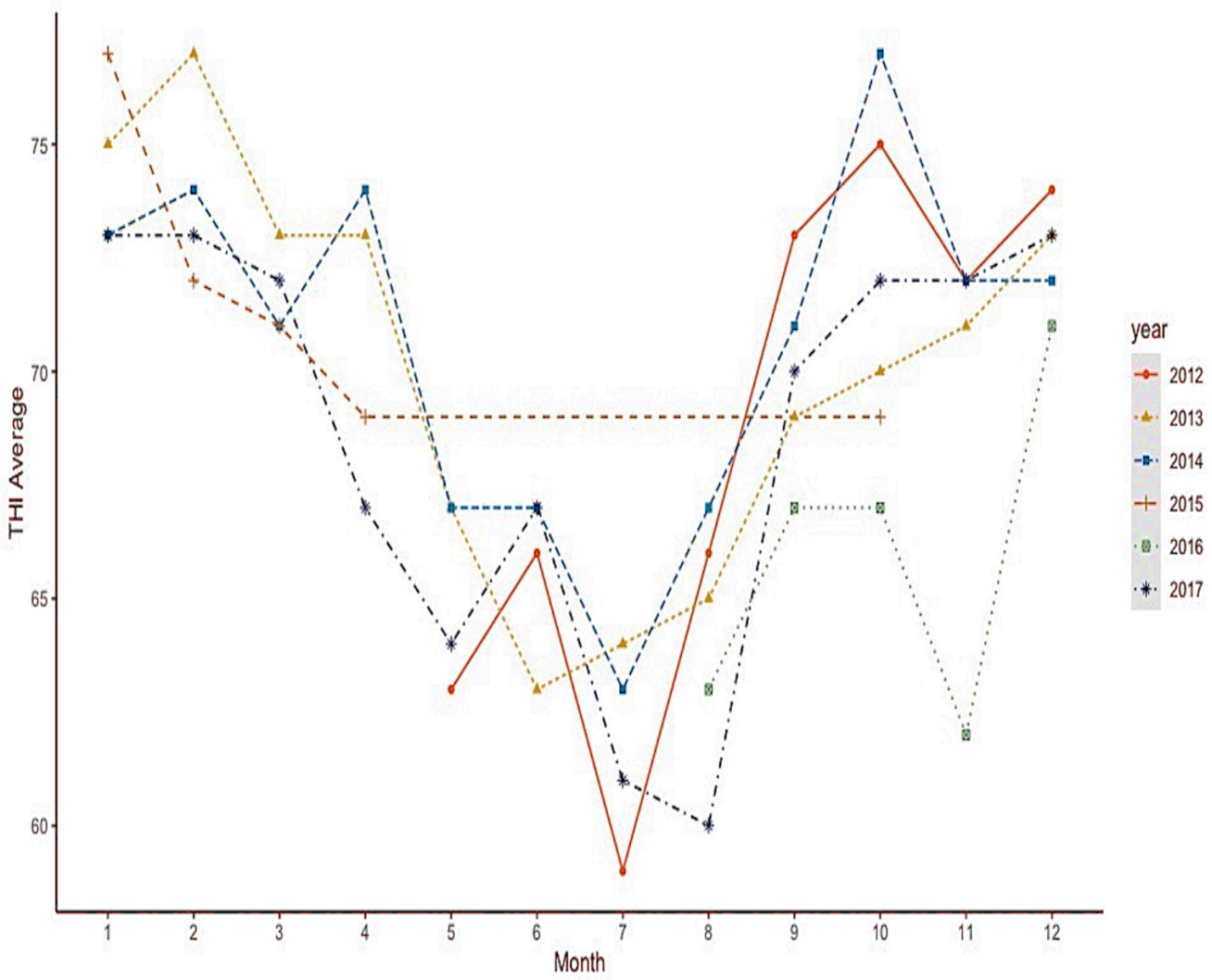
Zielke, L., Bortfeldt, R., Tetens, J., Thaller, G., & Brockmann, G. (2013). The role of obesity genes for milk fat yield in Holstein dairy cattle. In *10th World Congress on Genetics Applied to Livestock Production*.

4. FINAL CONSIDERATION

The study performed in Chapter 2 shows that test-day milk yield records combined with weather data from public stations provide a valuable source of information for evaluated the effects of heat stress in dairy cattle. Remarkable differences between estimates heritability under heat stress and thermal-neutro conditions indicated more sensitive for some milk FA unsaturated. Thus, it reinforced the hypothesis that milk FA derived from diet can be potential targets for capture physical changes (feed intake, water onsumption) affected to heat stress. Although milk FA and climate records are not relevant for breeding purposes, we believe that both information is a valuable tool to define the best bulls for the tropical conditions and could also support the management systems adoptions for mitigating heat stress issues.

In chapter 3, we demonstrated that latent variables for milk FA were similar in the genetic correlation between parities. However, our experimental design provides a relevant opportunity for investigating other phenotypes defined with biological roles and pathways in common. Thereby, multivariate approach could be a valuable tool for the analysis of quantitative and complex traits. Moreover, this method has a relevant statistical advantage to reduce large datasets from breeding programs. In our data, we were able to reduce the number of variables with redundant information. Whole-genome mapping results suggested that genes were different between lactations in most cases, except for only one chromosome window. Although most of them had no association with biological function in bovine milk FA, it would still provide novel candidate genes for further studies. We carried ou complementary analysis (enrichment analysis) of GWAS. The results indicated that all of the significant functional terms were of genes different than was revealed in the whole-genome scan. However, relevants genes were found related to fatty acids byosinthesis and lipid metabolism in this analysis.

SUPPLEMENTARY FIGURES



Supplementary figure 2.1. Distribution of THI averages monthly across six years.

Lactation	Database	ID	Functional Terms	Genes	P-value
1	GO	GO:0044539	Long-chain fatty acid import into cell	*ACSL3, ACSL1*	0.037
1	GO	GO:0004467	Long-chain fatty acid-CoA ligase activity	*ACSL3, ACSL1*	0.048
1	KEGG	bta00061	Fatty acid biosynthesis	*ACSL3, MCAT, ACSL1*	0.010
1	Msig	M26553	Catalysis of the reaction: ATP + medium-chain carboxylic acid + CoA = AMP + diphosphate + acyl-CoA	*ACSL3, ACSL1*	0.039
2	Mesh	D010738	Type C Phospholipases	*XDH, RGS7, EGFR*	0.037
2	Msig	M15245	The chemical reactions and pathways involving linoleic acid	*CYP2J2, ELOVL2, ACSL1*	0.004
2	Msig	M13605	The chemical reactions and pathways involving alpha-linolenic acid	*ELOVL2, ACSL1*	0.042
2	Msig	M23818	The chemical reactions and pathways resulting in the formation of a fatty-acyl-CoA	*SLC25A1, ELOVL2, ACSL1*	0.043
3	GO	GO:0009395	Phospholipid catabolic process	*ABHD12, ENPP2*	0.030
3	Mesh	D050356	Lipid Metabolism	*CPT2, CD36, ARF1, GABARAPL2*	0.041
3	Msig	M29050	The chemical reactions and pathways involving glucosylceramides	*GBA3, UGCG*	0.022
3	Msig	M22880	The chemical reactions and pathways resulting in the breakdown of fatty acids with a chain length of less than C6	*ACADS, MUT*	0.030
3	Msig	M12334	Involved in the maintenance of an internal steady state of fatty acid within an organism or cell	*DGAT2, MLXIPL*	0.049

Supplementary figure 2.2. List of database, functional terms, genes and P-values from gene-set analysis.

SUPPLEMENTARY TABLES

Supplementary Table 3.1. Standardized parameter estimates from the measurement model for first parity.

Latent variables	*N*	*Estimate*	*SE*	*Zvalue*	R^2
Early					
Sat	1,015	0.61	0.012	20.35	0.37
Poly	1,015	0.83	0.001	31.18	0.69
C18:0	1,015	0.81	0.003	30.20	0.66
C18:1	1,015	0.91	0.004	35.38	0.82
Middle					
Sat	957	0.70	0.013	23.69	0.49
Poly	957	0.77	0.001	26.79	0.59
C18:0	957	0.78	0.003	27.59	0.62
C18:1	957	0.93	0.004	34.96	0.87
Late					
Sat	1,110	0.67	0.013	23.53	0.44
Poly	1,110	0.76	0.001	28.27	0.58
C18:0	1,110	0.78	0.003	29.10	0.61
C18:1	1,110	0.91	0.004	35.54	0.83

Abreviatture: N: number of records, estimate: Factor loadings, SE: standard error, R^2: coefficient of variation.

Supplementary table 3.2. Standardized parameter estimates from the measurement model for second parity.

Latent variables	*N*	*Estimate*	*SE*	*Zvalue*	R^2
Early					
Sat	887	0.74	0.015	25.57	0.74
Poly	887	0.86	0.001	31.53	0.86
C18:0	887	0.87	0.004	32.27	0.87
C18:1	887	0.96	0.005	37.74	0.96
Middle					
Sat	856	0.74	0.015	24.54	0.55
Poly	856	0.83	0.001	29.11	0.70
C18:0	856	0.80	0.004	27.51	0.64
C18:1	856	0.94	0.004	34.87	0.89
Late					
Sat	734	0.80	0.015	25.64	0.65
Poly	734	0.79	0.001	25.35	0.63
C18:0	734	0.82	0.004	26.27	0.67
C18:1	734	0.85	0.005	33.06	0.91

Abreviatture: N: number of records, estimate: Factor loadings, SE: standard error, R^2: coefficient of variation

Supplementary table 3.3. Standardized parameter estimates from the measurement model for third parity.

Latent variables	*N*	*Estimate*	*SE*	*Zvalue*	R^2
Early					
Sat	487	0.69	0.021	16.92	0.48
Poly	487	0.88	0.002	24.48	0.78
C18:0	487	0.87	0.005	24.08	0.77
C18:1	487	0.93	0.007	26.42	0.86
Middle					
Sat	469	0.77	0.021	19.50	0.60
Poly	469	0.84	0.002	22.41	0.72
C18:0	469	0.85	0.005	22.53	0.72
C18:1	469	0.96	0.007	27.50	0.93
Late					
Sat	319	0.82	0.026	17.54	0.68
Poly	319	0.83	0.002	17.92	0.69
C18:0	319	0.89	0.006	19.82	0.79
C18:1	319	0.94	0.008	21.98	0.89

breviatture: N: number of records, estimate: Factor loadings, SE: standard error, R^2: coefficient of variation.