HINAYAH ROJAS DE OLIVEIRA

GENETIC AND GENOMIC ANALYSES OF LONGITUDINAL TRAITS IN DAIRY CATTLE

Thesis presented to the Animal Science Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements for degree of Doctor Scientiae.

VIÇOSA
MINAS GERAIS – BRAZIL
2018

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
O48g
2018

Oliveira, Hinayah Rojas de, 1989-
Genetic and genomic analyses of longitudinal traits in dairy cattle / Hinayah Rojas de Oliveira. – Viçosa, MG, 2018.
xvii, 250f. : il. (algumas color.) ; 29 cm.

Texto em inglês.
Inclui apêndice.
Orientador: Fabyano Fonseca e Silva.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Bovinos de leite - Genética. 2. Expressão gênica. 3. Leite - Produção. 4. Lactação. 5. Análise de regressão. I. Universidade Federal de Viçosa. Departamento de Zootecnia. Programa de Pós-Graduação em Zootecnia. II. Título.

CDD 22. ed. 636.2142

HINAYAH ROJAS DE OLIVEIRA

GENETIC AND GENOMIC ANALYSES OF LONGITUDINAL TRAITS IN DAIRY CATTLE

Thesis presented to the Animal Science Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements for degree of *Doctor Scientiae*.

APPROVED: December 6th, 2018.

Flávio Schramm Schenkel
(Co-adviser)

Daniela A. Lino

Daniela Andressa Lino Lourenço
(Co-adviser)

Camila Ferreira Azevedo

Camila Ferreira Azevedo

Domingos Sávio Queiroz

Fabyano Fonseca e Silva
(Adviser)

To all people from the “coca-cola” generation.

I dedicate.

ACKNOWLEDGEMENTS

I believe that each person appear in our life for a reason. They always teach us something important and, when they leave, we are no longer the same people that we were before. I was very lucky during my PhD. I met (and re-met) very special people, which did not only teach me about genetics and genomics, but also taught me how to be a better person and chase my dreams. They helped me to not give up, even when I was scared and did not believe in myself. I could have never done half of what I did in my PhD if I have not met them at some point of my path.

Thus, I would like to show here my thanks to some of the people that were part of these three wonderful years of my life.

First of all, I would like to say a very big thank you to my advisor, **Dr. Fabyano Fonseca e Silva**, for all support that he has given me since my Master's degree. Fabyano, you gave me a lot of opportunities that I could have never imagined. You were always there ready to help me in everything that I needed. You always helped me to achieve my dreams as if they were yours, and you believed in my potential even when I did not. With your "crazy" and "science-passionate" way of being you taught me how to like science, and how to be happy by helping others. You are the best advisor that I could have asked for, and I am very proud of being your student. Thank you very much for everything that you have done for me.

Dr. Flavio Schenkel, I could have never imagined that someone important like you would be so nice and kind. You showed to me that we do not need to give up on things that we like to achieve our professional dreams, and how important is to have a balance between our professional and personal life. Without even realizing it, you taught me how discipline, time and organization are important to keep us going in the right direction. With your paternal heart and kind words you made me realize that we do not need to rush, and that we always can make things better. You are an example of professional and person for me. Thank you very much for all opportunities that you gave me and, especially, thank you for making me feel welcome in your research group and in your house.

Dr. Daniela Lourenco, who could imagine that someone so young could be so smart and have achieved so much in life! Your exciting way of seeing things, your willingness to help everybody, and your efforts to make Georgia's research group united and happy had a great impact in my life. Thank you for always being so accessible to me, and for all hours that you spent with me trying to make things work and explaining concepts. You have no idea how much I have learned from you! I am very happy to have meet you and I really hope that we can work together again in the future. You are an amazing person and professional.

If it is true that guardian angels sometimes appear in our lives as friends, I think you are mine, **Dr. Luiz Fernando Brito**. You have been by my side since my undergrad, helping, teaching and cheering for me. I have never met someone that is so benevolent and so enthusiastic to see other people's achievements. You proved to me how far we can go if we work hard, and how life bring us good surprises when we do our best. But especially, you taught me how much easier our life is when we have someone that we can always count on. All our discussions/partnerships/projects/travels/beers/parties were fundamental in making me who I am today. So, thank you for always believing in my ideas, encouraging me and doing your best to help me in everything that I needed. For sure you were my "right hand" in this journey.

Dr. Marcos Vinícius Gualberto, it was a pleasure to have the opportunity of working and learning from you again, after my undergrad. I admire you a lot for everything that you have been doing for the improvement of the dairy cattle industry in Brazil, especially for the Gyr and Girolando breeds. Thank you very much for all your teachings, suggestions and help during my PhD, especially in the begging of my thesis.

Some people arrive without any intension and end up becoming a fundamental part of our history. And you are one of them, **Dr. Janusz Jamrozik**. Thank you very much for all your prompt replies, patience, and for always answering my e-mails with "1,000 questions" in the best way possible. In addition, thank you very much for all your great suggestions and comments in our studies and, mainly, for all your intriguing questions that made me think a lot from a different perspective.

Dr. John Cant, thank you very much for believing in my ideas and in this project, even without knowing me. Your support and knowledge were fundamental to finalize this thesis. Moreover, thank you for all your great suggestions and comments in our manuscripts, especially in the last chapters of my thesis.

I also would like to say thank you to **Dr. Camila Ferreira Azevedo** and **Dr. Domingos Sávio Queiroz**, for accepting being part of my defense committee, and for all their contributions to this thesis.

During my short stay in Uberaba, I made some very special friends that were very important in the begging of my PhD. Even though our study was not very well related to my thesis, I really appreciated the fact that I could learn more about different things, and I really enjoyed the time that we spent together. Thus, I would like to manifest here my sincere thank you to **Dr. Henrique Torres Ventura** (especially for the great opportunity), **Dr. Edson Vinícius Costa** (for all conversations and friendship), **Dr. Renata Veroneze** (for being amazing), and to **Mariana Alencar**, **Gleida Marques**, **Valeria Castro** and **Jacqueline Araujo** for the great weekends. I had a super fun time with you, guys! I hope we can repeat it in the future.

Of course I also have to thanks all my friends from the "vicious city" (Vicosa), where all dreams start. Over these 10 years living in Vicosa is impossible do not think about a lot of special people that are/were very important in my life. Thus, to represent all my "old friends" in Vicosa, I would like to say a special thank you to **Patrícia Abrahão** and **Bruno Carvalho**. It does not matter how physically far we are, our friendship and affection for each other is always the same. Thank you for being with me since my first year! For the friends from my PhD program, I have to say thanks to people from "Salinha do Melhoramento" and "Labtec". All conversations/discussions/courses/projects/GDMAmeetings/volleyballGames/barbecues/beers were very important to me. In special, I would like to thanks my good friends **José Teodoro Paiva**, **Hugo Teixeira Silva**, **Nadson Souza**, **Vinicius Junqueira**, **Alessandra Alves**, **Darlene Duarte**, **Delvan Alves**, **Felipe Gomes**, **Giovani Caetano**, **Ivan Carvalho**, **Laene Alcântara**, **Lais Brito**, **Leiri Barili**, **Leonardo Gloria**, **Pedro Freitas**, **Renata Henriques**, **Rodrigo Godinho**, **Rodrigo Reis**, **Sirlene Lazaro**, **Thaís Cristina**, and **Otavio Henrique Siqueira**. Thank you for being such amazing guys! At this point, I would like to emphasise my thank you to **Dr. Ricardo Frederico Euclydes**, **Dr. Robledo de Almeida Torres**, **Dr. Paulo Sávio Lopes** and **Dr. Marcelo Teixeira Rodrigues**, that have contributed to my training since my undergrad. Discretely I have being learning a lot from you, and I am very glad for the opportunities that you gave me over the years. Thank you very much for your teachings, patience and help.

And because life is a sweet surprise, I also had a chance to meet (and re-meet) really nice people during my stay in Guelph. Thus, I would like to show here my thanks to all people from "CGIL" and "Brazilian Mafia". In special, I would like to say a big thank you to **Aline Guarini**, **Fabieli Feitosa** and **Ivan Lange** (my best surprises!), **Adrien Butty**, **Agnese Balzani**, **Ana Lúcia Puerro**, **Aroa Suarez-Vega**, **Audrey Martin**, **Bayode Makanjuola**, **Carolina Reyes**, **Emhimad Abdalla**, **Erin Massender**, **Evert Varela**, **Gabriele Marras**, **Gerson Oliveira**, **Giovana Vargas**, **Hakimeh Emamgholi**, **Kristen Alves**, **Lucas Alcantara**, **Malane Muniz**, **Maria Amelia Menck**, **Pablo Fonseca**, **Riani Ananda**, **Samir Lahoucine**, **Saranya Narayana**, **Shadi Nayeri**, **Stephanie Lam**, **Tatiane Chud**, **Valdecy Cruz**, **Victor Pedrosa** and **Victoria Asselstine**. You made my days much better in Guelph (and in Croatia!). I also have to say thank you to **Mohamad Mansour**, **Paula Olivares**, **Frida Lopez**, **Elmer and Elvira Warkentin**, and to all **Schenkel family** (Sandra, Mariana, Daniel and Bianca), for receiving me so well in their lives and homes. Lastly in Guelph, I need to say thank you to **Bill Szkotnicki**, **Dr. Filippo Miglior**, **Dr. Mehdi Sargolzaei** and **Dr. Larry Schaeffer**, for their specific, but very important contributions during my PhD training. I am very grateful for your help and the opportunity to learn from you.

Another sweet surprise I had during my short stay in Athens. I could never have imagined being so well-welcomed in a place like I was at UGA. The way that everybody works, the lunches together, the spontaneous "coffee-breaks" for discussion, and the proximity between students and professors make the research group so united, that now I understand the mean of "together we will go further". Thus, I would like to say thank you for all of you that made me fell so welcomed. In special, I would like to say a big thank you to **Nedenia Stafuzza** and **Rafael Medeiros** (you are unbelievable guys!), **Amanda Maiorano**, **André Garcia**, **Ashley Ling**, **Breno Fragomeni**, **Heather Bradford**, **Ivan Pocrnic** and **Yvette Steyn**. In addition, I need to say a very special and big thank you to **Dr. Yutaka Masuda** and **Dr. Ignacy Misztal**, for all their help, patience, and availability, during my stay and after I left Athens. I am very glad for meeting you both in person, and I am very thankful for the opportunity of learning from you.

I also would like to say a big thank you for those who are my "safe haven": my family. If I did not had them in my life, I would never be able to accomplish anything. They are the most perfect imperfect-family that someone could ask for. Regardless of anything, they are always there for me, and without them even knowing, they are the ones that make me strong enough to chase my dreams. Thus, thank you very much **Rojas family**! I love you a lot, mom (**Sandra**), the best brother ever (**Murilo**), grandma (**Wanda**), Godmother (**Tania**), the more-siblings-than-cousins (**Mariana** and **Mauricio**), and, of course, the best Godson that I could have asked for (**Bento**). If I had the chance of choosing my family, I would choose all of you again!

And last but not least, I would like to say thanks to the "**Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)**", for the financial support that I received during my PhD. If I did not have the CAPES scholarship I would not be able to achieve this dream. In addition, I would like to say thank you to the "**Emerging Leaders in the Americas Program (ELAP)**", from the Government of Canada, and to the "**European Federation of Animal Science (EAAP)**", that allowed me to go further and show my study in different parts of the world.

“Opportunity dances with those already on the dance floor.”

H. Jackson Brown Jr.

BIOGRAPHY

Hinayah Rojas de Oliveira was born in Bragança Paulista, São Paulo, Brazil. In March 2009, she began her undergraduate studies in Animal Science at the Universidade Federal de Viçosa, located in Viçosa, Minas Gerais, Brazil.

Since the beginning of her Bachelor's degree, she was involved in several activities, including the roles of Trainee, Human Resources Director, President and Advisor in the EJZ Junior Consulting Enterprise; and Projeto Rondon, an extension activity organized by the Brazilian Government in conjunction with Universities to develop sustainable agricultural and environmental as well as cultural and social activities across the country. Hinayah also had the opportunity to study in California (USA) for a couple of months; to do a research internship at EMBRAPA Dairy Cattle (Juiz de Fora, Minas Gerais, Brazil) under the supervision of Dr. Marcos Vinícius Gualberto Barbosa da Silva; and to write weekly articles for a newspaper distributed in the São Paulo state.

She discovered her passion for Genetics and Animal Breeding early in her career, which enabled her to be granted a Scientific Initiation scholarship from the FAPEMIG Funding Agency for three consecutive years, under the joint supervision of Dr. Ricardo Frederico Euclydes and Dr. Robledo de Almeida Torres. In March 2014, Hinayah concluded her undergrad, with an honorable award from the Center of Agricultural Sciences from the Universidade Federal de Viçosa for her excellent academic performance in the Animal Science major courses.

In the same month, she received a CAPES scholarship and started her Master's program in Genetics and Breeding at the Universidade Federal de Viçosa, under the supervision of Dr. Fabyano Fonseca e Silva. During her Masters' degree, she participated as a Chief Financial Officer, in the Students Organization for Training in Genetics and Breeding (GENMELHOR), and as organizer in the Genetic and Animal Breeding Discussion Group (GDMA), from the Department of Animal Science, at the Universidade Federal de Viçosa. She successfully defended her M.Sc. thesis in July 2015.

The desire to expand her training in the area of Quantitative Genetics and Genomics motivated her to pursue a doctoral degree, starting in August 2015. Therefore, she received a CAPES scholarship to do her Ph.D in Genetics and Breeding at the same University, under the supervision of Dr. Fabyano Fonseca e Silva. During her PhD, she had various internship opportunities, including: 1) in the Brazilian Association of Zebu Breeders (ABCZ; Uberaba, Minas Gerais, Brazil), under the supervision of Dr. Henrique Torres Ventura; 2) at the Centre for Genetic

Improvement of Livestock (CGIL), University of Guelph (Guelph, Ontario, Canada), under the supervision of Drs. Flavio S. Schenkel and Luiz F. Brito; and, 3) at the University of Georgia (Athens, Georgia, USA), under the supervision of Dr. Daniela A. L. Lourenco. Due to her great dedication, she worked on several projects, from which the results are summarized in this thesis. Her Ph.D thesis was submitted to the defense committee and will be defended in December 2018.

TABLE OF CONTENTS

ABSTRACT

OLIVEIRA, Hinayah Rojas de, D.Sc., Universidade Federal de Viçosa, December, 2018. **Genetic and genomic analyses of longitudinal traits in dairy cattle.** Adviser: Fabyano Fonseca e Silva. Co-advisers: Daniela Andressa Lino Lourenço, Flavio Schramm Schenkel and Marcos Vinicius Gualberto Barbosa da Silva.

Traits with multiple phenotypic values taking over time are termed longitudinal traits, e.g., milk production. Despite of the great importance of analyzing these traits taking into account their time-dependent nature, the majority of studies on longitudinal traits have converted the repeated records for each animal into a single measure (e.g., average over all time points or accumulated yield), which does not allow any inference about the trait over time. Therefore, the general objective of this thesis was to better understand the genetic and genomic aspects of longitudinal traits over time in dairy cattle. Simulated and real datasets (from Brazilian Gyr and Canadian Ayrshire, Holstein and Jersey dairy cattle breeds) were used in this research. First, breeding values were predicted (EBVs) using a multiple-trait random regression model (RRM) combining Legendre orthogonal polynomials and linear B-splines to simultaneously describe the first and second lactation of Gyr Dairy cattle. Subsequently, genomic predictions, genome-wide association analyses were performed for milk, fat and protein yields, and somatic cell score from the first three lactations of the Canadian dairy breeds using different methodologies, including two-step and single-step genomic best linear unbiased prediction (GBLUP). The performance of the most used deregression methods for non-longitudinal traits for the deregression of cows' and bulls' EBVs for using in genomic evaluation of longitudinal traits was also evaluated, using RRMs and the Canadian Jersey data. In addition, the impact of including information from bulls and their daughters in the training population of multiple-step genomic evaluations was investigated using a simulated population. Combining different functions to model the fixed and random effects in multiple-trait RRMs seems to be a good alternative (based on the goodness-of-fit of model, breeding values and variance component estimates) for genetic modeling of lactation curves in dairy cattle, as shown here for Gyr cattle. Deregressed longitudinal EBVs obtained using well established methods of deregression for non-longitudinal traits can be used for genomic prediction of longitudinal traits. Furthermore, removing the parent average and the genotyped daughters' average from the deregressed EBVs can increase the reliability of genomic estimated breeding values (GEBVs). In Holstein, the reliability of GEBVs predicted using the RRM was in general lower than the reliability from the accumulated 305-d model when using the two-step GBLUP method, however, the RRM provided less biased GEBVs compared to the accumulated 305-d model. The use of single-step GBLUP to predict GEBVs for longitudinal traits based on RRMs increased the reliability and reduced bias of GEBVs compared to traditional parent average, in the Canadian Ayrshire, Holstein, and Jersey breeds. Different genomic regions associated with the analyzed traits were identified for different lactation stages, supporting differential gene control across lactation stages. For all Canadian breeds, the pattern of the effect of several single nucleotide polymorphisms associated with the analyzed longitudinal traits changed over time. In addition, prospective candidate genes with potential different patterns of expression over time were identified in putative chromosomal regions. The findings described in this thesis will contribute to advance the knowledge on the genomic expression and prediction of breeding values for longitudinal traits.

RESUMO

OLIVEIRA, Hinayah Rojas de, D.Sc., Universidade Federal de Viçosa, dezembro de 2018. **Análises genética e genômica de características longitudinais em gado de leite.** Orientador: Fabyano Fonseca e Silva. Coorientadores: Daniela Andressa Lino Lourenço, Flavio Schramm Schenkel e Marcos Vinicius Gualberto Barbosa da Silva.

Características com múltiplos valores fenotípicos registrados ao longo do tempo são denominadas características longitudinais, como por exemplo produção de leite. Apesar da grande importância de analisar essas características levando em conta o tempo, a maioria dos estudos sobre características longitudinais convertem os registros repetidos de cada animal em uma única estimativa (por exemplo, média de todos os tempos ou produção acumulada), o que não permite nenhuma inferência sobre a característica ao longo do tempo. Desta forma, o objetivo geral desta tese foi entender melhor os aspectos genéticos e genômicos de características longitudinais ao longo do tempo em bovinos leiteiros. Dados simulados e reais (do gado brasileiro Gir leiteiro e das raças leiteiras canadenses Ayrshire, Holandesa e Jersey) foram utilizados neste estudo. Primeiro, valores genéticos (EBVs) foram preditos usando um modelo multicaracterístico de regressão aleatória (RRM) combinando polinômios ortogonais de Legendre e B-splines lineares para descrever simultaneamente a primeira e segunda lactação do gado leiteiro Gir. Subsequentemente, predições genômicas e análises de associação genômica ampla e funcional foram realizadas para produção de leite, gordura e proteína, e escore de células somáticas nas três primeiras lactações das raças canadenses usando diferentes metodologias, como o melhor preditor linear não viesado genômico (GBLUP) em um único ou dois passos. O desempenho dos métodos de deregressão mais utilizados para características não longitudinais na deregressão dos EBVs de vacas e touros usados para avaliação genômica de características longitudinais também foram avaliados, usando RRMs e os dados da raça canadense Jersey. Além disso, o impacto da inclusão da informação de touros e suas filhas na população de treinamento em avaliações genômicas de múltiplos passos foi estudado usando uma população simulada. Combinar diferentes funções para modelar os efeitos fixos e aleatórios em RRMs multicaracteristicos parece ser uma alternativa viável (com base no ajuste do modelo, e nas estimativas de valores genéticos e componentes de variância) para modelagem genética das curvas de lactação em vacas leiteiras, como mostrado aqui para o gado Gir. A deregressão dos EBV longitudinais realizada utilizando métodos bem estabelecidos de deregressão para características não-longitudinais pode ser usada para predição genômica de características longitudinais. Além disso, remover a média dos pais e a média das filhas genotipadas do EBV deregredido pode aumentar a confiabilidade dos valores genômicos estimados (GEBVs). Na raça Holandesa, a confiabilidade dos GEBVs preditos usando o RRM foi em geral menor que a confiabilidade do modelo de produção acumulada, ao usar o método GBLUP em dois passos, no entanto, o RRM forneceu GEBVs menos viesados comparado ao modelo de produção acumulada. O uso do GBLUP em um único passo para predizer os GEBVs para características longitudinais baseado em RRMs aumentou a confiabilidade e reduziu o viés dos GEBVs comparado com a tradicional média dos pais, nas raças canadenses Ayrshire, Holandesa e Jersey. Diferentes regiões genômicas associadas às características analisadas foram identificadas para diferentes estágios da lactação, evidênciando o controle diferencial de genes ao longo dos estágios de lactação. Para todas as raças canadenses, o padrão do efeito de vários polimorfismos de nucleotídeo único associados com as características longitudinais analisadas mudou ao longo do tempo. Além disso, potenciais genes candidatos com diferentes padrões de expressão ao longo do tempo foram identificados em diferentes regiões cromossômicas. Os achados descritos nesta tese contribuirão para o avanço do conhecimento sobre a expressão genômica e predição de valores genéticos para características longitudinais.

CHAPTER 1: General introduction

1.1. Introduction

Dairy products provide valuable nutritional source of vitamins and minerals for millions of people around the globe. The world population is expected to reach 9.8 billion by 2050, which will be accompanied by an increased demand of 1,077 million tonnes of milk (FAO, 2017). Over the past century, substantial progress in milk production has been achieved especially due to genetic selection (Miglior et al., 2017). However, in order to meet the increasing demand for food production, more sophisticated selection methods need to be developed in order to fastener genetic progress. The most important milk related traits in dairy cattle (e.g., milk yield and milk constituents) are termed longitudinal traits, as they are measured over time. Longitudinal traits need to be analyzed using specific statistical methods to take into account the well-structured covariance pattern that exists among the repeated records. Over the past two decades, random regression models (**RRM**) have been an optimal method to analyze longitudinal traits across several livestock species. The main advantages of RRM compared to other statistical approaches rely on the fact that RRM enable fitting random genetic and environmental effects over time (Schaeffer, 2004), which consequently results in higher accuracy of estimated breeding values (**EBV**) when compared to other statistical approaches. Thus, using RRM in genetic evaluations of dairy cattle may enable accurate selection of animals with more desirable lactation curve patterns. For instance, cows with lower lactation peak, but with more persistent production, are likely to suffer less from negative energy balance and metabolic diseases, which can improve animal welfare and economic profitability at the farm level (Miglior et al., 2017).

Nowadays, the popularity of RRM has increased substantially in animal breeding, and the RRM have become more efficient through the development of alternative functions and advances in model structures to include, for exemple, multiple-trait analysis. A study in dairy goats showed that it is possible to combine different functions to describe different traits in the same multiple-trait RRM analysis to improve the the accuracy of both breeding values and genetic parameter estimates (Oliveira et al., 2016). However, to our best knowledge, there are no references about the flexibility of simultaneously using different functions to describe the fixed and random curves of different traits. In addition, it is still not well established what are the best methods to account for heterogeneity of residual variance in genetic evaluations under these conditions.

Despite the advancements and practical applications of RRM in breeding programs, there is still a lack of research investigating the appropriateness of using RRM in genomic evaluations. The majority of genomic predictions (and consequently association studies) reported in the literature for longitudinal traits have focused on the accumulated yield as the response variable in a multiple or single-step genomic best linear unbiased prediction (**GBLUP**) approach (e.g., Winkelman et al., 2015; Aliloo et al., 2016; Jenko et al., 2017). In brief, the multiple-step approach requires the estimation of the accumulated EBVs (i.e., EBV for the accumulated yield) based on traditional evaluation (using solely pedigree and phenotypic information) and the estimation of direct genomic value (**DGV**) using the accumulated EBVs from the previous step. Thereafter, the DGVs are combined with the parent average (**PA**) to generate the genomic EBV (**GEBV**). However, in order to avoid double-counting of information and double shrinkage of DGVs by using EBVs as pseudo-phenotypes in the second step, computation of deregressed EBVs (**dEBVs**) has been suggested to perform genomic predictions (Garrick et al., 2009).

There are several methods of EBV deregression described in the literature for non-longitudinal traits (e.g., Garrick et al., 2009; VanRaden et al., 2009; and Wiggans et al., 2011). However, there are no studies comparing the performance of these deregression methods for genomic prediction of longitudinal traits. Furthermore, the impact of using dEBVs from bulls and their daughters as response variables to estimate GEBVs in multiple-step genomic evaluations is still unknown, even though combining information of the dEBVs from bulls and their daughters has been routinely done for small-sized dairy cattle breeds.

The main limitations of using accumulated yield as the response variable for genomic predictions of longitudinal traits are the lack of inference over time and the non-use of the genetic variability observed for lactation patterns. Thus, using RRM in both steps of a multiple-step genomic prediction could be an efficient alternative to predict GEBVs over time, and therefore enable early selection of genotyped animals based on the pattern of the lactation curve without changing the currently method used for official genomic evaluations in several countries (Interbull, 2017). Nonetheless, no studies comparing the reliability and bias of breeding values obtained from the traditional BLUP, multiple-step GBLUP using RRM, and multiple-step GBLUP using the accumulated yield (i.e., 305-d model) were found in the literature.

When compared to the multiple-step GBLUP, studies have shown that simultaneously combining phenotypic records, pedigree, and genomic information in a single-step GBLUP can lead to more accurate and less biased GEBVs (e.g., Christensen and Lund, 2010; Kang et al., 2017).

Using the single-step GBLUP method based on RRMs would allow to predict GEBVs over time in a single step, which could increase the reliability of the genomic prediction for longitudinal traits, as shown by Kang et al. (2017) using a simulated dataset. Notwithstanding, the effect of including only genotypes from animals with accurate EBVs instead of all available genotypes when performing single-step GBLUP analyses, in terms of computational demand and reliability and bias of genomic predictions, is still unknown, especially when using complex models, such as multiple-trait RRM.

Regarding the genome-wide association studies, genes related to longitudinal traits might be differentially expressed over time, and this knowledge could enhance the performance of genomic evaluations, in terms of reliability and bias for specific times, and the understading of the underlying biology of these traits. In a study using German Holstein Friesian cows, Strucken et al. (2012) reported that the variance explained by certain loci changes across lactations. Moreover, it has been shown that the expression of genes related to milk traits in second-or-greater parity Holstein cows is affected by lactation stage (Bionaz and Loor, 2008). In addition, the underlying lactation biology and genetic mechanisms involved in the expression of milk traits might differ across breeds and populations. This highlights the need to investigate candidate genes affecting milk related traits throughout the lactation stages (e.g., beginning, middle and end of lactation) in different breeds. Identifying those differences may enable more efficient genetic selection, as well as the development of better management practices within the herd.

Das et al. (2011) showed that is possible to use RRM to exploit specific covariance functions to describe changes in the single nucleotide polymorphism (**SNP**) effects over time. By analyzing marker associations over time, regions with higher effects at specific stages are more likely to be identified, which could contribute to better detection of the genetic variation of longitudinal traits (Strucken et al., 2015). Currently, this approach is the most sophisticate and flexible tool to assess genomic-wide associations for longitudinal traits. However, despite the SNP availability and several genomic-wide association studies for milk related traits in dairy cattle, the use of RRM to estimate SNP effects over time remains underexplored, especially when considering different dairy cattle breeds.

1.2. Thesis outline

To better understand the alternatives for genetic and genomic evaluations of longitudinal traits, as well as the chronological development of RRMs, Chapter 2 provides a comprehensive literature review on the topic. In order to highlight the flexibility of RRM currently used for genetic evaluations, Chapter 3 focus on the implementation of multiple-trait RRM combining different functions to describe different traits, under a Bayesian approach. Additionally, Chapter 3 also presents the application of an unsupervised learning algorithm to define the classes of residual variances.

Before investigating genomic prediction of breeding values using RRM, some points that were unclear in the literature were further studied. Chapter 4 presents the performance of the most used deregression methods for the deregression of cows' and bulls' EBVs for genomic evaluation of longitudinal traits, while Chapter 5 focus on accessing the impact of including information from bulls and their daughters in the training population of multiple-step genomic evaluations. In addition, a new deregression method that accounts for the parent average and the genotyped daughters' average in the estimation of the deregressed bulls' EBVs was also proposed in the Chapter 5.

To evaluate the impact of different methodologies on genetic and genomic evaluations of longitudinal traits in dairy cattle, Chapter 6 compares the genetic predictions obtained from the traditional BLUP to the genomic predictions obtained from the two-step GBLUP approach, using both RRM and accumulated 305-d models. Furthermore, the impact of incorporating cows in the training and validation populations was also investigated in this chapter. Chapter 7 focus on using the ssGBLUP method to analyze longitudinal traits in the three main Canadian dairy cattle breeds, based on multiple-trait RRM. The effect of including only genotypes from animals with accurate EBVs in the analyses was also investigated in Chapter 7, with focus on computational demand and reliability and bias of genomic predictions.

To better understand the genetic mechanisms underlying the biology of longitudinal traits, Chapter 8 presents a two-step GWAS considering different lactation stages of the three main Canadian dairy cattle breeds, in order to identify prospective candidate genes and biological processes that are specific or shared among dairy cattle breeds and lactation stages. Chapter 9 focus on the implementation of ssGWAS using multiple-trait RRM to estimate SNP effects over time for various longitudinal traits, and to identify putative chromosomal regions and candidate genes with different potential expression pattern over time. Finally, Chapter 10 presents a general conclusion from this thesis.

1.3. Thesis objectives

The overarching objective of this thesis was to better understand the genetic and genomic aspects of longitudinal traits in dairy cattle. The main specific objectives of this thesis were to:

I. Present a comprehensive summary of the scientific contributions in the area of RRM applied to livestock breeding programs and review the applications of RRM incorporating genomic information;
II. Show the flexibility of the RRM currently used for genetic evaluation, combining Legendre orthogonal polynomials and linear B-splines to simultaneously describe the first and second lactation of Gyr dairy cattle under a multiple-trait RRM Bayesian framework;
III. Apply an unsupervised learning algorithm known as self-organizing map to define the classes of residual variances for all lactations analyzed in the same multiple-trait RRM;
IV. Investigate the performance of the most used deregression methods for non-longitudinal traits in the deregression of cows' and bulls' EBVs for genomic evaluation of test-day traits;
V. Investigate the impact of including information from bulls and their daughters in the training population of multiple-step genomic evaluations, and propose a new deregression method that accounts for parent average and the genotyped daughters' average in the estimation of the deregressed bulls' EBV;
VI. Compare the reliability and bias of breeding values obtained from the traditional BLUP, two-step GBLUP using RRM, and two-step GBLUP using accumulated yield (i.e., 305-d yield model);
VII. Investigate the reliability and bias of daily GEBVs estimated using ssGBLUP method to evaluate longitudinal traits from Canadian Ayrshire, Holstein, and Jersey breeds;
VIII. Evaluate the effect of including only genotypes from animals with accurate EBVs in the ssGBLUP, in terms of computational demand and reliability and bias of genomic predictions;
IX. Investigate prospective candidate genes and biological processes related to milk, fat, and protein yield, and somatic cells score that are exclusive or shared among different lactation stages of Canadian Ayrshire, Holstein and Jersey breeds;

X. Estimate SNP effects over time for various longitudinal traits of Canadian Ayrshire, Holstein and Jersey breeds, and to identify putative chromosomal regions and candidate genes with different potential expression pattern over time.

1.4. References

Aliloo, H., J.E. Pryce, O. González-Recio, B.G. Cocks, and B.J. Hayes. 2016. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. Genet. Sel. Evol. 48:8. doi:10.1186/s12711-016-0186-0.

Bionaz, M., and J.J. Loor. 2008. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J. Nutr. 138:1019–1024. doi:10.1093/jn/138.6.1019.

Christensen, O.F., and M.S. Lund. 2010. Genomic relationship matrix when some animals are not genotyped. Genet. Sel. Evol. 42:2. doi.org/10.1186/1297-9686-42-2

Das, K., J. Li, Z. Wang, C. Tong, G. Fu, Y. Li, M. Xu, K. Ahn, D. Mauger, R. Li, and R. Wu. 2011. A dynamic model for genome-wide association studies. Hum. Genet. 129:629–639. doi:10.1007/s00439-011-0960-6.

FAO. 2017. The state of food and agriculture: leveraging food systems for inclusive rural transformation. Rome, Food and Agriculture Organization. ISBN 978-92-5-109873-8.

Garrick, D.J., J.F. Taylor, and R.L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55. doi:10.1186/1297-9686-41-55.

Interbull. 2017. Description of national genomic evaluation systems. Accessed May. 02, 2018. http://www.interbull.org/ib/nationalgenoforms.

Jenko, J., G.R. Wiggans, T.A. Cooper, S.A.E. Eaglen, W.G. de. L. Luff, M. Bichard, R. Pong-Wong, and J.A. Woolliams. 2017. Cow genotyping strategies for genomic selection in a small dairy cattle population. J. Dairy Sci. 100:439–452. doi:10.3168/jds.2016-11479.

Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J.-F. Liu. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity (Edinb). 119:459–467. doi:10.1038/hdy.2016.91.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L.F., Martin, P. and Baes, C.F., 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci., 100(12), 10251-10271.

Oliveira, H.R., F.F. Silva, O.H.G.B.D. Siqueira, N.O. Souza, V.S. Junqueira, M.D. V. Resende, R.R.A. Borquis, and M.T. Rodrigues. 2016. Combining different functions to describe milk, fat, and protein yield in goats using Bayesian multiple-trait random regression models. J. Anim. Sci. 94:1865–1874. doi:10.2527/jas.2015-0150.

Schaeffer, L.R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35–45. doi:10.1016/S0301-6226(03)00151-9.

Strucken, E.M., R.H. Bortfeldt, D.J. de Koning, and G.A. Brockmann. 2012. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375–382. doi:10.1111/j.1365-2052.2011.02278.x.

Strucken, E.M., Y.C.S.M. Laurenson, and G.A. Brockmann. 2015. Go with the flow-biology and genetics of the lactation cycle. Front. Genet. 6:118. doi:10.3389/fgene.2015.00118.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F.S. Schenkel. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24. doi:10.3168/jds.2008-1514.

Wiggans, G.R., T.A. Cooper, P.M. VanRaden, and J.B. Cole. 2011. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. J. Dairy Sci. 94:6188–6193. doi:10.3168/jds.2011-4481.

Winkelman, A.M., D.L. Johnson, and B.L. Harris. 2015. Application of genomic evaluation to dairy cattle in New Zealand. J. Dairy Sci. 98:659–675. doi:10.3168/jds.2014-8560.

CHAPTER 2:

Literature review: Advances and applications of random regression models: from the genetics to the genomics

Hinayah Rojas de Oliveira[1,2], Luiz Fernando Brito[2,3], Daniela Andressa Lino Lourenco[4], Fabyano Fonseca e Silva[1], Janusz Jamrozik[2,5], Lawrence Raymond Schaeffer[2], and Flávio Schramm Schenkel[2]

[1] Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[2] Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

[3] Department of Animal Sciences, Purdue University, West Lafayette, Indiana, United States of America

[4] Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, United States of America

[5] Canadian Dairy Network, Guelph, Ontario, Canada

Journal of Dairy Science, 2018
(Invited review, submitted).

2.1. Abstract

An important goal in animal breeding is to improve longitudinal traits, i.e., traits recorded multiple times during an individual's lifetime or physiological cycle. Longitudinal traits were firstly genetically evaluated based on accumulated phenotypic expression and/or specific time points, or using repeatability models. Up to now, the genetic evaluation of longitudinal traits have mainly focused on using random regression models (**RRM**). The RRM enable fitting random genetic and environmental effects over time, which results in higher accuracy of estimated breeding values compared to other statistical approaches. In addition, RRM provide insights about temporal variation of biological processes and physiological implications underlying the studied traits. Despite the fact that genomic information have substantially contributed to speed up the rates of genetic progress for a variety of economically important traits in several livestock species, less attention has been given to longitudinal traits in the past years. However, including genomic information to evaluate longitudinal traits using RRM is a feasible alternative to yield more accurate selection and culling decisions, because it may enable the selection of young animals based on the complete pattern of the production curve with higher accuracy when compared to the use of traditional parent average. Moreover, RRM can be used to estimate single nucleotide polymorphism effects over time in genome-wide association studies. Thus, by analyzing marker associations over time, regions with higher effects at specific points in time are more likely to be identified. Despite all advances in applications of RRM in the genomic era, more research is still needed. Future research should provide better understanding of the temporal variation of biological processes and their physiological implications underlying the longitudinal traits.

Keywords: GEBV, genetic evaluation, lactation curves, longitudinal traits, test-day

2.2. Introduction

An important goal in animal breeding is to improve longitudinal traits, i.e., traits recorded multiple times during an individual's lifetime or physiological cycle, such as test-day milk yield (**TDMY**), egg production in birds and periodic body weight of growing animals (Mark, 2004; Lund et al., 2008; Schaeffer and Jamrozik, 2008). Before 1994, these traits were genetically evaluated based on accumulated production (e.g., milk production in 305 days), at a specific time point (e.g., yearling weight) or using repeatability models. After 1994, the genetic evaluation of longitudinal traits have mainly focused on random regression models (**RRM**). In brief, RRM enable fitting random genetic and environmental effects over time (Schaeffer, 2004), which results in higher accuracy of estimated breeding values (**EBV**) compared to other statistical models. In addition, RRM provide insights about temporal variation of biological processes and physiological implications underlying the studied traits (Strucken et al., 2015). Therefore, RRM generate relevant information to be exploited in breeding programs. For instance, selecting for an optimal shape of lactation curve in dairy species can reduce metabolic issues during the first weeks of lactation and the lactation peak, for example. In addition, RRM facilitate the selection for increased lactation persistency.

Genomic information has substantially contributed to speed up the rates of genetic progress for a variety of economically important traits in several livestock species (Goddard et al., 2010; Hayes and Goddard, 2010; Buch et al., 2012; Meuwissen et al., 2016). However, despite the great impact of genomics on single-record traits in animal breeding (e.g., García-Ruiz et al., 2016; Raszek et al., 2016; Miglior et al., 2017); plant breeding (e.g., Heffner et al., 2010; Pandey et al., 2016; Slater et al., 2016) and human genetics (e.g., de los Campos et al., 2013), less attention has been given to longitudinal traits in the past years.

Genomic RRM provide an opportunity to unravel genomic regions significantly associated with the traits of interest at specific time points, as well as the trajectory of biological processes during the animal's life or production cycle. Thus, by analyzing marker associations over time, regions with higher effects at specific stages are more likely to be identified, which could contribute to the better detection of the genetic variation of longitudinal traits (Strucken et al., 2015). Furthermore, using RRM in genomic predictions enable the selection of young animals based on the complete pattern of the production curve with higher accuracy when compared to the use of traditional parent average (Oliveira et al., 2018c). Therefore, the main objectives of this review are: 1) to present a comprehensive summary of the scientific contributions in the area of RRM applied to livestock breeding programs, with focus on dairy cattle; 2) to review the applications of RRM incorporating genomic information in a large number of livestock species; and, 3) to discuss the challenges and suggest how genomic information can be successfully incorporated into RRM.

2.3. Development history

In the genetic evaluation of dairy species, TDMY records were first used for genetic evaluation of accumulated 305-day milk yields (Voelker, 1981; Pander and Hill, 1993). In some countries such as Canada, initially the accumulated 305-day milk yield was estimated summing up the TDMY recorded every day during the lactation period. However, recording the amount of milk per cow every day was very costly, which made the countries to use weekly or monthly records in their evaluations. Therefore, the weekly or monthly TDMY were combined by linear interpolation, in order to produce the accumulated 305-day milk yield in the end (Schaeffer, 2016).

Over the years, several statistical methods have been proposed in order to accurately estimate the accumulated yield based on TDMY records (Lamb and McGilliard, 1967; Wiggans and Van Vleck, 1979; Wiggans and Powell, 1980; Wilmink, 1987; Weller, 1988; Trus and Buttazzoni, 1990). Among them, the most efficient methods allowed adjustments to first and last test day and projection of incomplete lactations up to 305 days without reducing the total genetic variance (VanRaden et al., 1991). Thus, the accumulated 305-day yield of each lactation was used as single trait, and each lactation was separately analyzed in a single-trait model (Ferris et al., 1985; Banos and Shook, 1990), or together in a multiple-trait analysis (Rothschild and Henderson, 1979; Tong et al., 1979; Strandberg and Danell, 1989; Short et al., 1990). Even though these methods have comparatively low computational demand (i.e., only one observation per animal per lactation), they do not provide information about the shape of the lactation curve. In addition, as several TDMY were combined into one single record per lactation, effects that were specific to individual TDMY were not taken into account, and projection factors were needed to estimate the accumulated 305-day yield from incomplete lactations (Ptak and Schaeffer, 1993).

Several alternatives were suggested to overcome these problems, such as analyzing the lactation curve parameters (coefficients) previously estimated in a first step (Shanks et al., 1981; Ferris et al., 1985; Ali and Schaeffer, 1987), and/or dividing the complete lactation into stages (e.g., months or trimesters of lactation) and analyzing these periods as separate traits in a multiple-trait analysis (Keown and Van Vleck, 1971; Auran, 1973; Schaeffer and Burnside, 1976; Lindgren et al., 1980; Agyemang et al., 1985). Attempts to use individual TDMY records as phenotypes in genetic evaluations were also evaluated in order to avoid loss of information and the approximations needed to project to 305-day lactations. In this case, TDMY records were considered as repeated measurements of the same trait, modeled using repeatability models (Pander and Hill, 1993; Ptak and Schaeffer, 1993; Carriedo et al., 1995; Swalve, 1995). Alternatively, each TDMY could also be considered as a different trait in a more complete and over-parameterized multiple-trait model approach, as suggested by Misztal et al. (2000).

Using both repeatability and multiple-trait models to analyze individual TDMY records presented considerable disadvantages. Firstly, evaluating TDMY records based on a repeatability model assumes that the genetic correlation between the records is equal to 1, which is unrealistic

(Ptak and Schaeffer, 1993; Vargas et al., 1998; Jensen, 2001). Secondly, assuming each TDMY record as a different trait generates a complex model with relatively small number of observations being analyzed for each trait. In order to decrease the number of traits in these multiple-trait models, studies defining trait as the test number after calving instead of the days in milk (**DIM**) were also performed (Danell, 1982; Meyer et al., 1989).

For genetic evaluation of other traits such as body weight, efforts were made to simplify the models and data collection, recognizing that the cost (in terms of time, labour, and disruption to the animal's daily life) of taking more measurements can be significant. For instance, this was done assuming specific standard ages to weight recording, such as birth, weaning and yearling (e.g., Kennedy and Henderson, 1975; Schaeffer and Wilton, 1981; Soller et al., 1984; Trus and Wilton, 1988). The other alternative was to analyze the growth curve parameters (coefficients) estimated in a preliminary phenotypic analysis (e.g., Brown et al., 1972; Kersey and Brinks, 1985; Meyer, 1995) instead of actual weights measured at different ages, or assuming weights as repeated measurements using a repeatability model, as pointed out by Meyer (1999).

In general, those alternatives used to model longitudinal traits did not take into account that the correlation between phenotypes recorded at closer times or ages was higher than the correlation between more distant periods, and that the genetic and environmental variances may change over time (Ptak and Schaeffer, 1993). Thus, because the covariance pattern among the repeated measurements of longitudinal traits is well structured, different statistical methods are required to better account for the trait structure compared to single-measurement traits. Nowadays, analyses of longitudinal data are carried out mainly based on RRM (e.g., Mota et al., 2018a; Padilha et al., 2018; Saghi et al., 2018; Yin and König, 2018; Zhao et al., 2018).

2.4. Random regression models

Random regression models are an extension of the repeatability models where the random genetic animal effect can vary for each time period (Lidauer et al., 2003), without greatly increasing model complexity, as occurs in the multiple-trait models. Random regression models were initially proposed by Henderson Jr. (1982) and Laird and Ware (1982) as an alternative to overcome the problem of over-parametrization in multiple-trait analyses. However, the potential for practical application was only presented a decade later by Schaeffer and Dekkers (1994), in which the authors used RRM to analyze TDMY records in dairy cattle. In addition, the first genetic parameters for TDMY from RRM were published by Jamrozik and Schaeffer (1997), and the first results of genetic evaluation for milk yield derived from RRM were presented by Jamrozik et al. (1997b).

According to Schaeffer (2004), a RRM should contain a set of fixed curves describing the phenotypic relationship among the various groups of individuals, and a set of random regressions, which should include those for the additive genetic and the permanent environmental effects. Consequently, the genetic and permanent environmental correlation between phenotypes at different times can be different from 1. In a general form, a RRM can be described as follows:

$$y_{ghij} = \sum_{p=1}^{P} b_{pg} z_{pg} + \sum_{q=1}^{Q} h_{qh} z_{qh} + \sum_{r=1}^{R} a_{ri} z_{ri} + \sum_{s=1}^{S} p_{si} z_{si} + e_{ghij}\,, \quad [1]$$

where y_{ghij} is the j^th^ repeated record of animal i (e.g., TDMY records, weight at different ages); b_{pg} is the pth fixed regression coefficient for the gth group; h_{qh} is the qth random regression coefficient for hth parity-herd-year-season of calving effect; a_{ri} is the rth random regression coefficient for additive genetic effect of the ith animal; p_{si} is the sth random regression coefficient for permanent environmental effect of the ith animal; e_{ghij} is the residual effect (which can have different variances depending on time); and z_{pg}, z_{qh}, z_{ri} and z_{si} are Legendre polynomial covariates related to the function used to describe time (e.g., DIM, age at measurement), assuming here the same function with possible different orders P, Q, R, and S (e.g., linear, quadratic, cubic). It is worth to note that the fixed regressions for the groups are important to account for time trends in the phenotypes. In dairy cattle, usually the group could be defined as year-month of calving, and/or the age-season of calving to account for time trends. In addition, it is important to note that besides the random regressions for the additive genetic and the permanent environmental effects, a random regression should be assumed for the parity-herd-year-season of calving effect in dairy cattle evaluations. The parity-herd-year-season of calving effect should be used to account for cows that share the same environmental effects throughout their lactation (i.e., contemporaries) instead of the fixed herd-year-season effect that was being wrongly used (Schaeffer, 2016). Therefore, as argued by Schaeffer (2018), herd-test-day effects are no longer (and should never have been) recommended for test-day models.

In theory, the covariates z can be any covariate, however, the covariates used in genetic evaluation of livestock species have been based on relatively simple polynomial functions that are known to fit the data reasonably well along time (Jensen, 2001). Numerous linear and nonlinear functions have been proposed for modeling the trajectory of effects included in the model. Among the most used linear functions are orthogonal polynomial functions, such as Legendre polynomials

(Kirkpatrick et al., 1990), and the smoothing polynomial functions, such as splines (Wegman and Wright, 1983; White et al., 1999) and B-splines (de Boor, 1980; Meyer, 2005) functions. Linear and nonlinear parametric functions have also been described to analyze milk related traits, such as by Wood (1967), Ali and Schaeffer (1987) and Wilmink (1987); and to describe growth rate traits, such as by Gompertz (1825), Brody and Lardy (1946) and von Bertalanffy (1957). Nonetheless, due to the difficulties to implement the nonlinear parametric functions as regressions in RRM in currently available software for genetic evaluation, these functions have being linearized before their use (e.g., Oliveira et al., 2016). A more in depth comparison and discussion on the functions currently used as regressions in RRM is presented latter in the section: "Comparison of Alternative *Functions to Model Trajectory Curves"*.

Henderson's Mixed Model Equations (Henderson, 1963, 1973; Henderson Jr., 1982) for RRM can be described as follows:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{A}^{-1}\otimes\mathbf{G}_0^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W}+\mathbf{I}\otimes\mathbf{P}_0^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \\ \hat{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}, \quad [2]$$

where **y** is the vector of observations, **X, Z** and **W** are the incidence matrices for the fixed (**b**), additive genetic random regression coefficients (**a**) and permanent environmental random regression coefficients (**p**), **A** is the additive genetic numerator relationship matrix based on pedigree information, and $\mathbf{I}$ is an identity matrix. The $\mathbf{G}_0$ and $\mathbf{P}_0$ are the (co)variance matrices of the additive genetic and permanent environmental random regression coefficients, respectively; and **R** is the (co)variance matrix of residual. Several studies have shown that better goodness of fit can be achieved assuming different residual variances for different periods or time intervals (e.g., Pereira et al., 2013; Flores and van der Werf, 2015).

Similar to single-record traits, the variance components of a RRM can be estimated using different methods described in the literature. Currently the most used methods are the restricted maximum likelihood (**REML**; e.g., Ghiasi and Carabaño, 2018; Hein et al., 2018; Sanchez et al., 2018) and Bayesian (e.g., Kheirabadi, 2018; Moretti et al., 2018; Nustad et al., 2018) methods. Considering the REML estimation method, the model assumptions for RRM showed in [1] can be described as:

$$\mathrm{E}[\mathbf{y}] = \mathbf{Xb} \text{ and } \mathrm{Var}\begin{bmatrix} \mathbf{a} \\ \mathbf{p} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\otimes\mathbf{G}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\otimes\mathbf{P}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R} \end{bmatrix}, \quad [3]$$

where **e** is the vector of residuals, and all other terms were previously defined in [2]. RRM implies different additive genetic and permanent environmental variances over time, because $\mathbf{G}_0$ and $\mathbf{P}_0$ contain the variances and covariances for the coefficients of the used function. The genetic ($\mathbf{\Sigma}$) and environmental ($\mathbf{\Phi}$) (co)variance matrices for time points can be obtained as:

$$\mathbf{\Sigma} = \mathbf{T}\mathbf{G}_0\mathbf{T}' \text{ and } \mathbf{\Phi} = \mathbf{T}\mathbf{P}_0\mathbf{T}' \text{ (assuming the same function for both factors),} \qquad [4]$$

where **T** is a matrix of independent covariates for all time points (e.g., ages or DIM), associated with the function used. Consequently, RRM can provide estimates of the genetic parameters and EBV for all analyzed time points. The vector of breeding values ($\mathbf{EBV}_j$) of the animal j including all analyzed times can be obtained as follow:

$$\mathbf{EBV}_j = \mathbf{T}\hat{\mathbf{a}}_j, \qquad [5]$$

where $\hat{\mathbf{a}}_j$ is the vector of estimated breeding values for the covariance function coefficients for the animal j.

Besides the RRM, there are several other models/methods to analyze longitudinal data taking into account all repeated records. Examples are the character process model (e.g., Pletcher and Geyer, 1999; David et al., 2007), the structured antedependence model (e.g., Jaffrézic et al., 2004; Zhao et al., 2005), the dynamic linear model (e.g., Forni et al., 2009; Das et al., 2011), and covariance functions. Due to a common misunderstanding between covariance functions and RRM, the next topic aims to summarize the comparison between them.

2.4.1. Covariance functions

Covariance functions are used to describe the covariance among records that are measured at different time periods or ages (Meyer and Hill, 1997). In brief, the covariance functions can reduce the rank of the covariance matrix in a way that they are still equivalent to covariance matrices for traits with many longitudinal records. There are several different covariance structures to account for covariances among random correlated effects of repeated records, such as autoregressive, compound symmetry and the most complex one, the unstructured (Sawalha et al., 2005). Additionally, several functions used in RRM to model the trajectory of genetic and environmental effects over time can also be used under a covariance function framework.

The first-order autoregressive covariance structure has only one additional parameter compared to the simple repeatability model, which implicitly assumes a compound symmetry structure, and allows for non-constant covariances (Wade and Quaas, 1993). Harville (1979) and

Quaas et al. (1984) proposed the use of an autoregressive structure to model covariance for random effects of repeated measures in animal breeding. Carvalheira et al. (1998), comparing genetic parameters estimated with models assuming or not an autoregressive covariance structure among permanent environmental effects within a lactation, reported that the autoregressive covariance structure for permanent environmental effects within lactation was effective for partitioning total variance and removing noise that would otherwise be confounded with genetic effects.

The coefficients of the covariance functions can be estimated directly from the dataset using REML (Meyer and Hill, 1997), or the generalized least squares (**GLS**) method (Kirkpatrick et al., 1990). However, if the same functions are used, models based on covariance functions are equivalent to RRM, as showed by Meyer and Hill (1997) and Van Der Werf et al. (1998). Models based on covariance functions and RRM can be computationally demanding, as they require appropriate modelling of the trajectories of genetic and environment effects, especially in the case of multiple-trait analysis.

2.4.2. Multiple-trait random regression models

The first genetic evaluation using multiple-trait RRM and the first multiple trait estimates of covariance components generated from multiple-trait RRM were presented by Jamrozik et al. (1997c) and Jamrozik et al. (1998b), respectively. In a multiple-trait RRM, the coefficients for the chosen polynomials are estimated in order to explain the variation of more than one trait simultaneously over time. The same RRM showed in [1] can be described under a multiple-trait approach considering N traits, as follow:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wp} + \mathbf{e}, \qquad [6]$$

i.e.:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_N \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_N \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_N \end{bmatrix} + \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_N \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_N \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_N \end{bmatrix}, \qquad [7]$$

where $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_N \end{bmatrix}'$, $\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_N \end{bmatrix}'$, $\mathbf{a} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_N \end{bmatrix}'$, $\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_N \end{bmatrix}'$ and $\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_N \end{bmatrix}'$ are the vectors of phenotypic records, fixed effects, additive genetic, permanent environmental and residual effects, respectively, sorted by trait (from trait 1 to N). Hence $\mathbf{y}_i$ is the vector of phenotypic records for the ith trait, $\mathbf{b}_i$ is the vector of fixed effects affecting the ith trait, $\mathbf{a}_i$ is the vector of additive genetic random regression coefficients for the animal effect on the ith trait, $\mathbf{p}_i$ is the vector of random regression coefficients for permanent environment effects on the ith trait and $\mathbf{e}_i$ is the vector of residuals for the ith trait.

The multiple-trait RRM is a powerful tool for joint genetic evaluation of several longitudinal traits. In brief, the multiple-trait RRM assures the advantages of multiple-trait analyses in relation to single-trait analyses, such as increased accuracy of estimate breeding values (Boligon et al., 2011). However, one specific advantage of multiple-trait RRM over single-trait RRM is the possibility to estimate genetic correlations between different traits over time. These genetic correlations over time enable the identification of the most feasible time periods to exploit indirect selection and correlated genetic responses aiming to improve the genetic gain in all traits analyzed, as suggested by Oliveira et al. (2016). Studies have reported the use of multiple-trait RRM in genetic evaluation in various species. For instance, successful application of multiple-trait RRM to perform joint genetic analysis was reported in dairy cattle (e.g., Pereira et al., 2013; Kheirabadi et al., 2014; Alam et al., 2015), beef cattle (e.g., Santana et al., 2015; Santana Jr et al., 2018), buffalo (e.g., Borquis et al., 2013), goats (e.g., Brito et al., 2017), and more recently in fish (e.g., Turra et al., 2018).

Until recently, the majority of studies performing genetic evaluations based on multiple-trait RRM had used the same function to model all traits, however, Meyer and Hill (1997) commented that, in some cases, it might be desired to fit different functions for different traits. Thus, a study in dairy goats showed that it is possible to combine different functions to describe different traits in the same multiple-trait RRM analysis (Oliveira et al., 2016). The referred authors reported that using different functions to model different traits might help to achieve a better fit for each trait, which can improve the accuracy of EBV and genetic parameter estimates. Application of multiple-trait RRM has become even more flexible, in terms of combination of different functions under the same RRM. A study in dairy cattle reported the possibility of using different functions to describe the fixed and random effects of each trait, as well as using different functions for different traits, under the same multiple-trait RRM (Oliveira et al., 2017a). In contrast to those authors, Jamrozik et al. (2001) suggest that the same function should be used to describe the random genetic and environmental effects of a trait, and that the selection of the function to describe the lactation curve should be determined by the goodness of fit at the phenotypic level.

2.4.3. Comparing alternative functions to model longitudinal trajectory

The first comparison of alternate functions to be used in RRM for TDMY was published by Jamrozik et al. (1997a). In the recent years, several studies have compared different functions in RRM for genetic analysis in various livestock species. For example, Bignardi et al. (2011) analyzed TDMY of Brazilian Holsteins and compared the use of Legendre orthogonal polynomials (Kirkpatrick et al., 1990) and B-spline functions (de Boor, 1980; Meyer, 2005) to the functions proposed by Ali and Schaeffer (1987) and Wilmink (1987). In summary, these authors concluded that RRM using B-splines and Legendre polynomials are more adequate to describe the covariance structure of the data. Studying Canadian Holsteins, Bohmanova et al. (2008) found smaller prediction errors and higher stability of EBVs by using spline functions compared to Legendre polynomials, which was in agreement with Bouallegue et al. (2015). On the other hand, Pereira et al. (2013) concluded that, despite the fact that linear splines have provided the smallest sum of residual variances over the lactation, RRM based on Legendre polynomials provides the best fit (according to the deviance information and Bayesian information criterion), and should be used to describe variations in milk yield of Brazilian Gyr cattle. This is in agreement to Laureano et al. (2014), who also concluded that Legendre polynomials are more adequate than B-spline functions to model milk yield, using Akaike's and Bayesian information criterion to compare the mentioned functions in Brazilian Holstein cattle.

For dairy goats, studies using spline functions (León et al., 2012; Thepparat et al., 2015) and parametric functions (e.g., Ali and Schaeffer's function; in González-Peña et al., 2012) to model the lactation curve have shown equal or better fit than the RRM based on Legendre polynomials. In relation to milk components from dairy goats, the majority of the studies were based on Legendre polynomials (Andonov et al., 2013), however, Oliveira et al. (2016) found that the model based on the Ali and Schaffer's function fitted better to describe protein yield in Saanen and Alpine goats. For beef cattle, the majority of studies using spline functions analyzed growth traits (Speidel et al., 2010). In general, these studies show that RRM based on spline functions is an alternative to substitute RRM based on Legendre polynomials, because similar EBV estimates are obtained using both functions (e.g., Iwaisaki et al., 2005; Robbins et al., 2005; Boligon et al., 2011). In addition, Bohmanova et al. (2005) found that spline functions are just as accurate as Legendre polynomials (within 0.2% accuracy change), using simulation to mimic beef cattle growth data. For sheep, Zamani et al. (2016) concluded that using RRM based on B-splines and Legendre polynomials to fit growth curves generated similar estimates of variance components as ratios to total phenotypic variance, but the models with Legendre polynomials showed better fitting properties.

Even though some divergent conclusions have been made about the best functions to use in RRM across different studies and traits, the results indicate that, in practice, the same bulls would be selected by both RRM using Legendre polynomials and spline functions (Pereira et al., 2013). Moreover, Jamrozik and Schaeffer (2002) showed that different comparison methods ranked differently RRM based on different functions. Thus, in order to define which function should be used in each case, advantages and disadvantages of each one of them should be taken into account. Non-orthogonal polynomials could produce coefficients that are highly correlated and this may yield to computational problems during estimation or solving of the mixed model equations. Thus, the orthogonal polynomials have been preferred among the polynomial functions, because they give the lowest correlations among the estimated regressions. The RRM based on Legendre polynomials tend to consider less coefficients than corresponding analysis using B-spline functions (Meyer, 2005), which may allow using more parsimonious models. However, RRM based on Legendre polynomials usually generate inflated variances at the extremes of the lactation curve, as demonstrated by Misztal (2006). Analysis using RRM based on spline functions also have the advantage of faster convergence over Legendre polynomials, which may be due to the fact that spline coefficients are sparser than their polynomial counterparts (Misztal, 2006). However, one of the most important questions when using spline functions is related to the number and location of knots (points at which independent linear segments are connected), which should be defined previous to the analysis (Jamrozik et al., 2010) or determined by a non-linear procedure. Although RRM based on parametric functions (e.g., Ali and Schaeffer and Wilmink) have shown poor fit when compared to Legendre and spline functions in the majority of studies, analyses based on these functions are of faster convergence compared to analyses based on polynomial functions, considering the same number of parameters when the number of observations is small (Oliveira et al., 2016).

2.4.4. Applications of random regression models

2.3.4.1. In the research field. The potential for practical applications of RRM was presented by Schaeffer and Dekkers (1994). RRM became more sophisticated over time through the development of alternative functions and advances in model structures. RRM have been used in a wide range of research areas, including genotype by environment interaction (**GxE**; e.g., Calus and Veerkamp, 2003; Oliveira et al., 2018a), multiple-trait analysis combining different functions to describe each trait (Oliveira et al., 2016, 2017a), and categorical traits (e.g., Averill et al., 2004; Oliveira et al., 2017b). Brügemann et al. (2013) showed that a threshold RRM could be applied to study genetic effects on binary traits. Yin et al. (2012) analyzed normally distributed and categorical traits in organic and low input dairy cattle herds and showed that RRM can be applied to data from small herds with a limited number of repeated measurements per cow. Furthermore, the effect of environmental sensitivity on production traits using RRM has also been investigated, where environmental sensitivity was based on herd characteristics (Calus and Veerkamp, 2003). Aguilar et al. (2010b) investigated genetic trends for milk yield under heat stress conditions using RRM accounting for heat stress, in which each parity was treated as a separate trait. For that, the authors performed a regression of milk yield on a function of both days in milk and temperature-humidity index (**THI**), calculated as proposed by Ravagnolo et al. (2000). The authors showed that the most heat-tolerant sires had lower production daughters. In addition, the authors drew important conclusions, such as the need to include records of later parities in this type of environmental sensitivity studies.

The popularity of RRM have increased substantially in animal breeding over the past years and is now considered the method of choice to genetically evaluate longitudinal traits in various livestock species, including dairy cattle (e.g., Kheirabadi, 2018; Köck et al., 2018; Padilha et al., 2018; Zetouni et al., 2018); beef cattle (e.g., Englishby et al., 2017; Liberona et al., 2018; Oliveira et al., 2018a); sheep (e.g., Dominik and Swan, 2018; Naderi, 2018; Saghi et al., 2018); goats (e.g., Oliveira et al., 2016; Brito et al., 2017; Ghiasi and Mokhtari, 2018); horses (e.g., Bartolomé et al., 2018); swine (e.g., Coyne et al., 2017; Huynh-Tran et al., 2017); poultry (e.g., Begli et al., 2016; Jun et al., 2017; Mookprom et al., 2017; Miyumo et al., 2018); quail (e.g., Abou Khadiga et al., 2017; Karami et al., 2017; Mota et al., 2018a); buffalo (e.g., Borquis et al., 2013; Ghavi Hossein-Zadeh et al., 2017); fish (e.g., He et al., 2017; Turra et al., 2018; Zhao et al., 2018); crab (e.g., Liu et al., 2018); and shrimp (e.g., Tan et al., 2017).

2.3.4.2. In the industry field. Australia was the first country in the world to use TDMY records in 1984 in the official genetic evaluations of dairy species, adjusting the accumulated yield for the test-day effect (Powell and Norman, 2006). However, the first country to implement RRM as the official model for genetic evaluation of animals was Canada, in 1999. The multiple-trait RRM initially used 12 traits (milk, fat and protein yields, and somatic cell score, in each of the first three lactations) and it was based on orthogonal Legendre polynomials. The RRM initially adopted by Canada are detailed in Schaeffer et al. (2000).

Canada is currently using two different multiple-trait RRM, one to analyze milk, fat and protein yields in each of the first three lactations; and another one to analyze somatic cell score in the first three lactations. The decision of analyzing somatic cell score separated from the production traits was officially made by the Canadian Dairy Network (CDN, Guelph, Ontario, Canada) in 2016, in order to improve the stability of published somatic cell score proofs (CDN, 2016). Both models, for the production traits and somatic cell score, include the effect of herd-test day-parity and DIM within lactation (from 5 to 305 days), and regression curves for year-season-region-parity-age of calving as fixed effect, the effect of herd-year of calving, additive genetic and permanent environment as random effects. All regressions are modeled through fourth order Legendre orthogonal polynomials (i.e., 5 covariates), and both models account for heterogeneity of residual variance, assuming 10 classes of residual variance for each parity (Interbull, 2018).

Many other countries have also adopted RRM as the official model for genetic evaluation in dairy cattle. Among them, Australia, Belgium, Croatia, Czech Republic, Denmark-Finland-Sweden, Estonia, Germany-Austria-Luxembourg, Italy, Japan, Latvia, Lithuania, New Zealand, Poland, Republic of Ireland, Slovak Republic, Switzerland, The Netherlands and United Kingdom, as reported by Interbull (2018). Although TDMY data have been collected in the United States of America (**USA**) since 1905 for herd management and have contributed to national genetic evaluations since 1936, RRM are still not used for genetic evaluations in that country. The official USA dairy cattle genetic evaluations still use accumulated yield due to a patent applied on RRM by Cornell University, in 1993. Thus, this patent has forestalled the use of RRM in the United States, except for regional and unofficial evaluations performed by the patent holders (Powell and Norman, 2006). In addition, it is important to note that the size of the USA dataset is very big, which would not allow RRM to be calculated efficiently.

2.4.5. Implementation of random regression models compared to other approaches

There are several advantages of using RRM to analyze longitudinal data in genetic evaluations compared to the use of other traditional models. One of them is the fact that RRM do not require that records are measured at the same time in all individuals, and consequently, they do not require a minimum number of records per animal (Schaeffer, 2004). In other words, all phenotypic information can be used for the genetic analyses. In addition, as previously discussed, the RRM enable the estimation of a smaller number of parameters when compared to other methodologies (e.g., multiple-trait models). This is even more advantageous when the main objective is to obtain genetic parameters and EBV for several traits analyzed simultaneously in a multiple-trait RRM approach.

The RRM also enable to account for changes in genetic variation in production over time, since covariance components can be estimated for all time points. Thus, genetic parameters and EBV can be estimated for numerous time points, including times in which phenotypes were not recorded but are within the measurable period range (White et al., 1999). In the dairy cattle context, RRM allow the prediction of the shape of lactation curves in the genetic evaluations and, therefore, genetically select animals with more desirable lactation curve patterns. The selection of animals based on the complete lactation curves takes into account the lactation biology and the physiological implications of different lactation curve patterns, which may help to decrease the energy deficiency experienced during early- and peak-lactation, as suggested by Jamrozik et al. (1998a) and Oliveira et al. (2018c). In this context, decreasing the energy deficiency can help to improve other important traits, such as health (especially avoiding metabolic diseases), welfare and fertility. In addition, expressing breeding values as curves of genetic merit allow for selection of animals for important traits that are defined based on a combination of different time points, such as lactation persistency (Cole and VanRaden, 2006; Siqueira et al., 2017; Padilha et al., 2018).

Another advantage of RRM compared to other models is that RRM allow a more precise adjustment for environmental effects in the course of the phenotypic curve, since RRM account for time-dependent variation over the lactation (Swalve, 2000). The more precise adjustment for environmental effects is due to the use of permanent environmental effect as a random regression, and the potential better adjustment for temporary environmental effects.

Comparing EBV predicted by random regression and repeatability models, several authors have concluded that the total genetic merit was more accurately predicted when using RRM (e.g., Andonov et al., 2013; Khanzadeh et al., 2013). Compared to the use of a linear model to analyze the accumulated yield, using RRM based on TDMY records avoided the use of extended records for culled cows and for lactations in progress, as well as the use of projection (for the entire lactation) factors, what can remove the errors in estimating the accumulated yield from the genetic analysis (Jensen, 2001). In agreement, Kistemaker (1997) reported that EBVs from RRM are more accurated than EBVs from accumulated 305-day models. Nevertheless, there are some issues related to the use of RRM. Firstly, the best method to estimate the prediction error (co)variances and the accuracies of the estimates generated from RRM is not known, even though some approximations have been reported in the literature (e.g., Jamrozik et al., 2000; Tier and Meyer, 2004). Secondly, it is still not well defined what is the best approach to summarize these different accuracies (estimated for each time point) in order to report a single value in official genetic evaluations.

Another limitation of RRM is related to the large amount of data to be analyzed, since several measurements are recorded for the same individual (Swalve, 2000; Jensen, 2001). Thus, the amount of data stored to be used in RRM is much bigger compared to other methodologies. In addition, RRM tend to have more parameters to estimate compared to traditional single-trait models, such as the accumulated yield and the repeatability models (Strabel and Misztal, 1999). The concern about the number of parameters to be estimated is mainly related to the computational demand, i.e., with the difficulties to achieve appropriate convergence in the MME (Swalve, 2000). Nowadays, computational algorithms to evaluate large datasets based on complex models are available (e.g., preconditioned conjugate gradient algorithm; Strandén and Lidauer, 1999; Tsuruta et al., 2001), which allow the iteration on data using blocks with different sizes as preconditioner to facilitate a joint estimation of the regression coefficients. Furthermore, over the past decades, there has been a great advancement in computing resources and these computing limitations might be partially overcome in the very near future.

Concerns about the amount of data and computational demand required by RRM analysis are even more important in the genomics era, in which a large number of genetic markers (e.g., Single-Nucleotide Polymorphisms; **SNP**) are included in the analysis to predict the realized relationship among individuals, enabling the implementation of genome-wide selection (**GWS**; Meuwissen et al., 2001).

2.5. An overview of genomic selection

The GWS, which started with the seminal studies of Nejati-Javaremi et al. (1997) and Meuwissen et al. (2001), has been implemented in various livestock breeding schemes and has been considered a landmark in modern livestock breeding programs (Boichard et al., 2016; Samorè and Fontanesi, 2016; Wiggans et al., 2017). Including genotypic information in the evaluations has increased the accuracy of breeding values in the range of 30-70% (VanRaden et al., 2009; Harris and Johnson, 2010), and the rates of genetic gain per year from 50-100% for productive traits (García-Ruiz et al., 2016). Especially for selection of candidates at a young age, the increase in accuracy enables a reduction in the generation interval and consequently, increases the genetic gain per time unit. In species with long generation intervals, such as dairy cattle, GWS reduces the costs with traditional progeny testing schemes (Schaeffer, 2006; König et al., 2009; VanRaden et al., 2009), which translates to saving millions of dollars worldwide every year.

The majority of livestock breeding programs implemented GWS initially based on a multiple-step genomic best linear unbiased prediction (**GBLUP**) approach, as reported by Interbull (2017) for dairy cattle. The prediction of Genomic Estimated Breeding Values (**GEBV**) based on the multiple-step GBLUP requires the prediction of traditional EBV and the subsequent use of pseudo-observations, typically either daughter yield deviations (**DYD**) or deregressed EBV (**dEBV**). Thereafter, the pseudo-observations are used as pseudo-phenotypes in a subsequent step to estimate the Direct Genomic Values (**DGV**; Gao et al., 2013; Koivula et al., 2015; Winkelman et al., 2015). In general, dEBVs are more often used compared to DYD because of their easy availability, especially when the training population is based on genotypes from different countries (Mäntysaari et al., 2011). Several methods have been suggested to estimate dEBVs, especially for non-longitudinal traits, such as Garrick et al. (2009), VanRaden et al. (2009) and Wiggans et al. (2011).

Usually the DGV from the GBLUP method are estimated based on a linear model, where an overall mean is assumed as a fixed effect and the marker effects are assumed as random. In the MME, the main difference is that the numerator relationship matrix (**A**) is replaced by a genomic relationship matrix (**G**). In brief, the **G** matrix reflects the observed identical-by-state relationship between genotyped individuals and it can be created in different ways (e.g., VanRaden, 2008; Makgahlela et al., 2013). The same DGV estimated by the GBLUP method can be obtained by using SNP estimates generated by the equivalent marker effects model (**SNP-BLUP**; Koivula et al., 2012). In order to increase the accuracy of genomic predictions, usually the DGV are blended to the Parent Average (**PA**) to generate the GEBV (VanRaden, 2008).

More recently, various studies have shown that simultaneously combining phenotypic records, pedigree and genomic information in a single-step can lead to more accurate and less biased GEBVs compared to the multiple-step GBLUP approach (Christensen and Lund, 2010; Kang et al., 2017). This method is known as single-step GBLUP (**ssGBLUP;** Misztal et al., 2009; Aguilar et al., 2010a; Christensen and Lund, 2010). Compared to the multiple-step GBLUP approach, the ssGBLUP seems to simplify the procedure, as only one analysis is performed and the same statistical model used in genetic evaluations can be used (Lourenco et al., 2014; Misztal et al., 2014). Moreover, the ssGBLUP avoids the estimation of deregressed EBVs (or daughter yield deviations) that are generally used as pseudo-phenotypes in the multiple-step GBLUP analysis. In general, the use of actual phenotypes in the ssGBLUP (instead of pseudo-phenotypes) may avoid estimation errors. In addition, the ssGBLUP method eliminates the problem of double-counting of information found in multiple-step evaluations when including genotypes from bulls and their daughters in the training population, when improper methods of EBV deregression are used (Uemoto et al., 2017).

In general, ssGBLUP replaces the **A** matrix used in the MME by an **H** matrix that combines the **A** and **G** relationship matrices (Misztal et al., 2009; Legarra et al., 2014). Usually the **H** is computationally demanding to create, however, its inverse has a simple structure (Aguilar et al., 2010a; Christensen and Lund, 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(\alpha\mathbf{G} - \beta\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix}, \qquad [5]$$

where **G** is the genomic relationship matrix, $\mathbf{A}^{-1}$ is the inverse of the traditional relationship matrix (**A**), $\mathbf{A}_{22}$ is the section of **A** related to genotyped animals, and $\mathbf{A}_{22}^{-1}$ is the section of $\mathbf{A}^{-1}$ related to genotyped animals. The α and β parameters are used to make **G** invertible or non-singular (Koivula et al., 2015), and the τ and ω parameters are used to account for the reduced genetic variance and for different depths of pedigree, respectively, in order to make $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ compatible (Misztal et al., 2017). Several studies have shown that adjusting $\mathbf{G}^{-1}$ towards their expected values in $\mathbf{A}_{22}^{-1}$ decreases the scaling problem, which can improve the accuracy and, simultaneously reduce the bias in the ssGBLUP evaluations (e.g., Vitezica et al., 2011; Christensen et al., 2012).

Even though RRM are routinely used in genetic evaluations of longitudinal traits in several countries (Interbull, 2018), usually GEBV are generated for the accumulated yield (e.g., Winkelman et al., 2015; Aliloo et al., 2016; Jenko et al., 2017) or for phenotypes taken at specific times (e.g., Martínez et al., 2017; Campos et al., 2018), which does not allow selecting animals for the complete curve pattern. In order to make inferences about the trajectory of the curves, some studies have proposed to estimate GEBV for curve parameters (coefficients) calculated in a first step (e.g., Lázaro et al., 2017). However, this approach gives limited inferences about the pattern of the phenotypic curve and might not have biological interpretation in some cases. Thus, including genomic information to evaluate longitudinal traits using RRM would be a more desirable alternative to yield more accurate selection and culling decisions.

2.6. Genomic predictions based on random regression models

Genomic predictions for various time points can be simultaneously done using RRM, which enable selection of breeding animals based on the complete pattern of the production curve using genomic information. There were two main approaches described in the literature to combine genomic information and RRM in order to estimate daily GEBVs: 1) using two-step GBLUP, where both steps are based on RRM (Oliveira et al., 2018c); and, 2) using ssGBLUP based on RRM (Koivula et al., 2015; Jattawa et al., 2016; Baba et al., 2017; Kang et al., 2018).

In the first approach, EBVs for each time point are estimated based on a RRM, then EBV are deregressed and used as pseudo-phenotypes to estimate the DGV for each time point, based on a second RRM (Oliveira et al., 2018c). In brief, the first RRM used by these authors is similar to that showed in equation [1], where several fixed and random effects are fitted. However, in the second RRM, only a fixed regression (mean) and the genomic random regression coefficients were fitted. Thus, the MME created for the second-step was simplified (i.e., there was no permanent environmental effect and homogeneity of residual variances was assumed). Additionally, the **A** matrix was replaced by a **G** matrix in the MME for RRM like in equation [2].

In the referred study, Oliveira et al. (2018c) reported that using RRM in both steps of a two-step GWS produces moderately accurate GEBVs over the lactation, but that the observed accuracy of GEBVs over the lactation curve is in general lower than the observed accuracy for 305-days accumulated yield. The authors related those findings to the fact that it is more difficult to predict GEBVs for each day in milk than for accumulated yield, which could have resulted in a larger total amount of estimation error generated by the RRM when compared to the use of the 305-d accumulated model. Nonetheless, the authors reported higher accuracy for GEBV compared to PA for all analyzed traits.

De-regression methods for longitudinal traits are complex, as they need to take into account the covariance between time points in the de-regression process. Thus, a recent study based on real Jersey cattle data compared the performance of three different de-regression methods that are usually applied to non-longitudinal traits (i.e., Garrick et al., 2009; VanRaden et al., 2009; Wiggans et al., 2011) in the de-regression of longitudinal traits for dairy cattle (Oliveira et al., 2018b). The authors concluded that deregressing longitudinal EBV using de-regression methods for non-longitudinal traits is feasible, and that the methods suggested by VanRaden et al. (2009) and Wiggans et al. (2011) yielded more accurate and less biased genomic predictions than the method suggested by Garrick et al. (2009).

For genomic predictions based on ssGBLUP, the same RRM used in genetic evaluations can be used (e.g., equation [1]). Thus, the MME created for ssGBLUP is similar to the MME showed in equation [2], with the only difference that the **A** matrix is replaced by the **H** matrix, constructed as showed in equation [5]. The possibility to use the same statistical models when incorporating genomic information via ssGBLUP is a great advantage of the ssGBLUP method (Lourenco et al., 2014; Misztal et al., 2014). This feature might justify the higher number of studies using ssGBLUP based on RRM (Koivula et al., 2015; Jattawa et al., 2016; Baba et al., 2017; Kang et al., 2018) compared to studies using the multiple-step approach (Oliveira et al., 2018c).

In general, studies using the ssGBLUP based on RRM concluded that this approach provides higher accuracies and less biased GEBV compared to other methods (Koivula et al., 2015; Jattawa et al., 2016; Kang et al., 2018). However, Koivula et al. (2015) showed that there is a great need to identify the best option to combine $\mathbf{A}^{-1}$ and $\mathbf{G}^{-1}$. In a simulation study, Kang et al. (2017) compared the performance of ssGBLUP and multiple-step GBLUP based on RRM and reported that the ssGBLUP method achieved the highest accuracy and unbiasedness under all simulated scenarios, indicating robust prediction ability in longitudinal trait analyses. Moreover, the authors showed that ssGBLUP had better persistency of accuracy over generations than the multiple-step GBLUP method (Kang et al., 2017). One year later, Kang et al. (2018) performed an ssGBLUP evaluation using a real dataset of Chinese Holsteins. The authors also concluded that ssGBLUP was feasible for implementation in breeding practice, and that its prediction performance (i.e., accuracy and unbiasedness) was superior to the RRM based on BLUP and the multiple-step GBLUP. In plants, Campbell et al. (2018) showed the utility of using RRM for genomic prediction of shoot growth trajectories in rice (Oryza sativa). In summary, these authors estimated 11.6% increase in prediction accuracy when using RRM, compared to a model that fitted each time point independently. Thus, they concluded that RRM can be effectively used to improve the accuracy of genomic prediction for longitudinal traits in rice.

Another interesting application of the combination of RRM and genomic information is for investigating GxE interactions via reaction norms, aiming to more accurately estimate GEBVs for selection candidates. Few studies report on the estimation of GEBVs for animals in different environments, using either a multiple-step (Silva et al., 2014) or single-step (Mota et al., 2016;

Oliveira et al., 2018a) approach. In this context, Silva et al. (2014) performed a sire evaluation for total number of pigs born and concluded that GxE interactions were important for the proper genomic evaluation of sires and that genomic information can increase the accuracy of selection across environments. Nonetheless, Mota et al. (2016), studying tick resistance in Hereford and Braford beef cattle, reported that genomic information does not lead to higher prediction accuracies in reaction norm models, suggesting that the benefits observed for combining genomic information and RRM are trait and species specific. Oliveira et al. (2018a) contrasted the use of traditional BLUP and ssGBLUP in reaction norms to evaluate the existence of G × E for yearling weight in Nellore cattle. The authors concluded that both methodologies were capable of identifying the G × E interaction. However, the ability to predict the animals' future performance was improved by 7.9% using ssGBLUP.

Besides the prediction of GEBVs, RRM can be also used to estimate SNP effects and the variances explained by them over time, in genome-wide association studies (**GWAS**).

2.7. Genome-wide association studies for longitudinal traits

Over the past decades, numerous GWAS have been performed using a large number of genetic markers (e.g., SNP) spread across the entire genome of species of interest, aiming to detect genetic variants associated with a particular phenotype. The main goals of such studies were to unravel the genetic architecture and underlying biology of the traits and to develop genetic tests and personalized treatments in humans (e.g., Gorlova et al., 2017; Ghouse et al., 2017; Turner et al., 2018), and propose more accurate genomic selection methods and breeding approaches in plants (e.g., Angelovici et al., 2017; Guo et al., 2018; Gyawali et al., 2018), and livestock species (e.g., Fang et al., 2017; Meng et al., 2017; Higgins et al., 2018). As previously discussed, a large number of economically important traits in livestock species are longitudinal traits, and similarly to single-record traits, an increased interest in estimating the effects of genomic markers for these traits has been observed (e.g., Ning et al., 2017; Soares et al., 2017; Mota et al., 2018b). This section discusses alternatives to perform GWAS for longitudinal traits, with a focus on RRM.

Different approaches have been used to perform GWAS for longitudinal traits. Until recently, the majority of studies have converted multiple phenotypic measurements for each animal into a single estimate (e.g., average over all time points and 305-day milk yield EBV (or dEBV)). These (pseudo-)phenotypes are used as response variables for the GWAS analyses (e.g., Meredith et al., 2012; Buitenhuis et al., 2014; Raven et al., 2014; Nayeri et al., 2016). However, various studies have shown that the additive polygenic effects of longitudinal traits are not constant during the entire phenotypic expression (e.g., growth or lactation, as exemplified by Strabel et al., 2005; Zavadilová et al., 2005; Berry et al., 2007; Brito et al., 2017). Furthermore, using milk production as an example, gene expression studies have indicated that the amount of RNA synthesised often varies across lactation (e.g., Bionaz and Loor, 2008), and specific genes (e.g., DGAT1) or genetic loci can be less pronounced or even reversed during the first stage of lactation (Strucken et al., 2011) or change from early, peak to late lactation (Strucken et al., 2012). Strucken et al. (2012) treated 10-day lactation intervals in the initial lactation stage and 305-day EBV as separated traits and they observed significant differences in allele effects across time periods. Along the same line, when studying growth traits, various studies have defined each weighing time as a different trait/phenotype to perform GWAS (e.g., Snelling et al., 2010; Lu et al., 2013; Qiao et al., 2015). In chickens, Xie et al. (2012) performed GWAS for body weights measured at hatching, 7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77, 84, and 90 days of age, and Yi et al. (2015) studied egg weight at 9 different points as separate traits. In these studies, functional analyses were also performed to identify candidate genes significantly associated with each time point (or trait, as defined by the authors).

Another alternative to generate pseudo-phenotypes for GWAS analyses of longitudinal traits considering the whole phenotypic trajectory is by using linear and nonlinear models, which summarize longitudinal records in a few parameters that usually have biological interpretation. These parameters are then used as phenotypes for the GWAS (or QTL detection). For instance, Rodriguez-Zas et al. (2002) used the parameters based on Morant and Gnanasakthy (1989) function to estimate the effects of a chosen set of markers on the parameters of the function and reported the presence of QTLs influencing the scale and shape parameters that describe the lactation pattern. A study carried out on German Holstein cows divergently selected for milk yield used Wilmink function parameters as phenotypes (Strucken et al., 2012) and also identified SNPs associated with lactation curve traits, such as lactation persistency. Later, Soares et al. (2017) proposed a multiple-trait GWAS of estimated growth curve parameters for scrotal circumference (**SC**) recorded over time (as indicator of fertility traits) in Brahman cattle. Five widely used nonlinear models (i.e., Brody and Lardy, 1946; von Bertalanffy, 1957; Richards, 1959; Nelder, 1961) were tested to fit SC records (measured at 6, 12, 18, and 24 months of age). The model parameters were used as pseudo-phenotypes for the subsequent GWAS analyses. A similar study was also performed by Crispim et al. (2015) using beef cattle growth data. An advantage of using the parameters of those functions as pseudo-phenotypes for the GWAS analyses is that they usually provide a direct description of a geometric property of the curve or have some biological interpretation (e.g., persistency) that can explain the underlying biology of the trait.

Another alternative approach for GWAS of longitudinal traits is Principal Component Analysis (**PCA**), in which the scores of a reduced number of Principle Components (**PC**) based on multiple variables (e.g., longitudinal records) can be used as a "single phenotype" when performing the GWAS analysis. As indicated by Macciotta et al. (2015), these model-free approaches are able to derive measures of lactation curve (or other longitudinal trait) traits without specific assumptions on data structure. For instance, Macciotta et al. (2015) performed a GWAS using the scores of the first 2 PCs from test-day records for milk yield, fat and protein percentages, and somatic cell score as the response variables for GWAS analyses.

Musolf et al. (2014), investigating human disease phenotypes, proposed a method for mapping genes with longitudinal phenotypes (e.g., severity or progression of a disease) using Bayesian Posterior probabilities. Their approach involved assigning individuals to trajectory subgroups and testing for associations with SNPs using the Bayesian Posterior probability (i.e., the probability of an individual belonging to a particular subgroup) of being in the clinically relevant subgroup as a quantitative trait (outcome) in GWAS.

The use of longitudinal models in GWAS studies, such as random regression, allows for SNPs having a heterogeneous effect across the trajectory to be characterized. Tetens et al. (2014) estimated deregressed EBVs for dry matter intake (**DMI**) from a RRM using third order Legendre polynomials and the daily deregressed EBVs for DIM 11, 30, 80, 130, and 180 were used as phenotypes for GWAS. They identified interesting lactation stage-specific candidate genes and pathways. Similarly, Pryce et al. (2010) performed a GWAS for different measures of lactation persistency on primiparous Holstein and Jersey cows based on RRM.

Wang et al. (2014b) proposed a two-step method for simultaneously testing the genetic association with multiple longitudinal traits aiming to detect pleiotropic effects in real and simulated data. In summary, the authors used a mixed effects model to estimate the random effects for each longitudinal trait in the first step, and then, in the second step, they performed a simultaneous association test. As conclusion, they comment that this method can efficiently detect the pleiotropic effects on multiple longitudinal traits.

As discussed by Szyda et al. (2014), the effects of particular genes on longitudinal traits were estimated using linkage analysis by Yang, (2006), Yang and Xu (2006), Yang et al. (2007), Lund et al. (2008), and Suchocki et al. (2013). The models proposed by Yang (2006) and Yang and Xu (2006) incorporate time-dependent fixed effects of QTL together with a random time-dependent residual effect, but no polygenic effect was considered. Yang et al. (2007) considered a fixed QTL effect with a polygenic component approximated by cofactors. Lund et al. (2008) and Suchocki et al. (2013) estimated individual additive effect curves underlying a random QTL effect together with a random additive polygenic effect. The association analysis described by Banos et al. (2008) considered the polygenic effect as time dependent and the candidate gene effect as independent of time. In Szyda et al. (2014), the additive effects of candidate genes were modelled as time-dependent and population-average fixed effects, similar to the approach described by Yang (2006) and Yang and Xu (2006). The authors reported changes in additive effects of specific markers throughout lactation.

The SNP effects on longitudinal traits can be obtained from GEBVs (e.g., Silva et al., 2014; Verardo et al., 2017; Mota et al., 2018b). For instance, Mota et al. (2018b) performed a candidate gene mapping study fitting genomic reaction norms in a single-step GWAS (**ssGWAS**; Wang et al., 2014a) in beef cattle, in which SNP effect patterns across environments were obtained. Howard et al. (2015) and Oliveira et al. (2018a) performed a ssGWAS on the animal polynomial coefficients (modeled by Legendre polynomials) for different longitudinal traits. Previous research have utilized RRM to characterize the effect of individual SNP across time using either simulated (Lund et al., 2008) or real data (Szyda et al., 2014) on a small number of markers. Characterizing SNP effects across a trajectory when the data are derived from dense SNP arrays is quite computationally demanding.

Random regression-based models are currently the most sophisticated and flexible tools to perform GWAS for longitudinal traits. Das et al. (2011) proposed a method based on random regression enabling to exploit, for instance, additive and dominant effects, and specific covariance functions to describe changes in the SNP effects over time. The authors generalized this method to a multiple-trait approach, making it even more powerful and applicable to situations involving more than one trait.

As discussed by Ning et al. (2017), RRM increases the power to detect QTL compared to the accumulated phenotypes, repeatability model and multiple-trait model in QTL mapping (Macgregor, 2005; Suchocki et al., 2013). The RRM was also suitable for QTL detecting in the presence of gene by environment interactions (Lillehammer et al., 2007). Ning et al. (2017) developed two models based on RRM to model the time-dependent SNP effect for the GWAS analysis, i.e., GWAS model treating each SNP as the covariate and GWAS model treating each SNP as the factor. The authors showed that the proposed models captured genetic differences that occurred during the analyzed time period, which increased the statistical power of QTL detection. Different from Das et al. (2011), Ning et al. (2017) divided the time-varied mean values for SNP genotypes into two parts, time-dependent population mean and SNP effects, instead of fitting them directly. Using simulated and real data, Ning et al. (2017) showed that the proposed methods based on longitudinal phenotypes outperformed GWAS methods that were based on deregressed proofs, both regarding control of false positive rate and power of QTL detection. A constraint of the models proposed by Ning et al. (2017) was the computational inefficiency, as the dimension of the mixed model equation was larger than other models. Subsequently, Ning et al. (2018) proposed an alternative approach, termed as rapid longitudinal GWAS method, also based on RRM, which uses Eigen decomposition to rotate the dense phenotypic covariance matrix into a diagonal matrix, thereby transforming the complex mixed linear model into weighted least squares analysis. Using simulated data, they showed that this method can control type I errors well and has higher power than a longitudinal GWAS method that does not include time-varied additive genetic effects. As mentioned by the authors, the orders of Legendre polynomials will influence the performance of the proposed longitudinal GWAS model.

Strategies to perform GWAS, which can fit the time-varied marker effects and directly use the raw longitudinal records are very advantageous, as they avoid the need to generate pseudo-phenotypes. In addition, as highlighted by Rodriguez-Zas et al. (2002) and Szyda et al. (2014), stage-specific QTLs might not be identified when using cumulative lactation record models since the effects of these QTL are likely to be diluted across the lactation.

Finally, if one focuses on few SNP of interest, the usual RRM can be augmented with a factor (SNP genotype) with either fixed or random regressions within each of the three genotypes. Either **A** or **H** matrices can be used to model the covariance among animal additive genetic effects for genotyped animals. This would give three trajectories, one for each SNP genotype, which can be plotted or tested for differences from each other. This is simple and existing software could be used, without substantially increasing computational time.

2.8. Conclusions

Genomic predictions for various points in time can be simultaneously done using RRM, which enables selection of breeding animals based on the complete pattern of the production curve using genomic information. Therefore, including genomic information to evaluate longitudinal traits using RRM is an alternative to yield more accurate selection and culling decisions. Moreover, RRM can be also used to estimate SNP effects and the variances explained by them over time in genome-wide association studies. By analyzing marker associations over time, one is more likely to identify regions with higher effects at specific time periods, which would contribute to explain better the observed phenotypic variation in longitudinal traits. Despite all advances in applications of RRM in the genomic era, more research is still needed. Future research should provide better understanding of the temporal variation of biological processes and their physiological implications underlying the longitudinal traits.

2.9. References

Abou Khadiga, G., B. Y. F. Mahmoud, G. S. Farahat, A. M. Emam, and E. A. El-Full. 2017. Genetic analysis of partial egg production records in Japanese quail using random regression models. Poult. Sci. 96:2569–2575. doi:10.3382/ps/pex081.

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010a. 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.

Aguilar, I., I. Misztal, and S. Tsuruta. 2010b. Short communication: Genetic trends of milk yield under heat stress for US Holsteins. J. Dairy Sci. 93:1754–1758. doi:10.3168/jds.2009-2756.

Agyemang, K., E. C. Clapp, and L. D. Van Vleck. 1985. Variance-covariance components associated with trimester yields of milk and fat and multiple trait sire evaluation for trimester yields. J. Dairy Sci. 68:1233–1240. doi:10.3168/jds.S0022-0302(85)80951-6.

Alam, M., C. I. Cho, T. J. Choi, B. Park, J. G. Choi, Y. H. Choy, S. S. Lee, and K. H. Cho. 2015. Estimation of genetic parameters for somatic cell scores of Holsteins using multi-trait lactation models in Korea. Asian-Australasian J. Anim. Sci. 28:303–310. doi:10.5713/ajas.13.0627.

Ali, T. E., and L. R. Schaeffer. 1987. Accounting for covariances among test day milk yields in dairy cows. Can. J. Anim. Sci. 67:637–644. doi:10.4141/cjas87-067.

Aliloo, H., J. E. Pryce, O. González-Recio, B. G. Cocks, and B. J. Hayes. 2016. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. Genet. Sel. Evol. 48:8. doi:10.1186/s12711-016-0186-0.

Andonov, S., J. Ødegård, M. Svendsen, T. Ådnøy, M. Vegara, and G. Klemetsdal. 2013. Comparison of random regression and repeatability models to predict breeding values from test-day records of Norwegian goats. J. Dairy Sci. 96:1834–1843. doi:10.3168/jds.2012-5910.

Angelovici, R., A. Batushansky, N. Deason, S. Gonzalez-Jorge, M. A. Gore, A. Fait, and D. DellaPenna. 2017. Network-guided GWAS improves identification of genes affecting free amino acids. Plant Physiol. 173:872-886. doi: 10.1104/pp.16.01287.

Auran, T. 1973. Studies on monthly and cumulative monthly milk yield records. Acta Agric. Scand. 23:189–199. doi:10.1080/00015127309433217.

Averill, T. A., R. Rekaya, and K. Weigel. 2004. Genetic analysis of male and female fertility using longitudinal binary data. J. Dairy Sci. 87:3947–3952. doi:10.3168/jds.S0022-0302(04)73534-1.

Baba, T., Y. Gotoh, S. Yamaguchi, S. Nakagawa, H. Abe, Y. Masuda, and T. Kawahara. 2017. Application of single-step genomic best linear unbiased prediction with a multiple-lactation random regression test-day model for Japanese Holsteins. Anim. Sci. J. 88:1226–1231. doi:10.1111/asj.12760.

Banos, G., and G. E. Shook. 1990. Genotype by environment interaction and genetic correlations among parities for somatic cell count and milk yield. J. Dairy Sci. 73:2563–2573. doi:10.3168/jds.S0022-0302(90)78942-4.

Banos, G., J. A. Woolliams, B. W. Woodward, A. B. Forbes, and M. P. Coffey. 2008. Impact of single nucleotide polymorphisms in Leptin, Leptin receptor, growth hormone receptor, and diacylglycerol acyltransferase (DGAT1) gene loci on milk production, feed, and body energy traits of UK dairy cows. J. Dairy Sci. 91:3190–3200. doi:10.3168/jds.2007-0930.

Bartolomé, E., A. Menéndez-Buxadera, A. Molina, and M. Valera. 2018. Plasticity effect of rider-horse interaction on genetic evaluations for Show Jumping discipline in sport horses. J. Anim. Breed. Genet. 135:138–148. doi:10.1111/jbg.12315.

Begli, H. E., R. V. Torshizi, A. A. Masoudi, A. Ehsani, and J. Jensen. 2016. Longitudinal analysis of body weight, feed intake and residual feed intake in F2 chickens. Livest. Sci. 184:28–34. doi:10.1016/j.livsci.2015.11.018.

von Bertalanffy, L. 1957. Quantitative laws in metabolism and growth. Q. Rev. Biol. 32:217–231. doi:10.1086/401873.

Bignardi, A. B., L. El Faro, R. A. A. Torres Júnior, V. L. Cardoso, P. F. Machado, and L. G. Albuquerque. 2011. Random regression models using different functions to model test-day milk yield of Brazilian Holstein cows. Genet. Mol. Res. 10:3565–3575. doi:10.4238/2011.October.31.4.

Bionaz, M., and J. J. Loor. 2008. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J. Nutr. 138:1019–1024. doi:10.1093/jn/138.6.1019.

Bohmanova, J., F. Miglior, J. Jamrozik, I. Misztal, and P. G. Sullivan. 2008. Comparison of random regression models with Legendre polynomials and linear Splines for production traits and somatic cell score of canadian Holstein cows. J. Dairy Sci. 91:3627–3638. doi:10.3168/jds.2007-0945.

Bohmanova, J., I. Misztal, and J. K. Bertrand. 2005. Studies on multiple trait and random regression models for genetic evaluation of beef cattle for growth. J. Anim. Sci. 83:62–67. doi:10.2527/2005.83162x.

Boichard, D., V. Ducrocq, P. Croiseau, and S. Fritz. 2016. Genomic selection in domestic animals: Principles, applications and perspectives. C. R. Biol. 339:274–277. doi:10.1016/j.crvi.2016.04.007.

Boligon, A. A., F. Baldi, M. E. Z. Mercadante, R. B. Lôbo, R. J. Pereira, and L. G. Albuquerque. 2011. Breeding value accuracy estimates for growth traits using random regression and multi-trait models in Nelore cattle. Genet. Mol. Res. 10:1227–1236. doi:10.4238/vol10-2gmr1087.

de Boor, C. R. J. 1980. A practical guide to Splines. Math. Comput. 34:325. doi:10.2307/2006241.

Borquis, R. R. A., F. R. A. Neto, F. Baldi, N. Hurtado-Lugo, G. M. F. de Camargo, M. Muñoz-Berrocal, and H. Tonhati. 2013. Multiple-trait random regression models for the estimation of genetic parameters for milk, fat, and protein yield in buffaloes. J. Dairy Sci. 96:5923–5932. doi:10.3168/jds.2012-6023.

Bouallegue, M., R. Steri, N. M'hamdi, and M. Ben Hamouda. 2015. Modelling of individual lactation curves of Tunisian Holstein-Friesian cows for milk yield, fat, and protein contents using parametric, orthogonal and spline models. J. Anim. Feed Sci. 24:11–18. doi:10.22358/jafs/65648/2015.

Brito, L. F., F. G. Silva, H. R. Oliveira, N. Souza, G. Caetano, E. V. Costa, G. R. O. Menezes, A. L. P. Melo, M. T. Rodrigues, and R. A. Torres. 2017. Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. Can. J. Anim. Sci. CJAS-2017-0019. doi:10.1139/CJAS-2017-0019.

Brody, S., and H. A. Lardy. 1946. Bioenergetics and growth. J. Phys. Chem. 50:168–169. doi:10.1021/j150446a008.

Brown, J. E., C. J. Brown, and W. T. Butts. 1972. A discussion of the genetic aspects of weight, mature weight and rate of maturing in Hereford and Angus cattle. J. Anim. Sci. 34:525–537. doi:10.2527/jas1972.344525x.

Brügemann, K., E. Gernand, U. U. von Borstel, and S. König. 2013. Application of random regression models to infer the genetic background and phenotypic trajectory of binary conception rate by alterations of temperature × humidity indices. Livest. Sci. 157:389–396. doi:10.1016/j.livsci.2013.08.009.

Buch, L. H., M. K. Sørensen, P. Berg, L. D. Pedersen, and A. C. Sørensen. 2012. Genomic selection strategies in dairy cattle: Strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. J. Anim. Breed. Genet. 129:138–151. doi:10.1111/j.1439-0388.2011.00947.x.

Buitenhuis, B., L. L. Janss, N. A. Poulsen, L. B. Larsen, M. K. Larsen, and P. Sørensen. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15:1112. doi:10.1186/1471-2164-15-1112.

Calus, M. P. L., and R. F. Veerkamp. 2003. Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model. J. Dairy Sci. 86:3756–3764. doi:10.3168/jds.S0022-0302(03)73982-4.

Campbell, M. T., H. Walia, and G. Morota. 2018. Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. bioRxiv 319897. doi:10.1101/319897.

Campos, G. S., F. A. Reimann, L. L. Cardoso, C. E. R. Ferreira, V. S. Junqueira, P. I. Schmidt, J. Braccini Neto, M. J. I. Yokoo, B. P. Sollero, A. A. Boligon, and F. F. Cardoso. 2018. Genomic prediction using different estimation methodology, blending and cross-validation techniques for growth traits and visual scores in Hereford and Braford cattle. J. Anim. Sci. 96:2579–2595. doi:10.1093/jas/sky175.

Carriedo, J. A., J. A. Baro, L. F. de la Fuente, and F. san Primitivo. 1995. Genetic parameters for milk yield in dairy sheep. J. Anim. Breed. Genet. 112:59–63. doi:10.1111/j.1439-0388.1995.tb00541.x.

Carvalheira, J. G. V., R. W. Blake, E. J. Pollak, R. L. Quaas, and C. V. Duran-Castro. 1998. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in United States Holstein herds. J. Dairy Sci. 81:2738–2751. doi:10.3168/jds.S0022-0302(98)75831-X.

CDN. 2016. Improving existing traits and adding exciting new ones. Canadian Dairy Network. Accessed Oct. 02, 2018. https://www.cdn.ca/document.php?id=442.

Christensen, O. F., and M. S. Lund. 2010. Genomic relationship matrix when some animals are not genotyped Genomic prediction models. Genet. Sel. Evol. 42.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. animal 6:1565–1571. doi:10.1017/S1751731112000742.

Cole, J. B., and P. M. VanRaden. 2006. Genetic evaluation and best prediction of lactation persistency. J. Dairy Sci. 89:2722–2728. doi:10.3168/jds.S0022-0302(06)72348-7.

Coyne, J. M., D. P. Berry, K. Matilainen, M. L. Sevon-Aimonen, E. A. Mantysaari, J. Juga, T. Serenius, and N. McHugh. 2017. Genetic co-variance functions for live weight, feed intake, and efficiency measures in growing pigs. J. Anim. Sci. 95:3822. doi:10.2527/jas2017.1408.

Crispim, A. C., M. J. Kelly, S. E. F. Guimarães, F. F. Silva, M. R. S. Fortes, R. R. Wenceslau, and S. Moore. 2015. Multi-trait GWAS and new candidate genes annotation for growth curve parameters in Brahman cattle. PLoS One 10:e0139906. doi:10.1371/journal.pone.0139906.

Danell, B. 1982. Studies on lactation yield and individual test-day yields of Swedish dairy cows. Acta Agric. Scand. 32:83–92. doi:10.1080/00015128209435734.

Das, K., J. Li, Z. Wang, C. Tong, G. Fu, Y. Li, M. Xu, K. Ahn, D. Mauger, R. Li, and R. Wu. 2011. A dynamic model for genome-wide association studies. Hum. Genet. 129:629–639. doi:10.1007/s00439-011-0960-6.

David, I., L. Bodin, G. Lagriffoul, E. Manfredi, and C. Robert-Granié. 2007. Character process model for semen volume in AI rams: evaluation of correlation structures for long and short-term environmental effects. Genet. Sel. Evol. 39:55. doi:10.1186/1297-9686-39-1-55.

Dominik, S., and A. A. Swan. 2018. Genetic and phenotypic parameters for reproduction, production and bodyweight traits in Australian fine-wool Merino sheep. Anim. Prod. Sci. 58:207. doi:10.1071/AN15738.

Englishby, T. M., K. L. Moore, D. P. Berry, M. P. Coffey, and G. Banos. 2017. Herd-specific random regression carcass profiles for beef cattle after adjustment for animal genetic merit. Meat Sci. 129:188–196. doi:10.1016/j.meatsci.2017.03.005.

Fang, L., G. Sahana, P. Ma, G. Su, Y. Yu, S. Zhang, S., M. S. Lund, and P. Sørensen. 2017. Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. BMC genomics 18:1-12. doi: 10.1186/s12864-017-4004-z.

Ferris, T. A., I. L. Mao, and C. R. Anderson. 1985. Selecting for lactation curve and milk yield in dairy cattle. J. Dairy Sci. 68:1438–1448. doi:10.3168/jds.S0022-0302(85)80981-4.

Flores, E. B., and J. van der Werf. 2015. Random regression test day models to estimate genetic parameters for milk yield and milk components in Philippine dairy buffaloes. J. Anim. Breed.

Genet. 132:289–300. doi:10.1111/jbg.12133.

Forni, S., D. Gianola, G. J. M. Rosa, and G. de los Campos. 2009. A dynamic linear model for genetic analysis of longitudinal traits1. J. Anim. Sci. 87:3845–3853. doi:10.2527/jas.2008-1514.

Gao, H., M. S. Lund, Y. Zhang, and G. Su. 2013. Accuracy of genomic prediction using different models and response variables in the Nordic Red cattle population. J. Anim. Breed. Genet. 130:n/a-n/a. doi:10.1111/jbg.12039.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc. Natl. Acad. Sci. 113:E3995–E4004. doi:10.1073/pnas.1519061113.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55. doi:10.1186/1297-9686-41-55.

Ghavi Hossein-Zadeh, N., M. A. Nazari, and A. A. Shadparvar. 2017. Genetic perspective of milk yield persistency in the first three lactations of Iranian buffaloes (Bubalus bubalis). J. Dairy Res. 84:434–439. doi:10.1017/S0022029917000498.

Ghiasi, H., and M. J. Carabaño. 2018. Eigenvector index for two female fertility traits based in random regression coefficients matrix in Holstein cows. Spanish J. Agric. Res. 16:e0401. doi:10.5424/sjar/2018161-12396.

Ghiasi, H., and M. S. Mokhtari. 2018. Possibility of modifying the growth trajectory in Raeini Cashmere goat. Trop. Anim. Health Prod. 1–6. doi:10.1007/s11250-018-1579-6.

Ghouse, J., M. W. Skov, R. S. Bigseth, G. Ahlberg, J. K. Kanters, and M. S. Olesen. 2017. Distinguishing pathogenic mutations from background genetic noise in cardiology: The use of large genome databases for genetic interpretation. Clin Genet, 93:459-466. doi: 10.1111/cge.13066.

Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2010. Genomic selection in livestock populations. Genet. Res. (Camb). 92:413–421. doi:10.1017/S0016672310000613.

Gompertz, B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philos. Trans. R. Soc. London 115:513–583. doi:10.1098/rstl.1825.0026.

González-Peña, D., J. Acosta, D. Guerra, N. González, M. Acosta, D. Sosa, and G. Torres-Hernández. 2012. Modeling of individual lactation curves for milk production in a population of Alpine goats in Cuba. Livest. Sci. 150:42–50. doi:10.1016/j.livsci.2012.07.026.

Gorlova, O. Y., E. I. Demidenko, C. I. Amos, and I. P. Gorlov. 2017. Downstream targets of GWAS-detected genes for breast, lung, and prostate and colon cancer converge to G1/S transition pathway. Hum. Mol. Gen. 26:1465-1471. doi: 10.1093/hmg/ddx050.

Guo, Z., G. Liu, M. S. Röder, J. C. Reif, M. W. Ganal, and T. Schnurbusch. 2018. Genome-wide association analyses of plant growth traits during the stem elongation phase in wheat. Plant Biotech J. 1-11. doi: 10.1111/pbi.12937.

Gyawali, S., S. Chao, S. S. Vaish, S. P. Singh, S. Rehman, S. R. Vishwakarma, and R. P. S. Verma. 2018. Genome wide association studies (GWAS) of spot blotch resistance at the seedling and the adult plant stages in a collection of spring barley. Mol Breeding, 38:1-14. doi: 10.1007/s11032-018-0815-0.

Harris, B. L., and D. L. Johnson. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J. Dairy Sci. 93:1243–1252. doi:10.3168/jds.2009-2619.

Harville, D. A. 1979. Recursive estimation using mixed linear models with autoregressive random effects. Pages 1-10 in Proc. of a Conference in Honor of CR Henderson, Cornell University, Ithaca, NY.

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

He, J., Y. Zhao, J. Zhao, J. Gao, D. Han, P. Xu, and R. Yang. 2017. Multivariate random regression analysis for body weight and main morphological traits in genetically improved farmed tilapia (Oreochromis niloticus). Genet. Sel. Evol. 49:80. doi:10.1186/s12711-017-0357-7.

Heffner, E. L., A. J. Lorenz, J. L. Jannink, and M. E. Sorrells. 2010. Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50:1681. doi:10.2135/cropsci2009.11.0662.

Hein, L., L. P. Sørensen, 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. doi:10.3168/JDS.2017-13225.

Henderson, C. R. 1963. Selection index and expected genetic advance. Stat. Genet. Plant Breed. 892:141–163.

Henderson, C. R. 1973. Sire evaluation and genetic trends. J. Anim. Sci. 1973:10–41. doi:10.1093/ansci/1973.Symposium.10.

Henderson Jr., C. R. 1982. Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. Biometrics 38:623. doi:10.2307/2530044.

Higgins, M. G., C. Fitzsimons, M. C. McClure, C. McKenna, S. Conroy, D. A. Kenny, M. McGee, S. M. Waters, and D. W. Morris. 2018. GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. Sci. Rep. 8:1-12.

Howard, J. T., S. Jiao, F. Tiezzi, Y. Huang, K. A. Gray, and C. Maltecca. 2015. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. BMC Genet. 16:59. doi:10.1186/s12863-015-0218-8.

Huynh-Tran, V. H., H. Gilbert, and I. David. 2017. Genetic structured antedependence and random regression models applied to the longitudinal feed conversion ratio in growing Large White pigs1. J. Anim. Sci. 95:4752–4763. doi:10.2527/jas2017.1864.

Interbull. 2017. Description of national genomic evaluation systems. Canada. Accessed May. 02, 2018. http://www.interbull.org/ib/nationalgenoforms.

Interbull. 2018. Description of national genetic evaluation systems. Accessed May, 02, 2018. http://www.interbull.org/ib/geforms.

Iwaisaki, H., S. Tsuruta, I. Misztal, and J.K. Bertrand. 2005. Genetic parameters estimated with multitrait and linear spline-random regression models using Gelbvieh early growth data1. J. Anim. Sci. 83:757–763. doi:10.2527/2005.834757x.

Jaffrézic, F., E. Venot, D. Laloë, A. Vinet, and G. Renand. 2004. Use of structured antedependence models for the genetic analysis of growth curves1. J. Anim. Sci. 82:3465–3473.

doi:10.2527/2004.82123465x.

Jamrozik, J., J. Bohmanova, and L. R. Schaeffer. 2010. Selection of locations of knots for linear splines in random regression test-day models. J. Anim. Breed. Genet. 127:87–92. doi:10.1111/j.1439-0388.2009.00829.x.

Jamrozik, J., D. Gianola, and L. R. Schaeffer. 2001. Bayesian estimation of genetic parameters for test day records in dairy cattle using linear hierarchical models. Livest. Prod. Sci. 71:223–240. doi:10.1016/S0301-6226(01)00186-5.

Jamrozik, J., G. Jansen, L. R. Schaeffer, and Z. Liu. 1998a. Analysis of persistency of lactation calculated from a random regression test day model. Interbull Bull. 17:64–69.

Jamrozik, J., G. J. Kistemaker, J. C. M. Dekkers, and L. R. Schaeffer. 1997a. Comparison of possible covariates for use in a random regression model for analyses of test day yields. J. Dairy Sci.. doi:10.3168/jds.S0022-0302(97)76210-6.

Jamrozik, J., L.. Schaeffer, and G.. Jansen. 2000. Approximate accuracies of prediction from random regression models. Livest. Prod. Sci. 66:85–92. doi:10.1016/S0301-6226(00)00158-5.

Jamrozik, J., and L.R. Schaeffer. 1997. Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. J. Dairy Sci.. doi:10.3168/jds.S0022-0302(97)75996-4.

Jamrozik, J., and L. R. Schaeffer. 2002. Bayesian comparison of random regression models for test-day yields in dairy cattle. Pages 21–24 in 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. Editions Quae, Montpellier, France.

Jamrozik, J., L. R. Schaeffer, and J. C. Dekkers. 1997b. Genetic evaluation of dairy cattle using test day yields and random regression model. J. Dairy Sci.. doi:10.3168/jds.S0022-0302(97)76050-8.

Jamrozik, J., L. R. Schaeffer, and F. Grignola. 1998b. Genetic parameters for production traits and somatic cell score of Canadian Holsteins with multiple trait random regression model. Pages 3-3 in 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia. Armidale, New England.

Jamrozik, J., L. R. Schaeffer, Z. Liu, and G. Jansen. 1997c. Multiple trait random regression test day model for production traits. Interbull Bull. 43.

Jattawa, D., M. A. Elzo, S. Koonawootrittriron, and T. Suwanasopee. 2016. Genomic-polygenic and polygenic evaluations for milk yield and fat percentage using random regression models with Legendre polynomials in a Thai multibreed dairy population. Livest. Sci. 188:133–141. doi:10.1016/j.livsci.2016.04.019.

Jenko, J., G. R. Wiggans, T. A. Cooper, S. A. E. Eaglen, W. G. de. L. Luff, M. Bichard, R. Pong-Wong, and J. A. Woolliams. 2017. Cow genotyping strategies for genomic selection in a small dairy cattle population. J. Dairy Sci. 100:439–452. doi:10.3168/jds.2016-11479.

Jensen, J. 2001. Genetic evaluation of dairy cattle using test-day models. J. Dairy Sci. 84:2803–2812. doi:10.3168/jds.S0022-0302(01)74736-4.

Jun, G., W. Kehua, Q. Liang, D. Taocun, S. Manman, M. Meng, and H. Yuping. 2017. Random regression analyses to model the longitudinal measurements of yolk proportions in the laying chicken. Poult. Sci. 96:3852–3857. doi:10.3382/ps/pex233.

Kang, H., C. Ning, L. Zhou, S. Zhang, Q. Yan, and J. F. Liu. 2018. Short communication: Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins. J. Dairy Sci. 101:1–7. doi: 10.3168/jds.2018-15090.

Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J. F. Liu. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity (Edinb). 119:459–467. doi:10.1038/hdy.2016.91.

Karami, K., S. Zerehdaran, B. Barzanooni, and E. Lotfi. 2017. Genetic analyses of partial egg production in Japanese quail using multi-trait random regression models. Br. Poult. Sci. 58:624–628. doi:10.1080/00071668.2017.1378423.

Kennedy, B. W., and C. R. Henderson. 1975. Components of variance of growth traits among Hereford and Aberdeen Angus calves. Can. J. Anim. Sci. 55:493–502. doi:10.4141/cjas75-061.

Keown, J. F., and L. D. Van Vleck. 1971. Selection on test-day fat percentage and milk production. J. Dairy Sci. 54:199–203. doi:10.3168/jds.S0022-0302(71)85812-5.

Kersey, R., and J. Brinks. 1985. Genetic and environmental aspects of the growth curve parameters in beef cows. J. Anim. Sci. 61:1431–1440.

Khanzadeh, H., N. Ghavi Hossein-Zadeh, M. Naserani, and B. Mohammad Nazari. 2013. Calculating daughter yield deviations for production traits in Holstein cattle using repeatability animal and random regression test day models. Livest. Sci. 157:408–413. doi:10.1016/j.livsci.2013.08.018.

Kheirabadi, K. 2018. Bayesian analysis of random regression models to model test-day somatic cell score of primiparous Holstein cattle in Iran. J. Appl. Anim. Res. 46:677–684. doi:10.1080/09712119.2017.1386107.

Kheirabadi, K., A. Rashidi, S. Alijani, and I. Imumorin. 2014. Modeling lactation curves and estimation of genetic parameters in Holstein cows using multiple-trait random regression models. Anim. Sci. J. 85:925–934. doi:10.1111/asj.12185.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979–93.

Kistemaker, G.J. 1997. The comparision of random regression test day models and a 305-day for evaluation of milk yield in dairy cattle. PhD Thesis. Univ. of Guelph, Guelph, Ontario, Canada.

Köck, A., M. Ledinek, L. Gruber, F. Steininger, B. Fuerst-Waltl, and C. Egger-Danner. 2018. Genetic analysis of efficiency traits in Austrian dairy cattle and their relationships with body condition score and lameness. J. Dairy Sci. 101:445–455. doi:10.3168/jds.2017-13281.

Koivula, M., I. Strandén, J. Pösö, G.P. Aamand, and E. A. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775–2784. doi:10.3168/jds.2014-8975.

Koivula, M., I. Strandén, G. Su, and E. A. Mäntysaari. 2012. Different methods to calculate genomic predictions—Comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). J. Dairy Sci. 95:4065–4073. doi:10.3168/jds.2011-4874.

König, S., H. Simianer, and A. Willam. 2009. Economic evaluation of genomic breeding programs.

J. Dairy Sci. 92:382–391. doi:10.3168/jds.2008-1310.

Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. Biometrics 38:963. doi:10.2307/2529876.

Lamb, R. C., and L. D. McGilliard. 1967. Ratio factors to estimate 305-day production from lactation records in progress. J. Dairy Sci. 50:1101–1108. doi:10.3168/jds.S0022-0302(67)87573-8.

Laureano, M. M. M., A. B. Bignardi, L. El Faro, V. L. Cardoso, H. Tonhati, and L. G. Albuquerque. 2014. Random regression models using different functions to model milk flow in dairy cows. Genet. Mol. Res. 13:7528–7541. doi:10.4238/2014.September.12.20.

Lázaro, S. F., N. Ibáñez-Escriche, L. Varona, F. F. Silva, L. C. Brito, S. E. F. Guimarães, and P. S. Lopes. 2017. Bayesian analysis of pig growth curves combining pedigree and genomic information. Livest. Sci. 201:34–40. doi:10.1016/j.livsci.2017.03.024.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single step, a general approach for genomic selection. Livest. Sci. 166:54–65. doi:10.1016/j.livsci.2014.04.029.

León, J. M., N. P. P. Macciotta, L. T. Gama, C. Barba, and J. V. Delgado. 2012. Characterization of the lactation curve in Murciano-Granadina dairy goats. Small Rumin. Res. 107:76–84. doi:10.1016/j.smallrumres.2012.05.012.

Liberona, J. S. D., J. M. Langdon, D. G. Riley, H. D. Blackburn, S. E. Speidel, B. C. Krehbiel, S. Sanders, and A. D. Herring. 2018. Heritability estimations for intramuscular fat in Hereford cattle using random regressions. J. Anim. Sci. 96:2–3. doi:10.1093/jas/sky027.004.

Lidauer, M., E. A. Mäntysaari, and I. Strandén. 2003. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. Livest. Prod. Sci. 79:73–86. doi:10.1016/S0301-6226(02)00142-2.

Lillehammer, M., J. Ødegård, and T. H. Meuwissen. 2007. Random regression models for detection of gene by environment interaction. Genet. Sel. Evol. 39:105. doi:10.1186/1297-9686-39-2-105.

Lindgren, N., J. Philipsson, and A. Elofson-Bernstedt. 1980. Studies on monthly protein records of individual cows. Acta Agric. Scand. 30:437–444. doi:10.1080/00015128009435291.

Liu, X., J. Zhao, Y. Zheng, N. Sun, X. Li, and R. Yang. 2018. Genetic analysis for absolute and relative growth of body weight and main morphological traits in chinese mitten crab, Eriocheir sinensis. J. World Aquac. Soc. 49:380–387. doi:10.1111/jwas.12518.

de los Campos, G., A. I. Vazquez, R. Fernando, Y. C. Klimentidis, and D. Sorensen. 2013. Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS Genet. 9:e1003608. doi:10.1371/journal.pgen.1003608.

Lourenco, D. A. L., I. Misztal, S. Tsuruta, I. Aguilar, E. Ezra, M. Ron, A. Shirak, and J. I. Weller. 2014. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. J. Dairy Sci. 97:1742–1752. doi:10.3168/jds.2013-6916.

Lu, D., S. Miller, M. Sargolzaei, M. Kelly, G. Vander Voort, T. Caldwell, Z. Wang, G. Plastow, and S. Moore. 2013. Genome-wide association analyses for growth and feed efficiency traits in beef cattle. J. Anim. Sci. 91:3612–3633. doi:10.2527/jas.2012-5716.

Lund, M. S., P. Sorensen, P. Madsen, and F. Jaffrézic. 2008. Detection and modelling of time-dependent QTL in animal populations. Genet. Sel. Evol. 40:177–194. doi:10.1051/gse:2007043.

Macciotta, N. P. P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. J. Dairy Sci. 98:8175–8185. doi:10.3168/jds.2015-9500.

Macgregor, S. 2005. Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. Genetics 171:1365–1376. doi:10.1534/genetics.105.043828.

Makgahlela, M. L., T. Knürr, G. P. Aamand, I. Strandén, and E. A. Mäntysaari. 2013. Single step evaluations using haplotype segments. Interbull Bull. 23–25.

Mäntysaari, E. A., M. Koivula, I. Strandén, J. Pösö, G. P. Aamand, and F. Coop. 2011. Estimation of GEBVs using deregressed individual cow breeding values. Interbull Bull. 19–24.

Mark, T. 2004. Applied genetic evaluations for production and functional traits in dairy cattle. J. Dairy Sci. 87:2641–2652. doi:10.3168/jds.S0022-0302(04)73390-1.

Martínez, R., D. Bejarano, Y. Gómez, R. Dasoneville, A. Jiménez, G. Even, J. Sölkner, and G. Mészáros. 2017. Genome-wide association study for birth, weaning and yearling weight in Colombian Brahman cattle. Genet. Mol. Biol. 40:453–459. doi:10.1590/1678-4685-gmb-2016-0017.

Meng, Q., K. Wang, X. Liu, H. Zhou, L. Xu, Z. Wang, and M. Fang. 2017. Identification of growth trait related genes in a Yorkshire purebred pig population by genome-wide association studies. Asian-Australas J. Anim. Sci. 30:462-469. doi: 10.5713/ajas.16.0548.

Meredith, B. K., F. J. Kearney, E. K. Finlay, D. G. Bradley, A. G. Fahey, D. P. Berry, and D. J. Lynn. 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet. 13:21. doi:10.1186/1471-2156-13-21.

Meuwissen, T., B. Hayes, and M. Goddard. 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front. 6:6–14. doi:10.2527/af.2016-0002.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157.

Meyer, B.K. 1999. Estimates of genetic and phenotypic covariance functions for postweaning growth and mature weight of beef cows. J. Anim. Breed. Genet. 116:181–205. doi:10.1046/j.1439-0388.1999.00193.x.

Meyer, K. 1995. Estimates of genetic parameters for mature weight of Australian beef cows and its relationship to early growth and skeletal measures. Livest. Prod. Sci. 44:125–137. doi:10.1016/0301-6226(95)00067-4.

Meyer, K. 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. Genet. Sel. Evol. 37:473–500. doi:10.1051/gse:2005012.

Meyer, K., H. U. Graser, and K. Hammond. 1989. Estimates of genetic parameters for first lactation test day production of Australian black and white cows. Livest. Prod. Sci. 21:177–199. doi:10.1016/0301-6226(89)90049-3.

Meyer, K., and W. G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or ‘repeated’ records by restricted maximum likelihood. Livest. Prod. Sci. 47:185–200. doi:10.1016/S0301-6226(96)01414-5.

Miglior, F., A. Fleming, F. Malchiodi, L. F. Brito, P. Martin, and C. F. Baes. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100:10251–10271. doi:10.3168/jds.2017-12968.

Misztal, I. 2006. Properties of random regression models using linear splines. J. Anim. Breed. Genet. 123:74–80. doi:10.1111/j.1439-0388.2006.00582.x.

Misztal, I., H. L. Bradford, D. A. L. Lourenco, S. Tsuruta, Y. Masuda, and A. Legarra. 2017. Studies on inflation of GEBV in single-step GBLUP for type. Interbull Bul. 38–42.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. doi:10.3168/jds.2009-2064.

Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. J. Dairy Sci. 97:3943–3952. doi:10.3168/jds.2013-7752.

Misztal, I., T. Strabel, J. Jamrozik, E. A. Mäntysaari, and T. H. E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. J. Dairy Sci. 83:1125–1134. doi:10.3168/jds.S0022-0302(00)74978-2.

Miyumo, S., C. B. Wasike, and A. K. Kahi. 2018. Genetic and phenotypic parameters for feed efficiency in indigenous chicken in Kenya. Livest. Sci. 207:91–97. doi:10.1016/j.livsci.2017.11.011.

Mookprom, S., W. Boonkum, S. Kunhareang, S. Siripanya, and M. Duangjinda. 2017. Genetic evaluation of egg production curve in Thai native chickens by random regression and spline models. Poult. Sci. 96:274–281. doi:10.3382/ps/pew326.

Morant, S. V., and A. Gnanasakthy. 1989. A new approach to the mathematical formulation of lactation curves. Anim. Prod. 49:151–162. doi:10.1017/S000335610003227X.

Moretti, R., M. P. G. de Rezende, S. Biffani, and R. Bozzi. 2018. Heritability and genetic correlations between rumination time and production traits in Holstein dairy cows during different lactation phases. J. Anim. Breed. Genet. 135:293–299. doi:10.1111/jbg.12346.

Mota, L. F. M., P. G. M. A. Martins, T. O. Littiere, L. R. A. Abreu, M. A. Silva, and C. M. Bonafé. 2018a. Genetic evaluation and selection response for growth in meat-type quail through random regression models using B-spline functions and Legendre polynomials. animal 12:667–674. doi:10.1017/S1751731117001951.

Mota, R. R., P. S. Lopes, R. J. Tempelman, F. F. Silva, I. Aguilar, C. C. G. Gomes, and F. F. Cardoso. 2016. Genome-enabled prediction for tick resistance in Hereford and Braford beef cattle via reaction norm models. J. Anim. Sci. 94:1834–1843. doi:10.2527/jas.2015-0194.

Mota, R. R., F. F. Silva, P. S. Lopes, R. J. Tempelman, B. P. Sollero, I. Aguilar, and F. F. Cardoso. 2018b. Analyses of reaction norms reveal new chromosome regions associated with tick resistance in cattle. animal 12:205–214. doi:10.1017/S1751731117001562.

Musolf, A., A. Q. Nato, D. Londono, L. Zhou, T. C. Matise, and D. Gordon. 2014. Mapping genes with longitudinal phenotypes via Bayesian posterior probabilities. BMC Proc. 8:S81. doi:10.1186/1753-6561-8-S1-S81.

Naderi, Y. 2018. Genetic evaluation and genetic trend of growth of Makouei sheep via random regression. J. Anim. Plant Sci. 28.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F. Schenkel, S. S. Moore, and

P. Stothard. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 17:75. doi:10.1186/s12863-016-0386-1.

Nejati-Javaremi, A., C. Smith, and J. P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection.. J. Anim. Sci. 75:1738. doi:10.2527/1997.7571738x.

Nelder, J. A. 1961. The fitting of a generalization of the logistic curve. Biometrics 17:89. doi:10.2307/2527498.

Ning, C., H. Kang, L. Zhou, D. Wang, H. Wang, A. Wang, J. Fu, S. Zhang, and J. Liu. 2017. Performance gains in genome-wide association studies for longitudinal traits via modeling time-varied effects. Sci. Rep. 7:590. doi:10.1038/s41598-017-00638-2.

Ning, C., D. Wang, X. Zheng, Q. Zhang, S. Zhang, R. Mrode, and J.-F. Liu. 2018. Eigen decomposition expedites longitudinal genome-wide association studies for milk production traits in Chinese Holstein. Genet. Sel. Evol. 50:12. doi:10.1186/s12711-018-0383-0.

Nustad, H. E., C. M. Page, A. H. Reiner, M. Zucknick, and M. LeBlanc. 2018. A Bayesian mixed modeling approach for estimating heritability. BMC Proc. 12:31. doi:10.1186/s12919-018-0131-z.

Oliveira, D. P., D. A. L. Lourenco, S. Tsuruta, I. Misztal, D. J. A. Santos, F. R. de Araújo Neto, R. R. Aspilcueta-Borquis, F. Baldi, R. Carvalheiro, G. M. F. de Camargo, L. G. Albuquerque, and H. Tonhati. 2018a. Reaction norm for yearling weight in beef cattle using single-step genomic evaluation1. J. Anim. Sci. 96:27–34. doi:10.1093/jas/skx006.

Oliveira, H. R., F. F. Silva, M. V. G. B. Silva, O. H. G. B. D. Siqueira, M. A. Machado, J. C. C. Panetto, L. S. Glória, and L. F. Brito. 2017a. Bayesian models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. Livest. Sci. 201:78–84. doi:10.1016/j.livsci.2017.05.007.

Oliveira, H. R., F. F. Silva, L. F. Brito, A. R. Guarini, J. Jamrozik, and F. S. Schenkel. 2018b. Comparing deregression methods for genomic prediction of test-day traits in dairy cattle. J. Anim. Breed. Genet. 135:97–106. doi:10.1111/jbg.12317.

Oliveira, H. R., F. F. Silva, O. H. G. B. D. Siqueira, N. O. Souza, V. S. Junqueira, M. D. V. Resende, R. R. A. Borquis, and M. T. Rodrigues. 2016. Combining different functions to describe milk, fat, and protein yield in goats using Bayesian multiple-trait random regression models1. J. Anim. Sci. 94:1865–1874. doi:10.2527/jas.2015-0150.

Oliveira, H. R., L. F. Brito, F. F. Silva, D. A. L. Lourenco, J. Jamrozik, and F. S. Schenkel. 2018c. Genomic prediction of lactation curves for milk, fat, protein and somatic cell score in Holstein cattle. J. Dairy Sci. 102:1-12.

Oliveira, L. T., C. M. Bonafé, F. F. Silva, H. T. Ventura, H. R. Oliveira, G. R. O. Menezes, M. D. V. Resende, and J.M.S. Viana. 2017b. Bayesian random regression threshold models for genetic evaluation of pregnancy probability in Red Sindhi heifers. Livest. Sci. 202:166–170. doi:10.1016/j.livsci.2017.06.005.

Padilha, A. H., E. P. M. Alfonzo, D. S. Daltro, H. A. L. Torres, J. B. Neto, and J. A. Cobuci. 2018. Genetic trends and genetic correlations between 305-day milk yield, persistency and somatic cell score of Holstein cows in Brazil using random regression model. Anim. Prod. Sci.. doi:10.1071/AN16835.

Pander, B., and W. Hill. 1993. Genetic evaluation of lactation yield from test day records on incomplete lactation. Livest. Prod. Sci. 37:23–36. doi:10.1016/0301-6226(93)90062-M.

Pandey, M. K., M. Roorkiwal, V. K. Singh, A. Ramalingam, H. Kudapa, M. Thudi, A. Chitikineni, A. Rathore, and R. K. Varshney. 2016. Emerging genomic tools for legume breeding: current status and future prospects. Front. Plant Sci. 7:455. doi:10.3389/fpls.2016.00455.

Pereira, R. J., A. B. Bignardi, L. El Faro, R. S. Verneque, A. E. Vercesi Filho, and L. G. Albuquerque. 2013. Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (Bos indicus) cattle. J. Dairy Sci. 96:565–574. doi:10.3168/jds.2011-5051.

Pletcher, S. D., and C. J. Geyer. 1999. The genetic analysis of age-dependent traits: Modeling the character process. Genetics. doi:10.1534/genetics.104.029181.

Powell, R. L., and H. D. Norman. 2006. Major advances in genetic evaluation techniques. J. Dairy Sci. 89:1337–1348. doi:10.3168/jds.S0022-0302(06)72201-9.

Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, and B. J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93:3331–3345. doi:10.3168/jds.2009-2893.

Ptak, E., and L. R. Schaeffer. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34:23–34. doi:10.1016/0301-6226(93)90033-E.

Qiao, R., J. Gao, Z. Zhang, L. Li, X. Xie, Y. Fan, L. Cui, J. Ma, H. Ai, J. Ren, and L. Huang. 2015. Genome-wide association analyses reveal significant loci and strong candidate genes for growth and fatness traits in two pig populations. Genet. Sel. Evol. 47:17. doi:10.1186/s12711-015-0089-5.

Quaas, R. L., R. D. Anderson, and A. R. Gilmour. 1984. Use of mixed models for prediction and for estimation of (co)variance components. Animal Genetics and Breeding Unit, Univ. of New England, New South Wales, Australia.

Raszek, M. M., L. L. Guan, and G. S. Plastow. 2016. Use of genomic tools to improve cattle health in the context of infectious diseases. Front. Genet. 7:30. doi:10.3389/fgene.2016.00030.

Raven, L. A., B. G. Cocks, and B. J. Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:62. doi:10.1186/1471-2164-15-62.

Richards, F. J. 1959. A flexible growth function for empirical use. J. Exp. Bot. 10:290–301. doi:10.1093/jxb/10.2.290.

Robbins, K. R., I. Misztal, and J. K. Bertrand. 2005. A practical longitudinal model for evaluating growth in Gelbvieh cattle. J. Anim. Sci. 83:29–33. doi:10.2527/2005.83129x.

Rodriguez-Zas, S. L., B. R. Southey, D. W. Heyen, and H. A. Lewin. 2002. Detection of quantitative trait loci influencing dairy traits using a model for longitudinal data. J. Dairy Sci. 85:2681–2691. doi:10.3168/jds.S0022-0302(02)74354-3.

Rothschild, M. F., and C. R. Henderson. 1979. Maximum likelihood estimates of parameters of first and second lactation milk records. J. Dairy Sci. 62:990–995. doi:10.3168/jds.S0022-0302(79)83360-3.

Saghi, D. A., A. R. Shahdadi, F. K. Borzelabad, and K. Mohammadi. 2018. Estimates of covariance functions for growth of Kordi sheep in Iran using random regression models. Small Rumin. Res. 162:69–76. doi:10.1016/j.smallrumres.2018.03.007.

Samorè, A. B., and L. Fontanesi. 2016. Genomic selection in pigs: state of the art and perspectives. Ital. J. Anim. Sci. 15:211–232. doi:10.1080/1828051X.2016.1172034.

Sanchez, M. P., M. El Jabri, S. Minéry, V. Wolf, E. Beuvier, C. Laithier, A. Delacroix-Buchet, M. Brochard, and D. Boichard. 2018. Genetic parameters for cheese-making properties and milk composition predicted from mid-infrared spectra in a large data set of Montbéliarde cows. J. Dairy Sci.. doi:10.3168/JDS.2018-14878.

Santana Jr., M. L., J. P. Eler, A. B. Bignardi, A. A. Boligon, and J. B. S. Ferraz. 2018. Genetic correlation between growth and reproductive performance of beef females depends on environment. Anim. Prod. Sci. 58:1201. doi:10.1071/AN16108.

Santana, M. L., J. P. Eler, A. B. Bignardi, and J. B. S. Ferraz. 2015. Two-trait random regression model to estimate the genetic association of scrotal circumference with female reproductive performance in Nelore cattle. Theriogenology 83:1534–1540. doi:10.1016/j.theriogenology.2015.02.003.

Sawalha, R. M., J. F. Keown, S. D. Kachman, and L. D. Van Vleck. 2005. Evaluation of autoregressive covariance structures for test-day records of Holstein cows: estimates of parameters. J. Dairy Sci. 88:2632–2642. doi:10.3168/jds.S0022-0302(05)72940-4.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35–45. doi:10.1016/S0301-6226(03)00151-9.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123:218–223. doi:10.1111/j.1439-0388.2006.00595.x.

Schaeffer, L. R. 2016. Random regression models. Univ. of Guelph, Guelph, Ontario, Canada. Available at http://animalbiosciences.uoguelph.ca/~lrs/BOOKS/rrmbook.pdf.

Schaeffer, L. R. 2018. Necessary changes to improve animal models. J. Anim. Breed. Genet.. doi:10.1111/jbg.12321.

Schaeffer, L. R., and E. B. Burnside. 1976. Estimating the shape of the lactation curve. Can. J. Anim. Sci. 56.

Schaeffer, L. R., and J. C. Dekkers. 1994. Random regressions in animal models for test-day production in dary cattle. Pages 443-446 in Proc. 5th World Congr. Genet. Appl. Livest. Prod. Guelph, Ontario, Canada.

Schaeffer, L. R., and J. Jamrozik. 2008. Random regression models: a longitudinal perspective. J. Anim. Breed. Genet. 125:145–146. doi:10.1111/j.1439-0388.2008.00748.x.

Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and J. Van Doormaal. 2000. Experience with a test-day model. J. Dairy Sci. 83:1135-1144. doi: 10.3168/jds.S0022-0302(00)74979-4.

Schaeffer, L. R., and J. W. Wilton. 1981. Estimation of variances and covariances for use in a multiple trait beef sire evaluation model. Can. J. Anim. Sci. 61:531-538.

Shanks, R. D., P. J. Berger, A. E. Freeman, and F. N. Dickinson. 1981. Genetic aspects of lactation curves. J. Dairy Sci. 64:1852–1860. doi:10.3168/jds.S0022-0302(81)82775-0.

Short, T. H., R. W. Blake, R. L. Quaas, and L. D. Van Vleck. 1990. Heterogeneous within-herd variance. 1. Genetic parameters for first and second lactation milk yields of grade Holstein cows. J. Dairy Sci. 73:3312–3320. doi:10.3168/jds.S0022-0302(90)79025-X.

Silva, F. F., H. A. Mulder, E. F. Knol, M. S. Lopes, S. E. F. Guimarães, P. S. Lopes, P. K. Mathur, J. M. S. Viana, and J. W. M. Bastiaansen. 2014. Sire evaluation for total number born in pigs using a genomic reaction norms approach1. J. Anim. Sci. 92:3825–3834. doi:10.2527/jas.2013-6486.

Siqueira, O. H. G. B. D., R. R. Mota, H. R. Oliveira, D. A. S. Duarte, L. S. Glória, M. T. Rodrigues, and F. F. Silva. 2017. Genetic evaluation of lactation persistency and total milk yield in dairy goats. Livest. Res. Rural Dev. 29.

Slater, A. T., N. O. I. Cogan, J. W. Forster, B. J. Hayes, and H. D. Daetwyler. 2016. Improving genetic gain with genomic selection in autotetraploid potato. Plant Genome 9:0. doi:10.3835/plantgenome2016.02.0021.

Snelling, W. M., M. F. Allan, J. W. Keele, L. A. Kuehn, T. McDaneld, T. P. L. Smith, T. S. Sonstegard, R. M. Thallman, and G. L. Bennett. 2010. Genome-wide association study of growth in crossbred beef cattle. J. Anim. Sci. 88:837–848. doi:10.2527/jas.2009-2257.

Soares, A. C. C., S. E. F. Guimarães, M. J. Kelly, M. R. S. Fortes, F. F. Silva, L. L. Verardo, R. Mota, and S. Moore. 2017. Multiple-trait genomewide mapping and gene network analysis for scrotal circumference growth curves in Brahman cattle1. J. Anim. Sci. 95:3331–3345. doi:10.2527/jas.2017.1409.

Soller, M., T. Brody, Y. Eitan, T. Agursky, and C. Wexler. 1984. Minimum weight for onset of sexual maturity in female chickens: heritability and phenotypic and genetic correlations with early growth rate. Poult. Sci. 63:2103–2113. doi:10.3382/ps.0632103.

Speidel, S. E., R. M. Enns, and D. H. Crews Jr. 2010. Genetic analysis of longitudinal data in beef cattle: a review. Genet. Mol. Res. 9:19–33. doi:10.4238/vol9-1gmr675.

Strabel, T., and I. Misztal. 1999. Genetic parameters for first and second lactation milk yields of Polish black and white cattle with random regression test-day models. J. Dairy Sci. 82:2805–2810. doi:10.3168/jds.S0022-0302(99)75538-4.

Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of random regression test-day models for Polish black and white cattle. J. Dairy Sci. 88:3688–3699. doi:10.3168/jds.S0022-0302(05)73055-1.

Strandberg, E., and B. Danell. 1989. Genetic and phenotypic parameters for production and days open in the first three lactations of Swedish dairy cattle. Acta Agric. Scand. 39:203–215. doi:10.1080/00015128909438513.

Strandén, I., and M. Lidauer. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. J. Dairy Sci. 82:2779–2787. doi:10.3168/jds.S0022-0302(99)75535-9.

Strucken, E. M., R. H. Bortfeldt, D. J. Koning, and G. A. Brockmann. 2012. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375–382. doi:10.1111/j.1365-2052.2011.02278.x.

Strucken, E. M., D. J. Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442–449. doi:10.3168/jds.2009-2932.

Strucken, E. M., Y. C. S. M. Laurenson, and G. A. Brockmann. 2015. Go with the flow-biology and genetics of the lactation cycle. Front. Genet. 6:118. doi:10.3389/fgene.2015.00118.

Suchocki, T., J. Szyda, and Q. Zhang. 2013. Modelling QTL effect on BTA06 using random regression test day models. J. Appl. Genet. 54:49–60. doi:10.1007/s13353-012-0114-0.

Swalve, H. H. 1995. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. J. Dairy Sci. 78:929–938. doi:10.3168/jds.S0022-0302(95)76708-X.

Swalve, H. H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. J. Dairy Sci. 83:1115–1124. doi:10.3168/jds.S0022-0302(00)74977-0.

Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on complex traits as variables over time. Anim. Genet. 45:322–328. doi:10.1111/age.12144.

Tan, J., S. Luan, K. Luo, J. Guan, W. Li, J. Sui, Z. Guo, S. Xu, and J. Kong. 2017. Heritability and genotype by environment interactions for growth and survival in Litopenaeus vannamei at low and high densities. Aquac. Res. 48:1430–1438. doi:10.1111/are.12978.

Tetens, J., G. Thaller, and N. Krattenmacher. 2014. Genetic and genomic dissection of dry matter intake at different lactation stages in primiparous Holstein cows. J. Dairy Sci. 97:520–531. doi:10.3168/jds.2013-7301.

Thepparat, M., W. Boonkum, M. Duangjinda, S. Tumwasorn, S. Nakavisut, and T. Thongchumroon. 2015. Genetic evaluation using random regression models with different covariance functions for test-day milk yield in an admixture population of Thailand goats. Anim. Sci. J. 86:655–660. doi:10.1111/asj.12349.

Tier, B., and K. Meyer. 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. J. Anim. Breed. Genet. 121:77–89. doi:10.1111/j.1439-0388.2003.00444.x.

Tong, A. K. W., B. W. Kennedy, and J. E. Moxley. 1979. Heritabilities and genetic correlations for the first three lactations from records subject to culling. J. Dairy Sci. 62:1784–1790. doi:10.3168/jds.S0022-0302(79)83497-9.

Trus, D., and L. Buttazzoni. 1990. A multiple trait approach to modelling the lactation curve. Pages 492–495 in Proc. 4th World Congr. Genet. Appl. Livest. Prod. Edinburgh, Scotland.

Trus, D., and J. W. Wilton. 1988. Genetic parameters for maternal traits in beef cattle. Can. J. Anim. Sci. 68:119–128. doi:10.4141/cjas88-011.

Tsuruta, S., I. Misztal, and I. Strandén. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications.. J. Anim. Sci. 79:1166. doi:10.2527/2001.7951166x.

Turner, A. W., D. Wong, C. N. Dreisbach, and C. L. Miller. 2018. GWAS reveal targets in vessel wall pathways to treat coronary artery disease. Front Cardiovasc Med, 5: 1-10. doi: 10.3389/fcvm.2018.00072.

Turra, E. M., A. F. A. Fernandes, E. R. Alvarenga, E. A. Teixeira, G. F. O. Alves, L. G. Manduca, T. W. Murphy, and M. A. Silva. 2018. Longitudinal analyses of correlated response efficiencies of fillet traits in Nile tilapia. animal 12:445–453. doi:10.1017/S1751731117001768.

Uemoto, Y., T. Osawa, and J. Saburi. 2017. Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. animal 11:382–393. doi:10.1017/S1751731116001762.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24. doi:10.3168/jds.2008-1514.

VanRaden, P. M., G. R. Wiggans, and C. A. Ernst. 1991. Expansion of projected lactation yield to stabilize genetic variance. J. Dairy Sci. 74:4344–4349. doi:10.3168/jds.S0022-0302(91)78630-X.

Vargas, B., E. Perez, and J. A. M. Van Arendonk. 1998. Analysis of test day yield data of Costa Rican dairy cattle. J. Dairy Sci. 81:255–263. doi:10.3168/jds.S0022-0302(98)75574-2.

Verardo, L. L., M. S. Lopes, P. Mathur, O. Madsen, F. F. Silva, M. A. M. Groenen, E. F. Knol, P. S. Lopes, and S. E. F. Guimarães. 2017. Gene networks for total number born in pigs across divergent environments. Mamm. Genome 28:426–435. doi:10.1007/s00335-017-9696-5.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb). 93:357–366. doi:10.1017/S001667231100022X.

Voelker, D. E. 1981. Dairy herd improvement associations. J. Dairy Sci. 64:1269–1277. doi:10.3168/jds.S0022-0302(81)82700-2.

Wade, K. M., and R. L. Quaas. 1993. Solutions to a system of equations involving a first-order autoregressive process. J. Dairy Sci. 76:3026–3032. doi:10.3168/jds.S0022-0302(93)77642-0.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W. M. Muir. 2014a. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front. Genet. 5:134. doi:10.3389/fgene.2014.00134.

Wang, W., Z. Feng1, S. B. Bull, and Z. Wang. 2014b. A 2-step strategy for detecting pleiotropic effects on multiple longitudinal traits. Front. Genet. 5:357. doi: 10.3389/fgene.2014.00357.

Wegman, E. J., and I. W. Wright. 1983. Splines in statistics. J. Am. Stat. Assoc. 78:351. doi:10.2307/2288640.

Weller, J. 1988. Inclusion of partial lactations in the genetic analysis of yield traits by differential weighting of records. J. Dairy Sci. 71:1873–1879. doi:10.3168/jds.S0022-0302(88)79757-X.

Van Der Werf, J. H. J., M. E. Goddard, and K. Meyer. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. J. Dairy Sci. 81:3300–3308. doi:10.3168/jds.S0022-0302(98)75895-3.

White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. 82:632–638. doi:10.3168/jds.S0022-0302(99)75277-X.

Wiggans, G. R., J. B. Cole, S. M. Hubbard, and T. S. Sonstegard. 2017. Genomic selection in dairy cattle: the USDA experience. Annu. Rev. Anim. Biosci. 5:309–327. doi:10.1146/annurev-animal-021815-111422.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, and J. B. Cole. 2011. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. J. Dairy Sci.

94:6188–6193. doi:10.3168/jds.2011-4481.

Wiggans, G. R., and R. L. Powell. 1980. Projection factors for milk and fat lactation records. Dairy Herd Improv. Lett. 56:1–17.

Wiggans, G. R., and L. D. Van Vleck. 1979. Extending partial lactation milk and fat records with a function of last-sample production. J. Dairy Sci. 62:316–325. doi:10.3168/jds.S0022-0302(79)83242-7.

Wilmink, J. B. M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. Livest. Prod. Sci. 16:335–348. doi:10.1016/0301-6226(87)90003-0.

Winkelman, A. M., D. L. Johnson, and B. L. Harris. 2015. Application of genomic evaluation to dairy cattle in New Zealand. J. Dairy Sci. 98:659–675. doi:10.3168/jds.2014-8560.

Wood, P. D. P. 1967. Algebraic model of the lactation curve in cattle. Nature 216:164–165. doi:10.1038/216164a0.

Xie, L., C. Luo, C. Zhang, R. Zhang, J. Tang, Q. Nie, L. Ma, X. Hu, N. Li, Y. Da, and X. Zhang. 2012. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. PLoS One 7:e30910. doi:10.1371/journal.pone.0030910.

Yang, R. 2006. Mapping quantitative trait loci for longitudinal traits in line crosses. Genetics 173:2339–2356. doi:10.1534/genetics.105.054775.

Yang, R., H. Gao, X. Wang, J. Zhang, Z. B. Zeng, and R. Wu. 2007. A semiparametric approach for composite functional mapping of dynamic quantitative traits. Genetics 177:1859–1870. doi:10.1534/genetics.107.077321.

Yang, R., and S. Xu. 2006. Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. Genetics 176:1169–1185. doi:10.1534/genetics.106.064279.

Yi, G., M. Shen, J. Yuan, C. Sun, Z. Duan, L. Qu, T. Dou, M. Ma, J. Lu, J. Guo, S. Chen, L. Qu, K. Wang, and N. Yang. 2015. Genome-wide association study dissects genetic architecture underlying longitudinal egg weights in chickens. BMC Genomics 16:746. doi:10.1186/s12864-015-1945-y.

Yin, T., B. Bapst, U. U. v. Borstel, H. Simianer, and S. König. 2012. Genetic parameters for gaussian and categorical traits in organic and low input dairy cattle herds based on random regression methodology. Livest. Sci. 147:159–169. doi:10.1016/j.livsci.2012.04.017.

Yin, T., and S. König. 2018. Genetic parameters for body weight from birth to calving and associations between weights with test-day, health, and female fertility traits. J. Dairy Sci. 101:2158–2170. doi:10.3168/jds.2017-13835.

Zamani, P., M. R. Moradi, D. Alipour, and F. Ghafouri-Kesbi. 2016. Combination of B-Spline and Legendre functions in random regression models to fit growth curve of Moghani sheep. Small Rumin. Res. 145:115–122. doi:10.1016/j.smallrumres.2016.11.006.

Zavadilová, L., J. Jamrozik, and L. R. Schaeffer. 2005. Genetic parameters for test-day model with random regressions for production traits of Czech Holstein cattle. Czech J. Anim. Sci. 50:142–154.

Zetouni, L., M. Kargo, E. Norberg, and J. Lassen. 2018. Genetic correlations between methane production and fertility, health, and body type traits in Danish Holstein cows. J. Dairy Sci. 101:2273–2280. doi:10.3168/jds.2017-13402.

Zhao, J., Y. Zhao, Z. Song, H. Liu, Y. Liu, and R. Yang. 2018. Genetic analysis of the main growth traits using random regression models in Japanese flounder (Paralichthys olivaceus). Aquac. Res. 49:1504–1511. doi:10.1111/are.13605.

Zhao, W., W. Hou, R.C. Littell, and R. Wu. 2005. Structured antedependence models for functional mapping of multiple longitudinal traits. Stat. Appl. Genet. Mol. Biol. 4. doi:10.2202/1544-6115.1136.

CHAPTER 3:

Bayesian models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle

Hinayah Rojas de Oliveira[1], Fabyano Fonseca e Silva[1], Marcos Vinícius Gualberto Barbosa da Silva[2,3], Otávio Henrique Gomes Barbosa Dias de Siqueira[1], Marco Antônio Machado[2,3], João Cláudio do Carmo Panetto[2], Leonardo Siqueira Glória[4], Luiz Fernando Brito[5]

[1] Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[2] Embrapa Gado de Leite – Empresa Brasileira de Pesquisa Agropecuária, Rua Eugênio do Nascimento, Juiz de Fora, Minas Gerais, Brazil

[3] Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), Lago Sul, Brasília, Distrito Federal, Brazil

[4] Department of Animal Science, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, Rio de Janeiro, Brazil

[5] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

Livestock Science, 201:78-84, 2017.

3.1. Abstract

We aimed to combine Legendre polynomials (**LEG**) and linear B-splines (**BSP**) to describe simultaneously the first and second lactation of Gyr dairy cattle under a multiple-trait random regression models (**MTRRM**) framework. Additionally, we propose the application of self-organizing map to define the classes of residual variances under these models. A total of 26,438 and 23,892 milk yield test-day records were used, respectively, for the first and second lactations of 3,253 Gyr cows. Two preliminary MTRRM analyses considering 10 residual classes were performed: the first one was based on LEG for systematic and random effects for both lactations; and the second one was based on BSP. Three classes were defined by using a self-organizing map: from 6 to 35; 36 to 185 and 186 to 305 days in milk. After definition of residual variance classes, a total of 16 MTRRM combining LEG and BSP were compared. The MTRRM based on BSP to describe the systematic effects of the first and second lactation, BSP to describe the random effects of the first lactation and LEG to describe the random effects of the second lactation (BSP-BSP-BSP-LEG) outperformed all other models. From the BSP-BSP-BSP-LEG model, heritability estimates for milk yield over time ranged from 0.1107 to 0.2902, and from 0.2036 to 0.3967, for the first and second lactation, respectively. In general, additive genetic correlation estimates between days in milk within each lactation and between lactations had medium magnitude (mean of genetic correlations were 0.6630, 0.6226 and 0.4749 for the first, second and between both lactations, respectively). We concluded that combining different functions under a MTRRM framework is a feasible alternative for genetic modeling of lactation curves in Gyr dairy cattle.

Keywords: Bos indicus, combined model, random regression, residual variances, self-organizing map, test-day records

3.2. Introduction

In tropical countries, dairy cattle are usually raised in pasture-based production systems, where the animals are often exposed to climatic variations. To support these conditions, Gyr dairy cattle (Bos indicus) has become popular because its robustness, adaptability and resistance to parasites (Pereira et al., 2013a; Pereira et al., 2013b). Currently, Brazilian Dairy Gyr Breeding Program contains about 330 proven bulls evaluated by using accumulated milk yield until 305 days from the first lactation of his daughters (Panetto et al., 2016). However, to increase the selection accuracy, it is important the development of feasible methodologies allowing to evaluate simultaneously more than one lactation based on test-day records.

For joint genetic evaluation considering test-day milk yield from different lactations, multiple-trait random regression models (**MTRRM**) have been proposed as one of the most powerful statistical tools (Van der Werf et al., 1998; Schaeffer, 2004). These models describe changes in the (co)variances of repeated measures for several traits, thus enabling to estimate genetic correlations between different lactations over time without the need of pre-adjustments (Meyer et al., 1989; Swalve, 2000). The MTRRM based on Legendre orthogonal polynomials and linear splines have been successfully applied to joint analysis of milk yield from different lactations in dairy cattle (Bohmanova et al., 2008; Hammami et al., 2008; Konstantinov, 2015). However, specifically for Gyr cattle breed, studies approaching MTRRM are scarce in literature.

A pioneer study in dairy goats that combined different functions to describe several traits under a Bayesian framework showed to be feasible to improve breeding values and genetic parameter estimates in MTRRM (Oliveira et al., 2016). However, it was not yet tested in dairy cattle. Furthermore, there are no references about using different functions to describe the systematic and random curves of each trait under a combined MTRRM approach, as well as statistical methods to point out for heterogeneity of residual variance under this condition. Thus, we aimed to combine Legendre orthogonal polynomials and linear B-splines to describe simultaneously the first and second lactation of Gyr dairy cattle under a MTRRM framework. Additionally, we propose for the first time in the literature the application of self-organizing map (Kohonen, 1990) to define the classes of residual variances for all lactations underling the same model.

3.3. Material and methods

Phenotypic data and systematic effects

After checking data consistency, a total of 26,438 and 23,892 milk yield test-day records were used, respectively, for the first and second lactations of 3,253 Gyr cows. The relationship matrix included 8,200 animals. Data were obtained from the National Archive of Animal Science, managed by the Brazilian National Center for Dairy Cattle Research (Embrapa Gado de Leite, Juiz de Fora-MG, Brazil).

Test-day records were obtained from the fifth day after calving and were truncated at 305 days of lactation. Only animals containing at least five records were included in the analyses. Additionally, it was also required that all phenotyped animals presented at least one record in the beginning (before 100 days in milk), middle (between 100 and 200 days in milk) and end of lactation (after 200 days in milk). Contemporary groups were defined by the combination of herd, year, month and day of milk yield test. Each group contained at least 5 animals. The age of the dam at calving was included as linear and quadratic covariate for both first and second lactations. Data records considered as outliers (less and higher than mean ± 2.5 standard deviation inside contemporary groups) were also discarded.

Residual variance classes

Two preliminary MTRRM analyses considering 10 residual classes (6 to 35, 36 to 65, 66 to 95, 96 to 125, 126 to 155, 156 to 185, 186 to 215, 216 to 245, 246 to 275, and 276 to 305 days in milk) were performed. The first one was based on fifth-order Legendre polynomial for systematic and random effects for both lactations; and the second one was based on linear B-spline with five knots (at 1, 45, 105, 205 and 305 days) for the same effects and lactations. The residual variance estimated for each class from each MTRRM is shown in Table S1 at supplementary material.

Cluster analysis was performed to reduce the number of residual classes. Three classes were defined by using an unsupervised learning algorithm referred as self-organizing map (Kohonen, 1990). The network architecture was composed by the input and output layers, linked by the weight vectors. Euclidean distances for all weight vectors were computed and used to obtain the best matching unit (BMU), as proposed by Kohonen (2013). At each iteration, the input data (estimated residual variance vectors) was divided into subsets that share the same BMU, and a Gaussian neighborhood function was applied as a smoothing kernel to update the map. Through this iterative method, the corresponding weight vectors to the BMU were updated and adjusted according to their neighboring neurons. After the convergence, the following three classes (groups) were obtained: from 6 to 35; 36 to 185 and 186 to 305 days in milk. This cluster analysis was implemented in the kohonen package (Wehrens and Buydens, 2007) of the R software (R Core Team, 2016).

Multiple-trait random regression models (MTRRM)

A total of 16 MTRRM based on fifth-order Legendre orthogonal polynomials (LEG) and linear B-splines (BSP) with five knots (at 1, 45, 105, 205 and 305 days) were compared. These models were denoted by: 1) LEG-LEG-LEG-LEG; 2) LEG-LEG-LEG-BSP; 3) LEG-LEG-BSP-LEG; 4) LEG-BSP-LEG-LEG; 5) BSP-LEG-LEG-LEG; 6) LEG-LEG-BSP-BSP; 7) LEG-BSP-

LEG-BSP; 8) BSP-LEG-LEG-BSP; 9) LEG-BSP-BSP-LEG; 10) BSP-LEG-BSP-LEG; 11) BSP-BSP-LEG-LEG; 12) LEG-BSP-BSP-BSP; 13) BSP-LEG-BSP-BSP; 14) BSP-BSP-LEG-BSP; 15) BSP-BSP-BSP-LEG and 16) BSP-BSP-BSP-BSP; where the first and third abbreviations refers to the function used to describe the systematic effects of milk yield for the first and second lactation, respectively; and the second and fourth abbreviations refers to the function used to describe the random effects of milk yield for the first and second lactation, respectively. Additive genetic and permanent environmental random effects were modeled by the same function, as suggested by Schaeffer (2016).

The Bayesian MTRRM is described as follow:

$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{Wp} + \mathbf{e}$,

where:

$\mathbf{y} = [\mathbf{y}_1 \quad \mathbf{y}_2]'$ is the vector of phenotypic records with $\mathbf{y}_1$ and $\mathbf{y}_2$ referring to milk yield in the first and second lactation, respectively. This vector of observed values is assumed as $\mathbf{y}|\boldsymbol{\beta},\mathbf{a},\mathbf{p},\mathbf{G}_0,\mathbf{R}_p,\mathbf{R}_0 \sim \mathrm{N}(\mathbf{X\beta}+\mathbf{Za}+\mathbf{Wp},\mathbf{R}_0 \otimes \mathbf{I})$, where $\mathbf{G}_0$, $\mathbf{R}_p$ and $\mathbf{R}_0$ are the genetic, permanent environmental and residual (co)variance matrices, respectively. The incidence matrices for $\boldsymbol{\beta}$, $\mathbf{a}$ and $\mathbf{p}$ are given by $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{W}$, respectively;

$\boldsymbol{\beta} = [\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2]'$ is the vector of systematic effects (contemporary groups and age of the dam at calving), with $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ referring to the first and second lactation, respectively. For this vector is assumed $\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\beta \otimes \mathbf{I})$, being $\boldsymbol{\Sigma}_\beta$ a known diagonal matrix with values 1e+10 (large variances) to represent vague prior knowledge;

$\mathbf{a} = [\mathbf{a}_1 \quad \mathbf{a}_2]'$ is the vector of additive genetic coefficients, with $\mathbf{a}_1$ and $\mathbf{a}_2$ referring to the first and second lactation, respectively. This vector is assumed as $\mathbf{a} \mid \mathbf{G}_0,\mathbf{A} \sim \mathrm{N}(0,\mathbf{G}_0 \otimes \mathbf{A})$, being $\mathbf{A}$ the additive relationship matrix and $\mathbf{G}_0$ the additive genetic covariance matrix. Furthermore, it was assumed that $\mathbf{G}_0$ follows an inverted Wishart distribution, IW (v_a, $\mathbf{V}_a$), with hyperparameters v_a=3 and $\mathbf{V}_a = \hat{\mathbf{G}}_0$;

$\mathbf{p} = [\mathbf{p}_1 \quad \mathbf{p}_2]'$ is the vector of permanent environmental coefficients, with $\mathbf{p}_1$ and $\mathbf{p}_2$ referring to the first and second lactation, respectively. This vector is assumed as $\mathbf{p} \mid \mathbf{R}_p \sim \mathrm{N}(0,\mathbf{R}_p \otimes \mathbf{I})$, being $\mathbf{I}$ the identity matrix and $\mathbf{R}_p$ the permanent environmental covariance matrix. Similarly as assumed for $\mathbf{G}_0$, $\mathbf{R}_p \sim \mathrm{IW}\left(v_p, \mathbf{V}_p\right)$, with hyperparameters v_p=3 and $\mathbf{V}_p = \hat{\mathbf{R}}_p$;

$\mathbf{e} = \left[\mathbf{e}_1 \quad \mathbf{e}_2\right]'$ is the residual vector, with $\mathbf{e}_1$ and $\mathbf{e}_2$ referring to the first and second lactation, respectively. It was assumed that the residual covariance matrix is distributed as $\mathbf{R}_0 \sim \mathrm{IW}\left(v_e, \mathbf{V}_e\right)$, with hyperparameters analogously to $\mathbf{G}_0$ and $\mathbf{R}_p$.

The (co)variance components were estimated by using the Gibbs Sampler algorithm, available in GIBBS3F90 software (Misztal et al., 2002). A MCMC chain length of 500,000 cycles, considering a burn-in period of 250,000 cycles, and a sampling interval (thin) of 10 cycles was used in all analyses. The convergence was verified through graphical analysis and Geweke criterion (Geweke, 1992) available in the package Bayesian Output Analysis (Smith, 2007) of the R software (R Core Team, 2016).

Bayesian model comparison

Bayes factor (BF) was used to assess the goodness-of-fit of all tested models. The BF representing the evidence in favor of the model 1 (M_1) in relation to model 2 (M_2) is given by:

$$\mathrm{BF} = \frac{P(D \mid M_1)}{P(D \mid M_2)},$$

where $P(D \mid M_1)$ and $P(D \mid M_2)$ are the marginal likelihoods of models 1 and 2, respectively. The BF higher than 100 is considered decisive (Kass and Raftery, 1995), and the interpretation is that the observed data are 100 times more likely to have been generated under model 1 than model 2.

Additionally, the best models also were contrasted by the percentage of individuals selected in common considering 1%, 2%, 5%, 10% and 50% of the individuals with higher breeding values for accumulated milk yield $({}_{ac}\mathbf{E\hat{B}V}_{ij})$, obtained as the sum of all $\mathbf{E\hat{B}V}_{ij}$. The vector of estimated breeding values ($\mathbf{E\hat{B}V}_{ij}$) for each lactation of each animal i at test-day j was obtained from the a posterior distribution of the best MTRRM as follow:

$$\mathbf{E\hat{B}V}_{ij} = \mathbf{T}_i \; \hat{\mathbf{a}}_i,$$

where $\hat{\mathbf{a}}_i$ is the vector of estimated breeding values for the coefficients for the animal i; and $\mathbf{T}_i$ is a matrix associated with the assumed function (Legendre and B-spline polynomials).

Genetic parameter estimates

Once identified the best MTRRM, the posterior marginal distribution samples for heritabilities throughout the lactation curves were obtained by:

$$h^{2\ (k)}_{j(L)} = \frac{\hat{\sigma}^{2\ (k)}_{a_{j(L)}}}{\hat{\sigma}^{2\ (k)}_{a_{j(L)}} + \hat{\sigma}^{2\ (k)}_{p_{j(L)}} + \hat{\sigma}^{2\ (k)}_{e_{(L)}}},$$

where $h^{2\ (k)}_{j(L)}$ is the heritability estimate for the lactation L (first and second) at test-day j at MCMC iteration k, $\hat{\sigma}^{2\ (k)}_{a_{j(L)}}$ is the additive genetic variance estimate, $\hat{\sigma}^{2\ (k)}_{p_{j(L)}}$ is the permanent environmental variance estimate, and $\hat{\sigma}^{2\ (k)}_{e_{(L)}}$ is the residual variance estimate, which depends on the classes assumed earlier by the self-organizing map.

The posterior marginal distribution samples for correlation between days in milk within each lactation and between lactations throughout days in milk, estimated from the best MTRRM, were obtained respectively by:

$$r^{(k)}_{L(j,j')} = \frac{\hat{\sigma}^{(k)}_{a_{L(j,j')}}}{\sqrt{\hat{\sigma}^{2\ (k)}_{a_{L(j)}}\ \hat{\sigma}^{2\ (k)}_{a_{L(j')}}}} \quad \text{and} \quad r^{(k)}_{j(L,L')} = \frac{\hat{\sigma}^{(k)}_{a_{j(L,L')}}}{\sqrt{\hat{\sigma}^{2\ (k)}_{a_{j(L)}}\ \hat{\sigma}^{2\ (k)}_{a_{j(L')}}}},$$

where $r^{(k)}_{L(j,j')}$ and $r^{(k)}_{j(L,L')}$ are the genetic correlation estimate between days in milk j and j' within each lactation L at MCMC iteration k; and between lactations L and L' at test-day j at MCMC iteration k, respectively. The $\hat{\sigma}^{(k)}_{a_{L(j,j')}}$ and $\hat{\sigma}^{(k)}_{a_{j(L,L')}}$ are the genetic covariance estimates, and $\hat{\sigma}^{2\ (k)}_{a_{L(j)}}$, $\hat{\sigma}^{2\ (k)}_{a_{j(L)}}$, $\hat{\sigma}^{2\ (k)}_{a_{L(j')}}$ and $\hat{\sigma}^{2\ (k)}_{a_{j(L')}}$ are the additive genetic variance estimates.

3.4. Results

Phenotypic data

The overall average phenotypic test-day milk yields with respective standard deviations were 11.81 kg (5.70) and 13.85 kg (7.16), for the first and second lactation, respectively. Higher lactation peak and higher decrease in milk production were observed for the second lactation compared to the first one. In general, the number of records remained approximately constant until 285 days in milk, and decreased after this time period (Figure 1).

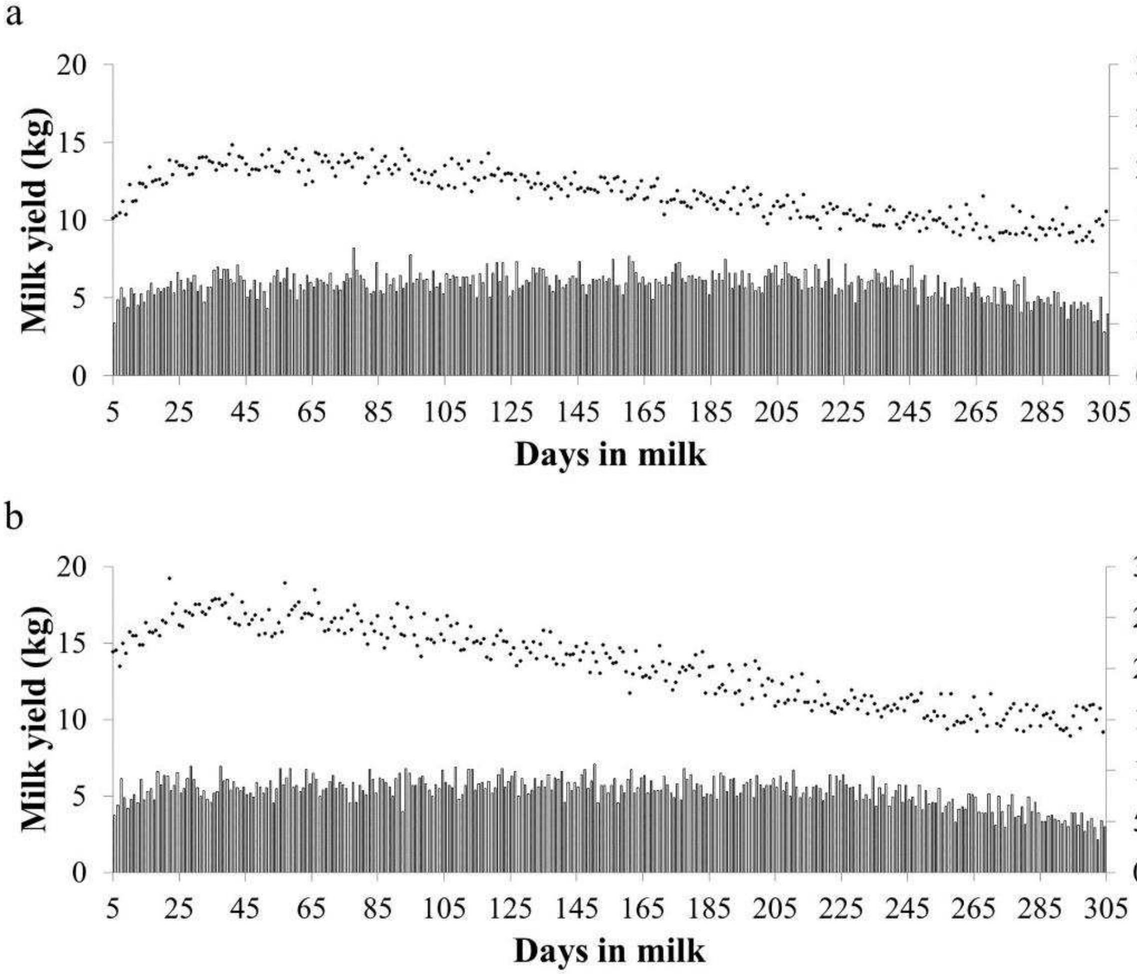

Figure 1. Number of records (bars) and average milk yield (points) according to days in milk, for the first (a) and second (b) lactation.

Bayesian model comparison

The combined MTRRM based on linear B-splines to describe the systematic effects of the first and second lactation, B-splines to describe the random effects of the first lactation and Legendre polynomials to describe the random effects of the second lactation (BSP-BSP-BSP-LEG, model 15) outperformed all other models (Table 1).

The second best model was the combined MTRRM based on linear B-splines to describe the systematic and random effects of the first lactation, and Legendre polynomials to describe the systematic and random effects of the second lactation (BSP-BSP-LEG-LEG, model 11). Although model 11 has been identified as the second one, the BF higher than 100 shows the clear superiority of the model 15 when compared to model 11, which makes the ranks provided by the models were different, mainly when the intensity of selection is high (Table 2).

Table 1. Bayes Factor (BF) comparing the best model (BSP-BSP-BSP-LEG, model 15)[1] with the other multiple-trait random regression models.

Model	**FIRST LACTATION**		**SECOND LACTATION**		**BF**
	Systematic	**Random[2]**	**Systematic**	**Random[2]**	
1	LEG	LEG	LEG	LEG	75,542
2	LEG	LEG	LEG	BSP	75,000
3	LEG	LEG	BSP	LEG	75,572
4	LEG	BSP	LEG	LEG	75,222
5	BSP	LEG	LEG	LEG	75,323
6	LEG	LEG	BSP	BSP	74,554
7	LEG	BSP	LEG	BSP	76,603
8	BSP	LEG	LEG	BSP	74,512
9	LEG	BSP	BSP	LEG	75,035
10	BSP	LEG	BSP	LEG	75,415
11	BSP	BSP	LEG	LEG	106.1
12	LEG	BSP	BSP	BSP	76,289
13	BSP	LEG	BSP	BSP	75,322
14	BSP	BSP	LEG	BSP	76,201
15	BSP	BSP	BSP	LEG	-
16	BSP	BSP	BSP	BSP	76,600

[1]Model 15 (BSP-BSP-BSP-LEG): based on B-splines (BSP) to describe the systematic effects of the first and second lactation, BSP to describe the random effects of the first lactation and Legendre polynomials (LEG) to describe the random effects of the second lactation; [2]Random: additive genetic and permanent environmental effects.

Table 2. Percentage of individuals selected in common, in increasing levels, when comparing the ranking provided by the model 15 (BSP-BSP-BSP-LEG) with the ranking provided by the model 11 (BSP-BSP-LEG-LEG).

Selection by:	**Percentage of individuals selected in common**				
	1%	**2%**	**5%**	**10%**	**50%**
First lactation	82.93%	89.02%	88.78%	92.68%	95.78%
Second lactation	87.80%	89.63%	92.93%	94.76%	96.32%
Both lactations	87.80%	89.02%	91.95%	93.90%	96.39%

Model 15 (BSP-BSP-BSP-LEG): based on B-splines to describe the systematic effects of the first and second lactation, B-splines to describe the random effects of the first lactation and Legendre polynomials to describe the random effects of the second lactation; Model 11 (BSP-BSP-LEG-LEG): based on B-splines to describe the systematic and random effects of the first lactation and Legendre polynomials to describe the systematic and random effects of the second lactation.

This result justifies the use of different functions to model the systematic and random effects in the combined MTRRM as a feasible alternative for genetic modeling of lactation curves. All other tested models presented similar and worse results when compared to model 15.

Genetic parameter estimates

Posterior means and 95% highest posterior density (HPD) intervals for the variance components estimated from the best MTRRM are shown in Figure 2.

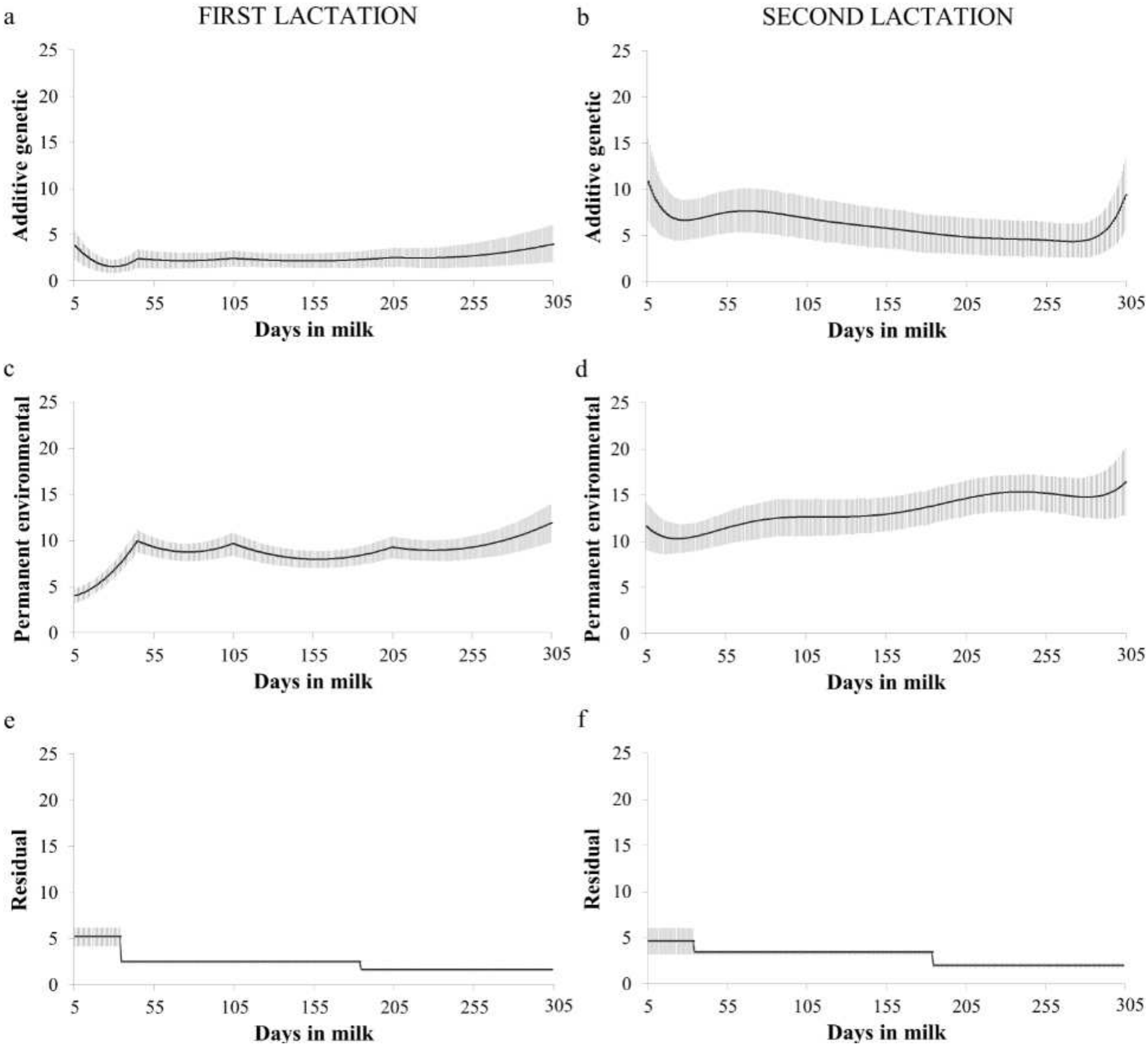

Figure 2. Posterior means and 95% highest posterior density intervals for additive genetic (a, b), permanent environmental (c, d), and residual (e, f) variances for the first (left) and second (right) lactation, estimated from the best model (model 15, based on B-splines to describe the systematic effects of the first and second lactation, B-splines to describe the random effects of the first lactation and Legendre polynomials to describe the random effects of the second lactation).

The posterior means of the additive genetic variances $(\hat{\sigma}_a^2)$ for the first lactation were lower than those estimated for the second one (Figure 2-a and 2-b). The pattern of the posterior means of permanent environmental variances $(\hat{\sigma}^2_{pe})$ were similar between lactations, since both tended to increase over time (Figure 2-c and 2-d). In general, $\hat{\sigma}^2_a$ and $\hat{\sigma}^2_{pe}$ of the second lactation shown higher HPD intervals compared to the first one. The posterior means and HPD intervals for the residual variance estimates ($\hat{\sigma}^2_e$) were approximately the same in both lactations (Figure 2-e and 2-f).

Figure 3 shows the heritability estimates (posterior means) for the first and second lactation throughout 305 days in milk estimated from the best MTRRM.

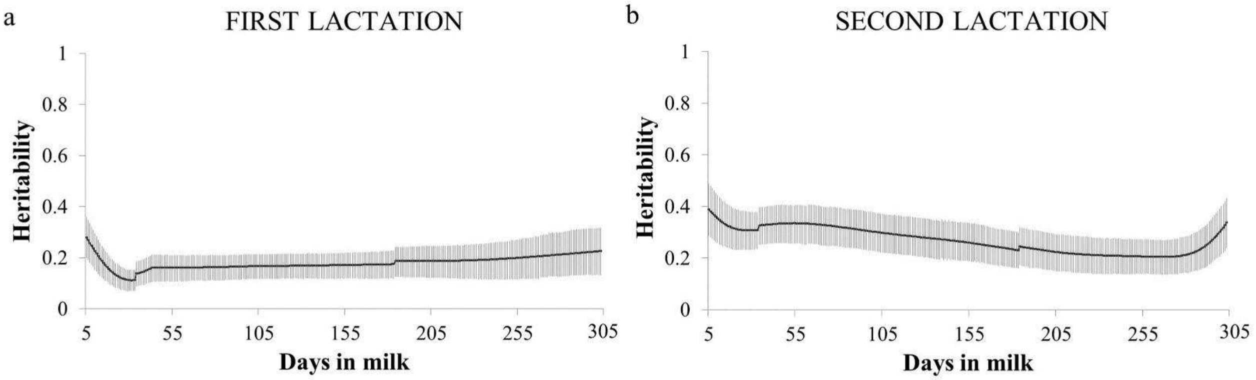

Figure 3. Posterior means and 95% highest posterior density intervals for heritabilities of milk yield in the first (a) and second (b) lactation, estimated from the best model (model 15, based on B-splines to describe the systematic effects of the first and second lactation, B-splines to describe the random effects of the first lactation and Legendre polynomials to describe the random effects of the second lactation).

The posterior means of heritability estimates for milk yield ranged from 0.1107 to 0.2902, and from 0.2036 to 0.3967, for the first and second lactation, respectively (Figure 3). In general, posterior mean estimates for additive genetic correlations between days in milk within each lactation had medium magnitude (mean of genetic correlations through days in milk were 0.6630 and 0.6226 for the first and second lactation, respectively), and they were higher between adjacent test-day records (Figure 4-a and 4-b). Negative correlations between the extreme points of the lactation curve were found out in both lactations (minimum genetic correlation found was -0.0533 and -0.1883 for the first and second lactation, respectively).

The genetic correlation between lactations showed mean of 0.4749 through days in milk, and had higher magnitude between the middle of both lactations. Negative values were found mainly between at the beginning of the first lactation and at the end of the second one (Figure 4-c).

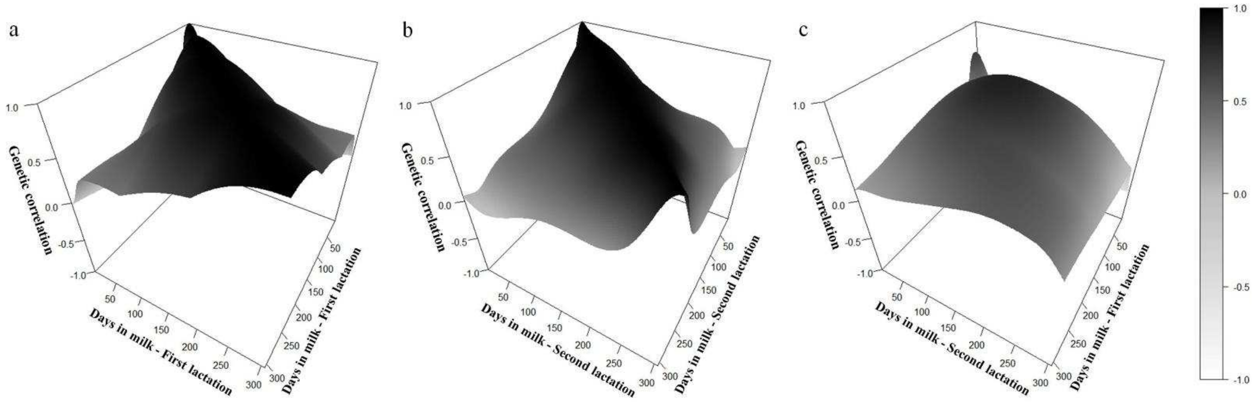

Figure 4. Posterior means for additive genetic correlations between days in milk inside the first (a) and second (b) lactations, and between different lactations (c), estimated from the best model (model 15, based on B-splines to describe the systematic effects of the first and second lactation, B-splines to describe the random effects of the first lactation and Legendre polynomials to describe the random effects of the second lactation).

3.5. Discussion

Phenotypic data

The overall average test-day milk yield (standard-deviation) was equal to 11.81 kg (5.70 kg) and 13.85 kg (7.16 kg) for the first and second lactation, respectively. Similar milk yield (11.05 kg) for the first lactation of Gyr dairy cattle was found by Gonzalez-Herrera et al. (2015). However, different milk yield for the first lactation were reported by Santana et al. (2015) and Pereira et al. (2013), which were equal to 13.28 kg and 9.81 kg, respectively. For the second lactation, Gonzalez-Herrera et al. (2015) reported lower milk yield (11.20 kg) compared to those obtained in this study.

Test-day milk yields shown different patterns over lactations, since the lactation peak and the decrease in milk production were higher at the second lactation (Figure 1). Due to the differences between the traits in the production levels and patterns throughout the lactation curve, it is important to consider specific systematic regression for each trait.

The number of records per day for both lactations tended to decrease after 285 days in milk (Figure 1) due to the occurrence of short lactations in Zebu dairy breeds when compared to European breeds (Pereira et al., 2013a).

Bayesian model comparison

The model 15 (BSP-BSP-BSP-LEG) was the best one according to BF (Table 1), followed by model 11 (BSP-BSP-LEG-LEG). Despite of the combined MTRRM based on different functions to fit systematic and random effects increase the complexity of the models in relation to the single MTRRM; it improved the fit quality. Nonetheless, the choice of the functions to compose the combined MTRRM need to be suitable to modeling the traits in study, otherwise, if the functions are wrongly chosen the combined MTRRM may present poor fit to the data; and consequently, provide worse perform than the simple MTRRM (based on the same functions for all traits and effects, models 1 and 16 in this study).

Comparing only the simple MTRRM, the model based on B-spline functions (BSP-BSP-BSP-BSP, model 16) presented worst fit than the simple MTRRM based on Legendre polynomials (LEG-LEG-LEG-LEG, model 1) maybe it is because B-spline functions tend to consider more coefficients than corresponding analyses with Legendre polynomial basis function (Meyer, 2005). Similar results in terms of single trait models based on these mentioned functions were found for the first lactation of Holstein cows (Bignardi et al., 2011) and for the first lactation of Gyr dairy cattle (Pereira et al., 2013a).

The estimated breeding values from the two best MTRRM were different, which may change the rank of the animals under selection. According to the ranks provided by the best models (Table 2), when increase the intensity of selection (especially under 10%), the animals selected in common by the model 15 and model 11 became more dissimilar, especially when selected only for the first lactation. This fact indicates that how higher the intensity of selection applied to the herd greater is the need of using an appropriate model, mainly when the animals are selected just based on the first lactation.

Genetic parameter estimates

The additive genetic variance estimate ($\hat{\sigma}_a^2$) obtained from the best model remained approximately constant over time for the first lactation, except in the lactation peak (around 30 days in milk), when it decreased; and after 260 days in milk, when it tended to increase (Figure 2-a).

For the second lactation, the decrease of the $\hat{\sigma}_a^2$ in the lactation peak and the increase after 260 days in milk (Figure 2-b) it was also observed, suggesting that the pattern of $\hat{\sigma}_a^2$ reduction of the in the lactation peak may be peculiar in Gyr dairy cattle. Moreover, the $\hat{\sigma}_a^2$ and $\hat{\sigma}_{pe}^2$ of the second lactation (Figure 2-a and 2-d) were inflated at the beginning and at the ending of the lactation, maybe due to Legendre polynomials (Misztal, 2006). Bignardi et al. (2011) also reported these inflated genetic variances estimated by Legendre polynomials for milk yield in dairy cattle. In general, similar to the results reported previously by Gonzalez-Herrera et al. (2015) in dairy Gyr, the $\hat{\sigma}_a^2$ were higher for the second lactation.

The patterns of the $\hat{\sigma}_{pe}^2$ were similar between lactations (Figure 2-c and 2-d), since both tended to increase over lactation. These results differ from those reported by Gonzalez-Herrera et al. (2015), who found that the $\hat{\sigma}_{pe}^2$ increased in first lactation and decreased after the seventh test day in the second lactation.

The $\hat{\sigma}_e^2$ estimated for milk yield in the first and second lactation (Figure 2-e and 2-f) were higher for the first class (5 to 35 days in milk) in relation to the second and third one, suggesting that the beginning of lactation is very susceptible to environmental factors in the first and second lactation of Gyr cattle. This pattern is similar to those reported by Pereira et al. (2010) and Pereira et al. (2013) for Gyr dairy cattle.

The heritability estimates showed different trends in the two lactations, especially after 35 days in milk, since in the first lactation they increased and in the second one they decreased after 35 days in milk (Figure 3). These estimates were higher for the second lactation when compared to the first one, suggesting that the inclusion of the second lactation in Gyr dairy cattle breeding programs can provide satisfactory genetic gains. In general, the heritability estimates for the first lactation were similar to those observed by Pereira et al. (2013) and Gonzalez-Herrera et al. (2015). For the second lactation, the estimates reported here were higher than those found by Gonzalez-Herrera et al. (2015).

As expected, the additive genetic correlation estimates between days in milk within each lactation were higher between adjacent test-day records (Figure 4-a and 4-b). Negative correlation estimates between the extreme points of the lactation curve were found in both lactations. Some authors try explain these negative values by the reduction of the number of observations in the end of Gyr cattle lactation, since this reduction can affect the genetic parameter estimates (Costa et al., 2005; Pereira et al., 2013a).

To date, reports about genetic correlation estimates of milk yield between the first and second lactation of dairy Gyr cattle are limited in literature. In the present study, the mean of genetic correlation through days in milk was positive and had medium magnitude (mean of 0.4749), indicating that the selection in one lactation can bring satisfactory gains for another one. However, it is necessary caution to select animals based just in the first lactation, since the genetic correlation between the begin of the first lactation and the end of the second one had negative values (Figure 4-c).

3.6. Conclusions

Combining different functions to model the systematic and random effects under a multiple-trait random regression model (MTRRM) approach is a feasible alternative for genetic modeling of lactation curves in Gyr dairy cattle. Based on the Bayes Factor and in the percentage of individuals selected in common, the combined MTRRM based on B-splines to describe the systematic effects of the first and second lactation, B-splines to describe the random effects of the first lactation and Legendre polynomials to describe the random effects of the second lactation was the best model to describe both lactations simultaneously. According to heritability and correlation estimates reported here, the inclusion of the second lactation in Brazilian dairy Gyr breeding programs can improve the genetic gain for milk yield in this breed.

3.7. Acknowledgments

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES). Marcos V. G. B. Silva was supported by the Embrapa (Brazil) SEG 02.09.07.008.00.00 “Genomic Selection in Dairy Cattle in Brazil,” CNPq PVE 407246/2013-4 “Genomic Selection in Dairy Gyr and Girolando Breeds,” and FAPEMIG CVZ PPM 00606/16 “Identification of Selection Signatures in Cattle Breeds from Next-generation Sequencing Data” appropriated projects.

3.8. References

Bignardi, A., El Faro, L., Torres Júnior, R., Cardoso, V., Machado, P., Albuquerque, L.G.d., 2011. Random regression models using different functions to model test-day milk yield of Brazilian Holstein cows. Genet. Mol. Res., 3565-3575.

Bohmanova, J., Miglior, F., Jamrozik, J., Misztal, I., Sullivan, P., 2008. Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. J. Dairy Sci. 91, 3627-3638.

Costa, C.N., Melo, C.d., Machado, C.H.C., Freitas, A.d., Packer, I.U., Cobuci, J., 2005. Estimation of genetic parameters for test day milk records of first lactation Gyr cows using repeatability and random regression animal models. R. Bras. Zootec 34, 1519-1530.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J., Dawid, A.P., Smith, A.F.M., ed. Oxford University Press, Oxford, UK, Pages 169–193 in Bayesian Statistics 4.

Gonzalez-Herrera, L., El Faro, L., Bignardi, A., Pereira, R., Machado, C., Albuquerque, L., 2015. Random regression analysis of test-day milk yields in the first and second lactations of Brazilian Gyr cows. Genet. Mol. Res. 14, 16497-16507.

Hammami, H., Rekik, B., Soyeurt, H., Gara, A.B., Gengler, N., 2008. Genetic parameters for Tunisian Holsteins using a test-day random regression model. J. Dairy Sci. 91, 2118-2126.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773-795.

Kohonen, T., 1990. The self-organizing map. In: Proceedings of the IEEE 78, 1464-1480.

Kohonen, T., 2013. Essentials of the self-organizing map. Neural Netw. 37, 52-65.

Konstantinov, K., 2015. Implementation of Multiple traits Multi Lactation Random Regression Test Day Model for Production Traits in Australia. Interbull Bulletin 49, 117-126.

Meyer, K., 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. Genet. Sel. Evol. 37, 473-500.

Meyer, K., Graser, H.-U., Hammond, K., 1989. Estimates of genetic parameters for first lactation test day production of Australian Black and White cows. Livest. Prod. Sci. 21, 177-199.

Misztal, I., 2006. Properties of random regression models using linear splines. J. Anim. Breed. Genet. 123, 74-80.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T. & Lee, D. 2002. BLUPF90 and related programs (BGF90). In: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France. Communication no. 28-07.

Oliveira, H.R., Silva, F.F, Siqueira, O., Souza, N., Junqueira, V., Resende, M., Borquis, R., Rodrigues, M., 2016. Combining different functions to describe milk, fat, and protein yield in goats using Bayesian multiple-trait random regression models. J. Anim. Sci. 94, 1865-1874.

Panetto, J.C.d.C., Verneque, R.d.S., Silva, M.V.G.B., Machado, M.A., Martins, M.F., Bruneli, F.Â.T., Peixoto, M.G.C.D., Santos, G.G., Arbex, W.A., Reis, D.R.d.L., Geraldo, C.C., Machado, C.H.C., Ventura, H.T., Pereira, M.A., Hortolani, B., Vercesi Filho, A.E., Maciel, R.d.S., Fernandes, A.R., 2016. Programa Nacional de Melhoramento do Gir Leiteiro. Sumário Brasileiro de Touros – Resultado do Teste de Progênie. 7a Prova de Pré-Seleção de Touros. Embrapa Gado de Leite. Documentos, 187, Juiz de Fora, MG, p. 86.

Pereira, R., Bignardi, A., El Faro, L., Verneque, R., Vercesi Filho, A., Albuquerque, L., 2013a. Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (Bos indicus) cattle. J. Dairy Sci. 96, 565-574.

Pereira, R.J., Ayres, D.R., El Faro, L., Vercesi Filho, A.E., da Silva Verneque, R., de Albuquerque, L.G., 2013b. Genetic parameters for production traits of dairy Gyr (Bos indicus)× Holstein cattle estimated with a random regression model. Livest. Sci. 158, 24-31.

Pereira, R.J., Lopes, P.S., Verneque, R.d.S., Santana Júnior, M.L., Lagrotta, M.R., Torres, R.d.A., 2010. Covariance functions for test-day milk yield in Gir cattle. Pesqui. Agropecu. Bras., 1303-1311.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Santana, M., Pereira, R., Bignardi, A., Vercesi Filho, A., Menendez-Buxadera, A., El Faro, L., 2015. Detrimental effect of selection for milk yield on genetic tolerance to heat stress in purebred Zebu cattle: Genetic parameters and trends. J. Dairy Sci. 98, 9035-9043.

Schaeffer, L., 2004. Application of random regression models in animal breeding. Livestock Production Science 86, 35-45.

Schaeffer, L.R., 2016. Random Regression Models.

Smith, B.J., 2007. boa: an R package for MCMC output convergence assessment and posterior inference. J. Stat. Softw. 21, 1-37.

Swalve, H., 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. Journal of Dairy Science 83, 1115-1124.

Van der Werf, J., Goddard, M., Meyer, K., 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. J. Dairy Sci. 81, 3300-3308.

Wehrens, R., Buydens, L.M., 2007. Self-and super-organizing maps in R: the Kohonen package. J. Stat. Softw. 21, 1-19.

3.9. Supplementary material

Table S1. Residual variance estimated for each class by using two multiple-trait random regression models: one based on fifth-order Legendre polynomial and other based on linear B-splines with five knots (at 1, 45, 105, 205 and 305 days).

[1]Class	Model 1		Model 2	
	[2]LEG_1	[2]LEG_2	[3]BSP_1	[3]BSP_2
1	3.5701	5.5451	4.0549	8.4083
2	2.8793	4.2135	1.8674	2.8644
3	2.2803	3.6741	2.5651	3.9881
4	2.4532	3.7735	1.9536	3.1888
5	1.9701	3.2889	2.1192	3.6473
6	2.3682	3.1041	2.5681	3.5519
7	1.9853	2.9731	1.6529	2.3429
8	1.7586	2.6917	1.9742	2.8643
9	1.8235	2.7501	2.0638	3.1059
10	1.9921	1.7791	2.5077	2.2585

[1]Residual variance classes: 1) 6 to 35; 2) 36 to 65; 3) 66 to 95; 4) 96 to 125; 5) 126 to 155; 6) 156 to 185; 7) 186 to 215; 8) 216 to 245; 9) 246 to 275 and 10) 276 to 305 days in milk; [2]LEG_1 and LEG_2 denotes respectively, the residual variance estimated for the first and second lactation by using the model based on Legendre polynomials; [3]BSP_1 and BSP_2 denotes respectively, the residual variance estimated for the first and second lactation by using the model based on linear B-splines.

CHAPTER 4:

Comparing deregression methods for genomic prediction of test-day traits in dairy cattle

Hinayah Rojas de Oliveira[1,2], Fabyano Fonseca e Silva[2], Luiz Fernando Brito[1], Aline Rocha Guarini[1], Janusz Jamrozik[1,3], Flávio Schramm Schenkel[1]

[1] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

[2] Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[3] Canadian Dairy Network, Guelph, Ontario, Canada

Journal of Animal Breeding and Genetics,
135(2):97-106, 2018.

4.1. Abstract

We aimed to investigate the performance of three de-regression methods (VanRaden, **VR**; Wiggans, **WG**; and Garrick, **GR**) of cows' and bulls' breeding values to be used as pseudo-phenotypes in the genomic evaluation of test-day dairy production traits. Three scenarios were considered within each de-regression method: 1) including only animals with reliability of estimated breeding value (**REL_{EBV}**) higher than the average of parent reliability (**REL_{PA}**) in the training and validation populations; 2) including only animals with REL_{EBV} higher than 0.50 in the training and REL_{EBV} higher than REL_{PA} in the validation population; and, 3) including only animals with REL_{EBV} higher than 0.50 in both training and validation populations. Individual random regression coefficients of lactation curves were predicted using the Genomic Best Linear Unbiased Prediction (**GBLUP**), considering either unweighted or weighted residual variances based on effective records contributions. In summary, VR and WG de-regression methods seemed more appropriate for genomic prediction of test-day traits without need for weighting in the genomic analysis, unless large differences in REL_{EBV} between training population animals exist.

Keywords: estimated breeding value, genomic value, Jersey, random regression, reliability

4.2. Introduction

Pseudo-phenotypes are commonly used as response variables in two-step genomic prediction (Meuwissen et al., 2001; VanRaden, 2008). In dairy cattle, when the raw phenotypic data is available, the yield deviations (**YD**, VanRaden & Wiggans, 1991) and the daughter yield deviations (**DYD**, VanRaden & Wiggans, 1991) are used to summarize the phenotypic performance for genomic analysis of cows and bulls, respectively. However, when YD and DYD are not available, estimated breeding values (**EBVs**) derived from genetic evaluations have been used (Calus et al., 2016).

In order to avoid double-counting of information and double shrinkage of the direct genomic values (**DGVs**) by using EBVs as pseudo-phenotypes, computation of deregressed EBVs (**dEBVs**) has been suggested as a feasible alternative to perform genomic predictions (Garrick et al., 2009). In general, the use of dEBV as response variable can contribute to reduce the bias and increase the reliability of DGVs in genomic prediction of breeding values when compared to the traditional EBVs (Guo et al., 2010, Ostersen et al., 2011). Methods of deregression reported in the literature vary from the simplest form of deregression, just dividing each EBV by its reliability (Goddard, 1985), to more sophisticated deregression procedures. The latter involves setting up the complete mixed model equations (**MME**) for all animals in the pedigree (Jairath et al., 1998). Thus, to simplify and avoid to reconstruct the complete MME, different strategies have been proposed (Garrick et al., 2009; VanRaden et al., 2009). In this way, the challenge is to find the most efficient deregression method that provides the best advantages of using dEBV as pseudo-phenotype in the second step of genomic prediction.

In a recent study based on simulated data, Calus et al. (2016) compared the performance of different deregression methods for non-longitudinal traits. These authors did not analyze the direct impact of deregression methods on the GEBVs, but showed that the majority of the methods yielded biased dEBVs when compared to EBVs, especially for cows. As an attempt to reduce the influence of bias of cow genetic evaluations and improve their contribution for the genomic selection accuracy, an adjustment to decrease the variance and mean of cow's EBVs and dEBVs was proposed for accumulated milk traits (Wiggans et al., 2011).

Although some countries are still using accumulated yields for genetic selection in dairy cattle, there are several advantages of using yields recorded on test-day for genetic evaluations (Schaeffer, 2004). Some of these advantages are more precise adjustment for environmental effects, accounting for the shape of lactation curves, and the possibility to evaluate animals with incomplete lactation through random regression models (**RRM**). Especially regarding genomic predictions, the application of RRM in both steps of two-step genomic selection could provide suitable opportunities to select genotyped animals based on the complete pattern of the production curve, without changing the genetic evaluation systems used in several countries. However, besides all progress achieved in the field of genomics in the last decade, to our knowledge, there are no reports evaluating deregression of EBVs for genomic predictions of test-day traits, especially in breeding programs with small training populations including cows' records. In this context, the objective of this study was to investigate the performance of the most used deregression methods for non-longitudinal traits in the deregression of cows' and bulls' EBVs for genomic evaluation of test-day traits in the Canadian Jersey breed.

4.3. Material and methods

The Canadian Dairy Network (**CDN**; Guelph, ON, Canada) provided the pedigree, genotypes and two official genetic evaluations for milk (**MY**), fat (**FY**) and protein (**PY**) yield and somatic cell score (**SCS**) from the first lactation of Jersey cows. The two official genetic evaluations were from December 2012 and December 2016, and contained the estimates of lactation curve random regression genetic coefficients and reliabilities for 134,631 and 160,241 animals, respectively. Details about the model used by CDN are presented in Interbull (2016a, 2016b).

In order to preserve the shape of the genetic lactation curves, EBVs for each day in milk (from five to 305 days) were derived using the genetic coefficients for each trait and, then, three different deregression methods were contrasted: 1) **VR** (VanRaden et al., 2009); 2) **WG** (Wiggans et al., 2011); and 3) **GR** (Garrick et al., 2009). The implementation of these single-trait deregression methods is described in the next section.

Deregression methods

VR method. It can be described as follows:

$$\mathrm{dEBV}_{\mathrm{ij}} = \mathrm{PA}_{\mathrm{ij}} + \frac{(\mathrm{EBV}_{\mathrm{ij}} - \mathrm{PA}_{\mathrm{ij}})}{\mathrm{R}_{\mathrm{i}}}, \qquad [1]$$

where $\mathrm{PA}_{\mathrm{ij}}$ and $\mathrm{EBV}_{\mathrm{ij}}$ are the parental average (**PA**) and the EBV for the animal i on day j, respectively, and R_{i} is the deregression factor. Since most of the animal breeding programs that use RRM just keep the average of reliability through the lactation for each animal, R_{i} was computed as follow:

$$\mathrm{R}_{\mathrm{i}} = \frac{\mathrm{ERC}_{(\mathrm{Ind-PA})_{\mathrm{i}}}}{\mathrm{ERC}_{(\mathrm{Ind-PA})_{\mathrm{i}}} + \mathrm{ERC}_{\mathrm{PA}_{\mathrm{i}}} + 1}, \qquad [2]$$

where effective record contributions (**ERC**) for the animal i, excluding PA information $(\mathrm{ERC}_{(\mathrm{Ind-PA})_{\mathrm{i}}})$, and for the PA $(\mathrm{ERC}_{\mathrm{PA}_{\mathrm{i}}})$ of the animal i were estimated based on (VanRaden & Wiggans, 1991):

$$\mathrm{ERC}_{(\mathrm{Ind-PA})_{\mathrm{i}}} = \left(\lambda \frac{\mathrm{REL}_{\mathrm{EBV}_{\mathrm{i}}}}{1-\mathrm{REL}_{\mathrm{EBV}_{\mathrm{i}}}}\right) - \mathrm{ERC}_{\mathrm{PA}_{\mathrm{i}}} \text{ and } \mathrm{ERC}_{\mathrm{PA}_{\mathrm{i}}} = \lambda \frac{\mathrm{REL}_{\mathrm{PA}_{\mathrm{i}}}}{1-\mathrm{REL}_{\mathrm{PA}_{\mathrm{i}}}}, \qquad [3]$$

in which $\lambda = (1-\mathrm{h}^2)/\mathrm{h}^2$, and $\mathrm{REL}_{\mathrm{EBV}_{\mathrm{i}}}$ and $\mathrm{REL}_{\mathrm{PA}_{\mathrm{i}}}$ are the average of reliabilities for the EBV and PA of the animal i for each analyzed trait, respectively. Heritabilities (**$\mathbf{h}^2$**) considered in all deregression methods were 0.37, 0.29, 0.31, 0.14 for MY, FY, PY and SCS, respectively (Interbull, 2016a, 2016b).

WG method. It consists in adjusting the EBVs and dEBVs of cows, estimated as in the VR method, in order to make them similar to those of bulls. In summary, for longitudinal traits, the following scheme was implemented:

1. Estimate $dEBV_{ij}$ for all cows and bulls based on equation [1] without adding the PA_{ij} effect;
2. Group the animals present in the official 2012 genetic evaluation run by similar REL_{EBV_i} (from 0.00 to 0.20, 0.20 to 0.40,…, 0.80 to 1.00);
3. Estimate the variance adjustment (VA_j) for each group of animals in each day j as standard deviation of $dEBV_j$ of bulls divided by the standard deviation of $dEBV_j$ of cows;
4. Estimate the mean adjustment (MA_j) for each group of animals in each day j as the difference between the average of $dEBV_j$ of bulls and cows after variance adjustment;
5. Average VA_j and MA_j across groups (all genotyped cows and dams of genotyped animals will receive the same adjustment for each day j);
6. Estimate the adjusted PA (PA_{adj}), and the adjusted dEBV for all genotyped bulls ($dEBV_{adj,bulls}$) and cows ($dEBV_{adj,cows}$) as follow:

$$PA_{adj_{ij}} = PA_{sire_{ij}} + \left[VA_j (PA_{dam_{ij}} - MA_j) \right], \qquad [4]$$

$$dEBV_{adj,bulls_{ij}} = dEBV_{ij} + PA_{adj_{ij}} \text{ and } dEBV_{adj,cows_{ij}} = VA_j (dEBV_{ij} - MA_j) + PA_{adj_{ij}};$$

where $PA_{sire_{ij}}$ and $PA_{dam_{ij}}$ are the fraction of PA from the sire and dam of the animal i in the day j, respectively.

7. Reverse the deregression to obtain the adjusted EBV (EBV_{adj}) for animal i in the day j as:

$$EBV_{adj_{ij}} = (1 - R_i) PA_{ij} + R_i (dEBV_{adj_{ij}}), \qquad [5]$$

where R_i was computed as described in equation (2).

GR method. This method is an approximation of the evaluation equations that allows for obtaining the dEBV free of the PA effects. In general, the equations to be solved for each animal are described as follows:

$$\begin{bmatrix} \mathbf{W}'_{PA}\mathbf{W}_{PA} + 4\lambda & -2\lambda \\ -2\lambda & \mathbf{W}'_{Ind-PA}\mathbf{W}_{Ind-PA} + 2\lambda \end{bmatrix} \begin{bmatrix} \mathbf{PA} \\ \mathbf{EBV} \end{bmatrix} = \begin{bmatrix} \mathbf{y}^*_{PA} \\ \mathbf{y}^*_{Ind-PA} \end{bmatrix}, \qquad [6]$$

where **PA** and **EBV** are the vectors of solutions for PA and EBV in each day, respectively; $\mathbf{y}^{*}_{\text{PA}}$ and $\mathbf{y}^{*}_{\text{Ind-PA}}$ are the amount of information equivalent to the right-hand-side elements due to PA and animal effect free of PA, respectively. The $\mathbf{W}'_{\text{PA}}\mathbf{W}_{\text{PA}}$ and $\mathbf{W}'_{\text{Ind-PA}}\mathbf{W}_{\text{Ind-PA}}$ represent the unknown information content of the PA and animal effect free of PA, respectively. These terms are estimated as follows:

$$\mathbf{W}'_{\text{PA}}\mathbf{W}_{\text{PA}} = \lambda(0.5\boldsymbol{\alpha}-4)+0.5\lambda\sqrt{(\boldsymbol{\alpha}^2+16/\boldsymbol{\delta})} \text{ and } \mathbf{W}'_{\text{Ind-PA}}\mathbf{W}_{\text{Ind-PA}} = \boldsymbol{\delta}\mathbf{W}'_{\text{PA}}\mathbf{W}_{\text{PA}} + 2\lambda(2\boldsymbol{\delta}-1) \quad [7]$$

where $\boldsymbol{\alpha} = 1/(0.5-\mathbf{REL_{PA}})$, $\boldsymbol{\delta} = (0.5-\mathbf{REL_{PA}})/(1-\mathbf{REL_{EBV}})$ and, $\lambda = (1-\text{h}^2)/\text{h}^2$, as previously defined. Programming codes in R (R Core Team, 2016), written to create the dEBVs by using the different methods of deregression are available as supplementary material.

Statistical analysis

In order to compare the efficiency of deregression methods, we have performed genomic predictions of breeding values using pseudo-phenotypes calculated based on each deregression method (described above). The training population included genotyped animals from 2012 official genetic evaluation provided by CDN. Genotyped individuals born after 2011 that were not included in the training population were used as validation population. Thus, animals in the training and validation populations were born between 1960 and 2010, and between 2011 and 2014, respectively.

Three scenarios were considered for each trait: 1) including only animals with REL_{EBV} higher than the average of REL_{PA} for the analyzed trait in the training and validation populations (SCEN1); 2) including only animals with REL_{EBV} higher than 0.50 in the training population and REL_{EBV} higher than the average REL_{PA} in the validation population (SCEN2); and, 3) including only animals with REL_{EBV} higher than 0.50 in the training and validation populations (SCEN3). Average of $\mathbf{REL_{PA}}$ in the training population was 0.35, 0.35, 0.34 and 0.31 for MY, FY, PY and SCS, respectively; and 0.34, 0.33, 0.33 and 0.27 for MY, FY, PY and SCS in the validation population, respectively. Genotyped animals that did not have own performance or progeny information ($\text{ERC}_{(\text{Ind-PA})_i} = 0$) were excluded from all scenarios. This is because they would not contribute to the marker effect estimation and could cause double-counting of information. In other words, an animal without own phenotype or without phenotype of daughters would have an EBV calculated solely based on PA and, therefore, this would be duplicated information in the analysis.

Genotype quality control. Originally, a total of 499 animals were genotyped using low density panels (3K, 7K, 8K or 15K), 57 animals were genotyped using medium density panels (20K or 30K), 1046 animals were genotyped using 50K panels, and 202 animals were genotyped using high density panels. All genotypes were imputed to 50K using FImpute software (Sargolzaei et al., 2011). Details about the accuracy of imputation for Canadian dairy cattle breeds can be found in Larmer et al. (2014). After the imputation process, single nucleotide polymorphisms (SNPs) with Mendelian conflicts, call rate less than 0.95, minor allele frequency (MAF) less than 0.01 and SNPs with departure from expected heterozygosity higher than 0.15 were removed using the PreGSF90 software (Aguilar et al., 2014). After the quality control, 38,166 SNPs on the 29 bovine autosomes remained in the analysis.

Random regression GBLUP. In order to contrast the performance of different deregression methods of cows' and bulls' breeding values for genomic prediction of test-day traits, individual random regression coefficients for each trait were predicted using GBLUP (VanRaden, 2008). The general RRM used for each trait can be described as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e}, \qquad [8]$$

where $\mathbf{y}$, $\mathbf{b}$, $\mathbf{g}$, and $\mathbf{e}$ are vectors of pseudo-phenotypes (deregressed EBVs for all days in milk, i.e., from five to 305), fixed regression coefficients, random regression coefficients, and residuals, respectively. The $\mathbf{b}$ and $\mathbf{g}$ effects were modelled by fourth order orthogonal Legendre polynomials (Kirkpatrick et al., 1990). The $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices for fixed and random effects, respectively. The model assumptions are described as:

$$\mathrm{E}[\mathbf{y}] = \mathbf{Xb} \text{ and } \mathrm{Var}\begin{bmatrix} \mathbf{g} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_0 \otimes \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{bmatrix}, \qquad [9]$$

where $\mathbf{G}_0$ and $\mathbf{R}$ are the additive genetic and residual variance-covariance matrices, respectively. Residual variance was initially assumed homogenous along lactation for all traits ($\mathbf{R} = \mathbf{I}\sigma_e^2$), since the official genetic evaluation carried out by CDN already accounted for heterogeneous variance (Interbull, 2016a, 2016b). The $\mathbf{G}$ is the genomic relationship matrix, constructed according to VanRaden (2008). The AIREMLF90 software (Misztal et al., 2002) was used to estimate the variance components and the solutions of the MME for the regression coefficients.

Weights. In general, there is a high variability of REL_{EBV} between animals in dairy cattle populations, especially between bulls and cows. Thus, considering this variability may yield to more reliable predictions (Vandenplas & Gengler, 2015). In order to verify if the performance of the deregression methods is influenced for this variability, all analyses were carried out considering no weight (first sub-scenario), and considering the ERC of an animal, excluding the PA information ($ERC_{(Ind-PA)}$), as weight for the inverse of residual variance matrix (second sub-scenario). Hereafter the first sub-scenario will be referred as "UW" and the second one as "WT", in the subscript of each scenario. For instance, the first scenario considering the second sub-scenario will be called $SCEN1_{WT}$. Finally, a total of 18 analyses were carried out for each trait (3 deregression methods × 3 scenarios × 2 sub-scenarios).

Direct genomic values. The vector of DGV solutions for each trait of each animal i was obtained as:

$$\mathbf{DGV}_i = \mathbf{T}\hat{\mathbf{g}}_i, \qquad [10]$$

where $\hat{\mathbf{g}}_i$ is the vector of predicted genomic random regression coefficients; and $\mathbf{T}$ is a matrix of orthogonal covariates associated with the fourth order Legendre polynomials.

Comparison of different deregression methods

The different deregression methods were compared by contrasting reliabilities and bias (slope of the linear regression of dEBV on DGV; b_{DGV}), considering all days in milk (i.e, from five to 305), for each trait in the validation population. Realized reliabilities of dEBV (r^2_{dEBV}) and DGV (r^2_{DGV}) were calculated as the square of Pearson correlation between dEBV and EBV, and between DGV and dEBV, respectively. To adjust r^2_{DGV} for the inaccuracy of dEBVs, adjusted realized reliabilities (r^2_{DGVadj}) were calculated dividing r^2_{DGV} by r^2_{dEBV}. Regression coefficients (b_{DGV}) were used as indicator of bias of DGV, which assumes that the dEBVs were unbiased.

To evaluate the influence of deregression methods on cows' and bulls' DGV separately, they were split into two datasets after the GBLUP, and average and standard deviation of DGV and b_{DGV} were calculated for cows (COWS), bulls (BULLS), and all animals combined (COMB).

4.4. Results

Descriptive statistics of the training population in the COWS, BULLS and COMB datasets are shown in Table 1. Similar descriptive statistics for the validation populations are shown in Table S1 at the supplementary material.

Table 1. Descriptive statistics of training population of the combined dataset (COMB), and of the dataset containing only cows (COWS) and only bulls (BULLS), considering minimum reliability of estimated breeding value ($\mathrm{REL}_{\mathrm{EBV}}$) equal to average of reliability of parent average, and considering minimum $\mathrm{REL}_{\mathrm{EBV}}$ of 0.50.

Scenario	**Traits**	**N**	**Mean (SD)**				
			$\mathrm{REL}_{\mathrm{EBV}}$	**EBV**	**$\mathrm{dEBV}_{\mathrm{VR}}$**	**$\mathrm{dEBV}_{\mathrm{WG}}$**	**$\mathrm{dEBV}_{\mathrm{GR}}$**
				COMB			
Average $\mathrm{REL}_{\mathrm{PA}}$	MY	1,257	0.68 (0.18)	8.23 (1.59)	8.40 (2.06)	7.76 (2.00)	8.53 (2.54)
	FY	1,226	0.66 (0.19)	0.38 (0.07)	0.39 (0.09)	0.35 (0.08)	0.39 (0.12)
	PY	1,249	0.65 (0.19)	0.26 (0.07)	0.27 (0.10)	0.24 (0.09)	0.27 (0.10)
	SCS	1,199	0.56 (0.20)	-1.09 (0.31)	-1.09 (0.47)	-1.06 (0.48)	-1.09 (0.93)
0.50	MY	1,001	0.74 (0.15)	8.14 (1.64)	8.25 (1.91)	7.60 (1.85)	8.32 (2.14)
	FY	944	0.73 (0.15)	0.38 (0.07)	0.39 (0.08)	0.34 (0.07)	0.39 (0.09)
	PY	944	0.73 (0.15)	0.26 (0.07)	0.27 (0.07)	0.24 (0.07)	0.27 (0.08)
	SCS	544	0.74 (0.16)	-1.08 (0.34)	-1.08 (0.38)	-1.06 (0.38)	-1.08 (0.48)
				BULLS			
Average $\mathrm{REL}_{\mathrm{PA}}$	MY	747	0.72 (0.21)	8.21 (1.64)	8.32 (2.05)	7.81 (2.01)	8.41 (2.51)
	FY	731	0.71 (0.21)	0.37 (0.07)	0.37 (0.08)	0.34 (0.08)	0.38 (0.11)
	PY	741	0.71 (0.22)	0.26 (0.07)	0.26 (0.08)	0.24 (0.08)	0.26 (0.10)
	SCS	721	0.63 (0.22)	-1.07 (0.33)	-1.06 (0.48)	-1.05 (0.48)	-1.05 (0.97)
0.50	MY	574	0.81 (0.15)	8.07 (1.71)	8.08 (1.89)	7.58 (1.85)	8.11 (2.06)
	FY	556	0.81 (0.15)	0.37 (0.07)	0.37 (0.08)	0.34 (0.07)	0.37 (0.08)
	PY	554	0.81 (0.15)	0.25 (0.07)	0.25 (0.07)	0.23 (0.07)	0.25 (0.08)
	SCS	450	0.78 (0.14)	-1.07 (0.35)	-1.07 (0.38)	-1.06 (0.38)	-1.08 (0.45)
				COWS			
Average $\mathrm{REL}_{\mathrm{PA}}$	MY	510	0.61 (0.10)	8.25 (1.52)	8.52 (2.07)	7.67 (1.99)	8.71 (2.57)
	FY	495	0.58 (0.09)	0.39 (0.06)	0.41 (0.09)	0.35 (0.09)	0.42 (0.13)
	PY	508	0.57 (0.09)	0.27 (0.06)	0.28 (0.12)	0.24 (0.11)	0.29 (0.10)
	SCS	478	0.45 (0.06)	-1.13 (0.28)	-1.13 (0.46)	-1.09 (0.47)	-1.16 (0.86)
0.50	MY	427	0.65 (0.06)	8.22 (1.53)	8.47 (1.94)	7.62 (1.86)	8.60 (2.22)
	FY	388	0.63 (0.05)	0.40 (0.06)	0.41 (0.08)	0.35 (0.07)	0.42 (0.10)
	PY	390	0.62 (0.05)	0.27 (0.06)	0.28 (0.07)	0.24 (0.07)	0.29 (0.08)
	SCS	94	0.54 (0.04)	-1.12 (0.28)	-1.11 (0.39)	-1.07 (0.40)	-1.11 (0.60)

Average $\mathrm{REL}_{\mathrm{PA}}$: minimum $\mathrm{REL}_{\mathrm{EBV}}$ equal to average of reliability of parent average. Average $\mathrm{REL}_{\mathrm{PA}}$ was estimated to be 0.35, 0.35, 0.34 and 0.31 for milk (MY), fat (FY), and protein (PY) yield, and somatic cell score (SCS), respectively. dEBV: deregressed EBV estimated by Van Raden ($\mathrm{dEBV}_{\mathrm{VR}}$), Wiggans ($\mathrm{dEBV}_{\mathrm{WG}}$), and Garrick ($\mathrm{dEBV}_{\mathrm{GR}}$) methods. N: number of animals.

Scenarios considering only animals with REL_{EBV} greater than 0.50 in the training population (SCEN2 and SCEN3, displayed as scenario 0.5 in Table 1) had approximately 20%, 23%, 24%, and 55% less animals compared to SCEN1 for MY, FY, PY and SCS, respectively. The proportion of cows in the training population was around 40% for almost all traits and scenarios (except for SCS in SCEN2 and SCEN3, in which the proportion of cows was around 17%). As expected, REL_{EBV} were higher for BULLS, and the mean of EBVs from each trait varied between scenarios according to direction of selection (SCEN1 had mean of EBV increased for MY, FY and PY, and decrease for SCS, when compared to SCEN2 and SCEN3). The EBVs from COWS were more favorable (higher for MY, FY and PY, and lower for SCS) than EBVs from BULLS, and the differences between them were higher in SCEN1 for SCS, and in SCEN2 and SCEN3 for MY, FY and PY. For all traits and scenarios, the WG allowed to reduce the difference between dEBV from COWS and BULLS when compared to VR and GR methods. Among all tested methods, GR showed consistently the highest standard deviation of dEBVs for all analyzed traits and scenarios. Meanwhile the average dEBV was similar for VR and GR methods and constantly lower for WG method across the scenarios.

Table 2 shows the r^2_{dEBV} , r^2_{DGV} , r^2_{DGVadj} , and b_{DGV} for each deregression method, considering all scenarios in the COMB dataset. According to unweighted analysis, r^2_{dEBV} and r^2_{DGV} in $SCEN3_{UW}$ were higher than $SCEN1_{UW}$ and $SCEN2_{UW}$. However, especially for $SCEN2_{UW}$ and $SCEN3_{UW}$, r^2_{DGVadj} was similar for MY, FY and PY. In general, r^2_{dEBV} , r^2_{DGV} and r^2_{DGVadj} were similar for VR and WG (within each scenario), and lower reliability values were found for GR method for all traits. The DGV was overestimated for all traits and methods, since b_{DGV} were significantly lower than 1. However, when comparing b_{DGV} estimates, the DGVs were less biased for minimum reliability of 0.5 in the reference and the average of REL_{PA} for validation population ($SCEN2_{UW}$). Considering $ERC_{(Ind-PA)}$ as weight increased r^2_{DGV} , r^2_{DGVadj} and b_{DGV} just for PY and SCS in $SCEN1_{WT}$, for VR and WG methods. For $SCEN2_{WT,}$ the b_{DGV} of WG method generate undesirably decrease when compared to $SCEN2_{UW}$. In addition, WG method only increased b_{DGV} for MY and SCS in the $SCEN2_{UW}$ and $SCEN3_{UW}$, and for MY and SCS in the $SCEN3_{WT}$.

Average DGV and b_{DGV} for BULLS and COWS, for unweighted analysis, are shown in Table 3.

Table 2. Realized reliabilities of deregressed estimated breeding value (r^2_{dEBV}), realized and adjusted realized reliabilities of DGV (r^2_{DGV} and r^2_{DGVadj}), and regression coefficients (b_{DGV}) for milk (MY), fat (FY), and protein (PY) yields, and somatic cell score (SCS) in the combined file.

Scenario	**Trait**	**Method**	r^2_{dEBV}	**Without weight**			$\mathrm{ERC}_{(\mathrm{Ind-PA})}$ **as weight**		
				r^2_{DGV}	r^2_{DGVadj}	b_{DGV} **± SE**	r^2_{DGV}	r^2_{DGVadj}	b_{DGV} **± SE**
Average REL$_{\mathrm{PA}}$ (SCEN1)	**MY**	**VR**	0.71	0.27	0.38	0.87 ± 0.003	0.27	0.38	0.87 ± 0.003
		WG	0.69	0.27	0.39	0.87 ± 0.003	0.27	0.39	0.87 ± 0.003
		GR	0.55	0.15	0.27	0.71 ± 0.004	0.15	0.27	0.72 ± 0.004
	FY	**VR**	0.77	0.22	0.28	0.72 ± 0.003	0.22	0.28	0.72 ± 0.003
		WG	0.73	0.20	0.27	0.70 ± 0.003	0.20	0.27	0.70 ± 0.003
		GR	0.55	0.08	0.14	0.49 ± 0.004	0.08	0.14	0.49 ± 0.004
	PY	**VR**	0.78	0.42	0.52	0.79 ± 0.002	0.46	0.59	0.88 ± 0.002
		WG	0.74	0.38	0.51	0.75 ± 0.002	0.43	0.58	0.85 ± 0.002
		GR	0.56	0.24	0.43	0.72 ± 0.003	0.24	0.43	0.74 ± 0.003
	SCS	**VR**	0.74	0.17	0.23	0.50 ± 0.003	0.18	0.24	0.54 ± 0.002
		WG	0.75	0.16	0.21	0.50 ± 0.003	0.17	0.23	0.54 ± 0.003
		GR	0.39	0.02	0.05	0.17 ± 0.003	0.02	0.05	0.17 ± 0.003
0.50 for training and average REL$_{\mathrm{PA}}$ for validation (SCEN2)	**MY**	**VR**	0.71	0.26	0.37	0.93 ± 0.003	0.26	0.37	0.93 ± 0.004
		WG	0.69	0.26	0.38	0.94 ± 0.004	0.26	0.38	0.90 ± 0.003
		GR	0.55	0.15	0.27	0.83 ± 0.005	0.15	0.27	0.83 ± 0.005
	FY	**VR**	0.77	0.24	0.31	0.86 ± 0.004	0.24	0.31	0.86 ± 0.004
		WG	0.73	0.22	0.30	0.86 ± 0.004	0.23	0.32	0.80 ± 0.003
		GR	0.55	0.10	0.19	0.71 ± 0.005	0.10	0.19	0.71 ± 0.005
	PY	**VR**	0.78	0.47	0.60	0.94 ± 0.002	0.47	0.60	0.94 ± 0.002
		WG	0.74	0.44	0.60	0.92 ± 0.002	0.44	0.60	0.88 ± 0.002
		GR	0.56	0.25	0.46	0.86 ± 0.003	0.25	0.46	0.86 ± 0.003
	SCS	**VR**	0.74	0.23	0.31	0.78 ± 0.003	0.23	0.61	0.78 ± 0.003
		WG	0.75	0.23	0.31	0.79 ± 0.003	0.23	0.31	0.78 ± 0.003
		GR	0.39	0.03	0.09	0.53 ± 0.01	0.03	0.09	0.53 ± 0.01
0.50 (SCEN3)	**MY**	**VR**	0.90	0.34	0.38	0.91 ± 0.003	0.34	0.38	0.91 ± 0.003
		WG	0.90	0.35	0.39	0.92 ± 0.003	0.35	0.39	0.92 ± 0.003
		GR	0.81	0.23	0.28	0.80 ± 0.004	0.23	0.28	0.80 ± 0.004
	FY	**VR**	0.88	0.28	0.32	0.84 ± 0.003	0.28	0.32	0.84 ± 0.003
		WG	0.86	0.26	0.30	0.84 ± 0.004	0.26	0.30	0.84 ± 0.004
		GR	0.72	0.14	0.19	0.69 ± 0.004	0.14	0.19	0.69 ± 0.004
	PY	**VR**	0.91	0.56	0.61	0.93 ± 0.002	0.56	0.61	0.93 ± 0.002
		WG	0.89	0.52	0.58	0.91 ± 0.002	0.52	0.58	0.91 ± 0.002
		GR	0.76	0.37	0.49	0.84 ± 0.003	0.37	0.49	0.84 ± 0.003
	SCS	**VR**	0.85	0.32	0.38	0.80 ± 0.01	0.33	0.39	0.80 ± 0.01
		WG	0.85	0.32	0.38	0.81 ± 0.01	0.32	0.38	0.81 ± 0.01
		GR	0.55	0.10	0.18	0.63 ± 0.01	0.10	0.18	0.63 ± 0.01

Average REL$_{\mathrm{PA}}$: minimum REL$_{\mathrm{EBV}}$ equal to average of reliability of parent average. Average REL$_{\mathrm{PA}}$ was estimated to be 0.35, 0.35, 0.34 and 0.31 for MY, FY, PY SCS, respectively. Deregression was based on Van Raden (VR), Wiggans (WG) and Garrick (GR) methods. SCEN: scenario.

Table 3. Average and standard deviation (SD) of direct genomic values (DGV) and regression coefficients (b_{DGV}) for milk (MY), fat (FY), and protein (PY) yield, and somatic cell score (SCS) in the dataset containing only bulls (BULLS) and only cows (COWS).

Scenario	**Trait**	**Method**	**BULLS**		**COWS**	
			DGV (SD)	**b_{DGV} ± SE**	**DGV (SD)**	**b_{DGV} ± SE**
Average REL_{PA} (SCEN1)	**MY**	**VR**	8.73 (1.31)	1.13 ± 0.01	8.53 (1.36)	0.86 ± 0.003
		WG	8.08 (1.26)	1.13 ± 0.01	7.91 (1.31)	0.85 ± 0.003
		GR	8.94 (1.48)	1.02 ± 0.02	8.69 (1.62)	0.70 ± 0.004
	FY	**VR**	0.41 (0.07)	0.74 ± 0.01	0.40 (0.06)	0.72 ± 0.003
		WG	0.37 (0.06)	0.70 ± 0.01	0.36 (0.05)	0.69 ± 0.003
		GR	0.42 (0.08)	0.57 ± 0.01	0.41 (0.07)	0.49 ± 0.004
	PY	**VR**	0.29 (0.06)	0.99 ± 0.01	0.28 (0.07)	0.77 ± 0.002
		WG	0.26 (0.06)	0.98 ± 0.01	0.25 (0.06)	0.74 ± 0.002
		GR	0.30 (0.07)	0.92 ±0.01	0.29 (0.07)	0.71 ± 0.003
	SCS	**VR**	-1.16 (0.33)	0.44 ± 0.01	-1.12 (0.31)	0.51 ± 0.003
		WG	-1.13 (0.33)	0.43 ± 0.01	-1.09 (0.31)	0.50 ± 0.003
		GR	-1.20 (0.63)	0.24 ± 0.02	-1.11 (0.55)	0.17 ± 0.003
0.50 for training and average REL_{PA} for validation (SCEN2)	**MY**	**VR**	8.73 (1.31)	1.13 ± 0.02	8.53 (1.36)	0.93 ± 0.004
		WG	8.08 (1.26)	1.13 ± 0.02	7.91 (1.31)	0.93 ± 0.004
		GR	8.94 (1.48)	1.04 ± 0.02	8.69 (1.62)	0.82 ± 0.005
	FY	**VR**	0.41 (0.07)	0.85 ± 0.01	0.40 (0.06)	0.86 ± 0.004
		WG	0.37 (0.06)	0.80 ± 0.01	0.36 (0.05)	0.86 ± 0.004
		GR	0.42 (0.08)	0.66 ± 0.02	0.41 (0.07)	0.72 ± 0.005
	PY	**VR**	0.29 (0.06)	1.06 ± 0.01	0.28 (0.07)	0.94 ± 0.002
		WG	0.26 (0.06)	1.06 ± 0.01	0.25 (0.06)	0.92 ± 0.002
		GR	0.30 (0.07)	0.98 ± 0.02	0.29 (0.07)	0.85 ± 0.003
	SCS	**VR**	-1.16 (0.33)	0.71 ± 0.02	-1.12 (0.31)	0.78 ± 0.003
		WG	-1.13 (0.33)	0.71 ± 0.02	-1.09 (0.31)	0.79 ± 0.004
		GR	-1.20 (0.63)	0.41 ± 0.04	-1.11 (0.55)	0.54 ± 0.01
0.50 (SCEN3)	**MY**	**VR**	8.70 (1.14)	1.10 ± 0.01	8.47 (1.25)	0.90 ± 0.003
		WG	8.05 (1.10)	1.12 ± 0.02	7.84 (1.19)	0.91 ± 0.003
		GR	8.87 (1.22)	1.02 ± 0.02	8.56 (1.39)	0.79 ± 0.004
	FY	**VR**	0.42 (0.05)	0.73 ± 0.02	0.40 (0.05)	0.84 ± 0.003
		WG	0.38 (0.05)	0.63 ± 0.02	0.36 (0.05)	0.84 ± 0.004
		GR	0.44 (0.05)	0.53 ± 0.02	0.41 (0.06)	0.69 ± 0.004
	PY	**VR**	0.29 (0.06)	0.97 ± 0.01	0.28 (0.06)	0.93 ± 0.002
		WG	0.26 (0.05)	0.96 ± 0.01	0.25 (0.06)	0.90 ± 0.002
		GR	0.30 (0.06)	0.87 ± 0.01	0.28 (0.06)	0.84 ± 0.003
	SCS	**VR**	-1.18 (0.22)	0.90 ± 0.01	-1.12 (0.24)	0.77 ± 0.01
		WG	-1.26 (0.22)	0.90 ± 0.01	-1.10 (0.24)	0.78 ± 0.01
		GR	-1.21 (0.24)	0.86 ± 0.02	-1.13 (0.27)	0.57 ± 0.01

Average REL_{PA}: minimum REL_{EBV} equal to average of reliability of parent average. Average REL_{PA} was estimated in 0.35, 0.35, 0.34 and 0.31 for MY, FY, PY SCS, respectively. Deregression of estimated breeding value was based on Van Raden (VR), Wiggans (WG) and Garrick (GR) methods. SE: standard error. SCEN: scenario.

For all traits, average DGV were more favorable (higher for MY, FY and PY, and lower for SCS) in BULLS compared to COWS dataset (Table 3). Regarding to b_{DGV}, DGV for MY tended to be underestimated for bulls and overestimated for cows, with all different methods and scenarios. In general, values of b_{DGV} close to 1 were observed for PY in BULLS, mainly from VR and WG methods. Similar b_{DGV} estimates were found for FY and SCS for BULLS and COWS datasets, within each deregression method, especially in $SCEN1_{UW}$ and $SCEN2_{UW}$.

4.5. Discussion

Reliable performance information and large size of the training population are crucial factors to ensure high reliability in the genomic predictions (Liu et al., 2011). Thus, in order to increase the size of training population in breeds with lower number of animals, females usually have been included in official genomic evaluations in several countries (Interbull, 2016c; Koivula et al., 2016). A lower proportion of cows was observed for SCS when the minimum REL_{EBV} was considered equal to 0.50, since most cows had REL_{EBV} lower than this assumed threshold. In addition, average EBVs from cows were higher (in absolute values) than EBVs from bulls for all analyzed traits, probably because only important cows have been genotyped in the Jersey breed. Thus, these cows might have some level of preferential treatment. Preferential treatment have been reported as one of the main causes of bias in cows evaluations (Dassonneville et al., 2012), since cows selected to be genotyped are important animals in the breed (i.e., dams or future dams of bulls with high EBV for economical important traits). Several studies have reported that preferential treatment increases linearly the bias of cow evaluations (Kuhn et al., 1994). However, adjustments have been developed to make these evaluations comparable with those of bulls. Besides its simple application, adjustments of WG method allowed to reduce the differences between dEBV from cows and bulls when compared to VR and GR methods (Table 1).

Efficiency of genomic predictions depends on the accuracy of estimated marker effects, which relies on the information contained in the dEBVs. In spite of the fact that dEBV is derived from EBV, they might substantially differ and the difference between them increases as the REL_{EBV} decreases. Thus, since the deregression process is less severe for animals with high REL_{EBV}, r^2_{dEBV} and r^2_{DGV} were higher in the scenario that just considered animals with REL_{EBV} higher than 0.50 in the training and validation populations (SCEN3). However, when adjusting r^2_{DGV} for the inaccuracy of dEBVs in the validation population (r^2_{DGVadj}), SCEN2 and SCEN3 were similar for MY, FY and

PY traits (Table 2). Especially for SCS, r^2_{DGVadj} estimated in the SCEN2 were lower than those estimated in the SCEN3, probably due to the large effect of reduced number of animals in the validation population when a minimum REL_{EBV} of 0.50 was considered (Table S1, at the supplementary material). Thus, our results suggest that the influence of REL_{EBV} in the r^2_{DGV} for the animals in the validation population can be easily adjusted using the r^2_{dEBV} . Nowadays, several authors have opted to report reliabilities adjusted for the inaccuracy of the dEBVs in genomic selection studies (Baba et al., 2016; Koivula et al., 2016).

Estimated r^2_{dEBV} , r^2_{DGV} and r^2_{DGVadj} were similar for VR and WG methods, suggesting that the adjustment made by WG had small effect in the reliabilities of genomic prediction for Jersey cattle. These results corroborate with those reported by Wiggans et al. (2011) for total milk, fat and protein yield in Jersey cattle, where small difference between realized reliability estimated using adjusted and unadjusted predicted transmitting abilities were found. Contrasting all analyzed methods, GR resulted in lower r^2_{dEBV} , r^2_{DGV} and r^2_{DGVadj} for all traits (Table 2), which is due to the strength of the deregression of EBVs in this method when compared to VR and WG. For this reason, the GR generated more variable dEBV for all traits (which is indicated by standard deviations in Table 1), suggesting that the deregression was too strong. These results corroborate with those reported by Song et al. (2017), which also presented higher variance for dEBVs and lower r^2_{dEBV} generated by the GR method. Calus et al. (2016) also concluded that the deregression by GR was too strong compared to VR method. Therefore, based on our findings, the use of the GR method is less suitable for test-day traits compared to the others methods investigated here.

The ability to reduce the bias in the genomic predictions is an important factor when comparing different deregression methods. In general, a necessary (but not sufficient) condition for unbiased prediction of a model is that the slope of the linear regression of dEBV on predicted DGV should be close to one, assuming unbiased dEBVs. In this context, we found that the DGV was overestimated in all tested deregression methods. This is likely a consequence that the analyzed traits are under selection pressure. When considering animals with lower REL_{EBV} in the training population further increased the overestimation. It means that increasing the size of training population using animals with low REL_{EBV} (in unweighted analysis) increases the bias, probably due to more heterogeneous amount of information used to generate the EBVs. The heterogeneity of information was larger for PY and SCS in SCEN1, which explains the best performance of these traits in the weighted analysis ($SCEN1_{WT}$). In agreement with Guo et al. (2010), the weighted analysis in scenarios that considered minimum REL_{EBV} of 0.50 in the training population ($SCEN2_{WT}$ and $SCEN3_{WT}$) did not result in accuracy improvement. The WG method in the $SCEN2_{WT}$ (Table 2) had the b_{DGV} significantly decreased when compared to $SCEN2_{UW}$. This suggests that for different minimum criteria for REL_{EBV} in the training and validation populations, the weights used should consider the adjustments made in dEBVs of cows.

It is important to highlight that, in despite of the great numerical properties of GR (Garrick et al., 2009), this method showed the worst performance among all comparisons in the present study. In addition, using weighted analysis did not improved the GR method (Table 2). Thus, the stronger deregression performed by GR method (as discussed earlier) requires the use of specific weights in the inverse of residual variance matrix ($\mathbf{R}^{-1}$). Garrick et al. (2009) have deduced the optimal weights to be considered in $\mathbf{R}^{-1}$ based on heritability, REL_{EBV} and the fraction of genetic variance not explained by markers. In this context, Song et al. (2017) tested the GR method considering the optimal weights based on different heritabilities and proportions of genetic variance not explained by markers. However, these authors found that, given a fixed heritability, increasing the proportion of genetic variance not explained by markers decreased the accuracy of genomic prediction and increased the bias. Thus, they concluded that weighting the $\mathbf{R}^{-1}$ matrix by the reliability of response variable is, generally, more stable. In our study, as the weights were just used to verify if the performance of the methods is changed when using weighted analysis, we have opted to use the same weight for all tested methods. For this reason, we used one of the simplest weights described in the literature, i.e. effective record contributions. However, additional studies contrasting the performance of different deregression methods considering different weights are necessary.

In contrast to Table 1, breeding values in the validation population were more favorable (higher for MY, FY and PY, and lower for SCS) for BULLS than COWS (Table S1 at the supplementary material, and Table 3). This is probably related to the higher selection intensity for young bulls when compared to young cows, which have resulted in a small number of bulls mainly in the validation population (Table S1, supplementary material).

4.6. Conclusions

Deregressed longitudinal EBVs obtained using well established methods of deregression for non-longitudinal traits seems a feasible alternative for genomic prediction of longitudinal traits. For all analyzed test-day traits, the VR and WG deregression methods yielded to similar results, being, however, more accurate and less biased than GR independently of the minimum REL_{EBV} considered. The scenario considering a minimum REL_{EBV} of 0.50 in the training and minimum REL_{EBV} equal to the average of REL_{PA} in the validation population allowed for obtaining high reliabilities of predicted DGV and the smallest prediction biases. Weighted residual analysis only showed benefits when differences in REL_{EBV} among animals in the training population were high.

4.7. Acknowledgements

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES). The authors gratefully acknowledge the financial support to this research from Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative.

4.8. References

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. In: Proceedings, 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada, August 2014. doi: 10.13140/2.1.4801.5045.

Baba, T., Gotoh, Y., Yamaguchi, S., Nakagawa, S., Abe, H., Masuda, Y., & Kawahara, T. (2016). Application of single-step genomic best linear unbiased prediction with a multiple-lactation random regression test-day model for Japanese Holsteins. Animal Science Journal. doi: 10.1111/asj.12760.

Calus, M., Vandenplas, J., Ten Napel, J., & Veerkamp, R. (2016). Validation of simultaneous deregression of cow and bull breeding values and derivation of appropriate weights. Journal of Dairy Science, 99(8), 6403-6419. doi: 10.3168/jds.2016-11028.

Dassonneville, R., Baur, A., Fritz, S., Boichard, D., & Ducrocq, V. (2012). Inclusion of cow records in genomic evaluations and impact on bias due to preferential treatment. Genetics Selection Evolution, 44(1), 40. doi: 10.1186/1297-9686-44-40.

Garrick, D. J., Taylor, J. F., & Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution, 41(1), 1. doi: 10.1186/1297-9686-41-55.

Goddard, M. (1985). A method of comparing sires evaluated in different countries. Livestock Production Science, 13(4), 321-331. doi: 10.1016/0301-6226(85)90024-7.

Guo, G., Lund, M. S., Zhang, Y., & Su, G. (2010). Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. Journal of Animal Breeding and Genetics, 127(6), 423-432. doi: 10.1111/j.1439-0388.2010.00878.x.

Interbull. (2016a). Description of national genetic evaluation system and trend validation for production traits. Canada. Retrieved from http://www.interbull.org/ib/geforms. Accessed 02.02.2017.

Interbull. (2016b). Description of national genetic evaluation system and trend validation for milk somatic cell and clinical mastitis. Canada. Retrieved from http://www.interbull.org/ib/geforms. Accessed 02.02.2017.

Interbull. (2016c). Description of national genomic evaluation systems. Retrieved from http://www.interbull.org/ib/nationalgenoforms. Accessed 02.02.2017.

Jairath, L., Dekkers, J., Schaeffer, L., Liu, Z., Burnside, E., & Kolstad, B. (1998). Genetic evaluation for herd life in Canada. Journal of Dairy Science, 81(2), 550-562. doi: 10.3168/jds.S0022-0302(98)75607-3.

Kirkpatrick, M., Lofsvold, D., & Bulmer, M. (1990). Analysis of the inheritance, selection and evolution of growth trajectories. Genetics, 124(4), 979-993.

Koivula, M., Strandén, I., Aamand, G. P., & Mäntysaari, E. A. (2016). Effect of cow reference group on validation reliability of genomic evaluation. Animal, 10(06), 1061-1066. doi: 10.1017/S1751731115002864.

Kuhn, M., Boettcher, P., & Freeman, A. (1994). Potential Biases in Predicted Transmitting Abilities of Females from Preferential Treatment. Journal of Dairy Science, 77(8), 2428-2437. doi: http://dx.doi.org/10.3168/jds.S0022-0302(94)77185-X.

Larmer, S., Sargolzaei, M., & Schenkel, F. (2014). Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. Journal of Dairy Science, 97(5), 3128-3141. doi: 10.3168/jds.2013-6826.

Liu, Z., Seefried, F. R., Reinhardt, F., Rensing, S., Thaller, G., & Reents, R. (2011). Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genetics Selection Evolution, 43(1), 1. doi: 10.1186/1297-9686-43-19.

Meuwissen, T., Hayes, B., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4), 1819-1829.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., & Lee, D. (2002). BLUPF90 and related programs (BGF90). In: Proceedings, 7th World Congress of Genetics Applied to Livestock Production, Montpellier, France, August 2002.

Ostersen, T., Christensen, O. F., Henryon, M., Nielsen, B., Su, G., & Madsen, P. (2011). Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. Genetics Selection Evolution, 43(1), 38. doi: 10.1186/1297-9686-43-38.

R Core Team, (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sargolzaei, M., Chesnais, J., & Schenkel, F. (2011). FImpute-An efficient imputation algorithm for dairy cattle populations. Journal of Dairy Science, 94(1), 421.

Schaeffer, L.R. (2004). Application of random regression models in animal breeding. Livestock Production Science, 86, 35-45. doi: https://doi.org/10.1016/S0301-6226(03)00151-9.

Song, H., Li, L., Zhang, Q., Zhang, S., & Ding, X. (2017). Accuracy and bias of genomic prediction with different de-regression methods. Animal, 1-7. doi:10.1017/S175173111700307X.

Vandenplas, J., & Gengler, N. (2015). Strategies for comparing and combining different genetic and genomic evaluations: A review. Livestock Science, 181, 121-130. doi: https://doi.org/10.1016/j.livsci.2015.09.012.

VanRaden, P. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science, 91(11), 4414-4423. doi: 10.3168/jds.2007-0980.

VanRaden, P., Van Tassell, C., Wiggans, G., Sonstegard, T., Schnabel, R., Taylor, J., & Schenkel, F. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science, 92(1), 16-24. doi: 10.3168/jds.2008-1514.

VanRaden, P., & Wiggans, G. (1991). Derivation, calculation, and use of national animal model information. Journal of Dairy Science, 74(8), 2737-2746. doi: 10.3168/jds.S0022-0302(91)78453-1

Wiggans, G., Cooper, T., VanRaden, P., & Cole, J. (2011). Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. Journal of Dairy Science, 94(12), 6188-6193. doi: 10.3168/jds.2011-4481.

4.9. Supplementary material

4.9.1. Programming codes for the different deregression methods

```
#Programming codes - Hinayah Rojas de Oliveira
Contact: holivier@uoguelph.ca

#Deregression VanRaden (2009)
h2 <- 0.37 #Heritability
kd <- (1-h2)/h2
#Effective record contributions - from parent average (ERCpa) and from animal (ERCind)
ERCpa <- kd*(data$REL_PA/(1-data$REL_PA)) #REL_PA: Reliability of parente average (PA)
ERCind <- (kd*(data$REL_EBV/(1-data$REL_EBV)))-ERCpa #REL_EBV: Reliability of EBV
#Deregressed EBV (dEBV)
R <- ERCind/(ERCind+ERCpa+1)
dEBV <- ((data$EBV - data$PA)/R)+data$PA
#Final file
dEBV_ok <- cbind(data, dEBV, R, ERCind)
#Drop animals without own/daugther information
dEBV_ok1 <- dEBV_ok[!dEBV_ok$ERCind==0,]

#Deregression Garrick (2009)
h2 <- 0.37
k <- (1-h2)/h2
alfa <- 1/(0.5-data$REL_PA)
omega <- (0.5-data$REL_PA)/(1-data$REL_EBV)
ZpaZpa <- k*(0.5*alfa-4) + 0.5*k*sqrt(alfa*alfa + 16/omega)
ZiZi <- omega*ZpaZpa + 2*k*(2*omega-1)
#Deregressed EBV
EBV_sPA <- (-2*k*data$PA)+((ZiZi+2*k)*data$EBV)
```

```
dEBV <- EBV_sPA/ZiZi
REL_dEBV <- 1-k/(ZiZi+k)
#Weight estimated by the Garrick method
c <- 0.1
weigth <- (1-h2)/((c+(1-REL_dEBV)/REL_dEBV)*h2)
#Final file
dEBV_ok <- cbind(data, dEBV, REL_dEBV, weigth)

#Deregression Wiggans (2011)
#Split bulls and cows in different files:
data_bulls <- data[data$sex=="M",]
data_cows <- data[data$sex=="F",]
#Bulls
h2 <- 0.37
kd <- (1-h2)/h2
ERCpa_bulls <- kd*(data_bulls$REL_PA/(1-data_bulls$REL_PA))
ERCind_bulls <- (kd*(data_bulls$REL_EBV/(1-data_bulls$REL_EBV))) - ERCpa_bulls
R_bulls <- ERCind_bulls/(ERCind_bulls+ERCpa_bulls+1)
dEBV_bulls <- ((data_bulls$EBV) - (data_bulls$PA))/R_bulls
dEBV_ok_bulls <- cbind(data_bulls,dEBV_bulls, R_bulls, ERCind_bulls)
dEBV_ok0_bulls <- dEBV_ok_bulls[order(dEBV_ok_bulls$ID, dEBV_ok_bulls$DIM),]
dEBV_ok1_bulls <- dEBV_ok0_bulls[!dEBV_ok0_bulls$ERCind_bulls==0,]
#Cows
h2 <- 0.37
kd <- (1-h2)/h2
ERCpa_cows <- kd*(data_cows$REL_PA/(1-data_cows$REL_PA))
ERCind_cows <- (kd*(data_cows$REL_EBV/(1-data_cows$REL_EBV))) - ERCpa_cows
R_cows <- ERCind_cows/(ERCind_cows+ERCpa_cows+1)
dEBV_cows <- ((data_cows$EBV) - (data_cows$PA))/R_cows
dEBV_ok_cows <- cbind(data_cows,dEBV_cows, R_cows, ERCind_cows)
dEBV_ok0_cows <- dEBV_ok_cows[order(dEBV_ok_cows$ID, dEBV_ok_cows$DIM),]
dEBV_ok1_cows <- dEBV_ok0_cows[!dEBV_ok0_cows$ERCind_cows==0,]
#Group animals by EBV reliability
#Cluster 1: from 0.5 to 0.6
dEBV_ok1_bulls_g1 <- dEBV_ok1_bulls[dEBV_ok1_bulls$REL_EBV > 0.4999 &
dEBV_ok1_bulls$REL_EBV < 0.6,]
dEBV_ok1_cows_g1 <- dEBV_ok1_cows[dEBV_ok1_cows$REL_EBV > 0.4999 &
dEBV_ok1_cows$REL_EBV < 0.6,]
#Variance adjustment (VA)
sd_bulls <- tapply(dEBV_ok1_bulls_g1$dEBV_bulls, dEBV_ok1_bulls_g1$DIM, sd)
sd_cows <- tapply(dEBV_ok1_cows_g1$dEBV_cows, dEBV_ok1_cows_g1$DIM, sd)
VA_g1 <- sd_bulls/sd_cows
VA_g1_ok <- cbind(as.matrix(VA_g1), as.matrix(5:305))
colnames(VA_g1_ok) <- c("VA", "DIM")
#Mean adjustment (MA)
dEBV_ok1_bulls_VA_g1 <- dEBV_ok1_bulls_g1
adj_cows_g1 <- merge(dEBV_ok1_cows_g1, VA_g1_ok, by.x="DIM", by.y="DIM")
adj_cows_g1_1 <- adj_cows_g1$dEBV_cows*adj_cows_g1$VA
dEBV_ok1_cows_VA_g1 <- cbind(adj_cows_g1, adj_cows_g1_1)
```

```
MA_g1_mean_cows <- tapply(dEBV_ok1_cows_VA_g1$adj_cows_g1_1,
dEBV_ok1_cows_VA_g1$DIM, mean)
MA_g1_mean_bulls <- tapply(dEBV_ok1_bulls_VA_g1$dEBV_bulls, dEBV_ok1_bulls_VA_g1$DIM,
mean)
MA_g1 <- MA_g1_mean_cows - MA_g1_mean_bulls
MA_g1_ok <- cbind(as.matrix(MA_g1), as.matrix(5:305))
colnames(MA_g1_ok) <- c("MA", "DIM")
#Cluster 2: from 0.6 to 0.7
dEBV_ok1_bulls_g2 <- dEBV_ok1_bulls[dEBV_ok1_bulls$REL_EBV > 0.5999 &
dEBV_ok1_bulls$REL_EBV < 0.7,]
dEBV_ok1_cows_g2 <- dEBV_ok1_cows[dEBV_ok1_cows$REL_EBV > 0.5999 &
dEBV_ok1_cows$REL_EBV < 0.7,]
#Variance adjustment (VA)
sd_bulls <- tapply(dEBV_ok1_bulls_g2$dEBV_bulls, dEBV_ok1_bulls_g2$DIM, sd)
sd_cows <- tapply(dEBV_ok1_cows_g2$dEBV_cows, dEBV_ok1_cows_g2$DIM, sd)
VA_g2 <- sd_bulls/sd_cows
VA_g2_ok <- cbind(as.matrix(VA_g2), as.matrix(5:305))
colnames(VA_g2_ok) <- c("VA", "DIM")
#Mean adjustment (MA)
dEBV_ok1_bulls_VA_g2 <- dEBV_ok1_bulls_g2
adj_cows_g2 <- merge(dEBV_ok1_cows_g2, VA_g2_ok, by.x="DIM", by.y="DIM")
adj_cows_g2_1 <- adj_cows_g2$dEBV_cows*adj_cows_g2$VA
dEBV_ok1_cows_VA_g2 <- cbind(adj_cows_g2, adj_cows_g2_1)
MA_g2_mean_cows <- tapply(dEBV_ok1_cows_VA_g2$adj_cows_g2_1,
dEBV_ok1_cows_VA_g2$DIM, mean)
MA_g2_mean_bulls <- tapply(dEBV_ok1_bulls_VA_g2$dEBV_bulls, dEBV_ok1_bulls_VA_g2$DIM,
mean)
MA_g2 <- MA_g2_mean_cows - MA_g2_mean_bulls
MA_g2_ok <- cbind(as.matrix(MA_g2), as.matrix(5:305))
colnames(MA_g2_ok) <- c("MA", "DIM")
#Cluster 3: from 0.7 to 0.8
dEBV_ok1_bulls_g3 <- dEBV_ok1_bulls[dEBV_ok1_bulls$REL_EBV > 0.6999 &
dEBV_ok1_bulls$REL_EBV < 0.8,]
dEBV_ok1_cows_g3 <- dEBV_ok1_cows[dEBV_ok1_cows$REL_EBV > 0.6999 &
dEBV_ok1_cows$REL_EBV < 0.8,]
#Variance adjustment (VA)
sd_bulls <- tapply(dEBV_ok1_bulls_g3$dEBV_bulls, dEBV_ok1_bulls_g3$DIM, sd)
sd_cows <- tapply(dEBV_ok1_cows_g3$dEBV_cows, dEBV_ok1_cows_g3$DIM, sd)
VA_g3 <- sd_bulls/sd_cows
VA_g3_ok <- cbind(as.matrix(VA_g3), as.matrix(5:305))
colnames(VA_g3_ok) <- c("VA", "DIM")
#Mean adjustment (MA)
dEBV_ok1_bulls_VA_g3 <- dEBV_ok1_bulls_g3
adj_cows_g3 <- merge(dEBV_ok1_cows_g3, VA_g3_ok, by.x="DIM", by.y="DIM")
adj_cows_g3_1 <- adj_cows_g3$dEBV_cows*adj_cows_g3$VA
dEBV_ok1_cows_VA_g3 <- cbind(adj_cows_g3, adj_cows_g3_1)
MA_g3_mean_cows <- tapply(dEBV_ok1_cows_VA_g3$adj_cows_g3_1,
dEBV_ok1_cows_VA_g3$DIM, mean)
```

```
MA_g3_mean_bulls <- tapply(dEBV_ok1_bulls_VA_g3$dEBV_bulls, dEBV_ok1_bulls_VA_g3$DIM,
mean)
MA_g3 <- MA_g3_mean_cows - MA_g3_mean_bulls
MA_g3_ok <- cbind(as.matrix(MA_g3), as.matrix(5:305))
colnames(MA_g3_ok) <- c("MA", "DIM")
#Cluster 4: from 0.8 to 0.9
dEBV_ok1_bulls_g4 <- dEBV_ok1_bulls[dEBV_ok1_bulls$REL_EBV > 0.7999 &
dEBV_ok1_bulls$REL_EBV < 1,]
dEBV_ok1_cows_g4 <- dEBV_ok1_cows[dEBV_ok1_cows$REL_EBV > 0.7999 &
dEBV_ok1_cows$REL_EBV < 1,]
#Variance adjustment (VA)
sd_bulls <- tapply(dEBV_ok1_bulls_g4$dEBV_bulls, dEBV_ok1_bulls_g4$DIM, sd)
sd_cows <- tapply(dEBV_ok1_cows_g4$dEBV_cows, dEBV_ok1_cows_g4$DIM, sd)
VA_g4 <- sd_bulls/sd_cows
VA_g4_ok <- cbind(as.matrix(VA_g4), as.matrix(5:305))
colnames(VA_g4_ok) <- c("VA", "DIM")
#Mean adjustment (MA)
dEBV_ok1_bulls_VA_g4 <- dEBV_ok1_bulls_g4
adj_cows_g4 <- merge(dEBV_ok1_cows_g4, VA_g4_ok, by.x="DIM", by.y="DIM")
adj_cows_g4_1 <- adj_cows_g4$dEBV_cows*adj_cows_g4$VA
dEBV_ok1_cows_VA_g4 <- cbind(adj_cows_g4, adj_cows_g4_1)
MA_g4_mean_cows <- tapply(dEBV_ok1_cows_VA_g4$adj_cows_g4_1,
dEBV_ok1_cows_VA_g4$DIM, mean)
MA_g4_mean_bulls <- tapply(dEBV_ok1_bulls_VA_g4$dEBV_bulls, dEBV_ok1_bulls_VA_g4$DIM,
mean)
MA_g4 <- MA_g4_mean_cows - MA_g4_mean_bulls
MA_g4_ok <- cbind(as.matrix(MA_g4), as.matrix(5:305))
colnames(MA_g4_ok) <- c("MA", "DIM")
#Cluster 5: from 0.9 to 1.0
dEBV_ok1_bulls_g5 <- dEBV_ok1_bulls[dEBV_ok1_bulls$REL_EBV > 0.8999 &
dEBV_ok1_bulls$REL_EBV <= 1,]
dEBV_ok1_cows_g5 <- dEBV_ok1_cows[dEBV_ok1_cows$REL_EBV > 0.8999 &
dEBV_ok1_cows$REL_EBV <= 1,]
#Variance adjustment (VA)
sd_bulls <- tapply(dEBV_ok1_bulls_g5$dEBV_bulls, dEBV_ok1_bulls_g5$DIM, sd)
sd_cows <- tapply(dEBV_ok1_cows_g5$dEBV_cows, dEBV_ok1_cows_g5$DIM, sd)
VA_g5 <- sd_bulls/sd_cows
VA_g5_ok <- cbind(as.matrix(VA_g5), as.matrix(5:305))
colnames(VA_g5_ok) <- c("VA", "DIM")
#Mean adjustment (MA)
dEBV_ok1_bulls_VA_g5 <- dEBV_ok1_bulls_g5
adj_cows_g5 <- merge(dEBV_ok1_cows_g5, VA_g5_ok, by.x="DIM", by.y="DIM")
adj_cows_g5_1 <- adj_cows_g5$dEBV_cows*adj_cows_g5$VA
dEBV_ok1_cows_VA_g5 <- cbind(adj_cows_g5, adj_cows_g5_1)
MA_g5_mean_cows <- tapply(dEBV_ok1_cows_VA_g5$adj_cows_g5_1,
dEBV_ok1_cows_VA_g5$DIM, mean)
MA_g5_mean_bulls <- tapply(dEBV_ok1_bulls_VA_g5$dEBV_bulls, dEBV_ok1_bulls_VA_g5$DIM,
mean)
```

```
MA_g5 <- MA_g5_mean_cows - MA_g5_mean_bulls
MA_g5_ok <- cbind(as.matrix(MA_g5), as.matrix(5:305))
colnames(MA_g5_ok) <- c("MA", "DIM")

#Final adjustment:
MA_completo <- as.data.frame(rbind(as.matrix(MA_g1_ok),
as.matrix(MA_g2_ok),as.matrix(MA_g3_ok), as.matrix(MA_g4_ok), as.matrix(MA_g5_ok)))
VA_completo <- as.data.frame(rbind(as.matrix(VA_g1_ok),
as.matrix(VA_g2_ok),as.matrix(VA_g3_ok), as.matrix(VA_g4_ok), as.matrix(VA_g5_ok)))
MA <- tapply(MA_completo$MA, MA_completo$DIM, mean)
VA <- tapply(VA_completo$VA, VA_completo$DIM, mean)
VA_ok <- cbind(as.matrix(VA), as.matrix(5:305))
colnames(VA_ok) <- c("VA", "DIM")
MA_ok <- cbind(as.matrix(MA), as.matrix(5:305))
colnames(MA_ok) <- c("MA", "DIM") #PS: VA and MA are populational parameters. The dataset used
as example here does not fit the reality. For our article, the mean VA and MA parameters of the WG
method ranged over days in milk from 0.94 to 0.96 and from 0.29 to 0.36 for MY; from 0.93 to 1.04
and from 0.02 to 0.03 for FY; from 0.96 to 0.99 and from 0.01 to 0.02 for PY; and from 0.94 to 1.10
and from -0.04 to -0.02 for SCS, respectively.

#Making adjusted PA and adjusted dEBV:
#Bulls:
dEBV_ok1_bulls_ok <- merge(dEBV_ok1_bulls, VA_ok, by.x="DIM", by.y="DIM")
dEBV_ok1_bulls_ok1 <- merge(dEBV_ok1_bulls_ok, MA_ok, by.x="DIM", by.y="DIM")
PAaj_bulls_moms <- dEBV_ok1_bulls_ok1$VA*((dEBV_ok1_bulls_ok1$EBV_mom/2)-
dEBV_ok1_bulls_ok1$MA)
PAaj_bulls <- ((dEBV_ok1_bulls_ok1$EBV_dad/2) + PAaj_bulls_moms)
dEBVaj_bulls <- dEBV_ok1_bulls_ok1$dEBV_bulls + PAaj_bulls
dEBVaj_bulls_ok <- cbind(dEBV_ok1_bulls_ok1, PAaj_bulls, dEBVaj_bulls)
colnames(dEBVaj_bulls_ok) <- c("DIM", "ID", "sex", "EBV", "EBV_dad", "EBV_mom", "PA", "REL_EBV",
"REL_PA", "dEBV", "R", "ERCind", "VA", "MA", "PAadj", "dEBVadj")
#Cows:
dEBV_ok1_cows_ok <- merge(dEBV_ok1_cows, VA_ok, by.x="DIM", by.y="DIM")
dEBV_ok1_cows_ok1 <- merge(dEBV_ok1_cows_ok, MA_ok, by.x="DIM", by.y="DIM")
PAaj_cows_moms <- dEBV_ok1_cows_ok1$VA*((dEBV_ok1_cows_ok1$EBV_mom/2)-
dEBV_ok1_cows_ok1$MA)
PAaj_cows <- ((dEBV_ok1_cows_ok1$EBV_dad/2) + PAaj_cows_moms)
dEBVaj_cows <- (dEBV_ok1_cows_ok1$VA*(dEBV_ok1_cows_ok1$dEBV_cows -
dEBV_ok1_cows_ok1$MA)) + PAaj_cows
dEBVaj_cows_ok <- cbind(dEBV_ok1_cows_ok1, PAaj_cows, dEBVaj_cows)
colnames(dEBVaj_cows_ok) <- c("DIM", "ID", "sex", "EBV", "EBV_dad", "EBV_mom", "PA",
"REL_EBV", "REL_PA", "dEBV", "R", "ERCind", "VA", "MA", "PAadj", "dEBVadj")
dEBV_all <- rbind(dEBVaj_bulls_ok, dEBVaj_cows_ok)
#Making adjusted EBV (EBVadj):
EBVadj <- ((1-dEBV_all$R)*dEBV_all$PA) + (dEBV_all$R*dEBV_all$dEBVadj)
#Final file
final <- cbind(dEBV_all, EBVadj)
final_ok <- final[order(final$ID, final$DIM),]
```

4.9.2. Exemple of the data file

```
ID sex DIM EBV EBV_dad EBV_mom PA REL_EBV REL_PA
1 F 5 5.142256009 5.173521082 5.141187673 5.157354377 0.58002 0.3332525
1 F 6 5.195831028 5.220897314 5.211923082 5.216410198 0.58002 0.3332525
1 F 7 5.247994738 5.26705428 5.280881063 5.273967672 0.58002 0.3332525
1 F 8 5.298764775 5.312007441 5.348083844 5.330045642 0.58002 0.3332525
1 F 9 5.348158677 5.355772169 5.413553524 5.384662847 0.58002 0.3332525
1 F 10 5.396193879 5.39836375 5.477312075 5.437837912 0.58002 0.3332525
1 F 11 5.442887718 5.439797379 5.539381341 5.48958936 0.58002 0.3332525
1 F 12 5.488257428 5.480088164 5.599783039 5.539935601 0.58002 0.3332525
1 F 13 5.532320146 5.519251125 5.658538756 5.588894941 0.58002 0.3332525
1 F 14 5.575092905 5.557301194 5.715669955 5.636485574 0.58002 0.3332525
1 F 15 5.616592641 5.594253213 5.771197968 5.68272559 0.58002 0.3332525
1 F 16 5.656836186 5.630121938 5.825144001 5.727632969 0.58002 0.3332525
1 F 17 5.695840276 5.664922036 5.877529131 5.771225584 0.58002 0.3332525
1 F 18 5.733621544 5.698668084 5.92837431 5.813521197 0.58002 0.3332525
1 F 19 5.770196521 5.731374574 5.977700358 5.854537466 0.58002 0.3332525
1 F 20 5.805581642 5.763055907 6.025527972 5.894291939 0.58002 0.3332525
2 M 5 7.839837075 7.003523998 7.772748138 7.388136068 0.65292 0.3520325
2 M 6 7.941161003 7.07087631 7.876011565 7.473443937 0.65292 0.3520325
2 M 7 8.040358947 7.1367259 7.976962381 7.556844141 0.65292 0.3520325
2 M 8 8.13745678 7.201091743 8.075629213 7.638360478 0.65292 0.3520325
2 M 9 8.23248023 7.263992699 8.172040527 7.718016613 0.65292 0.3520325
2 M 10 8.325454873 7.325447517 8.266224622 7.795836069 0.65292 0.3520325
2 M 11 8.416406143 7.385474835 8.358209634 7.871842234 0.65292 0.3520325
2 M 12 8.50535932 7.444093177 8.448023536 7.946058357 0.65292 0.3520325
2 M 13 8.592339541 7.501320959 8.535694137 8.018507548 0.65292 0.3520325
2 M 14 8.677371793 7.557176483 8.621249081 8.089212782 0.65292 0.3520325
3 M 200 9.625363507 8.275121696 7.523205407 7.899163551 0.95617 0.34936
3 M 201 9.624051972 8.272367971 7.521611126 7.896989548 0.95617 0.34936
3 M 202 9.622874281 8.269698093 7.520169857 7.894933975 0.95617 0.34936
3 M 203 9.621826934 8.267109251 7.518877894 7.892993572 0.95617 0.34936
3 M 204 9.620906282 8.26459852 7.517731383 7.891164951 0.95617 0.34936
3 M 205 9.620108523 8.262162865 7.516726322 7.889444594 0.95617 0.34936
3 M 206 9.619429705 8.259799141 7.515858561 7.887828851 0.95617 0.34936
3 M 207 9.618865724 8.257504087 7.515123802 7.886313944 0.95617 0.34936
3 M 208 9.618412326 8.255274334 7.514517596 7.884895965 0.95617 0.34936
3 M 209 9.618065104 8.2531064 7.51403535 7.883570875 0.95617 0.34936
3 M 210 9.617819502 8.25099669 7.513672321 7.882334505 0.95617 0.34936
4 M 100 8.279484731 9.330342472 8.436378545 8.883360509 0.78252 0.3528325
4 M 101 8.262550786 9.311177996 8.426215338 8.868696667 0.78252 0.3528325
4 M 102 8.245397746 9.291725173 8.415827297 8.853776235 0.78252 0.3528325
4 M 103 8.228038248 9.271998124 8.405225536 8.83861183 0.78252 0.3528325
4 M 104 8.210484758 9.252010792 8.394421023 8.823215908 0.78252 0.3528325
4 M 105 8.192749576 9.231776939 8.383424575 8.807600757 0.78252 0.3528325
4 M 106 8.174844831 9.211310146 8.372246862 8.791778504 0.78252 0.3528325
4 M 107 8.156782483 9.190623814 8.360898406 8.77576111 0.78252 0.3528325
4 M 108 8.138574322 9.169731163 8.349389579 8.759560371 0.78252 0.3528325
4 M 109 8.120231969 9.148645234 8.337730608 8.743187921 0.78252 0.3528325
```

Table S1. Descriptive statistics of validation population in the combined dataset (COMB), and in the dataset containing only bulls (BULLS) and only cows (COWS), considering minimum reliability of estimated breeding value ($\mathrm{REL}_{\mathrm{EBV}}$) equal to average of reliability of parent average, and considering minimum $\mathrm{REL}_{\mathrm{EBV}}$ of 0.50.

Scenario	Traits	N	Mean (SD)				
			$\mathrm{REL}_{\mathrm{EBV}}$	EBV	$\mathrm{dEBV}_{\mathrm{VR}}$	$\mathrm{dEBV}_{\mathrm{WG}}$	$\mathrm{dEBV}_{\mathrm{GR}}$
				COMB			
Average $\mathrm{REL}_{\mathrm{PA}}$	**MY**	552	0.65 (0.07)	8.42 (1.53)	8.31 (2.26)	7.50 (2.17)	8.27 (2.96)
	FY	552	0.61 (0.07)	0.40 (0.06)	0.40 (0.09)	0.35 (0.08)	0.40 (0.13)
	PY	552	0.60 (0.07)	0.28 (0.06)	0.27 (0.08)	0.23 (0.08)	0.027 (0.11)
	SCS	552	0.46 (0.05)	-1.15 (0.27)	-1.13 (0.38)	-1.09 (0.39)	-1.11 (0.75)
0.50	**MY**	515	0.66 (0.05)	8.42 (1.53)	8.39 (1.93)	7.57 (1.86)	8.40 (2.30)
	FY	503	0.63 (0.05)	0.40 (0.06)	0.40 (0.08)	0.35 (0.08)	0.39 (0.09)
	PY	503	0.62 (0.05)	0.28 (0.06)	0.27 (0.07)	0.24 (0.07)	0.27 (0.09)
	SCS	67	0.53 (0.06)	-1.14 (0.26)	-1.11 (0.33)	-1.07 (0.34)	-1.08 (0.53)
				BULLS			
Average $\mathrm{REL}_{\mathrm{PA}}$	**MY**	32	0.62 (0.16)	8.63 (1.47)	8.10 (2.37)	7.57 (2.35)	8.20 (3.09)
	FY	32	0.59 (0.16)	0.41 (0.06)	0.40 (0.09)	0.37 (0.09)	0.39 (0.13)
	PY	32	0.59 (0.16)	0.28 (0.07)	0.27 (0.09)	0.25 (0.09)	0.26 (0.12)
	SCS	32	0.49 (0.13)	-1.18 (0.24)	-1.15 (0.47)	-1.14 (0.47)	-1.14 (1.02)
0.50	**MY**	21	0.70 (0.14)	8.79 (1.50)	8.74 (1.86)	8.22 (1.83)	8.76 (2.21)
	FY	17	0.71 (0.13)	0.43 (0.06)	0.43 (0.07)	0.40 (0.07)	0.37 (0.08)
	PY	17	0.71 (0.13)	0.30 (0.06)	0.30 (0.07)	0.28 (0.07)	0.30 (0.08)
	SCS	11	0.65 (0.09)	-1.18 (0.25)	-1.20 (0.28)	-1.18 (0.28)	-1.23 (0.37)
				COWS			
Average $\mathrm{REL}_{\mathrm{PA}}$	**MY**	520	0.65 (0.06)	8.41 (1.53)	8.32 (2.26)	7.49 (2.16)	8.30 (2.95)
	FY	520	0.61 (0.06)	0.40 (0.06)	0.40 (0.09)	0.34 (0.08)	0.40 (0.13)
	PY	520	0.60 (0.06)	0.28 (0.06)	0.27 (0.08)	0.23 (0.08)	0.27 (0.11)
	SCS	520	0.45 (0.04)	-1.15 (0.27)	-1.12 (0.38)	-1.08 (0.39)	-1.11 (0.74)
0.50	**MY**	494	0.66 (0.04)	8.41 (1.53)	8.38 (1.93)	7.55 (1.85)	8.38 (2.30)
	FY	486	0.63 (0.04)	0.40 (0.06)	0.40 (0.08)	0.35 (0.08)	0.42 (0.10)
	PY	486	0.61 (0.04)	0.28 (0.06)	0.27 (0.07)	0.24 (0.07)	0.27 (0.09)
	SCS	56	0.51 (0.01)	-1.13 (0.26)	-1.09 (0.34)	-1.05 (0.35)	-1.05 (0.55)

Average $\mathrm{REL}_{\mathrm{PA}}$: minimum $\mathrm{REL}_{\mathrm{EBV}}$ equal to average of reliability of parent average. Average $\mathrm{REL}_{\mathrm{PA}}$ was estimated to be 0.34, 0.33, 0.33 and 0.27 for milk (MY), fat (FY), and protein (PY) yield, and somatic cell score (SCS), respectively. dEBV: deregressed EBV estimated by the VanRaden ($\mathrm{dEBV}_{\mathrm{VR}}$), Wiggans ($\mathrm{dEBV}_{\mathrm{WG}}$), and Garrick ($\mathrm{dEBV}_{\mathrm{GR}}$) methods. N: number of animals.

CHAPTER 5:

Accessing the impact of including information from bulls and their daughters in the training population of multiple-step genomic evaluations: a simulation study

Hinayah Rojas de Oliveira[1,2], Luiz Fernando Brito[2,3,] Mehdi Sargolzaei[2,4], Fabyano Fonseca e Silva[1], Janusz Jamrozik[2,5], Daniela Andressa Lino Lourenco[6] and Flávio Schramm Schenkel[2]

[1] Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil
[2] Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada
[3] Department of Animal Sciences, Purdue University, West Lafayette, Indiana, United States of America
[4] HiggsGene Solutions Inc., Guelph, Ontario, Canada
[5] Canadian Dairy Network, Guelph, Ontario, Canada
[6] Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, United States of America

Animal, 2018 (submitted).

5.1. Abstract

The objective of this study was to investigate the impact of accounting for parent average (PA) and genotyped daughters' average (GDA) in the estimation of deregressed Estimated Breeding Values (dEBVs) for use as pseudo-phenotypes in multiple-step genomic predictions. In addition, an alternative deregression method was proposed (NEW). A simulated dairy cattle dataset was used to compare eight genomic prediction scenarios defined based on the composition of the training population, i.e., number of bulls, bulls' daughters and cows not sired by the genotyped bulls. For all these scenarios, Genomic Estimated Breeding Values (GEBVs) were predicted using the Genomic Best Linear Unbiased Prediction and dEBVs estimated based on four methods: VR, that includes PA and GDA information in the dEBV; VRpa, that excludes PA information; and JA and NEW, which exclude PA and GDA information from the dEBVs using either all information available in the complete pedigree or only information from parents and genotyped daughters, respectively. The impact of accounting for PA and GDA in the dEBVs was evaluated by the Pearson correlation coefficient between dEBVs and True Breeding Values (TBVs) and the validation reliability and scaling bias of GEBVs based on TBVs. In general, dEBVs estimated by VR and NEW showed the highest and lowest correlation between dEBVs and TBVs, respectively, for the majority of genomic prediction scenarios. However, VR and NEW showed the lowest and highest GEBV reliabilities, respectively. Among all deregression methods, VR displayed the most consistent bias across different scenarios and yielded one of the least biased GEBV estimates. Our results indicated that removing PA and GDA information from dEBVs can increase the reliability of genomic predictions. The NEW method proposed here seems a feasible alternative to generate dEBVs to be used in multiple-step genomic evaluations, however, it will be necessary to re-scale GEBVs.

Keywords: daughter average, deregression, double-counting, GBLUP, parent average

5.2. Introduction

Genomic selection has driven modern cattle breeding and substantially contributed to fastener genetic progress for a variety of economically important traits by increasing the reliability of breeding values of young animals and, thus, reducing the generation interval (Meuwissen et al., 2016). Nowadays, the majority of dairy cattle breeding programs around the world have implemented genomic selection based on a multiple-step approach (Interbull, 2017). In brief, the prediction of Genomic Estimated Breeding Values (GEBVs) based on the multiple-step approach involves the following steps: (1) prediction of traditional Estimated Breeding Values (EBVs) for all animals included in the pedigree; (2) estimation of pseudo-phenotypes, such as deregressed EBVs (dEBVs); (3) prediction of Direct Genomic Values (DGVs) for the genotyped animals; and (4) blending of DGV and Parent Average (PA) or EBV to generate the GEBVs of young animals and old animals, respectively. Among the main factors that influence the reliability of GEBVs are the number of animals used in the training population (VanRaden et al., 2009), and the genetic relationship between the training population and the target animals (Wu et al., 2015; Uemoto et al., 2017).

In dairy cattle, the training populations are usually composed by progeny tested bulls, as these bulls have more reliable information of their genetic merit when compared to cows (Uemoto et al., 2017). However, for some dairy cattle breeds (or populations) the number of progeny tested bulls is limited, which may hamper the success of genomic selection. To increase the number of animals in the training population, cows have been genotyped and added to the training population (Koivula et al., 2016; Wiggans et al., 2011). In general, the majority of genotyped females are daughters of progeny tested bulls, which have EBVs for economically important production traits estimated based on the PA and on the phenotypic information from their daughters. Thus, if both bulls and their daughters are included in the training population of a multiple-step evaluation to estimate the marker effects, these markers are estimated using duplicate information, because the EBVs of a bull and their daughters are estimated based on the same phenotypes.

To avoid this double-counting of information, EBVs are usually deregressed prior to their use in genomic predictions (Garrick et al., 2009). Several methods for deregression of EBVs have been proposed (e.g., Garrick et al., 2009; VanRaden et al., 2009), however, the most used methods only remove the double-counting of information due to PA (Oliveira et al., 2018). Consequently, training populations used for multiple-step genomic evaluations in small size dairy cattle breeds frequently have a remaining double-counting of information, which may have a negative impact on genomic predictions. For instance, Uemoto et al. (2017) evaluated the effect of including related animals in the training population of Japanese Holstein and concluded that using unrelated animals in the training population would be preferable to attain lower prediction bias of GEBVs, which they suggest could be related to the noise in the deregression process. Nonetheless, to our knowledge, there are no studies in the literature investigating the need to account for PA and genotyped daughters' average (GDA; i.e., the contribution of genotyped daughters to the bull's EBV) information in the deregression methods when estimating pseudo-phenotypes to be used in genomic predictions when the training population includes information from both bulls and their daughters. The use of simulated data is an efficient way to make these comparisons, as the True Breeding Values (TBVs) are known and can be used to assess the results against the true values. Therefore, the objectives of this study were: 1) to investigate the impact of accounting for PA and GDA in the calculation of dEBVs used in multiple-step genomic evaluations, using a simulated dairy cattle population; and, 2) to propose a new deregression method that accounts for PA and GDA.

5.3. Material and methods

Data simulation

In order to investigate the need to account for PA and GDA when estimating dEBVs to be used in a multiple-step genomic evaluation, a data simulation was performed using the QMSim software (Sargolzaei and Schenkel, 2009). The simulation intended to mimic a small size dairy cattle population, because the possible double-counting of information due to the inclusion of bulls and their daughters in the training population is more likely to occur in genomic analyses where the number of progeny tested bulls is limited.

Population structure and phenotypes. In the first simulation step, a historical population was created to establish an initial level of linkage disequilibrium similar to that observed in dairy cattle populations. In the historical population, 1 400 unrelated individuals (700 males and 700 females) were mated randomly for 1 000 generations. During these 1 000 historical generations, the population size decreased from 1 400 to 400 individuals, with the same sex ratio, to mimic a "bottleneck" (population reduction). After the bottleneck, 40 generations were simulated to expand the historical population size up to 20 500 animals (500 males and 20 000 females). Subsequently, 100 males and 10 000 females with the highest breeding value estimates were selected from the last generation of the historical population and used as founders to simulate the recent population.

From this founder population, a total of 15 overlapping generations were simulated. Breeding animals were selected based on the EBVs estimated using an animal model and Best Linear Unbiased Prediction (BLUP), and the selected animals were randomly mated. From generation one to 15, the reproductive rate of females was set to one offspring per cow, which generated a pedigree file with 160 100 animals (10 100 founders plus 150 000 animals born from generation one to 15). The average sex ratio was fixed to 0.50, and the replacement ratio assumed for sires and dams was 0.60 and 0.20, respectively. Individuals from generation six to 14 were genotyped and used to create the training and validation populations used for further analyses. Similar strategies and population parameters have been reported in other simulation studies for dairy cattle (Jiménez-Montero et al., 2012; Plieschke et al., 2018).

Phenotypes were simulated to mimic an economically important trait in dairy cattle (e.g., milk yield) and, therefore, a single trait with heritability of 0.30 and phenotypic variance of 100 was assumed. In addition, it was considered to be a sex-limited trait (i.e., phenotypes were measured only on females). Total heritability due to the 1 979 simulated Quantitative Trait Loci (QTL) was assumed to be 0.27, i.e., 0.03 was assumed to be due to the remaining polygenic effect. To increase the precision of the simulation process, the analyses were replicated 10 times. Descriptive statistics for the simulated populations in the 10 replicates are shown in Supplementary Table S1.

Genotypes. In order to create a more realistic scenario, the simulated genome mimicked the real bovine genome. Thus, 29 autosomes with length identical to the bovine genome were defined based on the Bos_taurus_UMD_3.1.1 assembly, whose information was available at the National Center for Biotechnology Information website (NCBI, https://www.ncbi.nlm.nih.gov/genome/?term=cow). The number of Single Nucleotide Polymorphism (SNP) markers in each chromosome was defined according to the Illumina BovineSNP50v2 BeadChip (50K, Illumina Inc., San Diego, CA), which contained a total of 57 024 SNPs, as reported in the SNPchiMp v.3 database (http://bioinformatics.tecnoparco.org/SNPchimp/index.php/download/download-cow-data). The total number of simulated QTLs was 1 979, and they were distributed across chromosomes according to the Animal QTL Database (QTLdb, https://www.animalgenome.org/cgi-bin/QTLdb/BT/traitmap?trait_ID=1044) considering milk yield as the analyzed trait. Positions for SNP markers and QTLs in the simulation were considered evenly and randomly distributed in the genome, respectively.

It was assumed that all loci had the same number of alleles (two alleles) in the first historical generation, with equal marker allele frequencies. The number of QTL alleles in the first historical generation was assumed random and were sampled from a Uniform distribution considering two, three and four QTL alleles with equal frequency. Considering samples taken from a Uniform distribution in the first historical population allowed to account for the polygenic nature of the production traits, because the range of the QTL effects is limited in this case. The additive allelic effects were randomly sampled from a Gamma distribution with shape parameter equal to 0.4 (Meuwissen et al., 2001). The SNP markers and QTL positions were randomized within each chromosome in each one of the 10 replicates.

The rate of missing marker genotypes was defined as 0.01, and the rate of marker genotyping error was defined as 0.005. A recurrent mutation rate of 0.0001 for both SNP markers and QTLs was considered to establish mutation-drift equilibrium in the historical generations (Brito et al., 2011). Parameters used in the simulation process are summarized in Supplementary Tables S2 and S3.

Genomic prediction scenarios

The genotyped animals were split in two groups in order to create the training and validation populations used in the genomic predictions. In summary, genotypes and pseudo-phenotypes (dEBVs) were available only for animals in the training population, and they were used to estimate the SNP marker effects. For the validation population, only the genotypic information was included in the analyses and their GEBVs were predicted based on the marker effects estimated in the training population. To quantify the effect of removing the PA and GDA information from dEBV used in multiple-step genomic evaluations, different scenarios were created based on the composition of the training populations.

Training populations. Genotyped animals born from generation six to 13 were selected to create different training populations. The method used to select animals was: a) 100 bulls born from generation six to 12 and that had more than 70 daughters born from generation seven to 13 were randomly selected; b) for each one of these 100 bulls, 70 daughters were also randomly selected; c) other additional 300 bulls, born from generation six to 13, were randomly selected; d) 7 600 cows born from generation six to 13 that were non-daughters of the 400 selected bulls, were randomly selected. The average number of animals in the 10 replicates from where the selected animals were sampled from was 80 000, in which 39 921 were cows, daughters of 520 bulls. From these bulls, an average of 144 bulls were born from generation six to 12 and had more than 70 daughters in the training population.

From the selected animals, eight different scenarios (SCEN) were created as follows: SCEN1) only the 400 selected bulls were included in the training population; SCEN2) the 300 selected bulls with no daughters in the training population, the 100 selected bulls with daughters, and 10 daughters per bull (randomly selected from the 70 genotyped daughters), were included in the training population; SCEN3) the 300 selected bulls with no daughters, the 100 selected bulls with daughters, and the 70 daughters per bull were included in the training population; SCEN4) the 300 selected bulls with no daughters in the training population, the 100 selected bulls with daughters in the training population, and a different number of daughters per bull (that ranged from 10 to 70 daughters, randomly selected from the 70 genotyped daughters per bull), were included in the training population. To avoid the confounding effect between the inclusion of bulls and their genotyped daughters in the training population and the gain in GEBV reliability due to the increase in the population size, four extra scenarios were created based on the previous ones. These extra scenarios included cows non-daughters of training population bulls (non-daughters), in order to keep the same number of animals in all training populations (n = 8 000). The extra four scenarios can be described as: SCEN5) animals from SCEN1 plus 7 600 non-daughters; SCEN6) animals from SCEN2 plus 6 600 non-daughters; SCEN7) animals from SCEN3 plus 600 non-daughters; and, SCEN8) animals from SCEN4 plus 3 600 non-daughters. The non-daughters were randomly selected from the 7 600 cows non-daughters of bulls included in the training population used in SCEN5.

As a total of 10 replicates were performed in each scenario, it was required that in all replicates the sum of the randomly selected genotyped daughters included in SCEN4 and SCEN8 was equal to 4 000. Mean (± SD) of daughters in SCEN4 and SCEN8 was 40 (± 18.45) daughters, in all 10 replicates. A summary of the number of animals in the training population of each scenario is shown in Table 1.

Table 1. Number of animals included in the training population in each scenario.

Scenario	Number of bulls with no daughters	Number of bulls with daughters	Number of daughters per bull	Total number of daughters	Number of non daughters[1]	Total number of animals
SCEN1	400	0	0	0	0	400
SCEN2	300	100	10	1 000	0	1 400
SCEN3	300	100	70	7 000	0	7 400
SCEN4	300	100	10 to 70	4 000	0	4 400
SCEN5	400	0	0	0	7 600	8 000
SCEN6	300	100	10	1 000	6 600	8 000
SCEN7	300	100	70	7 000	600	8 000
SCEN8	300	100	10 to 70	4 000	3 600	8 000

SCEN = Scenario. [1] Cows non-daughters of training population bulls.

Validation population. A total of 60 genotyped bulls born in the generation 14, that had daughters in generation 15 and EBV reliability greater than 0.70, were used as validation population. The same validation population was used in all scenarios of each replicate.

Deregression methods

To evaluate the need to account for PA and GDA information to estimate dEBVs to be used as pseudo-phenotypes in genomic evaluations, when the training population includes information from both bulls and their daughters, different deregression methods were contrasted within each scenario. In summary, the deregression methods assessed this study included situations in which: 1) PA and GDA contributions were not removed; 2) PA contribution was removed, but GDA contribution was retained; and 3) PA and GDA contributions were removed from the dEBV.

Situation 1: PA and GDA contributions were not removed from the dEBV. To describe this situation the deregression method proposed by VanRaden et al. (2009) was used. Thus, the dEBV for each animal i ($\mathrm{dEBV_i}$) was estimated as:

$$\mathrm{dEBV_i} = \mathrm{PA_i} + \frac{(\mathrm{EBV_i} - \mathrm{PA_i})}{\mathrm{R_i}}, \qquad [1]$$

where $\mathrm{PA_i}$, $\mathrm{EBV_i}$ and $\mathrm{R_i}$ are the PA, the EBV and the deregression factor for the animal i, respectively. The $\mathrm{R_i}$ was computed as follows:

$$\mathrm{R_i} = \frac{\mathrm{ERC_{(Ind\text{-}PA)_i}}}{\mathrm{ERC_{(Ind\text{-}PA)_i}} + \mathrm{ERC_{(PA)_i}} + 1}, \qquad [2]$$

where effective record contributions (ERC) for the animal i excluding PA information $(\mathrm{ERC_{(Ind\text{-}PA)_i}})$, and for the PA $(\mathrm{ERC_{(PA)_i}})$ of the animal i were estimated based on VanRaden and Wiggans (1991), i.e.:

$$\mathrm{ERC_{(Ind\text{-}PA)_i}} = \left(\lambda \frac{\mathrm{REL_{EBV_i}}}{1 - \mathrm{REL_{EBV_i}}}\right) - \mathrm{ERC_{PA_i}} \text{ and,} \qquad [3]$$

$$\mathrm{ERC_{PA_i}} = \lambda \frac{\mathrm{REL_{PA_i}}}{1 - \mathrm{REL_{PA_i}}}, \qquad [4]$$

in which $\mathrm{REL_{EBV_i}}$ and $\mathrm{REL_{PA_i}}$ are the reliabilities for the EBV and PA of the animal i, respectively. The $\mathrm{REL_{EBV_i}}$ were calculated using the prediction error (co)variance matrix (**PEV**) estimated directly using the coefficient matrix ($\mathbf{C}^{22}$) extracted from the inverse of the left-hand-side of the mixed model equations (MME); and the $\mathrm{REL}_{\mathrm{PA}_i}$ were calculated as $\mathrm{REL}_{\mathrm{PA}_i} = \frac{\mathrm{REL}_{\mathrm{EBV}_{\mathrm{sire}}} + \mathrm{REL}_{\mathrm{EBV}_{\mathrm{dam}}}}{4}$. In all deregression methods λ was assumed as $\lambda = \frac{1-h^2}{h^2}$, in which a heritability (h^2) estimate of 0.30 was considered.

Situation 2: PA contribution was removed from the dEBV, but GDA was retained. The deregression method used here was similar to the previous one, with the exception that the PA component was removed from the dEBV (VanRaden et al., 2009):

$$\mathrm{dEBV}_i = \frac{(\mathrm{EBV}_i - \mathrm{PA}_i)}{R_i}, \qquad [5]$$

where PA_i, EBV_i and R_i were already described in Situation 1.

Situation 3: PA and GDA were removed from the dEBV. To investigate this situation two different deregression methods were compared. The first one was based on the deregression method proposed by Jansen (2008), in which an iterative algorithm is used to estimate the ERC and the dEBV for each animal, using information available in the complete pedigree (Situation 3a). The second one is a deregression method proposed in this study, which is an extension of the deregression method proposed by VanRaden et al. (2009) that allows to account for the information of genotyped daughters included in the training population in a simpler way (Situation 3b).

Situation 3a. The deregression method used here was developed by Jansen (2008). In summary, this deregression method aims to approximate the MME in order to compute dEBVs free of contribution of other animals as follows:

$$(\mathbf{D} + \mathbf{A}^{-1}\lambda)\{\mathbf{EBV}\} = 2\mathbf{Dy}, \qquad [6]$$

where **D** is a diagonal matrix, in which each diagonal element represents the ERC from one animal; $\mathbf{A}^{-1}$ is the inverse of the additive relationship matrix for animals in the pedigree; **EBV** is the vector of EBVs; and **y** is the vector of dEBVs. A different $\mathbf{A}^{-1}$ was created for each scenario previously described, since it was based on animals included in the training scenario evaluated plus their parents and relatives from up to 3 generations back. The λ was previously described in Situation 1.

Since **D** and **y** are unknown, the diagonal values of **D** are estimated iteratively given the known EBV reliabilities and the relationship between animals. In brief, the iterative algorithm can be described as:

$$\mathrm{REL}_{\mathrm{EBV}_i} = 1 - \frac{\lambda d^{ii}}{a_{ii}}, \qquad [7]$$

in which d^{ii} is the appropriate diagonal element of $(\mathbf{D} + \mathbf{A}^{-1}\lambda)^{-1}$, and a_{ii} is the corresponding diagonal element of **A**. Once the **D** have been solved, computation of **y** is straightforward using [6].

Situation 3b. The deregression method proposed here aimed to keep the simplicity of the method proposed by VanRaden et al. (2009) and extend its use for situations where bulls and their daughters are included in the training population. In this method, dEBVs from cows and bulls without genotyped daughters are computed as shown in [5]. However, dEBVs from bulls with genotyped daughters are adjusted to remove the information from these daughters. The dEBV from bulls with genotyped daughters can be calculated as:

$$\mathrm{dEBV}_{\mathrm{sire}_i} = \frac{\mathrm{EBV}_{\mathrm{sire}_i} - (\mathrm{PA}_{\mathrm{sire}_i} + \mathrm{GDA}_{\mathrm{sire}_i})}{\mathrm{R}_{\mathrm{sire}_i}}, \qquad [8]$$

where $\mathrm{dEBV}_{\mathrm{sire}_i}, \mathrm{EBV}_{\mathrm{sire}_i}, \mathrm{PA}_{\mathrm{sire}_i}, \mathrm{GDA}_{\mathrm{sire}_i}$ and $\mathrm{R}_{\mathrm{sire}_i}$ are the dEBV, EBV, PA, GDA and R for the bull i that has n genotyped daughters, respectively. The $\mathrm{GDA}_{\mathrm{sire}_i}$ was calculated as a weighted mean, in which the dEBVs from genotyped daughters of each bull i were weighted by their respective R. The dEBV and R from genotyped daughters used in the weighted mean were calculated as showed in the Situation 2 (i.e., in equations [5] and [2], respectively). The $\mathrm{R}_{\mathrm{sire}_i}$ was computed as follows:

$$\mathrm{R}_{\mathrm{sire}_i} = \frac{\mathrm{ERC}_{[\mathrm{Ind}-(\mathrm{PA}+\mathrm{GDA})]_{\mathrm{sire}_i}}}{\mathrm{ERC}_{[\mathrm{Ind}-(\mathrm{PA}+\mathrm{GDA})]_{\mathrm{sire}_i}} + \mathrm{ERC}_{(\mathrm{PA})_{\mathrm{sire}_i}} + \mathrm{ERC}_{(\mathrm{GDA})_{\mathrm{sire}_i}} + 1}, \qquad [9]$$

where $\mathrm{ERC}_{(\mathrm{PA})_{\mathrm{sire}_i}}$ is the $\mathrm{ERC}_{(\mathrm{PA})_i}$ for the bull i that has n genotyped daughters. The effective record contributions for the bulls i excluding PA and GDA information $(\mathrm{ERC}_{[\mathrm{Ind}-(\mathrm{PA}+\mathrm{GDA})]_{\mathrm{sire}_i}})$, and for the GDA $(\mathrm{ERC}_{(\mathrm{GDA})_{\mathrm{sire}_i}})$ were calculated as:

$$\mathrm{ERC}_{[\mathrm{Ind}\text{-}(\mathrm{PA}+\mathrm{GDA})]_{\mathrm{sire}_i}} = \left(\lambda \frac{\mathrm{REL}_{\mathrm{EBV}_{\mathrm{sire}_i}}}{1-\mathrm{REL}_{\mathrm{EBV}_{\mathrm{sire}_i}}} \right) - (\mathrm{ERC}_{(\mathrm{PA})_{\mathrm{sire}_i}} + \mathrm{ERC}_{(\mathrm{GDA})_{\mathrm{sire}_i}}) \quad \text{and,} \tag{10}$$

$$\mathrm{ERC}_{(\mathrm{GDA})_{\mathrm{sire}_i}} = \lambda \frac{\mathrm{REL}_{\mathrm{GDA}_{\mathrm{sire}_i}}}{1-\mathrm{REL}_{\mathrm{GDA}_{\mathrm{sire}_i}}} \ , \tag{11}$$

in which $\lambda = \frac{1-h^2}{h^2}$, and $\mathrm{REL}_{\mathrm{EBV}_{\mathrm{sire}_i}}$ is the $\mathrm{REL}_{\mathrm{EBV}_i}$ for the bull i. The $\mathrm{REL}_{\mathrm{GDA}_{\mathrm{sire}_i}}$ is the reliability for the GDA from bull i. Because bulls do not have own records for production traits in dairy cattle, we can assume that:

$$\mathrm{REL}_{\mathrm{GDA}_{\mathrm{sire}_i}} = \mathrm{REL}_{\mathrm{EBV}_{\mathrm{sire}_i}} - (\mathrm{REL}_{\mathrm{PA}_{\mathrm{sire}_i}} + \mathrm{REL}_{\mathrm{NGDA}_{\mathrm{sire}_i}}) \ , \tag{12}$$

in which $\mathrm{REL}_{\mathrm{PA}_{\mathrm{sire}_i}}$ is the $\mathrm{REL}_{\mathrm{PA}_i}$ for the bull i, and $\mathrm{REL}_{\mathrm{NGDA}_{\mathrm{sire}_i}}$ is the reliability for the non-genotyped daughters from the bull i. Considering that $\mathrm{REL}_{\mathrm{EBV}}$ of a bull, deriving from its daughters, is proportional to the number of daughters and their reliabilities, we can define that:

$$\mathrm{REL}_{\mathrm{GDA}_{\mathrm{sire}_i}} = \mu\, n_{\mathrm{GD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{GD}_{\mathrm{sire}_i}} \ \text{, and,} \tag{13}$$

$$\mathrm{REL}_{\mathrm{NGDA}_{\mathrm{sire}_i}} = \mu\, n_{\mathrm{NGD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{NGD}_{\mathrm{sire}_i}} \ , \tag{14}$$

i.e.:

$$\mathrm{REL}_{\mathrm{EBV}_{\mathrm{sire}_i}} - \mathrm{REL}_{\mathrm{PA}_{\mathrm{sire}_i}} = \mu(n_{\mathrm{GD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{GD}_{\mathrm{sire}_i}}) + (n_{\mathrm{NGD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{NGD}_{\mathrm{sire}_i}}) \ , \tag{15}$$

in which $n_{\mathrm{GD}_{\mathrm{sire}_i}}$ and $n_{\mathrm{NGD}_{\mathrm{sire}_i}}$ are the number of genotyped and non-genotyped daughters of the bull i, respectively; and $\bar{R}_{\mathrm{GD}_{\mathrm{sire}_i}}$ and $\bar{R}_{\mathrm{NGD}_{\mathrm{sire}_i}}$ are the average R calculated for the dEBVs from genotyped and non-genotyped daughters of the bull i, respectively. Due to the fact that genotyped and non-genotyped daughters contribute similarly to the EBV estimate of a bull, one can assume that:

$$\mathrm{REL}_{\mathrm{EBV}_{\mathrm{sire}_i}} - \mathrm{REL}_{\mathrm{PA}_{\mathrm{sire}_i}} = \left(k \times \frac{n_{\mathrm{GD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{GD}_{\mathrm{sire}_i}}}{n_{\mathrm{GD}_{\mathrm{sire}_i}} + n_{\mathrm{NGD}_{\mathrm{sire}_i}}} \right) + \left(k \times \frac{n_{\mathrm{NGD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{NGD}_{\mathrm{sire}_i}}}{n_{\mathrm{GD}_{\mathrm{sire}_i}} + n_{\mathrm{NGD}_{\mathrm{sire}_i}}} \right), \tag{16}$$

in which the only unknown term (k) can be calculated as:

$$k = \frac{(n_{\mathrm{GD}_{\mathrm{sire}_i}} + n_{\mathrm{NGD}_{\mathrm{sire}_i}}) \times (\mathrm{REL}_{\mathrm{EBV}_{\mathrm{sire}_i}} - \mathrm{REL}_{\mathrm{PA}_{\mathrm{sire}_i}})}{(n_{\mathrm{GD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{GD}_{\mathrm{sire}_i}}) + (n_{\mathrm{NGD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{NGD}_{\mathrm{sire}_i}})} \ , \tag{17}$$

and used to compute $\mathrm{REL}_{\mathrm{GDA}_{\mathrm{sire}_i}}$ as:

$$\mathrm{REL}_{\mathrm{GDA}_{\mathrm{sire}_i}} = k \times \left(\frac{n_{\mathrm{GD}_{\mathrm{sire}_i}} \times \bar{R}_{\mathrm{GD}_{\mathrm{sire}_i}}}{n_{\mathrm{GD}_{\mathrm{sire}_i}} + n_{\mathrm{NGD}_{\mathrm{sire}_i}}} \right). \qquad [18]$$

in which all terms were previously defined. Average of $n_{\mathrm{GD}_{\mathrm{sire}_i}}$, $n_{\mathrm{NGD}_{\mathrm{sire}_i}}$, $\bar{R}_{\mathrm{GD}_{\mathrm{sire}_i}}$, $\bar{R}_{\mathrm{NGD}_{\mathrm{sire}_i}}$, $\mathrm{REL}_{\mathrm{GDA}_{\mathrm{sire}_i}}$ and $\mathrm{REL}_{\mathrm{NGDA}_{\mathrm{sire}_i}}$ in the 10 replicates are shown in Supplementary Table S4.

In order to have the dEBVs estimated by all different methods in the same scale for subsequent comparison, average EBV was added to the dEBVs of each animal in the Situations 2, 3a and 3b. From now on, the deregression methods used in this study will be coded as VR, VRpa, JA and NEW, to describe the deregression methods defined in Situations 1, 2, 3a and 3b, respectively.

Genomic predictions

Genotype quality control. In all analyses, a quality control was performed to exclude possible incorrect SNP markers and/or non-informative SNP markers. Monomorphic SNPs, SNPs with Mendelian conflicts, call rate less than 0.95, minor allele frequency (MAF) less than 0.01 and with a difference between observed and expected heterozygosity frequency greater than 0.15 (Wiggans et al., 2009) were removed using the preGSF90 software (Misztal et al., 2014).

Statistical analyses. The genomic BLUP (GBLUP; VanRaden, 2008) was used to predict DGVs for animals in the training and validation populations. In matrix notation, the single-trait animal model used is described as follows:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}, \qquad [19]$$

where y is the vector of dEBVs; μ is the overall mean, **Z** is the incidence matrix for the random effects of the animals (**g**); and **e** is the vector of residuals. The model assumptions were:

$$E[\mathbf{y}] = \mu \text{ and } \mathrm{Var}\begin{bmatrix} \mathbf{g} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G}\sigma_g^2 & 0 \\ 0 & \mathbf{I}\sigma_e^2 \end{bmatrix}, \qquad [20]$$

where σ_g^2 and σ_e^2 are the genomic and residual variances for dEBVs, and **I** is an identity matrix. The genomic relationship matrix (**G**) was created as suggested by VanRaden (2008):

$$\mathbf{G} = 0.95 \frac{\mathbf{ZZ'}}{2\sum_{k=1}^{K} p_k(1-p_k)} + 0.05\mathbf{A}, \qquad [21]$$

where $\mathbf{Z} = \mathbf{M} - \mathbf{P}$, in which **M** contains the centered genotypes (i.e., -1, 0 and 1 to represent AA, Aa and aa, respectively), **P** contains allele frequency for SNP k (p_k) in its kth column, expressed as $2(p_k - 0.5)$; $2\sum_{k=1}^{K} p_k(1-p_k)$ is a scaling parameter. In order to guarantee that the **G** matrix was positive definite and accounted for the "unknown" polygenic effect, 5% of the traditional additive relationship matrix (**A**) was added to 95% of the **G** matrix. The polygenic effect assumed here was slightly different from the one simulated in order to mimic the reality, where the real polygenic effect of the trait is unknown. The AIREMLF90 software (Misztal et al., 2014) was used to obtain the variance components and the solutions of the MME in all analyses.

DGV theoretical reliabilities and GEBVs. Theoretical reliabilities of estimated DGVs for each animal i $(\mathrm{REL}_{\mathrm{DGV}_i})$ in the validation population was computed as:

$$\mathrm{REL}_{\mathrm{DGV}_i} = 1 - \left(\frac{\mathrm{PEV}_i}{\hat{\sigma}_g^2} \right), \qquad [22]$$

where PEV$_i$ is the prediction error variance for the animal i, and $\hat{\sigma}_g^2$ is the estimated genomic variance. The GEBV for each animal i (GEBV$_i$) in the validation population was calculated as:

$$\mathrm{GEBV}_i = \left(w_i \times \mathrm{PA}_i \right) + \left(\frac{\mathrm{REL}_{\mathrm{DGV}_i}}{\mathrm{REL}_{\mathrm{PA}_i} + \mathrm{REL}_{\mathrm{DGV}_i}} \times \mathrm{DGV}_i \right), \qquad [23]$$

in which $w_i = \frac{\mathrm{REL}_{\mathrm{PA}_i}}{\mathrm{REL}_{\mathrm{PA}_i} + \mathrm{REL}_{\mathrm{DGV}_i}}$ for the VR method, and $w_i = 1$ for the VRpa, JA and NEW methods. All the other terms were previously defined.

Evaluating the impact of accounting for PA and GDA in the calculation of dEBVs

The impact of accounting for PA and GDA in the calculation of dEBVs used in multiple-step genomic evaluations was evaluated based on both dEBVs and GEBVs.

Comparison of methods based on dEBVs. The different deregression methods were compared by contrasting means and standard deviations of TBVs, EBVs and dEBVs for all animals in the training population and solely for bulls with genotyped daughters. In order to assess dEBVs accuracy, Pearson correlation coefficients between TBVs and dEBVs were calculated. In addition, Pearson correlation coefficients between dEBVs estimated based on different methods were also calculated.

Comparison of methods based on GEBVs. The TBVs were used to estimate the reliability and scaling bias of genomic prediction in each scenario. Thus, validation reliability was calculated as the squared Pearson correlation coefficient between GEBVs and TBVs for individuals in the validation population. To evaluate the scaling bias of the genomic predictions, the regression coefficients (b) estimated using a linear regression of TBV on GEBV were used. For this propose, bias was defined as the deviation from 1, i.e., bias = b - 1. In order to report a single value of reliability and bias for each deregression method and scenario, average (± SD) of reliability and bias in the 10 replicates were calculated. Additionally, significant differences (p-value < 0.05) between reliabilities and biases from different deregression methods, within each analyzed scenario, were accessed using a paired t-test (Rosner, 1982).

5.4. Results

Comparing methods based on dEBVs

Similar overall means were observed within each scenario for TBVs, EBVs and dEBVs. Means estimated considering only bulls that had genotyped daughters were higher than means estimated considering all animals in the training population. As expected, there was a reduction in variability (i.e., decrease in the SDs) in the EBVs compared to TBVs. Considering all animals included in the training population, the deregression methods increased the variability of dEBVs (except in SCEN1, where only VR increased the variability of dEBVs compared to EBVs). Among all deregression methods, the JA presented less stable increase in dEBV variability across scenarios. When considering only bulls that had genotyped daughters, all deregression methods did not increase the EBV variability (Table 2).

Pearson correlation coefficients estimated between TBVs and dEBVs are shown in Table 3. In general, when the training population included only genotyped bulls (SCEN1), the correlation between dEBVs estimated by different deregression methods and between dEBVs and TBVs were high. However, the correlations decreased when non-daughters of genotyped bulls were included (SCEN5, for all animals). Among all deregression methods, dEBVs estimated by VR and JA had highest correlations with TBVs when non-genotyped (SCEN1 and SCEN5) or a small number of genotyped (SCEN2 and SCEN6) daughters where included. Although VR and JA had high correlation when a small number of genotyped daughters was added to the training population, including a large number of genotyped daughters (SCEN3 and SCEN7 when considering only bulls, and SCEN4 and SCEN8 for both all and only genotyped bulls) decreased the correlation between dEBVs calculated based on JA and TBVs, and between the JA and the other deregression methods. The NEW method presented the lowest correlation between dEBVs and TBVs when including a small number of genotyped daughters (SCEN2 and SCEN6), and the second smaller correlation coefficients when including a large number of genotyped daughters (SCEN3, SCEN4, SCEN7 and SCEN8).

Table 2. Mean and standard deviation (in brackets) of true breeding values (TBVs), breeding values (EBVs) and deregressed EBVs (dEBVs) in each scenario.

Animals	**Scenarios**	**TBV**	**EBV**	**dEBV**			
				VR	**VRpa**	**JA**	**NEW**
All	**SCEN1**	33.5 (7.85)	33.6 (7.60)	33.4 (8.04)	33.4 (5.81)	33.2 (5.88)	33.4 (5.81)
	SCEN2	30.6 (7.38)	30.6 (6.57)	30.3 (8.94)	30.3 (6.78)	28.0 (7.74)	30.3 (7.05)
	SCEN3	29.1 (6.85)	29.1 (5.78)	28.7 (9.00)	28.7 (6.92)	30.3 (19.54)	28.7 (6.97)
	SCEN4	29.7 (6.93)	29.6 (5.92)	29.2 (8.96)	29.3 (6.87)	27.1 (17.87)	29.3 (6.97)
	SCEN5	22.9 (7.16)	23.0 (6.12)	22.6 (8.79)	22.7 (6.68)	21.3 (6.96)	22.7 (6.68)
	SCEN6	23.8 (7.46)	23.8 (6.47)	23.5 (9.09)	23.5 (6.71)	22.0 (7.28)	23.5 (6.81)
	SCEN7	28.0 (7.05)	27.9 (5.97)	27.5 (9.20)	27.5 (7.01)	30.2 (21.65)	27.5 (7.08)
	SCEN8	26.4 (7.68)	26.4 (6.72)	26.0 (9.42)	26.0 (6.77)	24.0 (10.10)	26.0 (6.81)
Bulls	**SCEN1**	34.5 (5.41)	34.4 (5.19)	31.4 (4.70)	34.9 (2.36)	33.2 (2.80)	34.9 (2.36)
	SCEN2	34.5 (5.41)	34.4 (5.19)	31.2 (4.70)	33.0 (2.36)	33.6 (1.83)	37.3 (3.21)
	SCEN3	33.9 (5.42)	33.9 (5.18)	29.8 (4.56)	32.0 (2.35)	31.3 (2.41)	36.8 (2.38)
	SCEN4	34.5 (5.41)	34.4 (5.19)	31.2 (4.70)	32.4 (2.36)	58.1 (2.54)	37.2 (2.66)
	SCEN5	34.5 (5.41)	34.4 (5.19)	31.4 (4.70)	34.9 (2.36)	28.8 (2.59)	34.9 (2.36)
	SCEN6	34.5 (5.41)	34.4 (5.19)	30.5 (4.70)	26.6 (2.36)	29.2 (1.90)	32.4 (3.21)
	SCEN7	33.3 (5.32)	33.1 (5.09)	33.4 (5.17)	30.7 (2.38)	38.7 (2.22)	31.8 (2.37)
	SCEN8	34.4 (5.36)	34.4 (5.14)	30.4 (4.59)	29.2 (2.32)	40.3 (2.06)	32.4 (2.63)

All = all animals included in the training population; Bulls = only genotyped bulls that had genotyped daughters included in the training population; SCEN = scenario, as defined in Table 1.

[1] The dEBVs were estimated using the VanRaden et al. (2009) method, without removing and removing the parentage average (VR and VRpa, respectively), the Jansen (2008) method (JA), and the deregression method proposed in this study (NEW). Means and standard deviations showed here are from 10 replicates.

Table 3. Pearson correlation coefficients[1] between true (TBVs) and deregressed breeding values[2], in all analyzed scenarios.

		TBV	**VR**	**VRpa**	**JA**	**NEW**			**TBV**	**VR**	**VRpa**	**JA**	**NEW**
	TBV	1.00	0.96	0.74	0.93	0.74		**TBV**	1.00	0.73	0.32	0.77	0.32
	VR	0.85	1.00	0.78	0.96	0.78		**VR**	0.85	1.00	0.78	0.85	0.78
SCEN1	**VRpa**	0.65	0.59	1.00	0.86	1.00	**SCEN5**	**VRpa**	0.65	0.59	1.00	0.47	1.00
	JA	0.96	0.88	0.50	1.00	0.86		**JA**	0.95	0.86	0.43	1.00	0.47
	NEW	0.65	0.59	1.00	0.50	1.00		**NEW**	0.65	0.59	1.00	0.43	1.00
	TBV	1.00	0.79	0.47	0.83	0.48		**TBV**	1.00	0.75	0.31	0.79	0.33
	VR	0.85	1.00	0.81	0.89	0.81		**VR**	0.85	1.00	0.76	0.86	0.77
SCEN2	**VRpa**	0.65	0.59	1.00	0.60	0.98	**SCEN6**	**VRpa**	0.65	0.59	1.00	0.46	0.99
	JA	0.91	0.84	0.49	1.00	0.62		**JA**	0.91	0.85	0.53	1.00	0.49
	NEW	0.28	0.37	0.69	0.41	1.00		**NEW**	0.28	0.37	0.69	0.44	1.00
	TBV	1.00	0.71	0.35	0.12	0.35		**TBV**	1.00	0.72	0.33	0.13	0.33
	VR	0.85	1.00	0.82	0.14	0.82		**VR**	0.85	1.00	0.81	0.15	0.80
SCEN3	**VRpa**	0.65	0.59	1.00	0.10	1.00	**SCEN7**	**VRpa**	0.65	0.59	1.00	0.10	0.99
	JA	-0.28	-0.25	-0.19	1.00	0.12		**JA**	-0.30	-0.31	-0.16	1.00	0.11
	NEW	0.41	0.58	0.96	-0.17	1.00		**NEW**	0.39	0.41	0.96	-0.15	1.00
	TBV	1.00	0.73	0.36	0.43	0.37		**TBV**	1.00	0.76	0.30	0.68	0.30
	VR	0.85	1.00	0.82	0.48	0.81		**VR**	0.85	1.00	0.75	0.75	0.75
SCEN4	**VRpa**	0.65	0.59	1.00	0.31	0.99	**SCEN8**	**VRpa**	0.65	0.59	1.00	0.41	0.99
	JA	-0.22	-0.20	-0.19	1.00	0.33		**JA**	-0.34	-0.29	-0.13	1.00	0.43
	NEW	0.36	0.49	0.87	-0.14	1.00		**NEW**	0.35	0.49	0.87	-0.09	1.00

SCEN = Scenario, as defined in Table 1. [1]Pearson correlation coefficients were estimated considering all animals (above diagonal) and only genotyped bulls that had genotyped daughters (below diagonal) included in the training population. [2]Deregressed breeding values were estimated using the VanRaden et al. (2009) method, without removing and removing the parentage average (VR and VRpa, respectively), the Jansen (2008) method (JA), and the deregression method proposed in this study (NEW). Coefficients showed here are from 10 replicates.

Comparing methods based on the GEBVs

Reliability of GEBVs estimated in the different scenarios and using different deregression methods are shown in Figure 1. In general, including non-daughters in the training population (SCEN5 to SCEN8) did not increase reliability estimates compared to the scenarios that did not include those cows (SCEN1 to SCEN4). The NEW and VR methods were the most and the least (or they were among the most/least) reliable methods in all scenarios, respectively. The VRpa reliability was either similar to NEW or intermediate between VR and NEW methods across scenarios. The JA displayed large variability across scenarios, and it was generally less reliable than VRpa and NEW, and similar or better than VR across scenarios.

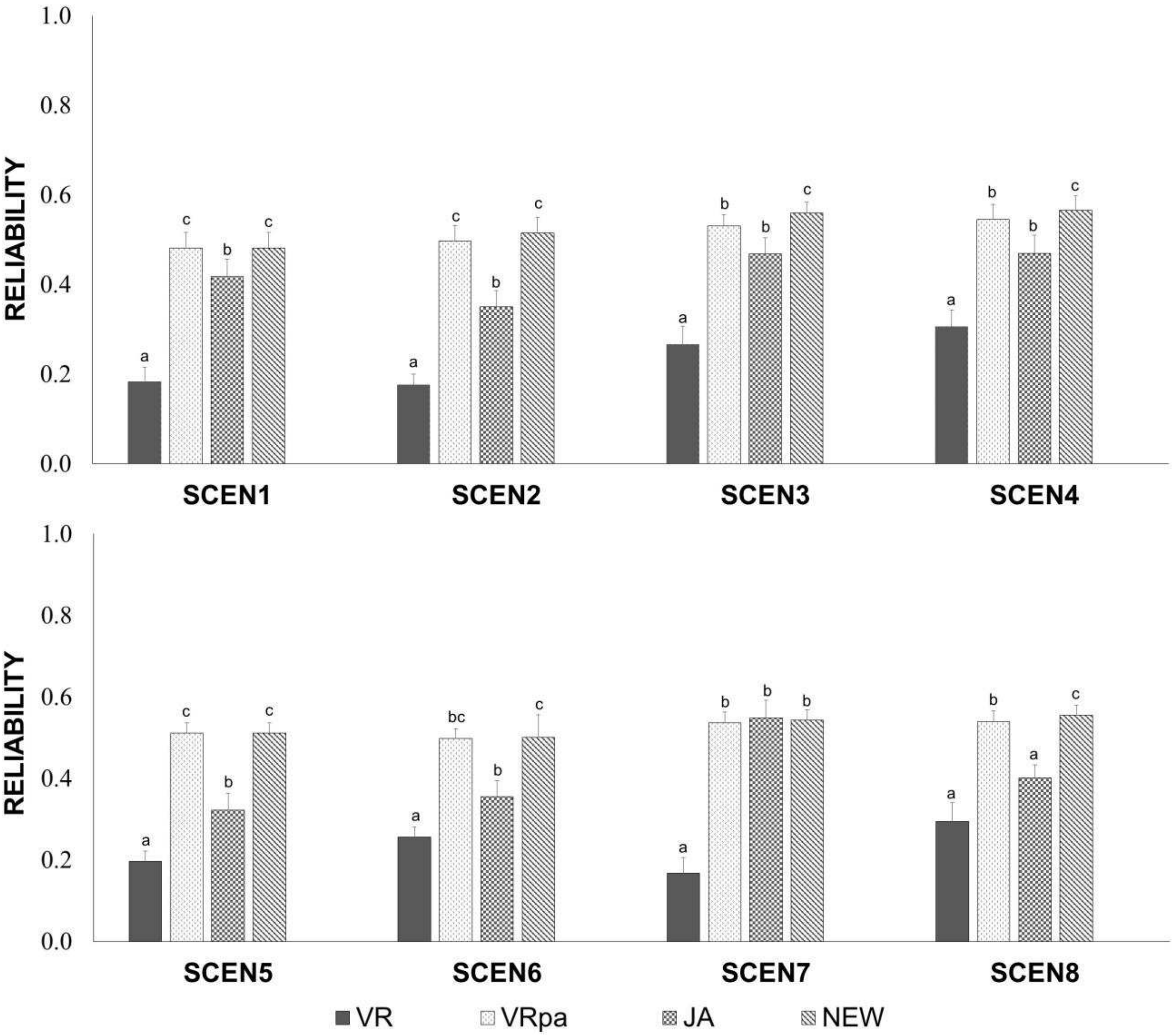

Figure 1. Average reliability and standard error of genomic breeding values estimated in the different scenarios (SCEN) from 10 replicates. Genomic breeding values were estimated considering as pseudo-phenotypes the deregressed breeding values estimated using the VanRaden et al. (2009) method, without removing and removing the parentage average (VR and VRpa, respectively), the Jansen (2008) method (JA), and the deregression method proposed in this study (NEW). Different letters, within each analyzed scenario, show significant difference (p-value < 0.05) between deregression methods.

Using dEBVs estimated by the different deregression methods generated different levels of scaling bias of genomic predictions across scenarios (Figure 2). In general, the bias was lower (i.e., closer to zero) when non-daughters were included in the training population (SCEN5 to SCEN8) than when they were not included (SCEN1 to SCEN4). Among all deregression methods, the VR yielded the most consistent bias estimates across scenarios (i.e., similar bias across scenarios), and it significantly showed one of the least biased genomic predictions. The VRpa, JA and NEW methods were mostly biased, showing large deflation of GEBVs across scenarios. However, when none or a small number of daughters were included in the training population (SCEN2, SCEN5 and SCEN6), the JA showed the lowest bias. In general, GEBVs based on dEBVs estimated by the VR were overestimated, while for the other methods the GEBVs were underestimated.

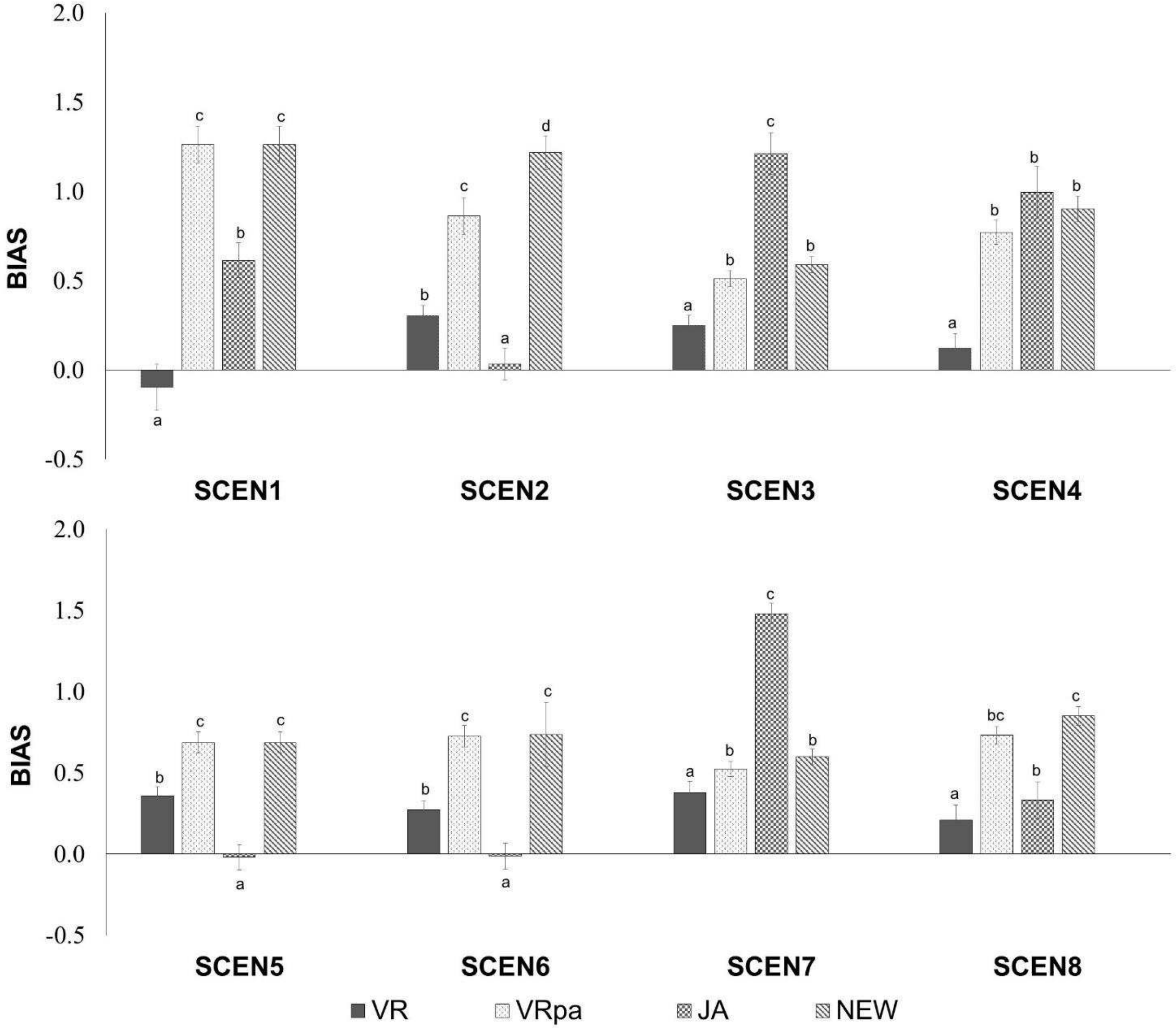

Figure 2 Average scaling bias and standard error of genomic breeding values in the different scenarios (SCEN) from 10 replicates. Genomic breeding values were estimated considering as pseudo-phenotypes the deregressed breeding values estimated using the VanRaden et al. (2009) method, without removing and removing the parentage average (VR and VRpa, respectively), the Jansen (2008) method (JA), and the deregression method proposed in this study (NEW). Bias was defined as the deviation from 1. Different letters, within each analyzed scenario, show significant difference (p-value < 0.05) between deregression methods.

5.5. Discussion

Usually EBVs are deregressed before their use in genomic evaluations to avoid a possible double-counting of information and to increase the variability of EBVs (Oliveira et al., 2018). However, an increase in the dEBV variability due to the deregression method was not observed when considering only bulls that had daughters included in the training population. This finding is likely due to the high REL_{EBV} estimated for these bulls, which had at least 70 daughters. As also reported in other studies (e.g., Song et al., 2017; Oliveira et al., 2018), high REL_{EBV} implies in a weaker deregression. In addition, the increase in dEBV variability is partially due to the PA, which can be observed in the difference of SD between the VR (that included PA in the dEBV) and the other methods that excluded PA. This is agreement with VanRaden (2001), who pointed out that variances are lower when removing PA from the estimates. Several studies in dairy cattle have used dEBVs removing PA (e.g., Butty et al., 2017; Song et al., 2017). Among the deregression methods, JA showed more irregular increase in dEBV variability between the scenarios, which might be due to the fact that JA takes into account information from all related animals to estimate the dEBVs (Jansen, 2008), not only from parents (such as in VR and VRpa) or parents and daughters (such as in NEW).

The success of genomic predictions depends on the reliability of GEBVs, which relies on the information contained in the dEBVs. In spite of the fact that the deregression process is less severe for animals with high REL_{EBV}, Pearson correlation coefficients estimated between TBVs and dEBVs were higher when considering only bulls (SCEN1) than bulls and non-daughters (SCEN5, for all animals). This is likely due to the fact that bulls usually have more reliable information compared to cows (Uemoto et al., 2017, Oliveira et al., 2018). In all scenarios, VR showed high correlation coefficients between TBV and dEBVs, which might be due to the inclusion of PA in the estimates. On the other hand, JA showed high Pearson correlation coefficients between TBV and dEBVs only when non-genotyped (SCEN1 and SCEN5) or a small number of genotyped (SCEN2 and SCEN6) daughters where included in the training population. In general, the NEW method yielded the lowest correlation coefficients, which is related to the fact that the NEW method removes not only the PA, but also the GDA from bull's dEBV. Thus, the amount of information available in the NEW method is substantially lower compared to the information available in dEBVs estimated using the other methods. Considering that, a higher proportion of the number of genotyped daughthers in relation to non-genotyped daughters for each bull is associated with lower correlations between TBV and dEBVs estimated using NEW method.

This may have great implications in the future, when a larger number of genotyped daughter per sire are available. In summary, our findings indicate that NEW and JA (especially when a large number of daughters is included in the training population) do, on average, a stronger deregression when compared to the other methods (Table 3).

Similarly to our findings, Wiggans et al. (2011) concluded that including genotypes and phenotypes from cows in the training population barely increased the reliability of genomic predictions. This might be related to the lower amount of information provided by a cow when compared to a single progeny-tested bull (Weller, 2016). For all scenarios evaluated in this study, genomic predictions based on dEBVs estimated based on the NEW method showed the highest reliabilities. This suggests that removing the double-counting of information due to the PA and DGA information using the NEW method is a feasible alternative to improve the performance of genomic predictions in a multiple-step evaluation. In general, JA showed lower reliabilities compared to the NEW and VRpa. However, when a larger number of daughters per bull was included in a training population that also included non-daughters (SCEN7), JA yielded similar prediction reliability compered to VRpa and NEW. This shows that the performance of JA is highly dependent of complexity of training population in terms of genetic relationships. Additionally, the use of dEBVs estimated based on VR had poor performance (based on reliability) for all scenarios, indicating that the double-counting of the information remaining after the deregression process might have given more weight to bulls with more genotyped daughters when estimating the GEBVs, which reflected in the reliability of genomic prediction. These findings may justify why other authors, such as Wu et al. (2015) and Uemoto et al. (2017), found that a closer relationship among training animals resulted in lower reliability estimates. On the other hand, Plieschke et al. (2018) concluded that it is possible to increase the reliability of genomic predictions including genotyped daughters in the training population. Their findings are related to the use of yield and daughter yield deviations as pseudo-phenotypes, which do not include the contribution of the PA (i.e., using yield deviations reduce the double-conting of information compared to the use of dEBV estimated from VR method). However, in disagreement with our results, Garrick et al. (2009) suggested that the double-counting of progeny’s performance has negligible impact on genomic predictions.

Including non-daughters in the training population (SCEN5 to SCEN8) reduced the scaling bias of genomic prediction compared to the scenarios that did not include them (SCEN1 to SCEN4), which is in agreement with Uemoto et al. (2017). The VR method showed one of the least biased genomic predictions and the most consistent bias across scenarios, which is associated with the fact that dEBVs generated by this method contained exactly the same information in all scenarios for all animals. On the other hand, methods that performed a stronger EBV deregression yielded more biased predictions. This shows the need to re-scale GEBVs when using dEBVs obtained by these methods. In general, GEBVs estimated by the VRpa, JA and NEW methods were underestimated, indicating that the predicted genetic merit are deflated.

5.6. Conclusions

This study suggests that removing PA and GDA information from dEBVs can increase the reliability of genomic prediction for small-sized dairy cattle populations. In addition, the results indicate that using stronger deregression methods (such as JA and NEW) may change the scale of GEBV. The NEW deregression method presented in this study is a feasible alternative to generate dEBVs to be used in multiple-step genomic evaluations. However, the resulting GEBVs need to be rescaled in order to minimize scaling bias.

5.7. Acknowledgements

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES). The authors gratefully acknowledge the financial support to this research from Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative.

5.8. References

Brito FV, Neto JB, Sargolzaei M, Cobuci JA and Schenkel FS 2011. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. BMC Genetics 12, 1-10.

Butty AM, Frischknecht M, Gredler B, Neuenschwander S, Moll J, Bieber A, Baes CF and Seefried FR 2017. Genetic and genomic analysis of hyperthelia in Brown Swiss cattle. Journal of Dairy Science, 100, 402-411.

Garrick DJ, Taylor JF and Fernando RL 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution 41, 1-8.

Interbull 2017. Description of national genomic evaluation systems. Accessed May 02, 2018. In, http://www.interbull.org/ib/nationalgenoforms.

Jansen G 2008. Computation of approximate EDC and DYD in general pedigrees. Proceedings of Dairy Cattle Breeding & Genetics Committee, September 2008, Guelph, Canada.

Jiménez-Montero JA, Gonzalez-Recio O and Alenda R 2012. Genotyping strategies for genomic selection in small dairy cattle populations. Animal 6, 1216-1224.

Koivula M, Strandén I, Aamand GP and Mäntysaari EA 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal 10, 1061-1066.

Meuwissen T, Hayes B and Goddard M 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829.

Meuwissen T, Hayes B and Goddard M 2016. Genomic selection: A paradigm shift in animal breeding. Animal Frontiers 6, 6-14.

Misztal I, Tsuruta S, Lourenco DAL, Aguilar I, Legarra A and Vitezica Z 2014. Manual for BLUPF90 family of programs. Athens: University of Georgia.

Oliveira H, Silva F, Brito L, Guarini A, Jamrozik J and Schenkel F 2018. Comparing deregression methods for genomic prediction of test-day traits in dairy cattle. Journal of Animal Breeding and Genetics 135, 97-106.

Plieschke L, Edel C, Pimentel EC, Emmerling R, Bennewitz J and Götz KU 2018. Genotyping of groups of cows to improve genomic breeding values of new traits. Journal of Animal Breeding and Genetics 135, 286-92.

Sargolzaei M and Schenkel FS 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics 25, 680-681.

Song H, Li L, Zhang Q, Zhang S and Ding X 2017. Accuracy and bias of genomic prediction with different de-regression methods. Animal, 12, 1111-1117.

Rosner B 1982. A generalization of the paired t-test. Applied Statistics 31, 9-13.

Uemoto Y, Osawa T and Saburi J 2017. Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. Animal 11, 382-393.

VanRaden P 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science 91, 4414-4423.

VanRaden P 2001. Methods to combine estimated breeding values obtained from separate sources. Journal of Dairy Science, 1, 47-55.

VanRaden P and Wiggans G 1991. Derivation, calculation, and use of national animal model information. Journal of Dairy Science 74, 2737-2746.

VanRaden P, Van Tassell C, Wiggans G, Sonstegard T, Schnabel R, Taylor J and Schenkel F 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science 92, 16-24.

Weller, J. I. 2016. Genomic selection in animals. John Wiley & Sons, Inc., Hoboken, New Jersey.

Wiggans G, Cooper T, VanRaden P and Cole J 2011. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. Journal of Dairy Science 94, 6188-6193.

Wiggans G, Sonstegard T, VanRaden P, Matukumalli L, Schnabel R, Taylor J, Schenkel F and Van Tassell C 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science 92, 3431-3436.

Wu X, Lund MS, Sun D, Zhang Q and Su G 2015. Impact of relationships between test and training animals and among training animals on reliability of genomic prediction. Journal of Animal Breeding and Genetics 132, 366-375.

5.9. Supplementary material

Table S1. Descriptive statistics[1] for the 10 replicates (Rep) of the simulated population.

	Rep1	**Rep2**	**Rep3**	**Rep4**	**Rep5**	**Rep6**	**Rep7**	**Rep8**	**Rep9**	**Rep10**
N	160 100	160 100	160 100	160 100	160 100	160 100	160 100	160 100	160 100	160 100
Females	84 674	85 165	85 269	84 782	84 860	85 093	84 994	84 712	85 089	84 674
Phen	22.28 (14.46)	22.27 (14.30)	22.89 (14.64)	21.62 (14.20)	22.28 (14.32)	20.46 (13.19)	21.31 (13.46)	20.05 (13.21)	20.78 (13.60)	22.23 (14.16)
TBV	21.74 (12.67)	22.41 (12.79)	23.11 (13.23)	21.66 (12.89)	22.51 (12.88)	20.40 (11.62)	21.43 (11.85)	19.89 (11.73)	20.75 (12.13)	22.41 (13.07)
EBV	21.50 (12.46)	22.16 (12.63)	22.85 (13.05)	21.40 (12.63)	22.18 (12.69)	20.17 (11.40)	21.14 (11.62)	19.63 (11.54)	20.51 (11.89)	22.15 (12.91)
$\mathbf{REL_{EBV}}$	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)	0.38 (0.08)

N = total number of animals. [1]Mean and standard deviation (in brackets) were estimated for phenotypes (Phen), true breeding values (TBV), breeding values estimated by the best linear unbiased predictor (EBV), and for EBV reliabilities (RELEBV).

Table S2. Summary of parameters used in the simulation process[1].

Population structure		
Historical population:		
	Generation 0	700 males and 700 females
	Generation 1 000	200 males and 200 females
	Generation 1 040	500 males and 20 000 females
Founder population:		
	Generation 0	100 males and 10 000 females
Recent generations:		
	Number of generations	15
	Selection system	Based on EBVs
	Breeding value estimation method	BLUP animal model
	Number of offspring per dam	1
	Sex ratio	0.50
	Mating system for selected animals	Random
	Replacement ratio for males	0.60
	Replacement ratio for females	0.20
Phenotypes		
	Trait type	Single trait
	Recorded on generations	Recent generations (1 to 15)
	Measurement on sex	Females
	Phenotypic variance	100
	Heritability	0.30
	Heritability due to polygenic effect	0.03
Genotypes		
	Recorded on generations	Recent generations (6 to 14)
	Number of chromosomes	29
	Total length	2 500 cM
	Total number of SNP markers	57 024
	SNP marker distribution	Evenly spaced
	Total number of QTL	1 979
	QTL distribution	Random
	Additive allelic effects for markers	Uniform distribution
	Additive allelic effects for QTL	Gamma distribution (shape = 0.40)
	Rate of missing marker genotypes	0.01
	Rate of marker genotyping error	0.005
	Rate of recurrent mutation	0.0001

EBV = Estimated Breeding Value; BLUP = Best Linear Unbiased Prediction; SNP = Single Nucleotide Polymorphism; QTL= Quantitative Trait Loci. [1]The whole simulation process was replicated 10 times.

Table S3. Length, number of single nucleotide polymorphisms (SNPs) and number of quantitative trait loci (QTLs) assumed for each simulated chromosome[1].

Chromosome	Length (cM)	Number of SNPs	Number of QTLs
1	158	3 647	30
2	137	3 022	26
3	121	2 777	34
4	120	2 754	23
5	121	2 447	46
6	119	2 713	88
7	112	2 540	30
8	113	2 592	13
9	105	2 208	20
10	104	2 387	18
11	107	2 500	32
12	91	1 893	14
13	84	2 022	42
14	84	1 935	1 022
15	85	1 916	24
16	81	1 864	43
17	75	1 741	23
18	66	1 524	20
19	64	1 570	36
20	72	1 682	184
21	71	1 607	53
22	61	1 432	13
23	52	1 180	44
24	62	1 412	7
25	42	1 091	21
26	51	1 241	22
27	45	1 066	18
28	46	1 076	15
29	51	1 185	18

[1] SNP markers and QTL positions were randomized within each chromosome in each one of the 10 replicates.

Table S4. Average of number of genotyped and non-genotyped daughters (n_{GD} and n_{NGD}, respectively), deregression factor estimated for the deregressed breeding values from genotyped and non-genotyped daughters ($\bar{R}_{GD}$ and $\bar{R}_{NGD}$, respectively), and reliability of genotyped and non-genotyped daughter average (REL_{GDA} and REL_{NGDA}, respectively), when including 10, 10 to 70 or 70 genotyped daughters per genotyped bull.

Scenario with[1]	n_{GD}	$\bar{R}_{GD}$	REL_{GDA}	n_{NGD}	$\bar{R}_{NGD}$	REL_{NGDA}
10 genotyped daughters	10	0.24	0.04	125	0.24	0.48
10 to 70 genotyped daughters	40	0.24	0.17	95	0.24	0.35
70 genotyped daughters	70	0.24	0.29	65	0.24	0.23

[1] Values showed here are an average of the 10 replicates.

CHAPTER 6:

Genomic prediction of lactation curves for milk, fat, protein and somatic cell score in Holstein cattle

Hinayah Rojas de Oliveira[1,2], Luiz Fernando Brito[1], Fabyano Fonseca e Silva[2], Daniela Andressa Lino Lourenco[3], Janusz Jamrozik[1,4], Flávio Schramm Schenkel[1]

[1] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

[2] Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[3] Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, United States of America

[4] Canadian Dairy Network, Guelph, Ontario, Canada

Journal of Dairy Science, 102:1-12, 2018.

6.1. Abstract

Application of random regression models (**RRM**) in a two-step genomic prediction might be a feasible way to select young animals based on the complete pattern of the lactation curve. In this context, the prediction reliability and bias of genomic estimated breeding value (**GEBV**) for milk, fat and protein yields, and somatic cell score over days in milk (**DIM**) using a two-step genomic approach were investigated. In addition, the effect of including cows in the training and validation populations was also investigated. Estimated breeding values for each DIM (from five to 305 days) from the first three lactations of Holstein animals were deregressed and used as pseudo-phenotypes in the second step. Individual additive genomic random regression coefficients for each trait were predicted using RRM and Genomic Best Linear Unbiased Prediction and further used to derive GEBV for each DIM. Theoretical reliabilities of GEBVs obtained by the RRM were slightly higher than theoretical reliabilities obtained by the accumulated yield up to 305 days (**P305**). However, validation reliabilities estimated for GEBV using P305 were higher than for GEBV using RRM. For all traits, higher theoretical and validation reliabilities were estimated when incorporating genomic information. Less biased GEBV estimates were found when using RRM compared to P305, and different validation reliability and bias patterns for GEBV over time were observed across traits and lactations. Including cows in the training population increased the theoretical reliabilities and bias of GEBV, nonetheless, the inclusion of cows in the validation population does not seem to affect the regression coefficients and the theoretical reliabilities. In summary, the use of RRM in two-step genomic prediction produced fairly accurate GEBVs over the entire lactation curve for all analyzed traits. Thus, selecting young animals based on the pattern of lactation curves seems to be a feasible alternative in genomic selection of Holstein cattle for milk production traits.

Keywords: days in milk, genomic breeding value, longitudinal trait, random regression model

6.2. Introduction

Genome-wide selection (**GWS**) is a landmark for modern livestock breeding programs, because it has the potential to increase the reliability of breeding values, especially for selection of candidates at a young age, enabling a reduction in the generation interval and, consequently, increasing the genetic gain per unit of time. Additionally, GWS also reduces the costs associated with progeny testing schemes that have been used in dairy cattle for a long time (Schaeffer, 2006).

There are several economically important traits in dairy cattle, such as milk production and composition traits, which are defined as longitudinal traits. Nowadays, the majority of genomic researches have focused on accumulated yield up to 305 days (**P305**) as the phenotype to predict genomic breeding values (Winkelman et al., 2015; Aliloo et al., 2016; Jenko et al., 2017). However, previous studies have reported advantages of using random regression models (**RRM**) over the P305 models to analyze longitudinal traits (Schaeffer et al., 2000; Jensen, 2001). In brief, P305 models are not able to account for genetic variation in production over time, and consequently, do not allow for selection of animals based on the complete pattern of the production curve.

The selection of animals based on the complete lactation curves would take into account the lactation biology and the physiological implications of different curve patterns. In a practical way, this may help to decrease the energy deficiency experienced during early- and peak-lactation, and, therefore, can improve health, welfare and fertility traits. In addition, RRM are also more advantageous compared to P305, because selection can be efficiently performed based on points of interest throughout the lactation curve. In order to overcome the limitations of P305 models, some studies have used the predicted P305 from traditional Best Linear Unbiased Prediction (**BLUP**) random regression genetic evaluation as phenotype for two-step GWS (Ding et al., 2013; Brown et al., 2016). The steps involved in the two-step GWS approach are: 1) estimation of breeding value (**EBV**) based on traditional evaluation and its de-regression; 2) estimation of direct genomic value (**DGV**) using deregressed EBVs from step 1 as phenotype for the genomic BLUP (**GBLUP**); subsequently, DGV can be combined with parent average (**PA**) into genomic estimated breeding value (**GEBV**). However, this strategy of only using RRM in the first-step genetic evaluation still does not allow making inferences over time, because only the total GEBV is predicted in the second step.

In order to predict GEBVs over time, recent studies have used RRM in single-step genomic evaluations (Koivula et al., 2015; Baba et al., 2016; Jattawa et al., 2016). However, the single-step GBLUP method (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010) is not currently used for official genomic evaluations in several countries, including Canada (Interbull, 2017). Thus, the application of RRM in both steps of a two-step GWS could provide a suitable alternative to select genotyped animals based on the pattern of the lactation curve. Therefore, the main goals of this study were to: 1) investigate the reliability and bias of GEBVs over DIM, in a two-step approach, for milk, fat and protein yields, and somatic cell score in the first three lactations of Holstein cattle; 2) compare the prediction reliability and bias of GEBVs from RRM to the prediction reliability and bias of GEBVs from P305 model; 3) compare the prediction reliability and bias of GEBVs obtained from RRM and from the P305 model to the prediction reliability and bias of PA. In addition, the impact of incorporating cows in the training and validation populations was also investigated.

6.3. Material and methods

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

Datasets

The Canadian Dairy Network (**CDN**; Guelph, ON, Canada) provided the pedigree, genotypes, EBVs for the individual genetic regression coefficients and reliabilities, from December 2012 and December 2016 genetic evaluation runs. The EBVs for the individual genetic regression coefficients were available for milk, fat and protein yields, and somatic cell score from the first three lactations of Holstein cattle. The only available reliabilities for each trait and animal were the average reliabilities over the first three lactations for the accumulated yields. The total number of animals in the official genetic evaluation in the 2012 and 2016 runs was 4,888,808 and 5,683,640, respectively. Traits were coded as **MY1**, **FY1**, **PY1**, **SCS1**, **MY2**, **FY2**, **PY2**, **SCS2**, **MY3**, **FY3**, **PY3** and **SCS3**; where MY, FY, PY and SCS are milk, fat and protein yields, and somatic cell score, respectively, and 1, 2 and 3 refer to the first, second and third lactation, respectively.

Phenotypes. EBVs for each lactation day were derived using the EBVs for genetic regression coefficients predicted by CDN based on two multiple-trait random regression test-day animal models (Schaeffer et al., 2000): one considering milk, fat and protein yields in the first three lactations; and another one considering somatic cell score in the first three lactations, as routinely performed by CDN. Details about the RRM used by CDN are fully presented in Interbull (2018a, b). In brief, both models included the fixed effects of herd-test day-parity and DIM (from 5 to 305 days), fixed regression curves for year-season-region-parity-age of calving effect, and random regression curves for herd-year of calving, additive genetic and permanent environment effects. Fixed and random regressions were modeled by fourth order Legendre orthogonal polynomials and heterogeneity of residual variance was taken into account by fitting 10 classes of residual variances in each parity. The pattern of the average EBVs over DIM for each trait is shown in Figure S1 (Supplementary Material).

EBVs from five to 305 DIM were used in the analyses. For all animals used in the second-step of GWS (i.e., the genotyped animals), the PA and the theoretical reliability of PA (**REL$_{PA}$**) were estimated as shown in VanRaden and Wiggans (1991), using the parent information available in the 2012 evaluation. To perform the analysis considering the traditional P305 model in the second-step of GWS, accumulated EBVs and PAs of each animal were obtained by summing up the EBVs and PAs of each day for each trait. Because a single theoretical reliability per animal for each trait was available from CDN, reliabilities of the EBVs for all DIM and for P305 were assumed the same.

In order to avoid double-counting of information and double shrinkage of the DGVs by using EBVs as pseudo-phenotypes, EBVs are usually deregressed before their use in genomic evaluations (Garrick et al., 2009; Oliveira et al., 2018). Therefore, deregressed daily and accumulated EBVs (**dEBVs**) were used as pseudo-phenotypes for each animal in the genomic evaluation of all traits. Since using well-established methods of deregression for non-longitudinal traits is a feasible alternative to obtain dEBVs used in the genomic prediction of longitudinal traits (Oliveira et al., 2018), dEBV$_i$ of animal i for each day and for the P305 were computed according to VanRaden et al. (2009):

$$\mathrm{dEBV_i} = \mathrm{PA_i} + \frac{(\mathrm{EBV_i} - \mathrm{PA_i})}{\mathrm{R_i}},$$

where PA$_i$ is the parent average for the animal i, EBV$_i$ is the estimated breeding value for the animal i, and $\mathrm{R_i}$ is the de-regression factor computed as follows:

$$\mathrm{R_i} = \frac{\mathrm{ERC_{(Ind-PA)_i}}}{\mathrm{ERC_{(Ind-PA)_i}} + \mathrm{ERC_{PA_i}} + 1},$$

where the effective record contributions for the animal i excluding the PA information $(\mathrm{ERC_{(Ind-PA)_i}})$ and for the PA ($\mathrm{ERC_{PA_i}}$) of the animal i were estimated based on VanRaden and Wiggans (1991) and Přibyl et al. (2013):

$$\mathrm{ERC_{(Ind-PA)_i}} = \left(\lambda \frac{\mathrm{REL_{EBV_i}}}{1-\mathrm{REL_{EBV_i}}}\right) - \mathrm{ERC_{PA_i}} \text{ and } \mathrm{ERC_{PA_i}} = \lambda \frac{\mathrm{REL_{PA_i}}}{1-\mathrm{REL_{PA_i}}},$$

in which $\lambda = (1-h^2)/h^2$, and REL_{EBV_i} and REL_{PA_i} are the theoretical reliabilities for EBV and PA of the animal i, respectively. Average heritabilities (**h²**) along the lactation curves considered for MY1, FY1, PY1, SCS1, MY2, FY2, PY2, SCS2, MY3, FY3, PY3 and SCS3 were estimated by CDN as 0.31, 0.27, 0.27, 0.14, 0.29, 0.26, 0.27, 0.18, 0.26, 0.24, 0.24 and 0.21, respectively (Interbull, 2018a, b). To avoid double-counting of information due to the fact that an animal without its own phenotype or without phenotypes of its daughters would have their own EBV calculated solely based on PA, animals that did not have their own performance or progeny information $(ERC_{(Ind-PA)_i} = 0)$ were excluded from the analyses.

Genotypes. Only animals with genotypic information were included in the second step of the GWS analysis. Animals were genotyped or accurately imputed to the Illumina BovineSNP50K BeadChip (Illumina, San Diego, CA, USA). The FImpute software (Sargolzaei et al., 2011) was used for imputation. Details about the accuracy of imputation for Canadian breeds can be found in Larmer et al. (2014).

After the genotype imputation process performed by CDN, single nucleotide polymorphisms (**SNPs**) with Mendelian conflicts, call rate less than 0.95, minor allele frequency (**MAF**) less than 0.01 and with a difference between observed and expected heterozygosity frequency higher than 0.15 (Wiggans et al., 2009) were removed using the preGSF90 software (Aguilar et al., 2014), version 1.10. A total of 43,556 SNPs distributed over the 29 bovine autosomes remained in the dataset after the quality control.

Statistical Analysis

Training and validation populations. Genomic predictions were performed for each trait separately. Genotyped animals included in the official Canadian evaluation performed in the 2012 genetic evaluation run were used as a training population; genotyped animals born after 2012, which were present in the 2016 genetic evaluation run, were used as a validation population. In addition, bulls born from 2007 to 2010 that only had daughters born after 2012, were also included in the validation population.

In order to investigate the effect of inclusion of cows in the training and validation populations, three scenarios were evaluated: 1) considering only bulls in the training and validation populations; 2) considering bulls and cows in the training population and only bulls in the validation population; and 3) considering bulls and cows in the training and validation populations. Based on preliminary analysis (results not shown), it was required that animals (bulls and cows) included in the training population to have a minimum REL_{EBV} of 0.65 for each trait. In addition, it was required that bulls and cows included in the validation population to have a minimum REL_{EBV} of 0.80 and 0.65 for each trait, respectively. A detailed description of all training and validation populations are shown in Table 1.

Table 1. Total number of animals (N) and bulls (N bulls), years of birth (YOB), mean and standard deviation (SD) of effective record contributions (ERC) and of reliability of estimated breeding values (REL_{EBV}), for the training and validation populations used in the genomic prediction of each trait.

Trait	**Parameters**	**Training**		**Validation**	
		^{1}Only bulls	**^{1}Bulls and cows**	**^{2}Only bulls**	**^{1}Bulls and cows**
MY	N	7,322	10,798	1,170	2,714
	N bulls	7,322	7,322	1,170	1,247
	^{3}YOB	1957 to 2008	1957 to 2009	2012 to 2013	2012 to 2013
	ERC (SD)	132.30 (744.7)	90.78 (616.15)	60.48 (155.2)	27.77 (105.81)
	REL_{EBV} (SD)	0.93 (0.12)	0.85 (0.12)	0.94 (0.01)	0.78 (0.14)
FY	N	7,325	12,034	1,168	2,436
	N bulls	7,325	7,325	1,168	1,251
	^{3}YOB	1957 to 2008	1957 to 2009	2012 to 2013	2012 to 2013
	ERC (SD)	166.43 (940.5)	102.93 (738.0)	75.39 (192.5)	38.28 (138.07)
	REL_{EBV} (SD)	0.93 (0.12)	0.83 (0.12)	0.94 (0.01)	0.81 (0.13)
PY	N	7,289	8,865	1,160	1,297
	N bulls	7,289	7,289	1,160	1,241
	^{3}YOB	1957 to 2008	1957 to 2008	2012 to 2012	2012 to 2012
	ERC (SD)	114.64 (815.6)	119.71 (741.5)	64.67 (163.4)	58.35 (155.63)
	REL_{EBV} (SD)	0.92 (0.10)	0.88 (0.10)	0.94 (0.02)	0.91 (0.08)
SCS	N	7,098	7,329	1,083	1,184
	N bulls	7,098	7,098	1,083	1,179
	^{3}YOB	1957 to 2008	1957 to 2008	2012 to 2012	2012 to 2012
	ERC (SD)	191.5 (1098.3)	185.9 (1081.3)	79.31 (177.1)	73.72 (170.37)
	REL_{EBV} (SD)	0.88 (0.06)	0.88 (0.06)	0.90 (0.02)	0.89 (0.06)

Traits: milk (MY), fat (FY) and protein (PY) yield, and somatic cells score (SCS). ^{1}Minimum REL_{EBV} equal to 0.65. ^{2}Minimum REL_{EBV} equal to 0.80. ^{3}YOB of cows and daughters of bulls in the validation population.

GBLUP P305. Accumulated DGVs for each animal were predicted using the accumulated dEBVs as pseudo-phenotypes based on the GBLUP method (VanRaden, 2008). The traditional single-trait animal model used for P305 is described as follow:

$$\mathbf{y}_{\mathrm{ac}} = \mathbf{1}\mu + \mathbf{W}_{\mathrm{ac}}\mathbf{g}_{\mathrm{ac}} + \mathbf{e}_{\mathrm{ac}},$$

where $\mathbf{y}_{\mathrm{ac}}$ is the vector of pseudo-phenotypic accumulated records; μ is the overall mean, $\mathbf{W}_{\mathrm{ac}}$ is the incidence matrix relating the random animal genomic effects ($\mathbf{g}_{\mathrm{ac}}$) to P305; and $\mathbf{e}_{\mathrm{ac}}$ is the vector of residuals for P305. The model assumptions were:

$$\mathrm{E}[\mathbf{y}_{\mathrm{ac}}] = \mu \text{ and } \mathrm{Var}\begin{bmatrix}\mathbf{g}_{\mathrm{ac}}\\ \mathbf{e}_{\mathrm{ac}}\end{bmatrix} = \begin{bmatrix}\mathbf{G}\sigma^2_{g_{\mathrm{ac}}} & 0\\ 0 & \mathbf{I}\sigma^2_{e_{\mathrm{ac}}}\end{bmatrix},$$

where $\sigma^2_{g_{\mathrm{ac}}}$ and $\sigma^2_{e_{\mathrm{ac}}}$ are the genomic and residual variances for P305, and $\mathbf{I}$ is the identity matrix. **G** is the genomic relationship matrix, which was created as in VanRaden (2008):

$$\mathbf{G} = 0.95\frac{\mathbf{ZZ'}}{2\sum_{k=1}^{K} p_k(1-p_k)} + 0.05\mathbf{A},$$

where $\mathbf{Z} = \mathbf{M} - \mathbf{P}$, in which **M** contains the centered genotypes (i.e., -1, 0 and 1 to represent AA, Aa and aa, respectively), **P** contains allele frequency for SNP k (p_k) in its kth column, expressed as $2(p_k - 0.5)$; $2\sum_{k=1}^{K} p_k(1-p_k)$ is a scaling parameter; and **A** is the traditional (pedigree-based) additive relationship matrix, which included a total of 39,918 animals. In order to guarantee that **G** matrix was positive definite, 5% of **A** was added to 95% of **G** (VanRaden, 2008).

GBLUP RRM. Genomic random regression coefficients for each animal were also predicted using the GBLUP method (VanRaden, 2008). In matrix notation, the RRM used for each trait was:

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

where y is the vector of pseudo-phenotypic records (dEBV for each DIM); **X** and **W** are the incidence matrices for the vectors of fixed regression coefficients (**b**) and genomic random regression coefficients (**g**), which were modelled by fourth order Legendre orthogonal polynomials (Kirkpatrick et al., 1990); and **e** is the vector of residuals. The model assumptions were:

$$\mathrm{E}[\mathbf{y}] = \mathbf{Xb} \text{ and } \mathrm{Var}\begin{bmatrix}\mathbf{g}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G}_0 \otimes \mathbf{G} & 0\\ 0 & \sigma_e^2 \otimes \mathbf{I}\end{bmatrix},$$

where $\mathbf{G}_0$ is the variance-covariance matrix of the genomic random regression coefficients; σ_e^2 is the residual variance, **G** is the genomic relationship matrix, and $\mathbf{I}$ is the identity matrix. The vector of estimated DGV for all lactation days of animal i and for each trait was obtained as follows:

$$\mathbf{DGV}_i = \mathbf{T}\hat{\mathbf{g}}_i,$$

where $\hat{\mathbf{g}}_\mathbf{i}$ is the vector of predicted genomic coefficients for each animal i; and $\mathbf{T}$ is a matrix of orthogonal covariates associated with the Legendre polynomial function. The AIREMLF90 software (Misztal et al., 2002) was used to obtain the variance components and the solutions of the mixed model equations (**MME**) for both P305 and RRM analyses.

GEBVs and theoretical reliabilities. The GEBV of each animal i, for each DIM and for P305 were calculated by combining DGV with PA information using a selection index weighted by their theoretical reliabilities (Hayes et al., 2009):

$$\mathrm{GEBV_i} = \mathrm{w_1 PA_i} + \mathrm{w_2 DGV_i},$$

in which w_1 and w_2 were given by:

$$\mathrm{w}_1 = \frac{\mathrm{REL_{PA_i}}}{\mathrm{REL_{PA_i}} + \mathrm{REL_{DGV_i}}} \text{ and } \mathrm{w}_2 = \frac{\mathrm{REL_{DGV_i}}}{\mathrm{REL_{PA_i}} + \mathrm{REL_{DGV_i}}},$$

where $\mathrm{REL_{DGV}}$ is the theoretical reliability of estimated DGVs. The $\mathrm{REL_{DGV}}$ for each animal in the validation population was computed based on its prediction error variance (**PEV**). For the P305 analyzes, the PEV matrix was estimated directly using the coefficient matrix ($\mathbf{C}^{22}$) extracted from the inverse of the left-hand-side of the MME. Specifically for the RRM analyzes, **PEV** for each animal i and for each trait was estimated as in Mrode (2014):

$$\mathbf{PEV}_i = \mathbf{T}'\mathbf{C}^{ii}\mathbf{T},$$

where $\mathbf{T}$ is a matrix of orthogonal covariates associated with the Legendre polynomial function, and $\mathbf{C}^{ii}$ is the subset of the inverse of MME corresponding to the genomic effect for the animal i. Then, the $\mathrm{REL_{DGV_i}}$ of the animal i for each DIM and for the P305 was computed as:

$$\mathrm{REL_{DGV_i}} = 1 - \left(\frac{\mathrm{PEV_i}}{\sigma_g^2}\right),$$

where σ_g^2 was the estimated genomic variance for each DIM or for P305.

The theoretical reliabilities of GEBVs (REL_{GEBV}) were approximated as:

$$REL_{GEBV_i} = w_1 REL_{PA_i} + w_2 REL_{DGV_i},$$

where all terms were as defined previously. It is worth to note that the theoretical reliabilities of EBV initially made available by CDN for each animal and trait were average reliabilities for the first three lactations and that the same animals were included in the genomic analysis of all three lactations for a same trait (e.g., MY1, MY2 and MY3). Therefore, the same amount of information was available for each animal and trait. Thus, in order to avoid presenting redundant information, a single average theoretical reliability will be reported for each trait, method and scenario for both PA and GEBV.

GEBV validation reliability and bias. In order to access the pattern of validation reliability and bias of GEBV over DIM, a proxy of the reliability of GEBV (validation reliability) obtained from the RRM for each trait was calculated as the squared Pearson correlation coefficient between GEBVs (based on the 2012 genetic evaluation) and EBVs (based on the 2016 genetic evaluation) for each DIM considering all individuals in the validation population. The validation reliabilities estimated in this study were not adjusted for the imperfect reliabilities of the 2016 EBVs, because only relatively accurate animals and with similar reliability estimates were included in the validation population. In addition, regression coefficients (b_1) were estimated using a linear regression of EBV on GEBV for each DIM (i.e., $EBV = b_0 + b_1 \times GEBV$), to assess the prediction bias (inflation or deflation of the GEBVs relative to EBVs) over DIM.

Reliability and bias of PAs and GEBVs obtained from RRM were compared to those obtained from the traditional P305 models, which were estimated using EBVs and GEBVs for the accumulated yield. To facilitate the comparison between P305 and RRM, daily PAs and daily GEBVs obtained from the RRM were summed up to generate accumulated PAs and accumulated GEBVs, respectively. Thereafter, validation reliability were estimated for PAs and GEBVs generated from both P305 and RRM, using the squared Pearson correlation coefficient between accumulated PAs and accumulated EBVs, and between accumulated GEBVs and accumulated EBVs, respectively. In addition, bias of PA and GEBV were estimated based on RRM and P305 models, using a linear regression of accumulated EBV on accumulated PA (b_1 obtained from

$\text{accumulated EBV} = b_0 + b_1 \times \text{accumulated PA}$) and of accumulated EBV on accumulated GEBV (b_1 obtained from $\text{accumulated EBV} = b_0 + b_1 \times \text{accumulated GEBV}$), respectively.

6.4. Results

GEBV theoretical reliabilities

Average of theoretical reliabilities of PA and GEBV obtained based on P305 and RRM, considering only bulls in the training and validation populations, bulls and cows in the training population and only bulls in the validation population, and bulls and cows in the training and validation populations are shown in Table 2. Theoretical reliabilities of GEBVs obtained by the RRM were slightly greater than theoretical reliabilities obtained by the P305. For all traits, higher theoretical reliabilities were estimated for GEBV compared to PA. Theoretical reliabilities of PA and GEBV were higher for MY, FY and PY compared to the ones observed for SCS. In general, a slightly increase in theoretical reliabilities of GEBV were observed when including cows in the training population. However, similar theoretical reliabilities of GEBV were observed when including cows in the validation population. In addition, similar reliabilities were also estimated for PA when including or not cows in both training and validation populations.

Table 2. Average (and standard deviation) of theoretical reliabilities of parent average (REL_{PA}) and genomic estimated breeding value (REL_{GEBV}) using the accumulated and random regression methods, considering only bulls in the training and validation populations (Bulls + Bulls), bulls and cows in the training population and only bulls in the validation population (Bulls and cows + Bulls), and bulls and cows in the training and validation populations (Bulls and cows + Bulls and cows).

Method	**Trait**	**Bulls + Bulls**		**Bulls and cows + Bulls**		**Bulls and cows + Bulls and cows**	
		REL_{PA}	**REL_{GEBV}**	**REL_{PA}**	**REL_{GEBV}**	**REL_{PA}**	**REL_{GEBV}**
Accumulated	**MY**	0.40 (0.047)	0.72 (0.011)	0.40 (0.047)	0.75 (0.013)	0.40 (0.043)	0.75 (0.011)
	FY	0.40 (0.048)	0.72 (0.011)	0.41 (0.048)	0.75 (0.012)	0.40 (0.044)	0.76 (0.011)
	PY	0.40 (0.047)	0.72 (0.011)	0.40 (0.047)	0.73 (0.014)	0.40 (0.048)	0.73 (0.011)
	SCS	0.37 (0.047)	0.71 (0.011)	0.37 (0.047)	0.71 (0.013)	0.37 (0.049)	0.71 (0.011)
Mean		0.40	0.72	0.40	0.74	0.40	0.74
Regression	**MY**	0.40 (0.049)	0.73 (0.011)	0.40 (0.049)	0.76 (0.012)	0.40 (0.049)	0.76 (0.012)
	FY	0.41 (0.049)	0.73 (0.011)	0.40 (0.049)	0.76 (0.049)	0.40 (0.049)	0.76 (0.011)
	PY	0.40 (0.048)	0.73 (0.011)	0.40 (0.048)	0.75 (0.013)	0.40 (0.048)	0.75 (0.013)
	SCS	0.37 (0.049)	0.72 (0.010)	0.37 (0.049)	0.73 (0.013)	0.37 (0.049)	0.73 (0.013)
Mean		0.39	0.73	0.39	0.75	0.39	0.75

Traits: milk (MY), fat (FY) and protein (PY) yield, and somatic cells score (SCS).

GEBV validation reliability and bias

Validation reliabilities estimated for PA and GEBV using P305 and RRM for all traits, are shown in Table 3.

Table 3. Validation reliabilities estimated for parent average (r^2_{PA}) and for genomic estimated breeding value (r^2_{GEBV}) using the accumulated and random regression methods, considering only bulls in the training and validation populations (Bulls + Bulls), bulls and cows in the training population and only bulls in the validation population (Bulls and cows + Bulls), and bulls and cows in the training and validation populations (Bulls and cows + Bulls and cows).

Method	Trait	Bulls + Bulls		Bulls and cows + Bulls		Bulls and cows + Bulls and cows	
		r^2_{PA}	r^2_{GEBV}	r^2_{PA}	r^2_{GEBV}	r^2_{PA}	r^2_{GEBV}
Accumulated	MY1	0.43	0.55	0.42	0.52	0.35	0.47
	MY2	0.41	0.49	0.40	0.48	0.34	0.44
	MY3	0.40	0.50	0.40	0.49	0.37	0.46
	FY1	0.40	0.55	0.40	0.50	0.36	0.45
	FY2	0.39	0.53	0.38	0.50	0.37	0.44
	FY3	0.41	0.56	0.40	0.53	0.40	0.46
	PY1	0.41	0.52	0.41	0.45	0.40	0.44
	PY2	0.44	0.52	0.42	0.44	0.43	0.42
	PY3	0.46	0.53	0.45	0.48	0.47	0.47
	SCS1	0.31	0.42	0.31	0.42	0.32	0.38
	SCS2	0.37	0.44	0.37	0.45	0.37	0.41
	SCS3	0.38	0.44	0.38	0.44	0.39	0.38
Mean		0.40	0.50	0.40	0.48	0.38	0.45
Regression	MY1	0.42	0.54	0.42	0.48	0.36	0.48
	MY2	0.40	0.49	0.39	0.44	0.34	0.44
	MY3	0.40	0.50	0.40	0.46	0.37	0.44
	FY1	0.41	0.52	0.40	0.44	0.36	0.45
	FY2	0.37	0.52	0.38	0.47	0.37	0.45
	FY3	0.40	0.55	0.39	0.50	0.40	0.47
	PY1	0.41	0.51	0.42	0.42	0.40	0.43
	PY2	0.43	0.50	0.44	0.41	0.43	0.42
	PY3	0.46	0.53	0.47	0.46	0.46	0.46
	SCS1	0.30	0.40	0.31	0.40	0.31	0.40
	SCS2	0.35	0.41	0.37	0.41	0.37	0.41
	SCS3	0.38	0.40	0.38	0.38	0.38	0.41
Mean		0.39	0.49	0.40	0.44	0.38	0.44

Traits: milk (MY1, MY2, MY3), fat (FY1, FY2, FY3) and protein (PY1, PY2, PY3) yield, and somatic cells score (SCS1, SCS2, SCS3) in the first, second and third lactation, respectively.

Similar validation reliabilities were estimated for PA using either P305 or RRM. However, generally reliabilities estimated for GEBV using P305 were higher than using RRM. Validation reliabilities estimated for GEBV were higher than validation reliabilities estimated for PA using both P305 and RRM. Higher validation reliabilities were estimated for PA and for GEBV when considering only bulls in the training and validation populations compared to the scenario where cows were included, using both P305 and RRM. Regression coefficients estimated using either P305 or RRM are shown in Table 4.

Table 4. Regression coefficients (± standard error) estimated for parent average (b_{PA}) and for genomic estimated breeding value (b_{GEBV}) using the accumulated and random regression methods, considering only bulls in the training and validation populations (Bulls + Bulls), bulls and cows in the training population and only bulls in the validation population (Bulls and cows + Bulls), and bulls and cows in the training and validation populations (Bulls and cows + Bulls and cows).

Method	**Trait**	**Bulls + Bulls**		**Bulls and cows + Bulls**		**Bulls and cows + Bulls and cows**	
		b_{PA}	b_{GEBV}	b_{PA}	b_{GEBV}	b_{PA}	b_{GEBV}
Accumulated	**MY1**	0.73±0.037	1.04±0.039	0.72±0.028	0.75±0.023	0.70±0.025	0.74±0.020
	MY2	0.71±0.041	1.01±0.033	0.72±0.029	0.76±0.025	0.72±0.026	0.75±0.022
	MY3	0.73±0.036	0.98±0.039	0.71±0.027	0.74±0.024	0.70±0.024	0.72±0.022
	FY1	0.66±0.040	0.98±0.037	0.64±0.025	0.69±0.023	0.64±0.022	0.69±0.019
	FY2	0.70±0.032	1.03±0.037	0.69±0.026	0.71±0.023	0.70±0.023	0.71±0.018
	FY3	0.71±0.031	1.03±0.035	0.71±0.026	0.73±0.023	0.70±0.023	0.73±0.019
	PY1	0.65±0.029	0.96±0.034	0.66±0.023	0.68±0.026	0.67±0.021	0.68±0.023
	PY2	0.72±0.032	0.98±0.036	0.71±0.024	0.71±0.027	0.70±0.024	0.70±0.023
	PY3	0.75±0.033	0.97±0.033	0.72±0.022	0.74±0.026	0.72±0.021	0.74±0.022
	SCS1	0.88±0.049	0.98±0.040	0.87±0.039	0.85±0.033	0.85±0.038	0.85±0.030
	SCS2	0.89±0.048	0.92±0.043	0.89±0.034	0.91±0.030	0.88±0.032	0.89±0.028
	SCS3	0.88±0.048	0.90±0.041	0.88±0.032	0.90±0.030	0.85±0.030	0.87±0.028
Mean		0.75	0.98	0.74	0.76	0.73	0.76
Regression	**MY1**	0.73±0.037	0.99±0.035	0.72±0.027	0.75±0.021	0.71±0.027	0.75±0.017
	MY2	0.73±0.038	0.94±0.037	0.73±0.028	0.77±0.022	0.71±0.027	0.75±0.019
	MY3	0.72±0.036	0.96±0.036	0.71±0.026	0.75±0.021	0.70±0.025	0.74±0.020
	FY1	0.69±0.041	0.91±0.035	0.74±0.024	0.76±0.024	0.67±0.022	0.72±0.022
	FY2	0.73±0.032	0.92±0.035	0.70±0.025	0.74±0.019	0.70±0.023	0.73±0.019
	FY3	0.71±0.031	0.93±0.034	0.71±0.023	0.78±0.019	0.71±0.021	0.77±0.019
	PY1	0.66±0.029	0.88±0.034	0.69±0.023	0.70±0.021	0.68±0.020	0.69±0.018
	PY2	0.75±0.032	0.88±0.035	0.70±0.023	0.71±0.022	0.71±0.021	0.71±0.020
	PY3	0.76±0.033	0.90±0.034	0.72±0.022	0.72±0.021	0.73±0.020	0.75±0.021
	SCS1	0.90±0.048	0.84±0.045	0.86±0.037	0.86±0.029	0.87±0.036	0.88±0.028
	SCS2	0.89±0.048	0.89±0.048	0.88±0.033	0.90±0.027	0.88±0.032	0.91±0.026
	SCS3	0.89±0.048	0.87±0.047	0.88±0.032	0.90±0.029	0.86±0.032	0.90±0.027
Mean		0.76	0.91	0.75	0.78	0.74	0.78

Traits: milk (MY1, MY2, MY3), fat (FY1, FY2, FY3) and protein (PY1, PY2, PY3) yield, and somatic cells score (SCS1, SCS2, SCS3) in the first, second and third lactation, respectively.

Overall, regression coefficients closer to 1.0 were found using the RRM method for both PA and GEBV than using P305. The most biased estimates (regression coefficients deviating from 1.0) were observed for PA using the P305 method. Less biased GEBVs were observed when considering only bulls in the training and validation populations compared to the scenario where cows were also included in the training population. Similar regression coefficients were estimated when including or not cows in the validation population.

Validation reliability and bias of GEBVs over DIM

Validation reliability and bias pattern of GEBVs over DIM for each trait, estimated using only bulls in the training and validation populations are shown in Figure 1. In summary, different validation reliability and bias patterns for the GEBVs along the lactation curve were observed for different traits and lactations.

6.5. Discussion

GEBV theoretical reliabilities

Theoretical reliabilities are related to the amount of information available in the MME. Therefore, the use of RRM to estimate daily DGVs (and consequently daily GEBVs), increased the amount of information available in the MME, which yielded a slightly increase in theoretical reliabilities of GEBVs obtained from the RRM when compared to P305. For both RRM and P305, including genomic information in the estimation of breeding values increased the theoretical reliabilities for all traits (Table 2), which is in agreement with other studies (e.g., VanRaden et al., 2009, Ding et al., 2013).

In general, theoretical reliabilities estimated in this study were high (greater than 0.37 for PA and 0.71 for GEBV). This is related to the reduction in the PEVs due to the use of deregressed EBVs as pseudo-phenotypes (compared to the use of actual phenotypes), and to the high reliability threshold used to select animals included in the training and validation populations. Lower theoretical reliabilities were obtained for SCS due to the lower heritability and the smaller number of animals in the **G** matrix for this trait when compared to the other traits, which consequently reflects in smaller amount of information. The reason for the smaller number of genotyped animals in the **G** matrix for SCS is the threshold for minimum reliability used to define the training and validation populations.

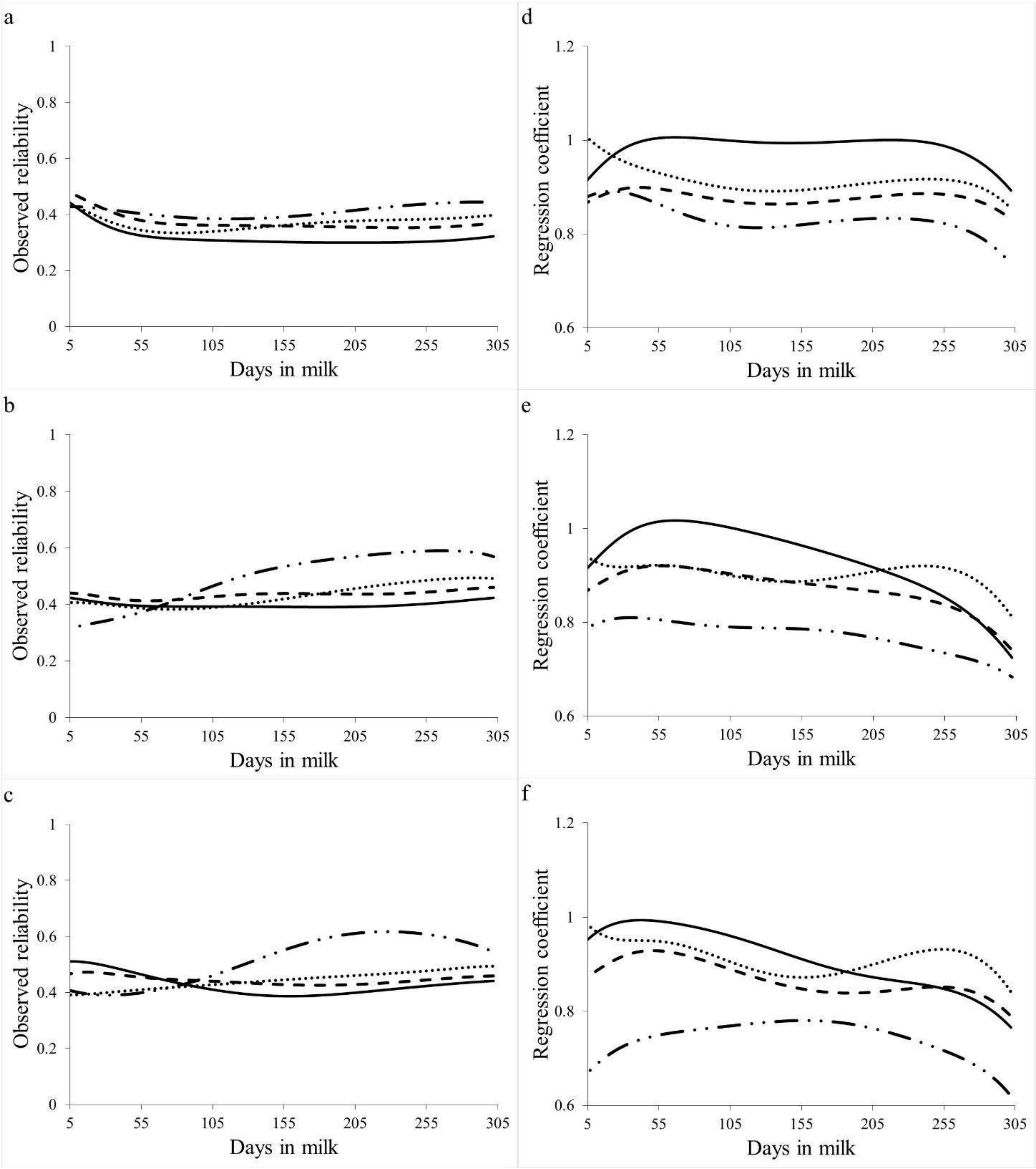

Figure 1. Validation reliability (a, b, c) and bias (regression coefficients; d, e, f) patterns of genomic estimated breeding values over days in milk, for milk (——), fat (• • •) and protein yield (- - -) and somatic cell score (▬ · · ▬) from the first (a, d), second (b, e) and third (c, f) lactation.

GEBV validation reliability and bias

Similar validation reliabilities estimated for PA were obtained using P305 and RRM (Table 3) because the accumulated PAs and EBVs were estimated based on daily PAs and EBVs, which makes them more highly correlated. Using RRM, the validation reliabilities of GEBVs for almost all traits were lower compared to P305 models. This may be expected, as it is more difficult to predict GEBVs for each DIM, resulting in larger total amount of estimation error when compared to the use of P305. In general, the inclusion of genomic information increased the validation reliability for almost all traits, because GEBVs take also into account the Mendelian sampling term and, therefore, they are a more useful estimate of the genetic merit of young animals for the lactation pattern compared to the PA. Validation reliabilities for GEBVs estimated in this study were similar to those reported in a study using single-step GBLUP methodology and a multiple-lactation random regression test-day model for genomic predictions in Japanese Holstein cattle (Baba et al., 2016).

Linear regression coefficients of EBV on GEBV lower than 1.0 may indicate overestimation of the GEBV. Regression coefficients closer to 1.0 were found using RRM (Table 4), likely because currently selection based on the pattern of the lactation curve in dairy cattle has been based solely on lactation persistency measurements, a more recent criterion than other milk production traits (e.g., accumulated yield). Regression coefficients of GEBVs estimated in this study were not compared to other studies using RRM, as they have used the single-step GBLUP approach and alternative scaling factors to combine the genomic (**G**) and pedigree-based additive (**A**) relationship matrices (Koivula et al., 2015; Baba et al., 2016). Scaling factors used to adjusting the **G** matrix toward their expected values in the **A** matrix have been shown to have a high impact on bias in single-step genomic evaluations (Vitezica et al., 2011; Christensen et al., 2012; Koivula et al., 2015).

Validation reliability and bias of GEBVs over DIM

Using the complete lactation curve to genetically select animals has been a complex task over the past decades. In order to simplify the genetic evaluation process in breeding programs around the world, the feasibility of selecting for specific lactation curve patterns or traits derived from it, such as lactation persistence and lactation peak, has been investigated. In addition, there is a great interest from the industry on deriving additional traits based on the lactation pattern and, for instance, its association with metabolic diseases (Miglior et al., 2017). Our study shows that genomics is a feasible alternative to accurately predict the lactation pattern of Holstein animals at an early age (without own phenotypes or progeny information), which would be useful when evaluating traits associated with the lactation pattern. The different validation reliability and bias patterns for the GEBVs along the lactation curve, for different traits and lactations, suggest that selection based on different lactation periods may have different usefulness (Figure 1). Furthermore, it implies that different genes might be expressed over DIM and, consequently, selection based on different lactation periods may have distinct genetic consequences on the lactation curve. Thus, this study indicates the need to re-evaluate statistical methods that have been used for selecting for lactation pattern (e.g., lactation persistency), since they were developed before the wide availability and use of genomic information (e.g., Rowlands et al., 1982; Cole and Null, 2009). The investigation of appropriate parameters to better summarize the lactation curve pattern when including genomic information may increase genetic progress for production traits, while contributing to improve health, welfare and fertility traits in dairy cattle. In addition, genome-wide association studies (**GWAS**) considering the whole lactation curve may help to better understand the genetic background of the biology of the lactation curve. Comparisons of results found here to other studies were not possible, as the studies found in the literature based on RRM have not reported reliabilities and bias over DIM (Koivula et al., 2015; Baba et al., 2016; Jattawa et al., 2016).

Impact of incorporating cows in the training and validation populations

The inclusion of cows in the training population yielded a slightly increase in the theoretical reliabilities of GEBVs (Table 2), which may be related to the increase in the amount of information available for the analyses. These findings are in agreement with previous studies that also showed that adding cows in the training population may significantly improve the theoretical reliabilities (Gao et al., 2015; Su et al., 2016). Although several studies have evaluated the effect of including cows in the training population of different dairy cattle breeds (e.g., Gao et al., 2015; Koivula et al., 2016; Su et al., 2016), to our best knowledge, there are no studies evaluating the effect of including cows in the validation population, especially under a RRM approach. Similar theoretical reliabilities were observed for GEBVs when including cows in the validation population, using both the P305 and RRM methods. In addition, similar theoretical reliabilities of PA were estimated when including or not cows in the training and validation populations.

A slightly increase in the validation reliabilities was observed when including only bulls in the training and validation populations compared to the inclusion of both bulls and cows in training population (Table 3). This may be related to the less reliable deregressed EBVs from cows compared to bulls (Uemoto et al., 2017). Reliable deregressed EBVs in the training population are required to accurately estimate the marker effects (and GEBVs), whereas reliable information in the validation population is required to validate the GEBVs properly. In general, about 10 additional cows are required to yield the same amount of information provided by a single progeny-tested bull (Weller, 2016). Nowadays, official genomic evaluations for the Holstein breed in Canada only include bulls in the training and validation populations (Interbull, 2017). In addition, various studies have included only bulls in the validation population (e.g., Gao et al, 2015; Koivula et al., 2015; Baba et al., 2016).

The inclusion of cows in the training population made the regression coefficients for the GEBVs to be lower than 1 and more similar to the regression coefficients observed for the PAs compared to when only bulls were included in the training population, in which the coefficients were all close to 1 (Table 4). This may be related to the difference of EBV scale between cows and bulls, which can be caused, for instance, by cow preferential treatment (Wiggans et al., 2011). To solve this scaling problem, official evaluations in USA consider some adjustment factors (Wiggans et al., 2011). No ad-hoc adjustments were used in our study. The inclusion of cows in the validation population does not seem to affect the regression coefficients estimated for PA and GEBVs (Table 4).

6.6. Conclusions

Validation reliability of RRM GEBVs was in general lower than the validation reliability for P305 model. Nevertheless, overall RRM provided less biased GEBVs compared to P305 model. More importantly, the use of RRM in two-step GWS produced fairly accurate GEBVs over the entire lactation curve for milk, fat and protein yields, and somatic cell score. Thus, selecting young animals based on the pattern of genomically predicted lactation curves seems feasible in Holstein cattle.

6.7. Acknowledgements

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES). The authors gratefully acknowledge the financial support to this research from Agriculture and Agri-Food Canada and by additional contributions from Dairy

Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative.

6.8. References

Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. 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(2):743-752. doi: 10.3168/jds.2009-2730.

Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. In Proc. 10th World Congr. Genet. Appl. Livest. Prod. Vancouver, Canada. doi: 10.13140/2.1.4801.5045.

Aliloo, H., J. E. Pryce, O. González-Recio, B. G. Cocks, and B. J. Hayes. 2016. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. Gen. Sel. Evol. 48(8):1-11. doi: 10.1186/s12711-016-0186-0.

Baba, T., Y. Gotoh, S. Yamaguchi, S. Nakagawa, H. Abe, Y. Masuda, and T. Kawahara. 2016. Application of single-step genomic best linear unbiased prediction with a multiple-lactation random regression test-day model for Japanese Holsteins. Animal Sci. J. 88(8):1226-1231. doi: 10.1111/asj.12760.

Brown, A., J. Ojango, J. Gibson, M. Coffey, M. Okeyo, and R. Mrode. 2016. Short communication: Genomic selection in a crossbred cattle population using data from the Dairy Genetics East Africa Project. J. Dairy Sci. 99(9):7308-7312. doi: http://dx.doi.org/10.3168/jds.2016-11083.

Christensen, O. F. and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Gen. Sel. Evol. 42(2):1-8. doi: https://doi.org/10.1186/1297-9686-42-2.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571. doi: 10.1017/S1751731112000742.

Cole, J. B., D. J. Null. 2009. Genetic evaluation of lactation persistency for five breeds of dairy cattle. J. Dairy Sci. 92(5):2248-2258. doi: https://doi.org/10.3168/jds.2008-1825.

Ding, X., Z. Zhang, X. Li, S. Wang, X. Wu, D. Sun, Y. Yu, J. Liu, Y. Wang, and Y. Zhang. 2013. Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. J. Dairy Sci. 96(8):5315-5323. doi: 10.3168/jds.2012-6194.

Gao, H., P. Madsen, U. Nielsen, G. Aamand, G. Su, K. Byskov, and J. Jensen. 2015. Including different groups of genotyped females for genomic prediction in a Nordic Jersey population. J. Dairy Sci. 98(12):9051-9059. doi: 10.3168/jds.2015-9947.

Garrick, D. J., J. F. Taylor, an R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Gen. Sel. Evol. 41(1):1-8. doi: 10.1186/1297-9686-41-55.

Hayes, B. J., P. J. Bowman, A. Chamberlain, and M. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92(2):433-443. doi: 10.3168/jds.2008-1646.

Interbull. 2017. Description of national genomic evaluation systems. Canada. Accessed May. 02, 2018. http://www.interbull.org/ib/nationalgenoforms.

Interbull. 2018a. Description of national genetic evaluation system and trend validation for production traits. Canada. Accessed May. 02, 2018. http://www.interbull.org/ib/geforms.

Interbull. 2018b. Description of national genetic evaluation system and trend validation for milk somatic cell and clinical mastitis. Canada. Accessed May. 02, 2018. http://www.interbull.org/ib/geforms

Jattawa, D., M. A. Elzo, S. Koonawootrittriron, and T. Suwanasopee. 2016. Genomic-polygenic and polygenic evaluations for milk yield and fat percentage using random regression models with Legendre polynomials in a Thai multibreed dairy population. Liv. Sci. 188:133-141. doi: https://doi.org/10.1016/j.livsci.2016.04.019.

Jenko, J., G. Wiggans, T. Cooper, S. Eaglen, W. d. L. Luff, M. Bichard, R. Pong-Wong, and J. Woolliams. 2017. Cow genotyping strategies for genomic selection in a small dairy cattle population. J. Dairy Sci. 100:439–452. doi: https://doi.org/10.3168/jds.2016-11479.

Jensen, J. 2001. Genetic evaluation of dairy cattle using test-day models. J. Dairy Sci. 84(12):2803-2812. doi: 10.3168/jds.S0022-0302(01)74736-4.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics. 124(4):979-993.

Koivula, M., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal, 10(06):1061-1066. doi: 10.1017/S1751731115002864.

Koivula, M., I. Strandén, J. Pösö, G. Aamand, and E. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98(4):2775-2784. doi: 10.3168/jds.2014-8975.

Larmer, S., M. Sargolzaei, and F. Schenkel. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. J. Dairy Sci. 97(5):3128-3141. doi: 10.3168/jds.2013-6826.

Miglior, F., A. Fleming, F. Malchiodi, L. F. Brito, P. Martin, and C. F. Baes. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100(12):10251-10271. doi: https://doi.org/10.3168/jds.2017-12968.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92(9):4648-4655. doi: 10.3168/jds.2009-2064.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. Lee. 2002. BLUPF90 and related programs (BGF90). Pages 21-22 in Proc. 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France.

Mrode, R. A. 2014. Linear models for the prediction of animal breeding values, 3rd ed. CAB Int., Wallingford, UK.

Oliveira, H. R., F. F. Silva, L. F. Brito, A. R. Guarini, J. Jamrozik, and F. S. Schenkel. 2018. Comparing deregression methods for genomic prediction of test-day traits in dairy cattle. J. Anim. Breed. Genet. 135(2):97-106. doi: 10.1111/jbg.12317.

Přibyl, J., P. Madsen, J. Bauer, J. Přibylová, M. Šimečková, L. Vostrý, and L. Zavadilová. 2013. Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. J. Dairy Sci. 96(3):1865-1873. doi: 10.3168/jds.2012-6157.

Rowlands, G., S. Lucey, and A. Russell. 1982. A comparison of different models of the lactation curve in dairy cattle. Animal Sci. 35(1):135-144. doi:10.1017/S0003356100000908

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2011. FImpute - An efficient imputation algorithm for dairy cattle populations. J. Dairy Sci. 94(E-Suppl. 1):421. (Abstr.)

Schaeffer, L. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Animal Breed. Gen. 123(4):218-223. doi: 10.1111/j.1439-0388.2006.00595.x.

Schaeffer, L., J. Jamrozik, G. Kistemaker, and J. Van Doormaal. 2000. Experience with a test-day model. J. Dairy Sci. 83(5):1135-1144. doi: https://doi.org/10.3168/jds.S0022-0302(00)74979-4.

Su, G., P. Ma, U. S. Nielsen, G. P. Aamand, G. Wiggans, B. Guldbrandtsen, and M. Lund. 2016. Sharing reference data and including cows in the reference population improve genomic predictions in Danish Jersey. Animal 10(6):1067-1075. doi: 10.1017/S1751731115001792.

Uemoto, Y., T. Osawa, and J. Saburi. 2017. Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. Animal, 11(3):382-393. doi: 10.1017/S1751731116001762.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11):4414–4423. doi: 10.3168/jds.2007-0980.

VanRaden, P., C. Van Tassell, G. Wiggans, T. Sonstegard, R. Schnabel, J. Taylor, and F. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92(1):16-24. doi: 10.3168/jds.2008-1514.

VanRaden, P. and G. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74(8):2737-2746. doi: 10.3168/jds.S0022-0302(91)78453-1.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb.) 93:357–366. doi:10.1017/S001667231100022X.

Weller, J. I. 2016. Genomic selection in animals. John Wiley & Sons, Inc., Hoboken, New Jersey.

Wiggans, G., T. Cooper, P. VanRaden, and J. Cole. 2011. Technical note: Adjustment of traditional cow evaluations to improve accuracy of genomic predictions. J. Dairy Sci. 94(12):6188-6193. doi: 10.3168/jds.2011-4481.

Wiggans, G., T. Sonstegard, P. VanRaden, L. Matukumalli, R. Schnabel, J. Taylor, F. Schenkel, and C. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92(7):3431-3436. doi: 10.3168/jds.2008-1758.

Winkelman, A., D. Johnson, and B. Harris. 2015. Application of genomic evaluation to dairy cattle in New Zealand. J. Dairy Sci. 98(1):659-675. doi: 10.3168/jds.2014-8560.

6.9. Supplementary material

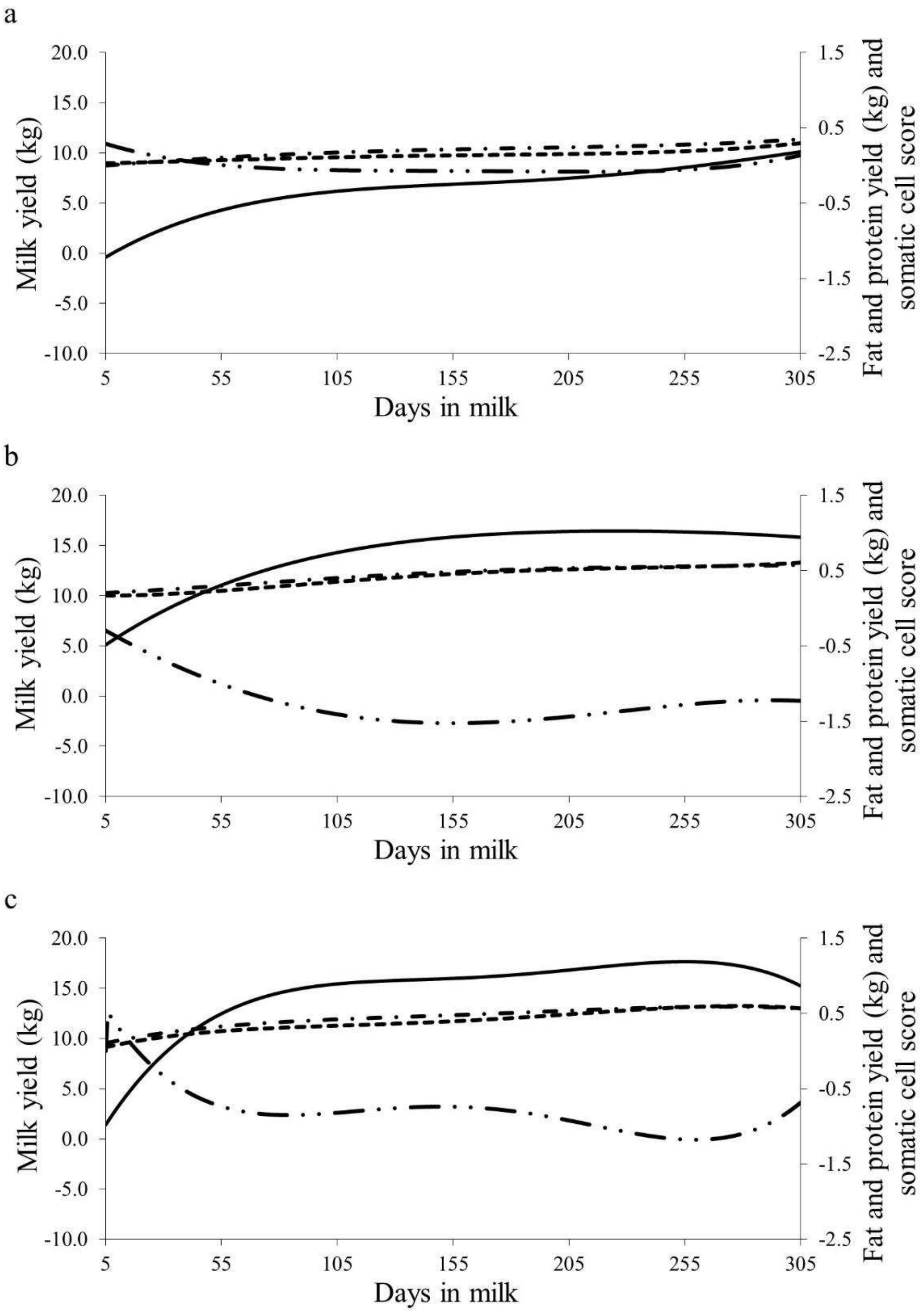

Figure S1. Pattern of the average estimated breeding values over days in milk for milk (——), fat (• • •) and protein yield (– – –) and somatic cell score (▬ · · ▬) from the first (a), second (b) and third (c) lactation.

CHAPTER 7:
Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle

Hinayah Rojas de Oliveira[1,2], Daniela Andressa Lino Lourenco[3], Yutaka Matsuda[3], Ignacy Misztal[3], Shogo Tsuruta[3], Janusz Jamrozik[1,4], Luiz Fernando Brito[1,5], Fabyano Fonseca e Silva[2], Flávio Schramm Schenkel[1]

[1] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

[2] Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[3] Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, United States of America

[4] Canadian Dairy Network, Guelph, Ontario, Canada

[5] Department of Animal Sciences, Purdue University, West Lafayette, Indiana, 47907, United States of America

Journal of Dairy Science, 2018 (submitted).

7.1. Abstract

Test-day traits are important for genetic evaluation in dairy cattle and are better modelled by multiple-trait Random Regression Models (**RRMs**). The reliability and bias of Genomic Estimated Breeding Values (**GEBVs**) estimated using multiple-trait RRMs via single-step Genomic Best Linear Unbiased Prediction (**ssGBLUP**) were investigated in the three major dairy cattle breeds in Canada, i.e., Ayrshire, Holstein, and Jersey. Individual additive genomic random regression coefficients for the test-day traits were predicted using two multiple-trait RRMs: 1) one for milk, fat and protein yields in the first, second and third lactations; and 2) one for somatic cell score in the first, second and third lactations. The predicted coefficients were used to derive GEBVs for each lactation day and, subsequently, the daily GEBVs were compared to the traditional daily parent averages obtained by BLUP. To assure compatibility between pedigree and genomic information for genotyped animals, different scaling factors for combining the inverse of genomic ($\mathbf{G}^{-1}$) and pedigree ($\mathbf{A}_{22}^{-1}$) relationship matrices were tested. In addition, the inclusion of only genotypes from animals with accurate breeding values (defined in preliminary analysis) was compared to the inclusion of all available genotypes in the analyzes. The ssGBLUP model led to considerably larger validation reliabilities than the BLUP model without genomic information. In general, scaling factors used to combine the $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ matrices had small influence on the validation reliabilities. However, greater impact was observed in the inflation of GEBVs. Less inflated GEBVs were obtained by the ssGBLUP compared to the parent average from traditional BLUP when using optimal scaling factors to combine the $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ matrices. Similar results were observed when including either all available genotypes or only genotypes from animals with accurate breeding values. These findings indicate that ssGBLUP using multiple-trait RRMs increases reliability and reduces bias of breeding values of young animals when compared to parent average from traditional BLUP in the Canadian Ayrshire, Holstein, and Jersey breeds.

Keywords: Ayrshire, Holstein, Jersey, longitudinal traits

7.2. Introduction

Random regression models (**RRMs**; Schaeffer, 2004) allow to take into account the well-structured covariance pattern among the repeated measurements in genetic analysis of test-day traits. The RRMs are especially useful under a multiple-trait approach, since it takes into account genetic and environmental correlations between different traits and/or lactations over days in milk (**DIM**), which enable the identification, for instance, of the time periods that would yield the most favorably correlated genetic responses (Oliveira et al., 2016). Including genomic information to evaluate test-day traits using multiple-trait RRMs may result in more accurate breeding values at a young age and therefore, facilitate the identification and selection of animals with better-shaped lactation curves.

Nowadays, the majority of the breeding programs have implemented genomic selection based on a multiple-step approach (Interbull, 2017). In brief, the prediction of Genomic Estimated Breeding Values (**GEBVs**) based on the multiple-step evaluation for test-day traits requires: (1) prediction of traditional Estimated Breeding Values (**EBVs**) for each day, based on RRMs; (2) estimation of the accumulated yield EBV and subsequent de-regression; (3) prediction of the accumulated Direct Genomic Values (**DGVs**); and (4) combination of accumulated Direct Genomic Values (**DGVs**) and Parent Average (**PA**) to generate accumulated GEBVs. Thus, even though RRMs have routinely been used in genetic evaluation of test-day traits in several countries (Interbull, 2018), usually GEBVs have been obtained for the accumulated yield, which does not allow selecting for lactation curve pattern. Furthermore, using RRM in all steps of the multiple-step evaluations to estimate GEBVs for all DIM does not seem to substantially increase the reliability of the genomic predictions when compared to the reliability of accumulated GEBVs (Oliveira et al., 2018).

Relevant studies have shown that simultaneously combining phenotypic records, pedigree, and genomic information in the single-step Genomic Best Linear Unbiased Prediction approach (**ssGBLUP**) can lead to more accurate and less biased GEBVs (Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010). In this context, Legarra et al. (2009) showed that genomic relationships are propagated to non-genotyped animals that are related to the genotyped ones through the **H** matrix, which combines the pedigree (**A**) and genomic (**G**) relationship matrices. However, sometimes **A** and **G** are not in the same scale (Misztal, 2017), which requires the use of optimal scaling factors to blend $\mathbf{G}^{-1}$ with the inverse of the pedigree relationship matrix for genotyped animals ($\mathbf{A}_{22}^{-1}$; Vitezica et al., 2011; Misztal et al., 2013). Previous studies already investigated the best combination of scaling factors for the evaluation of milk related traits using RRM in Nordic Red (Koivula et al., 2015) and Chinese (Kang et al., 2018) dairy cattle. However, ideal scaling factors can be population specific and are not available yet for the most important breeds in Canada.

The use of ssGBLUP based on RRMs enable the estimation of GEBVs for all DIM, which can increase the reliability of genomic predictions for test-day traits (Koivula et al., 2015; Kang et al., 2018). Despite the improvement in reliability, one of the main constrains for using ssGBLUP based on RRMs in national genomic evaluations is the computing time needed to analyze large datasets (Kang et al., 2017). Koivula et al. (2015) suggested that the number of genotyped animals may influence the number of iterations needed for ssGBLUP to converge, which reflects in the computing time. Thus, including only genotypes from animals with accurate EBVs in the analyses might optmize the computational efficiency of the ssGBLUP method. Notwithstanding, the effect of including only genotypes from animals with accurate EBVs instead of all available genotypes when performing ssGBLUP analyses on the GEBV preditions is still unknown, especially when using complex models, such as multiple-trait RRMs. Therefore, this study aimed to: a) investigate the reliability and bias of GEBVs predicted using ssGBLUP and test-day traits in Canadian Ayrshire, Holstein, and Jersey breeds; and b) evaluate the effect of including only genotypes from animals with accurate EBVs in the analyses.

7.3. Material and methods

Datasets and quality control

The Canadian Dairy Network (**CDN**; Guelph, ON, Canada) provided the December 2016 official pedigree, genotypes, and milk, fat and protein yields, and somatic cell count test-day records from the first three lactations of Ayrshire, Holstein, and Jersey breeds. Traits within each breed were coded as MY1, FY1, PY1, SCC1, MY2, FY2, PY2, SCC2, MY3, FY3, PY3 and SCC3, where MY, FY, PY and SCC refers to milk, fat and protein yields, and somatic cell count, respectively, and 1, 2 and 3 refers to the first, second and third lactations. To analyze all traits assuming a Gaussian distribution, SCC was converted to somatic cell score (**SCS**), using the standard definition implemented in the Canadian official genetic evaluations, which can be described as (Jamrozik et al., 2010; Jamrozik and Schaeffer, 2011; Do et al., 2018):

$$\mathrm{SCS} = \log_2(\mathrm{SCC}/100{,}000) + 3.$$

Thereafter, SCS have also been coded as SCS1, SCS2 and SCS3 to refer to the first, second, and third lactations, respectively. Outlier phenotypes for each trait (μ ± 3 SD, within each contemporary group, which were defined by the combination of herd and test-day), were removed from further analyses. Considering all traits together, a total of 8,260 (0.30%), 84,040 (0.12%), and 13,185 (0.71%) test-day records were identified as outliers (and consequently removed from the analyses) for the Ayrshire, Holstein and Jersey breeds, respectively. In addition, it was required that each contemporary group had at least 3 animals. This quality control filtering excluded 71,740

(2.85%), 444,365 (0.61%) and 151,711 (7.18%) test-day records for Ayrshire, Holstein and Jersey breeds, respectively. No restriction was imposed on the minimum number of records per lactation per animal. Test-day trait means for all breeds are shown in Table S1, in the Supplementary Material.

The number of animals genotyped with: low-density (< 20K) SNP panels was 300, 2,270 and 202; with medium-density (> 20K and < 60K) SNP panels was 944, 17,105, and 1,157; and with high-density (> 60K) SNP panels was 749, 2,459, and 158, for Ayrshire, Holstein and Jersey, respectively. All analyses in this study were performed using 50K SNP panel. Therefore, only single nucleotide polymorphisms (**SNPs**) present on the 50K panel were kept for the animals genotyped using high density panels. In addition, animals genotyped with low- and medium-density panels were imputed by CDN to 50K using the FImpute software (Sargolzaei et al., 2011), for each breed separately. For the imputation, SNPs were not filtered for minor allele frequency (**MAF**), because they were also used for determining extent of linkage disequilibrium (**LD**) and gametic phase consistency. After the genotype imputation process performed by CDN, SNPs with Mendelian conflicts, call rate less than 0.90, MAF less than 0.05 and with a difference between observed and expected heterozygosity higher than 0.15 (Wiggans et al., 2009) were removed using the preGSF90 software (Aguilar et al., 2014; Misztal et al., 2014). The total number of SNPs on the 29 bovine autosomes (**BTA**) that remained after the quality control was 38,096, 40,658 and 34,500, for the Ayrshire, Holstein and Jersey breeds, respectively.

After the quality control, a reduced dataset was created from the full dataset (i.e., from the dataset that included all information up to December 2016) by excluding the last 4 years. Therefore, the reduced dataset included all phenotypic information available until 2012 and it was used to predict GEBVs using ssGBLUP, and EBVs (or PAs for validation animals) using the traditional BLUP. The full dataset was used to predict current EBVs, which were used as a benchmark to validate GEBVs and PAs obtained from the reduced dataset for validation animals by assessing the reliability and bias of genomic predictions. Validation animals were defined as genotyped bulls with minimum trait EBV reliability of 0.65 for the Ayrshire and Jersey breeds, and 0.80 for the Holstein breed. In addition, validation bulls were required to be born from 2007 to 2010 and to have only daughters born after 2012. A total of 97, 1,160, and 88 bulls were used as validation animals for Ayrshire, Holstein and Jersey, respectively, in the multiple-trait analyses of MY, FY and PY. For SCS, a total of 87, 1,083, and 82 bulls were used as validation animals for Ayrshire, Holstein and Jersey, respectively. We have opted for using small validation populations of animals with more reliable trait EBVs in this study due to the small benefit of including animals with less reliable EBVs to increase the validation population's size, as reported by Oliveira et al. (2018). Details on the approach used to evaluate the prediction reliability and bias for the validation animals are given in the "Prediction Reliability and Bias" section.

Animals that were not considered as validation bulls and that were born from 1957 to 2010 were included in the training population. It is worth noting that the term "training population" is used here to refer to all non-validation animals, however, not all of them have both phenotype and genotype information. To evaluate the effect of including in the training population genotypes from animals with accurate EBVs only, two scenarios were created based on the training animals: 1) including only genotypes from animals with accurate EBVs (i.e., genotypes were included in the analyses when the EBV reliability for PY or SCS, estimated in a preliminary analyses, was higher than 0.50 for the Ayrshire and Jersey breeds and higher than 0.65 for the Holstein breed); and 2) including genotypes from all genotyped animals. The preliminary analyses were conducted using animal model and the traditional BLUP method, based on the same RRMs used in this study (the model is described in details in the "Statistical Analyses" section). Thereafter, the EBV reliabilities were estimated as an approximation of reliability of 305d yield, as described in Jamrozik et al. (2000). Animals that did not satisfy the criteria to be included in the validation or training populations were excluded from the analyses. Average EBV reliabilities, SD, and the corresponding 95% confidence interval for the first and second scenarios estimated in the preliminary analyses are shown in Table S2 (Supplementary Material). The total number of training animals and test-day records in the reduced and full datasets, and the number of genotyped animals in each scenario, for the different analyses and breeds, are given in Table 1.

It is worth mentioning that the number of test-day records and animals in the pedigree file of the Holstein breed used in this study are higher than the number used in Kang et al. (2018), who analyzed Chinese Holstein (2.8 million test-day records and 0.3 million animals in the pedigree file). The datasets used by Koivula et al. (2015), 85 million test-day records and 5.1 million animals in the pedigree file, are similar to the Holstein datasets used in the current study. However, Koivula et al. (2015) studied Nordic Red Dairy cattle, while in this study the focus was on the three main dairy cattle breeds in Canada, i.e. Ayrshire, Holstein and Jersey. In addition, Koivula et al. (2015) and Kang et al. (2018) did not analyzed SCS.

Table 1. Total number of animals included in the final pedigree file, number of test-day records (phenotypes) in the reduced and full datasets, and number of genotyped animals in both analyzed scenarios.

Trait[1]			**Ayrshire**	**Holstein**	**Jersey**
MY, FY and PY	**Pedigree[2]**	**Reduced**	204,429	5,085,542	157,718
		Full	221,533	5,976,711	185,737
	Phenotypes[3]	**Reduced**	2,143,941	59,724,786	1,353,185
		Full	2,446,662	72,363,090	1,688,987
	Genotypes[4,5]	**Scenario 1**	1,716 (542)	8,865 (7,370)	917 (547)
		Scenario 2	1,827 (634)	20,663 (8,517)	1,427 (849)
SCS	**Pedigree[2]**	**Reduced**	195,099	4,983,668	154,123
		Full	212,141	5,868,988	182,198
	Phenotypes[3]	**Reduced**	1,856,634	54,532,924	1,211,483
		Full	2,153,882	66,752,684	1,541,168
	Genotypes[4,5]	**Scenario 1**	1,141 (526)	7,329 (7,098)	524 (430)
		Scenario 2	1,827 (634)	20,663 (8,517)	1,427 (849)

[1]Milk (MY), fat (FY), and protein (PY) yields in the 1st, 2nd, and 3rd lactations were analyzed separately from somatic cell score (SCS) in the 1st, 2nd, and 3rd lactations. [2]Pedigree and [3]Phenoyopes are shown for the reduced dataset (all available information excluding the last 4 years), and for the full dataset. [4]Genotype scenarios: 1) including genotypes from animals with accurate estimated breeding values only; 2) including genotypes from all genotyped animals. [5]Genotypes: numbers inside brackets show the number of genotyped bulls.

Statistical analyses

Individual additive genetic random regression coefficients for each trait were predicted by the traditional BLUP method, using two different multiple-trait RRMs: one considering MY, FY and PY in the first three lactations; and another considering SCS in the first three lactations. The RRMs (Schaeffer et al., 2000) used in this study were similar to the multiple-trait RRMs used in official genetic evaluations performed by CDN in Canada (Interbull. 2018). In general, the multiple-trait RRM used for each breed can be described as:

$$y_{ijklm} = HTD_{ijl} + \sum_{o=1}^{5} \beta_{ojl} z_{ojkl} + \sum_{o=1}^{5} \delta_{ojl} z_{ojkl} + \sum_{o=1}^{5} \alpha_{ojkl} z_{ojkl} + \sum_{o=1}^{5} \rho_{ojkl} z_{ojkl} + e_{ijklm} ,$$

where y_{ijklm} is the mth phenotypic test-day record of animal k, for the trait l (i.e., MY, FY, PY or SCS), in the lactation j (i.e., first, second or third lactation); HTD_{ijl} is the effect of the ith herd-test day for the trait l in the lactation j; β_{ojl} is the fixed regression coefficient for age-parity-season of calving effect for the trait l in the lactation j; δ_{ojl} is the random regression coefficient for herd-year of calving effect for the trait l in the lactation j; α_{ojkl} is the random regression coefficient for the animal additive genetic effect for the animal k, trait l in the lactation j; ρ_{ojkl} is the random regression coefficient for the permanent environmental effect for the animal k, trait l in the lactation j; z_{ojkl} are the oth covariate related to the fifth order Legendre orthogonal polynomials (Kirkpatrick et al., 1990) for animal k, trait l and lactation j, and e_{ijklm} is the residual effect. The model assumptions were:

$$\mathrm{E}[\mathbf{y}] = \mathbf{X}\boldsymbol{\beta} \text{ and } \mathrm{Var}\begin{bmatrix}\boldsymbol{\delta}\\ \boldsymbol{\alpha}\\ \boldsymbol{\rho}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{I}\otimes\mathbf{HY}_0 & 0 & 0 & 0\\ 0 & \mathbf{A}\otimes\mathbf{G}_0 & 0 & 0\\ 0 & 0 & \mathbf{I}\otimes\mathbf{P}_0 & 0\\ 0 & 0 & 0 & \mathbf{R}\otimes\mathbf{R}_0\end{bmatrix},$$

where **y** is the vector of trait phenotypes sorted by test-day and trait/lactation within each animal, **X** is the incidence matrix for the fixed effects in the vector $\boldsymbol{\beta}$; $\mathbf{HY}_0$, $\mathbf{G}_0$ and $\mathbf{P}_0$ are the variance-covariance matrices among traits due to herd-year of calving, additive genetic and permanent environmental random effects, respectively; and, $\mathbf{R}_0$ is the variance-covariance matrix of residual effects among traits. The I is an identity matrix, **A** is the numerator relationship matrix obtained from pedigree information, and **R** is a diagonal matrix of residual variances sorted by test-day and trait/lactation, modeling the heterogeneity of residual variance in 10 different classes of DIM (first: 5 to 35 DIM; second: 36 to 65 DIM; third: 66 to 95 DIM; and so on).

In the ssGBLUP, the **A** was replaced by the **H** matrix, which combines the pedigree and genomic information. Usually the **H** is computationally demanding to compute, however its inverse has a simple structure (Aguilar et al., 2010; Christensen and Lund, 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix}0 & 0\\ 0 & \tau(0.95\mathbf{G}\text{-}0.05\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1}\end{bmatrix},$$

where $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix (calculated using the first method presented in VanRaden, 2008), $\mathbf{A}^{-1}$ is the inverse of the traditional relationship matrix, and $\mathbf{A}_{22}^{-1}$ is the inverse of the section of **A** related to genotyped animals. The $\mathbf{A}^{-1}$ included animals up to 10 generations back, and it accounted for inbreeding coefficients. In order to make **G** invertible, 0.05 of $\mathbf{A}_{22}^{-1}$ was added to 0.95 of **G**. The τ and ω parameters (i.e., scaling factors) were used to account for the reduced genetic variance and for different depths of pedigree, respectively, to make $\mathbf{G}^{-1}$ compatible with $\mathbf{A}_{22}^{-1}$ and also $\mathbf{A}^{-1}$. To evaluate the effect of including only genotypes from animals with accurate EBVs in the analyzes or genotypes from all animals, no scaling factors were used to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ in these two scenarios. After the definition of the best scenario with respect to the inclusion of genotyped animals in training population (based on validation reliabilities and regression coefficients), a total of three different values were tested for τ (i.e., 1.0, 1.5 and 2.0) and five for ω (i.e., 0.6, 0.7, 0.8, 0.9 and 1.0). Values for τ and ω were defined based on the literature, and the best combination of τ and ω parameters for each analyse of each breed was chosen according to the validation reliabilities and regression coefficients (bias).

The additive genetic (estimated by BLUP) and genomic (estimated by ssGBLUP) random regression coefficients were used to derive EBVs and GEBVs for each DIM, respectively. The vectors of EBVs and GEBVs for all DIM of each animal were obtained as follows:

$\mathbf{EBV}_{\mathrm{jkl}} = \mathbf{T}\hat{\boldsymbol{\alpha}}_{\mathrm{jkl}}$ and $\mathbf{GEBV}_{\mathrm{jkl}} = \mathbf{T}\hat{\boldsymbol{\delta}}_{\mathrm{jkl}}$,

where $\hat{\boldsymbol{\alpha}}_{\mathrm{jkl}}$ is the vector of predicted additive genetic coefficients for each animal k, trait l and lactation j; $\hat{\boldsymbol{\delta}}_{\mathrm{jkl}}$ is the vector of predicted genomic coefficients for each animal k, trait l and lactation j; and $\mathbf{T}$ is a matrix of independent covariates for every DIM (ranging from 5 to 305 days), associated with the Legendre polynomial function.

The BLUP90IOD software (Misztal et al., 2002; Misztal et al., 2014), which is based on iteration on data and on the Preconditioned Conjugate Gradient algorithm (**PCG**; Strandén and Lidauer, 1999; Tsuruta et al., 2001), was used to obtain the solutions of the mixed model equations for all analyses. The PCG algorithm converged when the relative difference between the right-hand and left-hand sides of the mixed model equations (**MME**) for all equations describing the genetic animal effects was smaller than 10^{-12}. The variance components used in this study for both BLUP and ssGBLUP analyses were those estimated in the December 2016 official genetic evaluation performed by CDN. Several ssGBLUP studies have also used the same variance components as the traditional BLUP evaluations (e.g., Koivula et al., 2012, Kang et al., 2017; Kang et al., 2018). The average heritabilities over DIM for MY1, FY1, PY1, SCS1, MY2, FY2, PY2, SCS2, MY3, FY3, PY3 and SCS3, as estimated by CDN, are shown in Table S3 (Supplementary Material).

Prediction reliability and bias

A proxy of the validation reliability for each trait was calculated as the squared Pearson correlation coefficient between daily GEBVs estimated based on the reduced dataset and daily EBVs estimated based on the full dataset, considering only animals in the validation populations. To assess the genomic prediction bias (i.e., inflation or deflation of the GEBVs relative to EBVs), regression coefficients (b_1) were estimated using a linear regression of the daily EBVs estimated based on the full dataset on the daily GEBVs estimated based on the reduced dataset (i.e., $\mathrm{EBV} = b_0 + b_1 \times \mathrm{GEBV}$), considering only animals in the validation population and all DIM together. To compare prediction reliability and bias of GEBVs to those of PA from traditional genetic evaluation, validation reliabilities and regression coefficients were also calculated using daily PAs and daily EBVs for the animals in the validation population, estimated based on the reduced and full datasets, respectively (i.e., squared Pearson correlation coefficient between EBVs and PAs, and b_1 was obtained from $\mathrm{EBV} = b_0 + b_1 \times \mathrm{PA}$).

Computational demand

To evaluate the feasibility of the use of ssGBLUP based on RRM, the computational demand was also investigated. The number of rounds (iterations) used by the PCG algorithm until the convergence, the average number of cores (processors), and the total computing time were recorded for each analysis. Total computing time was estimated as the amount of central processing unit (CPU) time spent within the process regarding all cores used. In other words, the amount of CPU time spent in user-mode code plus the amount of CPU time spent in the kernel, divided by the average number of cores that were used in the analysis. For the ssGBLUP analyses considering scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$, only the computational demand for the optimal scaling factor of each analysis was reported.

7.4. Results

Genotypes from animals with accurate EBVs only vs. from all genotypes

Overall, similar validation reliabilities and regression coefficients (bias) were observed when considering genotypes from only animals with accurate EBVs or from all genotyped animals (Tables 2 and 3). Including all available genotypes slightly increased the reliability for the Holstein and Jersey breeds. However, for the Ayrshire breed a small decrease in reliability was observed, as shown in Table 2.

Table 2. Validation reliabilities for genomic estimated breeding values using genotypes from animals with accurate estimated breeding values (EBVs) only ($r^2_{\mathrm{GEBV}_{\mathrm{acc}}}$) and from all ($r^2_{\mathrm{GEBV}_{\mathrm{all}}}$) genotyped animals for Ayrshire, Holstein, and Jersey breeds.

Trait	Ayrshire		Holstein		Jersey	
	$r^2_{\mathrm{GEBV}_{\mathrm{acc}}}$	$r^2_{\mathrm{GEBV}_{\mathrm{all}}}$	$r^2_{\mathrm{GEBV}_{\mathrm{acc}}}$	$r^2_{\mathrm{GEBV}_{\mathrm{all}}}$	$r^2_{\mathrm{GEBV}_{\mathrm{acc}}}$	$r^2_{\mathrm{GEBV}_{\mathrm{all}}}$
MY1	0.40	0.39	0.62	0.65	0.56	0.58
MY2	0.45	0.44	0.58	0.60	0.48	0.50
MY3	0.48	0.47	0.57	0.59	0.46	0.48
FY1	0.45	0.43	0.57	0.58	0.62	0.62
FY2	0.44	0.44	0.57	0.58	0.60	0.59
FY3	0.46	0.45	0.58	0.59	0.55	0.54
PY1	0.55	0.54	0.66	0.67	0.67	0.68
PY2	0.49	0.48	0.64	0.65	0.61	0.61
PY3	0.51	0.49	0.64	0.64	0.56	0.58
SCS1	0.42	0.43	0.42	0.45	0.28	0.31
SCS2	0.50	0.47	0.46	0.48	0.26	0.28
SCS3	0.49	0.45	0.44	0.46	0.36	0.37

Traits: milk (MY1, MY2, MY3), fat (FY1, FY2, FY3) and protein (PY1, PY2, PY3) yield, and somatic cell score (SCS1, SCS2, SCS3) in the first, second and third lactation, respectively. $r^2_{\mathrm{GEBV}_{\mathrm{acc}}}$: included genotypes from animals with accurate EBVs only (EBV reliability for PY or SCS higher than 0.50 for the Ayrshire and Jersey breeds, and higher than 0.65 for the Holstein breed).

Table 3. Regression coefficients (± SE) for genomic estimated breeding values using genotypes from animals with accurate estimated breeding values (EBVs) only ($b_{\mathrm{GEBV}_{\mathrm{acc}}}$) and from all ($b_{\mathrm{GEBV}_{\mathrm{all}}}$) genotyped animals for the Ayrshire, Holstein and Jersey breeds.

Trait	Ayrshire		Holstein		Jersey	
	$b_{\mathrm{GEBV}_{\mathrm{acc}}}$	$b_{\mathrm{GEBV}_{\mathrm{all}}}$	$b_{\mathrm{GEBV}_{\mathrm{acc}}}$	$b_{\mathrm{GEBV}_{\mathrm{all}}}$	$b_{\mathrm{GEBV}_{\mathrm{acc}}}$	$b_{\mathrm{GEBV}_{\mathrm{all}}}$
MY1	0.65±0.005	0.64±0.005	0.76±0.001	0.75±0.001	0.89±0.005	0.89±0.005
MY2	0.67±0.004	0.67±0.004	0.73±0.001	0.73±0.001	0.83±0.005	0.83±0.005
MY3	0.67±0.004	0.67±0.004	0.73±0.001	0.71±0.001	0.78±0.005	0.79±0.005
FY1	0.66±0.004	0.65±0.004	0.64±0.001	0.62±0.001	0.99±0.005	0.98±0.005
FY2	0.67±0.004	0.68±0.004	0.66±0.001	0.65±0.001	0.96±0.005	0.94±0.005
FY3	0.67±0.004	0.67±0.004	0.68±0.001	0.66±0.001	0.92±0.005	0.90±0.005
PY1	0.70±0.004	0.70±0.004	0.68±0.001	0.67±0.001	0.89±0.004	0.88±0.004
PY2	0.70±0.004	0.70±0.004	0.69±0.001	0.69±0.001	0.86±0.004	0.85±0.004
PY3	0.69±0.004	0.69±0.004	0.69±0.001	0.68±0.001	0.81±0.004	0.80±0.004
SCS1	0.97±0.007	0.95±0.007	0.68±0.001	0.70±0.001	0.76±0.008	0.77±0.007
SCS2	0.93±0.006	0.87±0.006	0.71±0.001	0.73±0.001	0.74±0.008	0.76±0.008
SCS3	0.90±0.005	0.84±0.006	0.70±0.001	0.71±0.001	0.79±0.007	0.83±0.007

Traits: milk (MY1, MY2, MY3), fat (FY1, FY2, FY3) and protein (PY1, PY2, PY3) yield, and somatic cell score (SCS1, SCS2, SCS3) in the first, second and third lactation, respectively. $b_{\mathrm{GEBV}_{\mathrm{acc}}}$: included genotypes from animals with accurate EBVs only (EBV reliability for PY or SCS higher than 0.50 for the Ayrshire and Jersey breeds, and higher than 0.65 for the Holstein breed).

With respect to the regression coefficients (Table 3), a slight increase in bias of GEBV (i.e., regression coefficients deviating from 1.0) were observed using all available genotypes for SCS in the Ayrshire breed (average regression coefficients considering only accurate and all animals were 0.93 and 0.89, respectively), MY, FY and PY in the Holstein breed (0.74 and 0.73 for MY; 0.66 and 0.64 for FY; and 0.69 and 0.68 for PY, respectively), and for FY and PY in the Jersey breed (0.96 and 0.94 for FY; and 0.85 and 0.84 for PY, respectively). The GEBVs were less biased for SCS when using all available genotypes in the Holstein and Jersey breeds (average regression coefficients considering only accurate and all animals were 0.70 and 0.71 in Holstein; and 0.76 and 0.79 in Jersey, respectively). No difference in average regression coefficients were observed for MY, FY and PY in the Ayrshire breed, and for MY in the Jersey breed (Table 3) using genotypes from all or accurate animals only.

ssGBLUP vs. traditional evaluation of test-day traits and impact of scaling factors

For almost all traits and breeds, the ssGBLUP yielded considerably higher validation reliabilities compared to the model without genomic information, even when using no scaling factors (default; $\tau_{1.0}\omega_{1.0}$) to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ (Tables S4, S5 and S6 at the Supplementary Material). However, regression coefficients deviating from 1.0 were obtained for the default ssGBLUP when compared to PA for Ayrshire and Jersey (except for FY in Jersey, where the PA usually had regression coefficients deviating from 1.0 for all lactations). On the other hand, regression coefficients obtained for the default ssGBLUP were closer to 1.0 when compared to PA for the Holstein breed (except for SCS2 and SCS3).

Scaling factors used to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ had, in general, small influence in the validation reliabilities. However, greater effect was observed in the regression coefficients. Less biased GEBVs were obtained from the ssGBLUP method compared to the PA when scaling factors were applied to the $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ for all traits and breeds (Tables S4, S5 and S6 at the Supplementary Material). The scaling factors used to combine $\mathbf{G}^{-1}$ (τ) and $\mathbf{A}_{22}^{-1}$ (ω) that produced the least biased predictions and the highest validation reliabilities were different for each breed and trait. For genomic evaluation of MY, FY and PY, $\tau = 2.0$ and $\omega = 0.6$ for the Ayrshire, $\tau = 1.5$ and $\omega = 0.6$ for the Holstein, and $\tau = 1.5$ and $\omega = 0.9$ for the Jersey breed yielded the least biased predictions and the highest validation reliabilities. For SCS, the optimal scaling factors were $\tau = 1.0$ and $\omega =$

0.9 for the Ayrshire, τ = 2.0 and ω = 0.6 for the Holstein, and τ = 1.0 and ω = 0.6 for the Jersey breed. Validation reliabilities and regression coefficients estimated using the optimal scaling factors for each breed are shown in Table 4.

Table 4. Validation reliabilities (r^2) and regression coefficients ($\hat{b}_1 \pm \mathrm{SE}$) for genomic breeding value estimated using all available genotypes and assuming the optimal[1] scaling factors (τ and ω) to blend $\mathbf{A}_{22}^{-1}$ and $\mathbf{G}^{-1}$ for Ayrshire, Holstein and Jersey breeds.

Breed	**Lactation**	**Milk yield**		**Fat yield**		**Protein yield**		**Somatic cell score**	
		$\mathbf{r}^2$	$\hat{\mathbf{b}}_1 \pm \mathbf{SE}$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1 \pm \mathbf{SE}$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1 \pm \mathbf{SE}$	$\mathbf{r}^2$	$\hat{\mathbf{b}}_1 \pm \mathbf{SE}$
Ayrshire	**First**	0.40	0.85±0.006	0.48	0.89±0.005	0.59	0.91±0.004	0.46	1.06±0.007
	Second	0.46	0.90±0.006	0.48	0.90±0.005	0.52	0.89±0.005	0.50	0.97±0.006
	Third	0.49	0.91±0.005	0.49	0.91±0.005	0.53	0.88±0.005	0.49	0.96±0.006
Holstein	**First**	0.66	0.88±0.001	0.59	0.75±0.001	0.69	0.77±0.001	0.44	0.94±0.002
	Second	0.61	0.86±0.001	0.60	0.78±0.001	0.67	0.79±0.001	0.50	0.97±0.001
	Third	0.61	0.85±0.001	0.61	0.80±0.001	0.66	0.78±0.001	0.48	0.96±0.002
Jersey	**First**	0.60	1.00±0.005	0.64	1.12±0.005	0.70	0.99±0.004	0.33	0.97±0.009
	Second	0.52	0.95±0.005	0.61	1.08±0.005	0.63	0.96±0.004	0.29	0.98±0.010
	Third	0.51	0.91±0.005	0.56	1.04±0.006	0.59	0.91±0.005	0.40	1.04±0.008

τ: scaling factor used for genomic relationship matrix ($\mathbf{G}^{-1}$); ω: scaling factor used for pedigree relationship matrix ($\mathbf{A}_{22}^{-1}$).[1]For genomic evaluation of milk, fat and protein yields the optimal scaling factors were: τ = 2.0 and ω = 0.6 for Ayrshire, τ = 1.5 and ω = 0.6 for Holstein, and τ = 1.5 and ω = 0.9 for Jersey were used. For genomic evaluation of somatic cell score the optimal scaling factors were: τ = 1.0 and ω = 0.9 for Ayrshire, τ = 2.0 and ω = 0.6 for Holstein, and τ = 1.0 and ω = 0.6 for Jersey were used.

The validation reliability patterns of GEBVs over days in milk, for each trait and breed (using all genotyped animals and the optimal scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ for each analysis), are shown in Figure 1. In general, slightly different validation reliability patterns were observed among traits, lactations and breeds.

Computational demand

A server with 512G memory and 88 cores (Intel(R) Xeon(R) CPU E5-2699 v4 @ 2.20GHz) was used in all analysis. The total number of rounds to achieve the convergence, the average number of cores, and the total CPU time used for each analysis is shown in Table 5. As expected, including genomic information increased the number of rounds, the number of cores, and the CPU time to complete the analyses. The increase in the number of rounds, cores and CPU time was higher when including all genotypes available compared to including genotypes only for animals with accurate EBVs. However, when comparing ssGBLUP using or not using scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$, a reduction in the total CPU time was observed when including the optimal scaling factors in all analyses (i.e., GEBVallw and GEBVall in the Table 5).

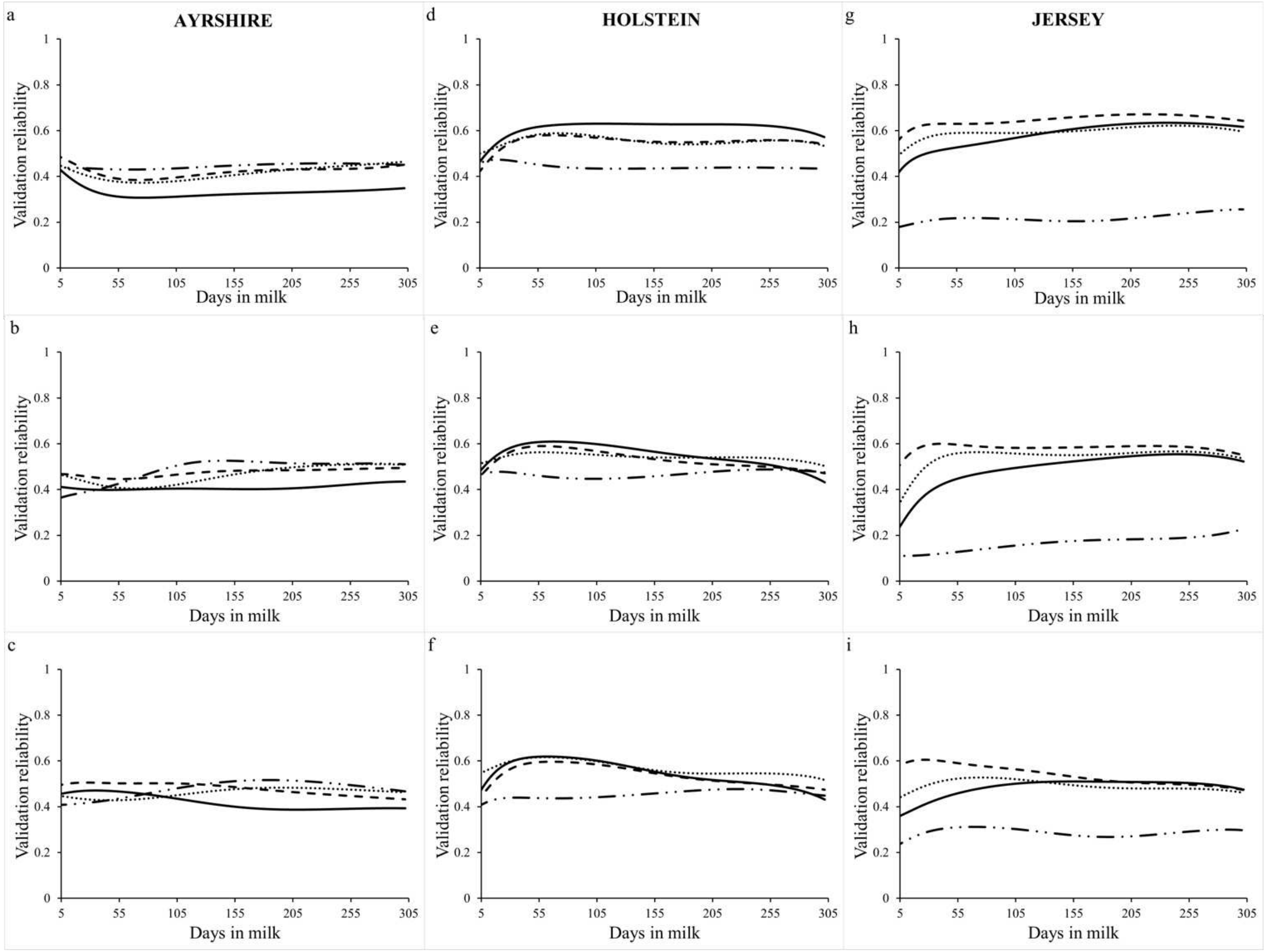

Figure 1. Pattern of validation reliabilities of genomic estimated breeding values through days in milk, using all genotyped animals, for milk (▬▬), fat (• • •) and protein yield (▬ ▬ ▬) and somatic cell score (▬ · · ▬) from the first (a, d and g), second (b, e and h) and third (c, f and i) lactations of Ayrshire, Holstein and Jersey breeds, respectively (considering the scaling factors to combine $\mathbf{G}^{-1}$ (τ) and $\mathbf{A}_{22}^{-1}$ (ω) that produced the least biased predictions and the highest validation reliabilities for each breed). For genomic evaluation of milk, fat and protein yields: $\tau = 2.0$ and $\omega = 0.6$ for Ayrshire, $\tau = 1.5$ and $\omega = 0.6$ for Holstein, and $\tau = 1.5$ and $\omega = 0.9$ for Jersey were used. For genomic evaluation of somatic cell score: $\tau = 1.0$ and $\omega = 0.9$ for Ayrshire, $\tau = 2.0$ and $\omega = 0.6$ for Holstein, and $\tau = 1.0$ and $\omega = 0.6$ for Jersey were used.

Table 5. Total number of rounds to achieve the convergence, average number of cores, and total central processing unit (CPU) time used for each analyses of each breed.

Computational parameter	Analyses	MY, FY and PY			SCS		
		Ayrshire	Holstein	Jersey	Ayrshire	Holstein	Jersey
Number of rounds	**EBV_r**	1,162	1,251	2,365	1,184	1,188	916
	$GEBV_{acc}$	1,643	2,317	2,776	1,319	2,027	977
	$GEBV_{all}$	2,365	3,140	4,670	2,454	3,162	1,626
	$GEBV_{allw}$	2,126	3,164	4,457	2,277	3,843	1,146
Average number of cores	**EBV_r**	5.28	3.59	6.91	8.13	3.10	6.84
	$GEBV_{acc}$	9.39	6.67	8.81	11.04	3.77	11.84
	$GEBV_{all}$	10.15	11.81	10.83	11.32	6.43	13.31
	$GEBV_{allw}$	10.69	12.67	11.38	12.88	7.52	13.80
Total CPU time (days)	**EBV_r**	0.15	5.10	0.20	0.05	1.26	0.03
	$GEBV_{acc}$	0.22	10.39	0.25	0.05	2.20	0.03
	$GEBV_{all}$	0.33	16.08	0.41	0.11	3.50	0.06
	$GEBV_{allw}$	0.30	15.90	0.38	0.10	3.16	0.04

Computational parameters were recorded for analyses conducted based on the reduced dataset (all information available until 2012) in order to estimate: breeding value (EBV_r) and genomic breeding values using genotypes from animals with accurate EBVs only ($GEBV_{acc}$), using all genotypes available ($GEBV_{all}$), and using all genotypes and the optimal scaling factors to combine genomic and pedigree information ($GEBV_{allw}$). The optimal scaling factors assumed for genomic evaluation of milk (MY), fat (FY) and protein (PY) yields were: $\tau = 2.0$ and $\omega = 0.6$ for Ayrshire, $\tau = 1.5$ and $\omega = 0.6$ for Holstein, and $\tau = 1.5$ and $\omega = 0.9$ for Jersey. For genomic evaluation of somatic cell score (SCS), the optimal scaling factors were: $\tau = 1.0$ and $\omega = 0.9$ for Ayrshire, $\tau = 2.0$ and $\omega = 0.6$ for Holstein, and $\tau = 1.0$ and $\omega = 0.6$ for Jersey. For $GEBV_{acc}$ and $GEBV_{all}$, both τ and ω were defined as 1.0 for all traits and breeds.

7.5. Discussion

Genotypes from animals with accurate EBVs only vs. from all genotypes

The changes in validation reliabilities due to the inclusion of all available genotypes are linked to the proportion of new genotypes included in the analysis (Table 1) and how accurate the EBVs of the additional animals included in the analyses are. For instance, in the genomic evaluation of MY, FY, and PY in the Holstein and Jersey breeds, approximately 133% and 56% new genotypes were included when considering all available genotypes in the population, respectively. For the evaluation of SCS, these proportions increased to approximately 182% for Holstein and 172% for Jersey. In both cases, the accuracy of the EBVs and proportion of genotypes included were enough to increase the validation reliabilities (Table 2). These findings are in agreement with Shabalina et al. (2017), who concluded that information from culled animals improved the accuracy of GEBVs based on ssGBLUP and with Patry and Ducrocq (2011), who reported that it is necessary to include information from all genotyped animals in the ssGBLUP evaluation.

For the analysis of MY, FY, and PY in the Ayrshire breed, only 6% new genotypes were included when considering all available genotypes in the analyses (second scenario), which did not contribute to increase the validation reliability. In fact, a slight decrease in the validation reliabilities for the Ayrshire breed was observed, probably due to the inclusion of a low number of animals with extremely low EBV reliability, which increased the amount of noise in the analyses (average reliability for the 1,716 genotyped animals with accurate EBVs and for the 111 additional genotyped animals was 0.75 and 0.42 for MY, 0.72 and 0.39 for FY, and 0.72 and 0.39 for PY, respectively). Average EBV reliability for all traits and breeds, in both scenarios, are shown in Table S2 (Supplementary Material). There was also no increase in the validation reliabilities estimated for the evaluation of SCS in the Ayrshire breed, even with approximately 60% new genotypes included in the second scenario. This might be due to the lower heritability for SCS compared to the heritability estimates for MY, FY and PY. In the case of lower heritability traits (e.g., SCS) the number of animals required to provide additional increase in prediction reliability should be greater than for high heritability traits. The minimum proportion of genotyped animals with low EBV accuracy needed to improve validation reliabilities for different traits should be determined in future studies, especially for small sized populations such as Ayrshire and Jersey breeds.

Regression coefficients lower than 1.0 were observed for both analyzed scenarios (Table 3), which indicates that the genomic predictions are inflated, and that the predicted differences in the genetic merit are biased upwards compared to the observed future performance. Regression coefficients deviating from 1.0 were also observed in the second scenario, including all available genotypes, for the majority of traits and breeds. This bias might originate from the use of animals with low accurate EBVs, which might introduce more noise into the analysis. In addition, pedigree depth for those animals may also contribute to increase the bias, if they have incomplete pedigree (Misztal, 2017). Koivula et al. (2018) investigated the effect of including genotypes of animals without progeny in the evaluation of Nordic Red Dairy cattle and concluded that the inclusion of genotypes of the culled bull calves may increase bias in the ssGBLUP evaluation.

The inclusion of all available genotypes is a reasonable principle and follows the BLUP theory, where all information used for making selection decisions should be accounted in the evaluation (Henderson, 1984). Thus, as the performance of ssGBLUP evaluations that included genotypes from only animals with accurate EBVs and from all genotyped animals were similar, all genomic information available was further used to define the most appropriate scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ in this study, for all analyzed breeds.

ssGBLUP vs. traditional evaluation of test-day traits and impact of scaling factors

Validation reliabilities estimated for GEBVs were generally higher than the validation reliabilities estimated for PAs, even when using no scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ (Tables S4, S5 and S6 at the Supplementary Material). These results are in agreement with other studies reported in the literature, such as Koivula et al. (2015) and Baba et al. (2016) that also concluded that ssGBLUP based on RRM increased the validation reliabilities compared to the traditional genetic evaluations based on RRM. Regression coefficients obtained for the default ssGBLUP were closer to 1.0 when compared to PA for all traits in the Holstein breed (except for SCS2 and SCS3), and for FY in the Jersey breed. This might be explained by the higher selection pressure assumed for these traits in these breeds, and suggests that SCS2 and SCS3 were included in the selection index latter than SCS1 in the genetic evaluation of Canadian Holstein. However, for all traits in the Ayrshire breed, for SCS2 and SCS3 in Holstein breed, and for MY, PY and SCS in the Jersey breed, regression coefficients were found to deviate more from 1.0 for GEBV than for PA.

The $\mathbf{H}^{-1}$ matrix used in ssGBLUP (as presented in Aguilar et al., 2010) assumes the correct model, which is only approximated in practice. Thus, additional parameters are usually required for a better model fit. Tsuruta et al. (2011) used τ and ω parameters multiplied by $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$, respectively; and they concluded that ω improved convergence and reduced bias in ssGBLUP analysis. Misztal et al. (2013) suggested that optimal scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ decrease the possible inflation of GEBV estimated by ssGBLUP. In fact, regression coefficients closer to 1.0 were obtained in the current study by ssGBLUP using scaling factors compared to the traditional genetic evaluation (Tables S4, S5 and S6 at the Supplementary Material). This suggests that GEBVs obtained based on ssGBLUP may be less biased compared to PAs. However, it is essential to determine the best scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$, as different optimal scaling factors were obtained for the different breeds and analyses. This is most likely due to differences in genotyping strategies, pedigree depth, and genetic architecture of the traits. The use of a genomic relationship matrix that weights markers according to the trait being analyzed (e.g., VanRaden, 2008; Makgahlela et al., 2013) may better account for differences in genetic architecture of the traits and, consequently, increase reliability and decrease bias, especially for small populations (Lourenco et al., 2017), such as Canadian Ayrshire and Jersey breeds. Additionally, Meuwissen et al. (2011) showed that relationships estimated by markers in **G** contain estimation errors, and that a better estimate may be obtained by regressing **G** back to **A**, which maximized the accuracy and removed the bias in the GEBVs. However, these authors commented that this method is computationally much more demanding than traditional methods used to construct **G**.

In general, changes in the τ parameter seems to have a small effect on the regression coefficients, unlike changes in the ω parameter (Tables S4, S5 and S6 at the Supplementary Material). Decreasing ω increases the importance of the pedigree information in the GEBV prediction, and it is also dependent on the completeness of the pedigree in each breed. Misztal et al. (2013) and Misztal (2017) related the ω parameter to incomplete pedigree and unaccounted inbreeding effect in $\mathbf{A}_{22}^{-1}$, respectively. These findings support the conclusions that ssGBLUP evaluations are inflated when the pedigree is deep, but incomplete (Misztal, 2017). Thus, one option to reduce the inflation of GEBVs estimated by ssGBLUP may be using shallower and more complete pedigree, especially for studies in which $\mathbf{A}_{22}^{-1}$ does not account for inbreeding coefficients. In this study, ω lower than 1.0 yielded less biased predictions for all breeds. Regarding to the τ parameter, genomic evaluation of MY, FY and PY for all breeds and SCS for the Holstein breed were less biased and led to higher validation reliabilities when τ was higher than 1.0. When τ is higher than 1.0, the genetic variance is reduced in order to make $\mathbf{G}^{-1}$ to have approximately the same variance as $\mathbf{A}^{-1}$. This suggests that the genetic variance in genotyped animals is reduced and that for unknown animals an inbreeding level of 0 is assumed instead of that of their contemporaries. Contrarily, for the genomic evaluation of SCS in the Ayrshire and Jersey breeds, the optimal τ was equal to 1.0. This indicates that selection has not been very intense for these traits (i.e., SCS1, SCS2 and SCS3) in the Ayrshire and Jersey breeds. Additional information about the theoretical background of τ and ω can be found in Martini et al. (2018).

Regardless the fact that the use of scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ have decreased bias for all traits compared to the ssGBLUP with no scaling (Tables S4, S5 and S6 at the Supplementary Material), both traditional and genomic evaluations generated mostly biased breeding values for the majority of traits analyzed in this study. Mäntysaari and Koivula (2012) showed that several factors may contribute to the bias in genomic predictions, such as the non-random selection of validation bulls. Thus, currently Interbull requires national GEBVs to be validated by the GEBV test before their inclusion in the international comparisons (Interbull, 2016). In summary, the GEBV test evaluates the bias in the genomic evaluations and the improvement in accuracy from the use of GEBVs compared to EBVs (Interbull, 2016). In our study, gains in reliabilities were obtained when genomic information was included, even when no scaling factors to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ were used (Tables S4, S5 and S6 at the Supplementary Material). However, the need for re-scaling the GEBVs and EBVs should be evaluated to meet the Interbull's bias requirements if ssGBLUP become a standard procedure for official evaluations in Canada, especially for the Ayrshire and Holstein breeds.

The current implementation of ssGBLUP does not allow different values for τ and ω for each trait in multiple-trait models. Therefore, the optimal values in this study were those that yielded high validation reliabilities and less biased regression coefficients across all traits analyzed. The pattern of GEBV reliabilities through DIM considering the optimal scaling factors for each analysis was slightly different across traits, lactations, and breeds (Figure 1), suggesting that there may be DIM for which higher genetic gain is possible compared to others. Similar pattern of GEBV reliabilities through DIM were obtained by Oliveira et al. (2018) using RRM in both steps of a two-step genomic prediction for the Holstein breed. These results may be relevant for traits that summarize the lactation curve pattern, such as lactation persistency. Thus, new studies to define the optimal way to measure lactation persistency should be done taking into account the genomic information.

Computational demand

The current study shows that incorporating genomic information into ssGBLUP based on RRM increases the computational demand for the evaluation of MY, FY, PY and SCS. The inclusion of genomic information affects convergence because the variance structure of genotyped animals in the $\mathbf{G}^{-1}$ is less diagonally dominant than in $\mathbf{A}^{-1}$ (Koivula et al., 2015), which requires more iterations to achieve convergence and, consequently, increase the CPU time. However, despite its high computational requirements, the ssGBLUP method seems to be suitable to analyze test-day traits of Canadian Ayrshire, Holstein and Jersey breeds, since all analyses achieved the convergence criteria.

It is important to notice that this was the first time that programs from the blupf90 family (Misztal et al., 2002; Misztal et al., 2014) were applied to RRM in ssGBLUP. Therefore, the version of BLUP90IOD software used in this study was not fully optimized for RRM using genomic information. Although it uses a different number of threads to read the data and pedigree in parallel, most of the computing time was spent on accessing the data from the disk (i.e., interation on data). Thus, further improvements may reduce computational time, such as the use of a more efficient parallel computing/multi-threading, and converting from iteration on data in the disk to in the memory. Another optimization could be the development of an appropriate preconditioner for RRM. An initial test of BLUP90IOD using a block preconditioner with dimension of the polynomial order times the number of traits was able to reduce the computing time from 16.08 to 4.19 days, using all genotypes available for the Holstein breed (results not shown). Improvements to reduce computational time are required especially to allow weekly ssGBLUP evaluations.

To our best knowledge, no studies comparing the computational demand of ssGBLUP to multiple-step GBLUP are available in the literature for RRM. Among the reasons is the fact that RRM have not been routinely used in multiple-step genomic evaluations, because it does not seem to increase the reliability of the genomic predictions compared to the use of accumulated yield (Oliveira et al., 2018). However, in a simulation study using RRM, Kang et al. (2017) showed that ssGBLUP was more accurate, less biased and had better persistency of accuracy over generations than multiple-step GBLUP.

7.6. Conclusions

Our findings indicate that including all available genotypes or only genotypes from animals with accurate EBVs yields similar genomic predictions in terms of daily validation reliability and bias, especially for populations with small number of genotyped animals. The use of ssGBLUP to predict genomic breeding values for test-day traits based on random regression models increases reliability and reduces bias of breeding values compared to traditional parentage average from BLUP in the Canadian Ayrshire, Holstein, and Jersey breeds. However, the scaling factors used to combine $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ should be carefully chosen.

7.7. Acknowledgements

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES). The authors gratefully acknowledge Dr. Ignacio Aguilar for sharing the PCG algorithm for random regression models. In addition, the authors acknowledge the financial support to this research provided by the Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative.

7.8. References

Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. 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.

Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. In Proc. 10th World Congr. Genet. Appl. Livest. Prod. Vancouver, Canada. doi: 10.13140/2.1.4801.5045.

Baba, T., Y. Gotoh, S. Yamaguchi, S. Nakagawa, H. Abe, Y. Masuda, and T. Kawahara. 2016. Application of single-step genomic best linear unbiased prediction with a multiple-lactation random regression test-day model for Japanese Holsteins. Animal Sci. J. 88:1226-1231. doi: 10.1111/asj.12760.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Gen. Sel. Evol. 42:1-8. doi: https://doi.org/10.1186/1297-9686-42-2.

Do, D. N., A. Fleming, F. Schenkel, F. Miglior, X. Zhao, and E. M. Ibeagha-Awemu. 2018. Genetic parameters of milk cholesterol content in Holstein cattle. Can. J. Anim. Sci. doi: 10.1139/CJAS-2018-0010.

Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. Can. Catal. Publ. Data. Univ. Guelph, Guelph, ON, Canada.

Interbull. 2016. Interbull CoP - Appendix VIII - Interbull validation test for genomic evaluations – GEBV test. Accessed Sept. 27, 2018. http://interbull.org/ib/cop_appendix8.

Interbull. 2017. Description of national genomic evaluation systems. Canada. Accessed May. 02, 2018. http://www.interbull.org/ib/nationalgenoforms.

Interbull. 2018. Description of national genetic evaluation systems. Accessed May, 02, 2018. http://www.interbull.org/ib/geforms.

Jamrozik, J., J. Bohmanova, and L. R. Schaeffer. 2010. Relationships between milk yield and somatic cell score in Canadian Holsteins from simultaneous and recursive random regression models. J. Dairy Sci. 93:1216-1233. doi: 10.3168/jds.2009-2585.

Jamrozik, J., and L. R. Schaeffer. 2011. Alternative parameterizations of the multiple-trait random regression model for milk yield and somatic cell score via recursive links between phenotypes. J. Anim. Breed. Genet. 128:258-266. doi: 10.1111/j.1439-0388.2011.00918.x.

Jamrozik, J., L. R. Schaeffer, and G. B. Jansen. 2000. Approximate accuracies of prediction from random regression models. Livest. Prod. Sci. 66:85-92. doi: https://doi.org/10.1016/S0301-6226(00)00158-5.

Kang, H., C. Ning, L. Zhou, S. Zhang, Q. Yan, and J. F. Liu. 2018. Short communication: Single-step genomic evaluation of milk production traits using multiple-trait random regression model in Chinese Holsteins. J. Dairy Sci.101:1-7. doi: 10.3168/jds.2018-15090.

Kang, H., L. Zhou, R. Mrode, Q. Zhang, and J. F. Liu. 2017. Incorporating single-step strategy into random regression model to enhance genomic prediction of longitudinal trait. Heredity 119:459-467. doi: 10.1038/hdy.2016.91.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979-993.

Koivula, M., I. Strandén, G. Aamand, and E. Mäntysaari. 2018. Reducing bias in the dairy cattle single-step genomic evaluation by ignoring bulls without progeny. J. Anim. Breed. Genet. 135:107-115. doi: 10.1111/jbg.12318.

Koivula, M., I. Strandén, G. Su, and E. A. Mäntysaari. 2012. Different methods to calculate genomic predictions - Comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). J. Dairy Sci. 95:4065–4073. doi: 10.3168/jds.2011-4874.

Koivula, M., I. Strandén, J. Pösö, G. Aamand, and E. Mäntysaari. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775-2784. doi: 10.3168/jds.2014-8975.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656-4663. doi: 10.3168/jds.2009-2061.

Lourenco, D. A. L., B. O. Fragomeni, H. L. Bradford, I. R. Menezes, J. B. S. Ferraz, I. Aguilar, S. Tsuruta, and I. Misztal. 2017. Implications of SNP weighting on single-step genomic predictions for different reference population sizes. J. Anim. Breed. Genet. 134:463-471. doi: 10.1111/jbg.12288.

Makgahlela, M. L., T. Knürr, G. Aamand, I. Stranden, and E. Mäntyasaari. 2013. Single step evaluations using haplotype segments. Interbull Bull. 47:217-221.

Mäntysaari, E. A., and M. Koivula. 2012. GEBV Validation Test Revisited. Interbull Bull. 45:1-5.

Martini, J. W., M. F. Schrauf, C. A. Garcia-Baccino, E. C. Pimentel, S. Munilla, A. Rogberg-Muñoz, R. J. C. Cantet, C. Reimer, N. Gao, V. Wimmer, and H. Simianer. 2018. The effect of the H−1 scaling factors τ and ω on the structure of H in the single-step procedure. Gen. Sel. Evol. 50:1-9. doi: https://doi.org/10.1186/s12711-018-0386-x.

Meuwissen, T. H., T. Luan, and J. A. Woolliams. 2011. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. J. Anim. Breed. Genet. 128:429-439. doi: 10.1111/j.1439-0388.2011.00966.x.

Misztal, I. 2017. Studies on inflation of GEBV in single-step GBLUP for type. Interbull Bull. 51:38-42.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648-4655. doi: 10.3168/jds.2009-2064.

Misztal, I., S. Tsuruta, D. A. L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. Manual for BLUPF90 family of programs. Athens: University of Georgia.

Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra, P. VanRaden, and T. Lawlor. 2013. Methods to approximate reliabilities in single-step genomic evaluation. J. Dairy Sci. 96:647-654. doi: 10.3168/jds.2012-5656.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. Lee. 2002. BLUPF90 and related programs (BGF90). Pages 21-22 in Proc. 7th World Congr. Genet. Appl. Livest. Prod. Montpellier, France.

Oliveira, H. R., F. F. Silva, O. H. G. B. D. Siqueira, N. O. Souza, V. S. Junqueira, M. D. Resende, R. R. Borquis, and M. T. Rodrigues. 2016. Combining different functions to describe milk, fat, and protein yield in goats using Bayesian multiple-trait random regression models. J. Anim. Sci. 94:1865-1874. doi: 10.2527/jas.2015-0150.

Oliveira, H. R., L. F. Brito, F. F. Silva, D. A. L. Lourenco, J. Jamrozik, and F. S. Schenkel. 2018. Genomic prediction of lactation curves for milk, fat, protein and somatic cell score in Holstein cattle. J. Dairy Sci. 102:1-12.

Patry, C., and V. Ducrocq. 2011. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. J. Dairy Sci. 94:1011-1020. doi: 10.3168/jds.2010-3804.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2011. FImpute - An efficient imputation algorithm for dairy cattle populations. J. Dairy Sci. 94(E-Suppl. 1):421. (Abstr.)

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Sci. 86:35-45. doi: https://doi.org/10.1016/S0301-6226(03)00151-9.

Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and J. Van Doormaal. 2000. Experience with a test-day model. J. Dairy Sci. 83:1135-1144. doi: https://doi.org/10.3168/jds.S0022-0302(00)74979-4.

Shabalina, T., E. C. G. Pimentel, C. Edel, L. Plieschke, R. Emmerling, and K. U. Götz. 2017. The role of genotypes from animals without phenotypes in single-step genomic evaluations. J. Dairy Sci. 100:8277-8281. doi: 10.3168/jds.2017-12734.

Strandén, I., and M. Lidauer. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. J. Dairy Sci. 82:2779-2787.

Tsuruta, S., I. Aguilar, I. Misztal, and T. J. Lawlor. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198–4204. doi: 10.3168/jds.2011-4256.

Tsuruta, S., I. Misztal, and I. Stranden. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. J. Anim. Sci. 79:1166-1172.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi: 10.3168/jds.2007-0980.

Vitezica, Z., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. 93:357-366. doi: 10.1017/S001667231100022X.

Wiggans, G., T. Sonstegard, P. VanRaden, L. Matukumalli, R. Schnabel, J. Taylor, F. Schenkel, and C. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92:3431-3436. doi: 10.3168/jds.2008-1758.

7.9. Supplementary material

Table S1. Test-day trait means (and standard deviation) for milk (kg), fat (kg) and protein (kg) yields, and somatic cells score in the Ayrshire, Holstein and Jersey breeds.

Trait	Ayrshire	Holstein	Jersey
MY1	20.76 (5.32)	26.97 (6.52)	17.65 (4.84)
MY2	23.62 (7.59)	31.41 (9.00)	20.54 (6.23)
MY3	24.90 (8.16)	33.15 (9.58)	21.95 (6.63)
FY1	0.84 (0.21)	1.00 (0.25)	0.86 (0.23)
FY2	0.94 (0.30)	1.16 (0.34)	1.01 (0.30)
FY3	0.98 (0.33)	1.23 (0.38)	1.07 (0.32)
PY1	0.69 (0.16)	0.87 (0.20)	0.67 (0.16)
PY2	0.79 (0.23)	1.01 (0.25)	0.79 (0.21)
PY3	0.83 (0.24)	1.05 (0.27)	0.84 (0.22)
SCS1	2.40 (1.62)	2.37 (1.66)	3.10 (1.49)
SCS2	2.84 (1.75)	2.74 (1.78)	3.07 (1.63)
SCS3	3.10 (1.83)	3.04 (1.87)	3.29 (1.73)

Traits within each breed were coded as MY1, FY1, PY1, SCS1, MY2, FY2, PY2, SCS2, MY3, FY3, PY3 and SCS3; where MY, FY, PY and SCS refers to milk, fat and protein yields, and somatic cells score, respectively, and 1, 2 and 3 refers to the first, second and third lactations.

Table S2. Average reliabilities of estimated breeding values (and standard deviations; REL_{EBV}), and the 95% confidence interval (95% CI), for genotyped animals with accurate estimated breeding values only (Scenario 1) and all genotyped animals (Scenario 2).

Breed	Trait[1]	Scenario 1		Scenario 2	
		REL_{EBV}	95% CI	REL_{EBV}	95% CI
Ayrshire	**MY**	0.75 (0.12)	0.744-0.756	0.73 (0.14)	0.724-0.736
	FY	0.72 (0.14)	0.713-0.727	0.70 (0.15)	0.693-0.707
	PY	0.72 (0.13)	0.714-0.726	0.70 (0.15)	0.693-0.707
	SCS	0.68 (0.17)	0.670-0.690	0.59 (0.18)	0.582-0.598
Holstein	**MY**	0.89 (0.09)	0.888-0.892	0.71 (0.18)	0.708-0.712
	FY	0.89 (0.09)	0.888-0.892	0.72 (0.18)	0.718-0.722
	PY	0.88 (0.10)	0.878-0.882	0.69 (0.19)	0.687-0.693
	SCS	0.82 (0.06)	0.819-0.821	0.61 (0.21)	0.607-0.613
Jersey	**MY**	0.75 (0.14)	0.741-0.759	0.62 (0.23)	0.608-0.632
	FY	0.73 (0.15)	0.720-0.740	0.60 (0.23)	0.588-0.612
	PY	0.73 (0.15)	0.720-0.740	0.60 (0.23)	0.588-0.612
	SCS	0.74 (0.16)	0.726-0.754	0.50 (0.22)	0.489-0.511

[1] Milk (MY), fat (FY), and protein (PY) yields in 1st, 2nd, and 3rd lactations were analyzed separated from somatic cell score (SCS). Animals were considered to have accurate EBVs when the reliability of estimated breeding values for PY or SCS, estimated in preliminary analysis, was higher than 0.50 for the Ayrshire and Jersey breeds and higher than 0.65 for the Holstein breed.

Table S3. Average heritabilities (and standard deviations) across days in milk for milk, fat and protein yields, and somatic cells score in the Ayrshire, Holstein and Jersey breeds.

Trait	**Ayrshire**	**Holstein**	**Jersey**
MY1	0.29 (0.05)	0.26 (0.04)	0.31 (0.05)
MY2	0.27 (0.07)	0.24 (0.06)	0.27 (0.06)
MY3	0.25 (0.08)	0.21 (0.07)	0.26 (0.06)
FY1	0.20 (0.04)	0.20 (0.04)	0.22 (0.04)
FY2	0.19 (0.08)	0.19 (0.07)	0.20 (0.06)
FY3	0.18 (0.09)	0.18 (0.08)	0.19 (0.07)
PY1	0.23 (0.05)	0.21 (0.03)	0.25 (0.05)
PY2	0.24 (0.07)	0.21 (0.05)	0.24 (0.06)
PY3	0.23 (0.09)	0.19 (0.07)	0.23 (0.06)
SCS1	0.10 (0.01)	0.10 (0.01)	0.10 (0.01)
SCS2	0.14 (0.04)	0.14 (0.02)	0.12 (0.02)
SCS3	0.13 (0.05)	0.17 (0.01)	0.12 (0.01)

Traits within each breed were coded as MY1, FY1, PY1, SCS1, MY2, FY2, PY2, SCS2, MY3, FY3, PY3 and SCS3; where MY, FY, PY and SCS refers to milk, fat and protein yields, and somatic cells score, respectively, and 1, 2 and 3 refers to the first, second and third lactations.

Table S4. Validation reliabilities (r^2) and regression coefficients ($\hat{b}_1 \pm SE$) for parentage average (PA) and genomic breeding value estimated using all available genotypes and assuming different τ and ω for Ayrshire.

Lactation	Method	Milk yield		Fat yield		Protein yield		Somatic cell score	
		r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$
First	**PA**	0.32	0.70±0.006	0.41	0.72±0.005	0.53	0.78±0.004	0.42	1.10±0.008
	$\tau_{1.0}\omega_{0.6}$	0.40	0.81±0.006	0.47	0.82±0.005	0.58	0.85±0.004	0.48	1.22±0.008
	$\tau_{1.0}\omega_{0.7}$	0.40	0.78±0.005	0.47	0.80±0.005	0.58	0.83±0.004	0.47	1.18±0.008
	$\tau_{1.0}\omega_{0.8}$	0.40	0.75±0.005	0.46	0.76±0.005	0.57	0.79±0.004	0.47	1.13±0.007
	$\tau_{1.0}\omega_{0.9}$	0.40	0.71±0.005	0.45	0.72±0.005	0.56	0.75±0.004	0.46	1.06±0.007
	$\tau_{1.0}\omega_{1.0}$	0.39	0.64±0.005	0.43	0.65±0.004	0.54	0.70±0.004	0.43	0.95±0.007
	$\tau_{1.5}\omega_{0.6}$	0.40	0.83±0.006	0.48	0.86±0.005	0.59	0.88±0.004	0.48	1.28±0.008
	$\tau_{1.5}\omega_{0.7}$	0.40	0.81±0.006	0.48	0.84±0.005	0.59	0.86±0.004	0.48	1.24±0.008
	$\tau_{1.5}\omega_{0.8}$	0.40	0.78±0.005	0.47	0.81±0.005	0.58	0.83±0.004	0.47	1.20±0.008
	$\tau_{1.5}\omega_{0.9}$	0.40	0.74±0.005	0.47	0.77±0.005	0.57	0.80±0.004	0.47	1.14±0.007
	$\tau_{1.5}\omega_{1.0}$	0.39	0.70±0.005	0.45	0.72±0.005	0.56	0.76±0.004	0.45	1.06±0.007
	$\tau_{2.0}\omega_{0.6}$	0.40	0.85±0.006	0.48	0.89±0.005	0.59	0.91±0.004	0.48	1.33±0.008
	$\tau_{2.0}\omega_{0.7}$	0.40	0.83±0.006	0.48	0.87±0.005	0.59	0.89±0.004	0.48	1.30±0.008
	$\tau_{2.0}\omega_{0.8}$	0.40	0.81±0.006	0.48	0.84±0.005	0.59	0.87±0.004	0.48	1.26±0.008
	$\tau_{2.0}\omega_{0.9}$	0.39	0.77±0.006	0.47	0.81±0.005	0.58	0.84±0.004	0.47	1.21±0.008
	$\tau_{2.0}\omega_{1.0}$	0.39	0.73±0.005	0.46	0.77±0.005	0.57	0.80±0.004	0.46	1.14±0.007
Second	**PA**	0.38	0.74±0.005	0.40	0.74±0.005	0.44	0.74±0.005	0.50	1.10±0.007
	$\tau_{1.0}\omega_{0.6}$	0.46	0.85±0.005	0.48	0.85±0.005	0.52	0.84±0.005	0.53	1.13±0.006
	$\tau_{1.0}\omega_{0.7}$	0.46	0.82±0.005	0.48	0.82±0.005	0.52	0.82±0.005	0.53	1.09±0.006
	$\tau_{1.0}\omega_{0.8}$	0.46	0.79±0.005	0.47	0.79±0.005	0.51	0.79±0.004	0.52	1.04±0.006
	$\tau_{1.0}\omega_{0.9}$	0.46	0.74±0.005	0.46	0.74±0.005	0.50	0.76±0.004	0.50	0.97±0.006
	$\tau_{1.0}\omega_{1.0}$	0.44	0.67±0.004	0.44	0.68±0.004	0.48	0.70±0.004	0.47	0.87±0.006
	$\tau_{1.5}\omega_{0.6}$	0.47	0.88±0.005	0.48	0.88±0.005	0.52	0.87±0.005	0.53	1.17±0.007
	$\tau_{1.5}\omega_{0.7}$	0.47	0.85±0.005	0.48	0.86±0.005	0.52	0.85±0.005	0.53	1.14±0.007
	$\tau_{1.5}\omega_{0.8}$	0.46	0.82±0.005	0.48	0.83±0.005	0.52	0.83±0.005	0.52	1.10±0.006
	$\tau_{1.5}\omega_{0.9}$	0.46	0.78±0.005	0.47	0.80±0.005	0.51	0.80±0.004	0.51	1.05±0.006
	$\tau_{1.5}\omega_{1.0}$	0.45	0.73±0.005	0.46	0.75±0.005	0.50	0.76±0.004	0.49	0.97±0.006
	$\tau_{2.0}\omega_{0.6}$	0.46	0.90±0.006	0.48	0.90±0.005	0.52	0.89±0.005	0.53	1.21±0.007
	$\tau_{2.0}\omega_{0.7}$	0.46	0.88±0.005	0.48	0.88±0.005	0.52	0.88±0.005	0.53	1.18±0.007
	$\tau_{2.0}\omega_{0.8}$	0.46	0.85±0.005	0.48	0.86±0.005	0.52	0.86±0.005	0.53	1.15±0.007
	$\tau_{2.0}\omega_{0.9}$	0.46	0.82±0.005	0.48	0.83±0.005	0.52	0.83±0.005	0.52	1.10±0.006
	$\tau_{2.0}\omega_{1.0}$	0.46	0.78±0.005	0.47	0.80±0.005	0.51	0.80±0.004	0.50	1.04±0.006
Third	**PA**	0.42	0.75±0.005	0.42	0.74±0.005	0.46	0.75±0.005	0.50	1.13±0.007
	$\tau_{1.0}\omega_{0.6}$	0.49	0.85±0.005	0.49	0.84±0.005	0.53	0.83±0.005	0.52	1.13±0.007
	$\tau_{1.0}\omega_{0.7}$	0.49	0.82±0.005	0.48	0.82±0.005	0.53	0.81±0.004	0.51	1.09±0.006
	$\tau_{1.0}\omega_{0.8}$	0.49	0.79±0.005	0.48	0.79±0.005	0.52	0.79±0.004	0.50	1.03±0.006
	$\tau_{1.0}\omega_{0.9}$	0.49	0.74±0.004	0.47	0.74±0.004	0.51	0.75±0.004	0.49	0.96±0.006
	$\tau_{1.0}\omega_{1.0}$	0.47	0.67±0.004	0.45	0.67±0.004	0.49	0.69±0.004	0.45	0.84±0.006
	$\tau_{1.5}\omega_{0.6}$	0.49	0.88±0.005	0.49	0.88±0.005	0.53	0.86±0.005	0.52	1.18±0.007
	$\tau_{1.5}\omega_{0.7}$	0.49	0.86±0.005	0.49	0.86±0.005	0.53	0.84±0.005	0.51	1.14±0.007
	$\tau_{1.5}\omega_{0.8}$	0.49	0.83±0.005	0.49	0.83±0.005	0.53	0.82±0.004	0.51	1.09±0.007
	$\tau_{1.5}\omega_{0.9}$	0.49	0.79±0.005	0.49	0.80±0.005	0.53	0.79±0.004	0.49	1.03±0.006
	$\tau_{1.5}\omega_{1.0}$	0.49	0.74±0.004	0.48	0.75±0.004	0.52	0.75±0.004	0.47	0.95±0.006
	$\tau_{2.0}\omega_{0.6}$	0.49	0.91±0.005	0.49	0.91±0.005	0.53	0.88±0.005	0.52	1.21±0.007
	$\tau_{2.0}\omega_{0.7}$	0.49	0.89±0.005	0.49	0.89±0.005	0.53	0.87±0.005	0.51	1.18±0.007
	$\tau_{2.0}\omega_{0.8}$	0.49	0.86±0.005	0.49	0.87±0.005	0.53	0.85±0.005	0.51	1.14±0.007
	$\tau_{2.0}\omega_{0.9}$	0.49	0.82±0.005	0.49	0.84±0.005	0.53	0.82±0.004	0.50	1.09±0.007
	$\tau_{2.0}\omega_{1.0}$	0.49	0.78±0.005	0.49	0.80±0.005	0.52	0.79±0.004	0.48	1.02±0.006

τ: scaling factor for genomic relationship matrix ($\mathbf{G}^{-1}$); ω: scaling factor used for pedigree relationship matrix ($\mathbf{A}_{22}^{-1}$).

Table S5. Validation reliabilities (r^2) and regression coefficients ($\hat{b}_1 \pm SE$) for parentage average (PA) and genomic breeding value estimated using all available genotypes and assuming different τ and ω for Holstein.

Lactation	Method	Milk yield		Fat yield		Protein yield		Somatic cell score	
		r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$
First	PA	0.33	0.61±0.001	0.30	0.47±0.001	0.46	0.58±0.001	0.20	0.68±0.001
	$\tau_{1.0}\omega_{0.6}$	0.65	0.85±0.001	0.57	0.72±0.001	0.68	0.75±0.001	0.45	0.92±0.002
	$\tau_{1.0}\omega_{0.7}$	0.65	0.83±0.001	0.58	0.70±0.001	0.68	0.74±0.001	0.45	0.88±0.002
	$\tau_{1.0}\omega_{0.8}$	0.65	0.81±0.001	0.58	0.68±0.001	0.68	0.72±0.001	0.45	0.84±0.002
	$\tau_{1.0}\omega_{0.9}$	0.65	0.78±0.001	0.58	0.65±0.001	0.68	0.70±0.001	0.45	0.78±0.001
	$\tau_{1.0}\omega_{1.0}$	0.65	0.75±0.001	0.58	0.62±0.001	0.67	0.67±0.001	0.45	0.70±0.001
	$\tau_{1.5}\omega_{0.6}$	0.66	0.88±0.001	0.59	0.75±0.001	0.69	0.77±0.001	0.38	0.88±0.002
	$\tau_{1.5}\omega_{0.7}$	0.66	0.86±0.001	0.59	0.73±0.001	0.69	0.76±0.001	0.38	0.85±0.002
	$\tau_{1.5}\omega_{0.8}$	0.66	0.84±0.001	0.60	0.71±0.001	0.69	0.75±0.001	0.39	0.82±0.002
	$\tau_{1.5}\omega_{0.9}$	0.66	0.81±0.001	0.60	0.69±0.001	0.69	0.73±0.001	0.40	0.77±0.002
	$\tau_{1.5}\omega_{1.0}$	0.66	0.78±0.001	0.60	0.66±0.001	0.68	0.71±0.001	0.45	0.74±0.002
	$\tau_{2.0}\omega_{0.6}$	0.65	0.85±0.001	0.57	0.72±0.001	0.68	0.75±0.001	0.44	0.94±0.002
	$\tau_{2.0}\omega_{0.7}$	0.67	0.88±0.001	0.60	0.76±0.001	0.69	0.78±0.001	0.45	0.91±0.002
	$\tau_{2.0}\omega_{0.8}$	0.67	0.86±0.001	0.60	0.74±0.001	0.69	0.77±0.001	0.45	0.87±0.002
	$\tau_{2.0}\omega_{0.9}$	0.66	0.84±0.001	0.61	0.72±0.001	0.69	0.75±0.001	0.45	0.83±0.002
	$\tau_{2.0}\omega_{1.0}$	0.65	0.85±0.001	0.57	0.72±0.001	0.68	0.75±0.001	0.45	0.77±0.001
Second	PA	0.31	0.63±0.001	0.30	0.52±0.001	0.45	0.62±0.001	0.29	0.88±0.002
	$\tau_{1.0}\omega_{0.6}$	0.60	0.84±0.001	0.58	0.75±0.001	0.66	0.77±0.001	0.50	0.96±0.002
	$\tau_{1.0}\omega_{0.7}$	0.60	0.82±0.001	0.58	0.73±0.001	0.66	0.75±0.001	0.50	0.92±0.002
	$\tau_{1.0}\omega_{0.8}$	0.61	0.80±0.001	0.58	0.71±0.001	0.66	0.74±0.001	0.50	0.88±0.002
	$\tau_{1.0}\omega_{0.9}$	0.60	0.77±0.001	0.58	0.69±0.001	0.66	0.72±0.001	0.50	0.82±0.001
	$\tau_{1.0}\omega_{1.0}$	0.60	0.73±0.001	0.58	0.65±0.001	0.65	0.69±0.001	0.48	0.73±0.001
	$\tau_{1.5}\omega_{0.6}$	0.61	0.86±0.001	0.60	0.78±0.001	0.67	0.79±0.001	0.45	0.95±0.002
	$\tau_{1.5}\omega_{0.7}$	0.61	0.84±0.001	0.60	0.76±0.001	0.67	0.77±0.001	0.45	0.92±0.002
	$\tau_{1.5}\omega_{0.8}$	0.61	0.82±0.001	0.60	0.74±0.001	0.67	0.76±0.001	0.46	0.88±0.002
	$\tau_{1.5}\omega_{0.9}$	0.61	0.79±0.001	0.60	0.72±0.001	0.67	0.74±0.001	0.46	0.83±0.001
	$\tau_{1.5}\omega_{1.0}$	0.61	0.76±0.001	0.59	0.69±0.001	0.66	0.72±0.001	0.49	0.77±0.001
	$\tau_{2.0}\omega_{0.6}$	0.60	0.84±0.001	0.58	0.75±0.001	0.66	0.77±0.001	0.50	0.97±0.001
	$\tau_{2.0}\omega_{0.7}$	0.62	0.86±0.001	0.61	0.78±0.001	0.67	0.79±0.001	0.50	0.94±0.001
	$\tau_{2.0}\omega_{0.8}$	0.62	0.84±0.001	0.61	0.77±0.001	0.67	0.78±0.001	0.50	0.91±0.001
	$\tau_{2.0}\omega_{0.9}$	0.62	0.82±0.001	0.61	0.74±0.001	0.67	0.76±0.001	0.50	0.86±0.001
	$\tau_{2.0}\omega_{1.0}$	0.60	0.84±0.001	0.58	0.75±0.001	0.66	0.77±0.001	0.49	0.80±0.001
Third	PA	0.33	0.63±0.001	0.32	0.55±0.001	0.45	0.62±0.000	0.28	0.86±0.002
	$\tau_{1.0}\omega_{0.6}$	0.59	0.83±0.001	0.59	0.77±0.001	0.65	0.76±0.001	0.48	0.95±0.002
	$\tau_{1.0}\omega_{0.7}$	0.60	0.81±0.001	0.60	0.75±0.001	0.65	0.74±0.001	0.48	0.91±0.002
	$\tau_{1.0}\omega_{0.8}$	0.60	0.79±0.001	0.60	0.73±0.001	0.65	0.73±0.001	0.48	0.86±0.001
	$\tau_{1.0}\omega_{0.9}$	0.60	0.76±0.001	0.60	0.70±0.001	0.65	0.71±0.001	0.47	0.80±0.001
	$\tau_{1.0}\omega_{1.0}$	0.59	0.71±0.001	0.59	0.66±0.001	0.64	0.68±0.001	0.46	0.71±0.001
	$\tau_{1.5}\omega_{0.6}$	0.61	0.85±0.001	0.61	0.80±0.001	0.66	0.78±0.001	0.43	0.97±0.002
	$\tau_{1.5}\omega_{0.7}$	0.61	0.83±0.001	0.61	0.78±0.001	0.66	0.76±0.001	0.44	0.93±0.002
	$\tau_{1.5}\omega_{0.8}$	0.61	0.81±0.001	0.61	0.76±0.001	0.66	0.75±0.001	0.44	0.89±0.002
	$\tau_{1.5}\omega_{0.9}$	0.61	0.79±0.001	0.61	0.74±0.001	0.66	0.73±0.001	0.44	0.83±0.002
	$\tau_{1.5}\omega_{1.0}$	0.60	0.75±0.001	0.61	0.70±0.001	0.66	0.71±0.001	0.47	0.75±0.002
	$\tau_{2.0}\omega_{0.6}$	0.59	0.83±0.001	0.59	0.77±0.001	0.65	0.76±0.001	0.48	0.96±0.002
	$\tau_{2.0}\omega_{0.7}$	0.61	0.84±0.001	0.62	0.80±0.001	0.67	0.78±0.001	0.48	0.93±0.002
	$\tau_{2.0}\omega_{0.8}$	0.61	0.83±0.001	0.62	0.78±0.001	0.67	0.77±0.001	0.48	0.89±0.002
	$\tau_{2.0}\omega_{0.9}$	0.61	0.81±0.001	0.62	0.76±0.001	0.67	0.75±0.001	0.48	0.84±0.002
	$\tau_{2.0}\omega_{1.0}$	0.59	0.83±0.001	0.59	0.77±0.001	0.65	0.76±0.001	0.47	0.78±0.002

τ: scaling factor for genomic relationship matrix ($\mathbf{G}^{-1}$); ω: scaling factor used for pedigree relationship matrix ($\mathbf{A}_{22}^{-1}$).

Table S6. Validation reliabilities (r^2) and regression coefficients ($\hat{b}_1 \pm SE$) for parentage average (PA) and genomic breeding value estimated using all available genotypes and assuming different τ and ω for Jersey.

Lactation	Method	Milk yield		Fat yield		Protein yield		Somatic cell score	
		r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$	r^2	$\hat{b}_1 \pm SE$
First	**PA**	0.46	0.93±0.006	0.42	0.91±0.006	0.55	0.90±0.005	0.32	0.87±0.008
	$\tau_{1.0}\omega_{0.6}$	0.59	1.02±0.005	0.61	1.13±0.005	0.68	1.00±0.004	0.33	0.97±0.009
	$\tau_{1.0}\omega_{0.7}$	0.59	0.99±0.005	0.61	1.10±0.005	0.68	0.98±0.004	0.33	0.93±0.008
	$\tau_{1.0}\omega_{0.8}$	0.58	0.96±0.005	0.62	1.07±0.005	0.68	0.95±0.004	0.33	0.88±0.008
	$\tau_{1.0}\omega_{0.9}$	0.58	0.93±0.005	0.62	1.03±0.005	0.68	0.92±0.004	0.32	0.83±0.008
	$\tau_{1.0}\omega_{1.0}$	0.58	0.89±0.005	0.62	0.98±0.005	0.68	0.88±0.004	0.31	0.77±0.007
	$\tau_{1.5}\omega_{0.6}$	0.60	1.06±0.005	0.63	1.20±0.006	0.69	1.05±0.004	0.33	1.03±0.009
	$\tau_{1.5}\omega_{0.7}$	0.60	1.05±0.005	0.63	1.18±0.005	0.69	1.04±0.004	0.32	1.00±0.009
	$\tau_{1.5}\omega_{0.8}$	0.60	1.02±0.005	0.64	1.15±0.005	0.69	1.01±0.004	0.32	0.97±0.009
	$\tau_{1.5}\omega_{0.9}$	0.60	1.00±0.005	0.64	1.12±0.005	0.70	0.99±0.004	0.32	0.92±0.008
	$\tau_{1.5}\omega_{1.0}$	0.60	0.97±0.005	0.64	1.08±0.005	0.70	0.96±0.004	0.31	0.87±0.008
	$\tau_{2.0}\omega_{0.6}$	0.60	1.10±0.005	0.63	1.25±0.006	0.70	1.10±0.005	0.32	1.09±0.010
	$\tau_{2.0}\omega_{0.7}$	0.61	1.09±0.005	0.64	1.23±0.006	0.70	1.08±0.004	0.32	1.06±0.010
	$\tau_{2.0}\omega_{0.8}$	0.61	1.07±0.005	0.64	1.21±0.005	0.70	1.06±0.004	0.32	1.03±0.010
	$\tau_{2.0}\omega_{0.9}$	0.61	1.05±0.005	0.65	1.18±0.005	0.70	1.04±0.004	0.31	1.00±0.009
	$\tau_{2.0}\omega_{1.0}$	0.61	1.02±0.005	0.65	1.15±0.005	0.70	1.02±0.004	0.31	0.95±0.009
Second	**PA**	0.39	0.89±0.007	0.41	0.90±0.007	0.48	0.87±0.005	0.29	0.93±0.009
	$\tau_{1.0}\omega_{0.6}$	0.51	0.97±0.006	0.59	1.10±0.006	0.61	0.97±0.005	0.29	0.98±0.010
	$\tau_{1.0}\omega_{0.7}$	0.51	0.95±0.006	0.59	1.07±0.005	0.62	0.95±0.005	0.29	0.94±0.009
	$\tau_{1.0}\omega_{0.8}$	0.51	0.91±0.005	0.59	1.04±0.005	0.62	0.93±0.004	0.29	0.90±0.009
	$\tau_{1.0}\omega_{0.9}$	0.50	0.88±0.005	0.59	1.00±0.005	0.62	0.89±0.004	0.28	0.84±0.008
	$\tau_{1.0}\omega_{1.0}$	0.50	0.83±0.005	0.59	0.94±0.005	0.61	0.85±0.004	0.28	0.76±0.008
	$\tau_{1.5}\omega_{0.6}$	0.52	1.02±0.006	0.60	1.16±0.006	0.63	1.03±0.005	0.29	1.05±0.010
	$\tau_{1.5}\omega_{0.7}$	0.52	1.00±0.006	0.61	1.14±0.006	0.63	1.01±0.005	0.29	1.02±0.010
	$\tau_{1.5}\omega_{0.8}$	0.52	0.97±0.006	0.61	1.11±0.005	0.63	0.99±0.005	0.29	0.98±0.010
	$\tau_{1.5}\omega_{0.9}$	0.52	0.95±0.005	0.61	1.08±0.005	0.63	0.96±0.004	0.29	0.93±0.009
	$\tau_{1.5}\omega_{1.0}$	0.51	0.91±0.005	0.61	1.05±0.005	0.63	0.93±0.004	0.28	0.87±0.009
	$\tau_{2.0}\omega_{0.6}$	0.52	1.05±0.006	0.61	1.21±0.006	0.63	1.07±0.005	0.29	1.11±0.011
	$\tau_{2.0}\omega_{0.7}$	0.52	1.04±0.006	0.61	1.19±0.006	0.63	1.05±0.005	0.29	1.08±0.011
	$\tau_{2.0}\omega_{0.8}$	0.53	1.02±0.006	0.61	1.17±0.006	0.64	1.03±0.005	0.29	1.05±0.010
	$\tau_{2.0}\omega_{0.9}$	0.52	1.00±0.006	0.62	1.14±0.005	0.64	1.01±0.005	0.29	1.00±0.010
	$\tau_{2.0}\omega_{1.0}$	0.52	0.97±0.006	0.62	1.11±0.005	0.64	0.99±0.004	0.28	0.95±0.009
Third	**PA**	0.39	0.89±0.007	0.40	0.90±0.007	0.46	0.86±0.006	0.42	0.96±0.009
	$\tau_{1.0}\omega_{0.6}$	0.50	0.95±0.006	0.55	1.07±0.006	0.58	0.93±0.005	0.40	1.04±0.008
	$\tau_{1.0}\omega_{0.7}$	0.50	0.92±0.005	0.55	1.04±0.006	0.58	0.91±0.005	0.40	1.01±0.008
	$\tau_{1.0}\omega_{0.8}$	0.50	0.88±0.005	0.55	1.00±0.005	0.58	0.88±0.004	0.39	0.96±0.008
	$\tau_{1.0}\omega_{0.9}$	0.49	0.84±0.005	0.55	0.96±0.005	0.58	0.85±0.004	0.38	0.91±0.007
	$\tau_{1.0}\omega_{1.0}$	0.48	0.79±0.005	0.54	0.90±0.005	0.58	0.80±0.004	0.37	0.83±0.007
	$\tau_{1.5}\omega_{0.6}$	0.51	0.99±0.006	0.56	1.12±0.006	0.59	0.98±0.005	0.38	1.10±0.009
	$\tau_{1.5}\omega_{0.7}$	0.51	0.97±0.006	0.56	1.10±0.006	0.59	0.96±0.005	0.38	1.07±0.009
	$\tau_{1.5}\omega_{0.8}$	0.51	0.94±0.006	0.56	1.07±0.006	0.59	0.94±0.005	0.38	1.03±0.008
	$\tau_{1.5}\omega_{0.9}$	0.51	0.91±0.005	0.56	1.04±0.006	0.59	0.91±0.005	0.38	0.99±0.008
	$\tau_{1.5}\omega_{1.0}$	0.50	0.87±0.005	0.56	1.00±0.005	0.59	0.88±0.004	0.37	0.93±0.008
	$\tau_{2.0}\omega_{0.6}$	0.52	1.03±0.006	0.56	1.17±0.006	0.59	1.02±0.005	0.37	1.15±0.009
	$\tau_{2.0}\omega_{0.7}$	0.52	1.01±0.006	0.56	1.15±0.006	0.59	1.00±0.005	0.37	1.12±0.009
	$\tau_{2.0}\omega_{0.8}$	0.52	0.99±0.006	0.57	1.12±0.006	0.60	0.98±0.005	0.37	1.09±0.009
	$\tau_{2.0}\omega_{0.9}$	0.51	0.96±0.006	0.57	1.10±0.006	0.60	0.96±0.005	0.37	1.05±0.009
	$\tau_{2.0}\omega_{1.0}$	0.51	0.93±0.005	0.57	1.07±0.006	0.60	0.93±0.005	0.37	1.00±0.008

τ: scaling factor for genomic relationship matrix ($\mathbf{G}^{-1}$); ω: scaling factor used for pedigree relationship matrix ($\mathbf{A}_{22}^{-1}$).

CHAPTER 8:

Genome-wide association for milk yield and milk composition traits in different lactation stages of Ayrshire, Holstein and Jersey dairy cattle

Hinayah Rojas de Oliveira[1,2], John Cant[1], Luiz Fernando Brito[1,3], Fabieli Loise Braga Feitosa[4], Pablo Augusto de Souza Fonseca[5], Janusz Jamrozik[1,6], Fabyano Fonseca e Silva[2], Daniela Andressa Lino Lourenco[7], Flávio Schramm Schenkel[1]

[1] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

[2] Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[3] Department of Animal Sciences, Purdue University, West Lafayette, Indiana, United States of America

[4] Department of Animal Sciences, Faculdade de Ciências Agrárias e Veterinárias, São Paulo, Brazil

[5] Departament of General Biology, Universidade Federal de Minas Gerais, Minas Gerais, Brazil

[6] Canadian Dairy Network, Ontario, Canada

[7] Department of Animal and Dairy Science, University of Georgia, Georgia, United States of America

8.1. Abstract

We performed a genomic-wide association studies (**GWAS**) for milk (**MY**), fat (**FY**), and protein (**PY**) yields, and somatic cell score (**SCS**) based on lactation stages in the first three parities of Canadian Ayrshire, Holstein and Jersey cattle. The GWAS were performed considering three different lactation stages for each trait and parity: 1) from 5 to 95; 2) from 96 to 215; and 3) from 216 to 305 days in milk. Effects of single nucleotide polymorphisms (**SNP**) for each lactation stage, trait, parity and breed were estimated by back-solving the direct breeding values estimated using genomic best linear unbiased prediction and random regression models. In order to identify important genomic regions related to the analyzed lactation stages, traits, parities and breeds, moving windows (SNP-by-SNP) of 20 adjacent SNPs explaining more than 0.30% of total genetic variance were selected for further analyses of candidate genes. A lower number of windows with relatively higher proportion of the explained genetic variance was found in the Holstein breed compared to the Ayrshire and Jersey breeds. Genomic regions associated with the analyzed traits were located on 12, 8 and 15 different chromosomes for Ayrshire, Holstein and Jersey breeds, respectively. Especially for Holstein breed, many of the identified candidate genes corroborated with previous reports in the literature. However, well-known genes with major effects on milk production traits (e.g., diacylglycerol O-acyltransferase 1) presented contrasting results among lactation stages, traits and parities of different breeds. Therefore, we concluded that there is evidence of differential set of candidate genes across breeds, traits, parities and lactation stages.

Keywords: fat yield, longitudinal trait, milk composition, protein yield, somatic cell score

8.2. Introduction

Milk yield (**MY**) and milk composition, including fat (**FY**) and protein (**PY**) yields, and somatic cell score (**SCS**), are complex traits influenced by management practices, environmental conditions, animal physiological stage (e.g., age, stage of lactation), and genetic merit of the animals. For instance, adding omega-3 fatty acids and conjugated linoleic acid in the dairy cows' diet can change the milk fatty acid profile and FY (Bauman et al., 2006; Bionaz and Loor, 2008a). In addition, performing a proper dry-cow therapy and adjusting milking machines with adequate suction pressure can substantially decrease the average SCS in the herd (Nyman et al., 2007; Oliveira et al., 2015). However, it is well established that all those mentioned traits are under hormonal control. In this sense, milk synthesis is directly influenced by the complementary action of several hormones, such as prolactin, somatotropin, corticosteroids, insulin, growth and thyroid (Baumgard et al., 2017), which are genetically regulated (Lucey et al., 2017; Osorio et al., 2016).

Genes related to MY and milk composition might be differentially expressed during lactation and across parities. Strucken et al. (2012) reported that the variance explained by certain loci changes from lactation to lactation in German Holstein Friesian cows. The mammary gland of first parity cows are still in development, and therefore, gene expression in this period might differ. Furthermore, there are studies showing that the genetic influence on milk traits vary throughout lactation (Strabel et al., 2005; Zavadilová et al., 2005; Brito et al., 2017). Therefore, a better understanding of the genetic and biological mechanisms associated with MY and milk composition traits in different lactation stages (e.g., beginning, middle and end of lactation) and across parities is crucial to enable more efficient genetic selection for these traits, as well as better management practices.

In spite of the fact that the majority of studies in dairy cattle are performed based on the Holstein breed, the scientific knowledge concerning lactation biology of the bovine mammary gland have advanced substantially over the past decades (Akers, 2017; Almeida and Eckersall, 2018). For instance, studies have shown that the expression of genes related to milk traits, such as ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 is affected by lactation stage in Holstein cows of second or greater parity (Bionaz and Loor, 2008b). In addition, Strucken et al. (2012b) performed a genome-wide association study (**GWAS**) for milk production traits in Holstein Friesians cows, defining the traits based on 10-day intervals over the first 60 lactation days and reported significant changes in effect sizes over lactation. However, the underlying lactation biology and genes associated with milk traits might differ across breeds and populations, highlighting the need to investigate candidate genes affecting milk traits throughout the lactation stages and parities in different breeds.

In this context, GWAS is essential to identify genetic variants and genomic regions associated with complex traits, which may contribute to understanding the genetic architecture and underlying biology of complex traits in different breeds. Therefore, the main objectives of this study were to: 1) perform GWAS for MY, FY, PY and SCS based on lactation stages in the first three parities of Canadian Ayrshire, Holstein and Jersey animals; 2) investigate prospective candidate genes related to the mentioned traits in each lactation stage; and 3) identify genes (or genomic regions) and biological processes that are shared among dairy cattle breeds, lactation stages and parities.

8.3. Material and methods

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

Datasets

The Canadian Dairy Network (**CDN**; Guelph, ON, Canada) provided the pedigree, genotypes and the December/2012 official genetic evaluation datasets, containing the estimates of the random regression genetic coefficients and reliabilities for MY, FY, PY and SCS from the first three parities of Ayrshire, Holstein and Jersey breeds. In brief, the random regression genetic coefficients and reliabilities were calculated by CDN using two multiple-trait random regression test-day models (**RRM**; Schaeffer et al., 2000): one considering MY, FY and PY in the first three parities; and the other one considering SCS in the first three parities. Both models included the fixed effects of herd-test-day-parity and days in milk (**DIM**; from 5 to 305 days), fixed regression curves for year-season-region-parity-age of calving effect, and random regression curves for herd-year of calving, additive genetic and permanent environmental effects. Fixed and random regressions were modeled through fourth order Legendre orthogonal polynomials (Kirkpatrick et al., 1990) and heterogeneity of residual variance was taken into account by fitting 10 different classes in each parity. More details about the RRM used by CDN, as well as the estimated variance components and heritabilities, are presented in Interbull (2018a,b). The number of animals included in the December/2012 official genetic evaluations was 188,045, 4,888,808 and 134,631, for the Ayrshire, Holstein and Jersey breeds, respectively.

Genotypes. Animals were initially genotyped with or imputed to the Illumina BovineSNP50K BeadChip (Illumina, San Diego, CA, USA) using the FImpute software (Sargolzaei et al., 2011). Details about the genotyping quality control and imputation accuracy are presented in Larmer et al. (2014). After the genotype imputation process performed by CDN, single nucleotide polymorphisms (**SNP**) with Mendelian conflicts, call rate less than 0.95, minor allele frequency (**MAF**) less than 0.01 and with a difference between observed and expected heterozygosity frequency higher than 0.15 (Wiggans et al., 2009) were removed using the preGSf90 software (Aguilar et al., 2014). The number of SNPs on the 29 bovine autosomes (**BTA**) that remained after quality control was 41,517, 43,556 and 38,194, for the Ayrshire, Holstein and Jersey breeds, respectively. Only animals with minimum reliability of estimated breeding values

(**EBV**) equal to 0.65 for Holstein and 0.50 for Ayrshire and Jersey breeds were included in the genomic analysis. The number of animals that remained in the analyses, birth year ranges and average reliability of EBV, for each breed and trait, are shown in Table 1.

Table 1. Total number of genotyped animals (N, i.e., bulls and cows) and bulls (N bulls), years of birth (YOB), mean and standard deviation (SD) of reliability of estimated breeding values (REL_{EBV}), for each breed and trait.

Trait	Parameters	^{1}Ayrshire	^{2}Holstein	^{1}Jersey
MY	N	1,748	10,798	973
	N bulls	526	7,322	546
	YOB	1961 to 2010	1957 to 2009	1960 to 2010
	REL_{EBV} (SD)	0.74 (0.13)	0.85 (0.12)	0.74 (0.15)
FY	N	1,715	12,034	917
	N bulls	537	7,325	529
	YOB	1961 to 2010	1957 to 2009	1960 to 2010
	REL_{EBV} (SD)	0.72 (0.14)	0.83 (0.12)	0.73 (0.15)
PY	N	1,716	8,865	917
	N bulls	537	7,289	527
	YOB	1961 to 2010	1957 to 2008	1960 to 2010
	REL_{EBV} (SD)	0.72 (0.13)	0.88 (0.10)	0.73 (0.15)
SCS	N	1,141	7,329	524
	N bulls	526	7,098	430
	YOB	1961 to 2009	1957 to 2008	1960 to 2008
	REL_{EBV} (SD)	0.68 (0.17)	0.88 (0.06)	0.74 (0.16)

Traits: milk (MY), fat (FY) and protein (PY) yield, and somatic cells score (SCS). ^{1}Minimum REL_{EBV} equal to 0.50; ^{2}Minimum REL_{EBV} equal to 0.65.

Prediction of Daily Direct Genomic Values. The steps to estimate daily direct genomic values (**DGV**) are fully presented in Oliveira et al. (2018). In brief, genomic random regression coefficients were predicted using RRM and the Genomic Best Linear Unbiased Prediction approach (**GBLUP**; VanRaden, 2008). For this purpose, random regression genetic coefficients were used to calculate EBV from 5 to 305 DIM. Thereafter, these daily EBVs were deregressed according to VanRaden et al. (2009) method and posteriorly used as pseudo-phenotypes in the GBLUP analyses. These GBLUP analyses were performed using single-trait RRMs (i.e., GBLUP analyses were performed individually for each trait and parity within each breed), as follow:

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

where y is the vector of pseudo-phenotypes (i.e., deregressed daily EBVs); $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices for the vectors of fixed ($\mathbf{b}$) and random ($\mathbf{g}$) regression coefficients, which were modelled by fourth order Legendre orthogonal polynomial (Kirkpatrick et al., 1990); and **e** is the vector of residuals. The model assumptions were:

$$\mathrm{E}[\mathbf{y}] = \mathbf{Xb} \text{ and } \mathrm{Var}\begin{bmatrix}\mathbf{g}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G}\otimes\mathbf{G}_0 & 0\\ 0 & \mathbf{I}\sigma_e^2\end{bmatrix},$$

where $\mathbf{G}_0$ is the variance-covariance matrix of the genomic random regression coefficients; σ_e^2 is the residual variance, I is an identity matrix, and **G** is the genomic relationship matrix created as proposed by VanRaden (2008):

$$\mathbf{G} = 0.95\frac{\mathbf{WW}'}{2\sum_{k=1}^{K} p_k(1-p_k)} + 0.05\mathbf{A},$$

where $\mathbf{W} = \mathbf{M} - \mathbf{P}$, in which **M** contains the centered genotypes (i.e., -1, 0 and 1 to represent AA, Aa and aa, respectively), **P** contains the allele frequency for SNP k (p_k) in its kth column, expressed as $2(p_k - 0.5)$; $2\sum_{k=1}^{K} p_k(1-p_k)$ is a scaling parameter; and **A** is the traditional (pedigree-based) additive relationship matrix. In brief, adding 0.05 of **A** guarantees that **G** is positive definite (VanRaden, 2008). The AIREMLF90 software (Misztal et al., 2002) was used to obtain the solutions of the mixed model equations for all GBLUP analyses.

The vector of DGV of each animal i, which included daily DGVs from all DIM (from 5 to 305 days) for each trait and parity within each breed, was obtained by multiplying the vector of predicted genomic random regression coefficients by the matrix of Legendre orthogonal polynomial covariates, i.e.:

$$\mathbf{DGV}_i = \mathbf{T}\hat{\mathbf{g}}_i,$$

where $\hat{\mathbf{g}}_i$ is the vector of predicted genomic coefficients for each animal i; and $\mathbf{T}$ is a matrix of orthogonal covariates associated with the Legendre orthogonal polynomial functions.

GWAS analyses

The GWAS analyses were performed for all traits, parities and breeds considering three different lactation stages: 1) from 5 to 95 DIM, representing the ascending production stage and lactation peak; 2) from 96 to 215 DIM, representing the lactation persistency phase; and 3) from 216 to 305 DIM, representing the production decline at the end of lactation. Therefore, DGV for each lactation stage of each animal i (for each trait and parity) was obtained summing up the daily DGV solutions of the specific DIM, i.e.:

$$\mathrm{DGV1_i = D\hat{G}V_{i_5} + D\hat{G}V_{i_6} + \ldots + D\hat{G}V_{i_{95}}}\ ,$$

$$\mathrm{DGV2_i = D\hat{G}V_{i_{96}} + D\hat{G}V_{i_{97}} + \ldots + D\hat{G}V_{i_{215}}}\ \text{, and}$$

$$\mathrm{DGV3_i = D\hat{G}V_{i_{216}} + D\hat{G}V_{i_{217}} + \ldots + D\hat{G}V_{i_{305}}}\ ,$$

where $\mathrm{DGV1_i}$, $\mathrm{DGV2_i}$ and $\mathrm{DGV3_i}$ are the DGVs for the first, second and third lactation stage of the animal i, obtained by summing up the DGVs from 5 to 95, 96 to 215, and 216 to 305 DIM, respectively.

The SNP effects for each lactation stage were estimated individually for each trait, parity and breed using the postGSf90 software (Aguilar et al., 2014). In summary, the SNP effects were estimated as follows:

$$\mathbf{\hat{u}1 = M'[MM']^{-1}D\hat{G}V1}\ ,$$

$$\mathbf{\hat{u}2 = M'[MM']^{-1}D\hat{G}V2}\ \text{, and}$$

$$\mathbf{\hat{u}3 = M'[MM']^{-1}D\hat{G}V3}\ ,$$

in which $\mathbf{\hat{u}1}$, $\mathbf{\hat{u}2}$, and $\mathbf{\hat{u}3}$ are the vectors of estimated SNP effects for the first, second and third lactation stage, respectively; $\mathbf{M}$ is the already mentioned matrix that contains the centered genotypes, and $\mathbf{D\hat{G}V1}$, $\mathbf{D\hat{G}V2}$, and $\mathbf{D\hat{G}V3}$ are the vectors of estimated DGVs, which include the DGVs from all animals of the same trait, parity and breed, for the first, second and third lactation stage, respectively.

In order to identify important genomic regions related to the analyzed traits, moving windows (SNP-by-SNP) of 20 adjacent SNPs explaining more than 0.30% of total genetic variance (calculated by summing the variance explained by each of 20 SNPs) were selected for further analyses. The threshold of 0.30% was chosen according to other studies investigating complex traits in cattle (e.g., Soares et al., 2017; Silva et al., 2017), and to allow the retrieval of important windows for all analyzed traits. In addition, genomic windows were defined based on 20 adjacent SNPs to avoid the large noise and the absence of peaks due to the use of small and large window sizes, respectively (Fragomeni et al., 2014). After selecting the important genomic windows, a search for non-overlapping regions was performed. Hence, only the non-overlapping windows that explained the highest variances for each lactation stage, trait, parity and breed were retained aiming to avoid double-counting of SNP effects (Fragomeni et al., 2014).

Linkage disequilibrium. To understand and compare the GWAS results from each breed, we estimated the extension of linkage disequilibrium (**LD**; Hill and Robertson, 1968). For computational ease, LD was determined only at distances lower than 0.50 Mb for all syntenic marker pairs. The LD between each two evaluated loci was calculated using PLINK 2.0 software (Chang et al., 2015) as follows:

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

where $D = f(AB) - f(A)f(B)$, and f(AB), f(A), f(a), f(B) and f(b) are observed frequencies of AB, A, a, B, and b, respectively.

Identification of positional candidate genes

In order to identify possible candidate genes associated with MY and milk composition traits, genes that were located within the most important non-overlapping genomic windows (i.e., between the start and end of genomic coordinates) were further investigated. Positional candidate genes were annotated using the Biomart tool (Kinsella et al., 2011) embedded in the Ensembl Genes database version 93 (http://useast.ensembl.org/index.html). The bovine genome UMD 3.1 was used as the reference genome.

Functional prioritization and candidate genes identification

In order to identify the candidate genes and, consequently, to reduce the number of candidate genes likely associated with the studied traits, a functional prioritization analysis was performed using the MeSH ORA tool (Tsuyuzaki et al., 2015), available in the BIOCONDUCTOR package of the R software (R Development Core Team, 2011). The MeSH ORA performs an enrichment analysis based on Medical Subject Headings (**MeSH**) vocabulary to retrieve statistically overrepresented annotations (**ORA**), i.e., annotations that appear more frequently in a selected gene group relative to their occurrence by chance in a set of reference genes (Nelson et al., 2004). Nowadays, MeSH analysis has proven to be useful for deducing meaning from sets of genes across several species, including cattle, swine, horses and chicken (Morota et al., 2015, 2016). The statistical significance of enrichment analyses was assessed by hypergeometric test (Adams and Skopek, 1987), in which the p-value was given by:

$$p = \sum_{i=k}^{\min(M,n)} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}} ,$$

where N is the total number of genes that were analyzed for each trait (reference genes), M is the total number of selected genes, n is the total number of genes in the MeSH term under study, k is the number of selected genes that belong to the MeSH term under study, and $\binom{y}{x} = \frac{y!}{x!}(y-x)!$ is the binomial coefficient. In order to increase the power of hypergeometric test for the smaller-sized breeds (i.e., Ayrshire and Jersey), and to reduce the number of false positive results, all lactation stages, parities and breeds from a same trait were analyzed together in the MeSH ORA.

Subsequently, selected genes retrieved from the significant (p-value < 0.01) MeSH terms were considered as candidate genes for the lactation stage, parity and breed that they were uncovered. Thereafter, to understand the biological processes related to the studied traits and define the most likely genes, complete gene functions were obtained from the National Center for Biotechnology Information database (**NCBI**, www.ncbi.nlm.nih.gov/gene/) and Animal QTL Database (**QTLdb**, www.animalgenome.org/cgi-bin/QTLdb/BT/index). The most likely candidate genes that were identified in each lactation stage, trait and breed were selected for further discussion in this paper.

8.4. Results and discussion

Identification of important genomic windows and linkage disequilibrium

Descriptive statistics of the important genomic windows, in each lactation stage, parity and breed, are summarized in Tables 2 to 5, for MY, FY, PY and SCS. The location of each important window as well as the proportion of genetic variance explained by each one of them are shown in Tables S1 to S4 (Supplementary Material) for MY, FY, PY and SCS. The number of important genomic windows was higher for Jersey (705), followed by Ayrshire (337) and Holstein (118).

Table 2. Number of important genomic windows (N), mean, standard deviation (SD), maximum (Max) and minimum (Min) proportion of genetic variance explained (%) for milk yield in each lactation stage, parity and breed.

Parity	Breed	Stage	N	Mean	SD	Max	Min
First	**AY**	**1**	8	0.43	0.07	0.52	0.31
		2	9	0.39	0.08	0.57	0.31
		3	12	0.34	0.04	0.44	0.30
	HO	**1**	5	1.30	1.99	4.86	0.34
		2	5	2.12	3.70	8.74	0.33
		3	5	1.73	2.98	7.07	0.31
	JE	**1**	19	0.39	0.11	0.76	0.31
		2	13	0.40	0.10	0.69	0.30
		3	13	0.41	0.12	0.62	0.31
Second	**AY**	**1**	11	0.36	0.06	0.49	0.31
		2	9	0.39	0.06	0.49	0.32
		3	9	0.34	0.05	0.43	0.30
	HO	**1**	5	1.81	2.93	7.05	0.33
		2	4	2.61	4.13	8.80	0.37
		3	1	2.63	-	2.63	2.63
	JE	**1**	24	0.38	0.09	0.65	0.30
		2	12	0.40	0.10	0.60	0.30
		3	16	0.38	0.09	0.65	0.31
Third	**AY**	**1**	11	0.36	0.06	0.52	0.31
		2	10	0.36	0.06	0.51	0.30
		3	8	0.36	0.06	0.44	0.30
	HO	**1**	4	1.82	2.70	5.87	0.39
		2	5	2.06	3.54	8.38	0.31
		3	1	2.33	-	2.33	2.33
	JE	**1**	18	0.41	0.11	0.71	0.30
		2	16	0.38	0.09	0.58	0.31
		3	13	0.39	0.10	0.67	0.32

Breeds: Ayrshire (AY), Holstein (HO) and Jersey (JE). Traits: milk (MY), fat (FY) and protein (PY) yield, and somatic cells score (SCS). Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

The higher number of important windows detected for the Jersey breed is likely due to the fixed windows size and to the higher LD found in this breed, which is a direct consequence of its smaller effective population size compared to the other breeds (Stachowicz et al., 2011). Average extension of LD for markers separated by up to 0.50 Mb was 0.214, 0.223 and 0.284, for the Ayrshire, Holstein and Jersey breeds, respectively. This indicates that the probability of finding the same Quantitative Trait Locus (**QTL**) in two consecutive non-overlapping windows is higher for the Jersey breed, since the windows' size was limited to 20 consecutive SNPs in all three breeds. Therefore, using larger windows size may be beneficial for future GWAS on the Jersey breed. Levels of LD found in this study were consistent with the literature (Sargolzaei et al., 2008; Larmer et al., 2014).

Table 3. Number of important genomic windows (N), mean, standard deviation (SD), maximum (Max) and minimum (Min) proportion of genetic variance explained (%) for fat yield in each lactation stage, parity and breed.

Parity	Breed	Stage	N	Mean	SD	Max	Min
First	**AY**	**1**	11	0.35	0.04	0.44	0.30
		2	11	0.35	0.05	0.46	0.30
		3	8	0.36	0.04	0.42	0.30
	HO	**1**	4	2.23	3.03	7.46	0.34
		2	5	2.48	3.84	10.15	0.36
		3	7	1.60	2.76	8.33	0.31
	JE	**1**	13	0.46	0.17	0.88	0.31
		2	15	0.39	0.08	0.61	0.32
		3	15	0.38	0.07	0.58	0.31
Second	**AY**	**1**	8	0.38	0.05	0.44	0.33
		2	8	0.36	0.08	0.52	0.30
		3	7	0.37	0.06	0.47	0.31
	HO	**1**	2	4.13	4.92	7.61	0.65
		2	3	3.22	4.74	8.69	0.33
		3	2	3.34	4.03	6.19	0.49
	JE	**1**	16	0.38	0.11	0.68	0.30
		2	20	0.36	0.06	0.53	0.30
		3	16	0.38	0.07	0.54	0.30
Third	**AY**	**1**	9	0.38	0.06	0.48	0.31
		2	6	0.39	0.10	0.58	0.32
		3	8	0.36	0.05	0.45	0.30
	HO	**1**	2	5.35	6.62	10.03	0.67
		2	3	3.81	5.74	10.44	0.33
		3	2	3.43	4.21	6.41	0.45
	JE	**1**	19	0.39	0.12	0.73	0.30
		2	21	0.36	0.06	0.52	0.30
		3	19	0.36	0.05	0.50	0.30

Breeds: Ayrshire (AY), Holstein (HO) and Jersey (JE). Traits: milk (MY), fat (FY) and protein (PY) yield, and somatic cells score (SCS). Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

Table 4. Number of important genomic windows (N), mean, standard deviation (SD), maximum (Max) and minimum (Min) proportion of genetic variance explained (%) for protein yield in each lactation stage, parity and breed.

Parity	Breed	Stage	N	Mean	SD	Max	Min
First	**AY**	**1**	9	0.38	0.06	0.49	0.31
		2	9	0.38	0.05	0.45	0.30
		3	10	0.35	0.07	0.51	0.30
	HO	**1**	3	0.72	0.64	1.46	0.32
		2	4	0.75	0.80	1.94	0.33
		3	4	0.63	0.59	1.52	0.31
	JE	**1**	18	0.41	0.08	0.53	0.30
		2	20	0.39	0.09	0.58	0.30
		3	22	0.38	0.08	0.58	0.31
Second	**AY**	**1**	10	0.33	0.03	0.41	0.30
		2	8	0.36	0.06	0.47	0.30
		3	10	0.35	0.05	0.43	0.31
	HO	**1**	3	0.84	0.84	1.80	0.34
		2	2	0.94	0.81	1.51	0.37
		3	1	0.31	-	0.31	0.31
	JE	**1**	20	0.40	0.07	0.54	0.32
		2	13	0.43	0.10	0.59	0.31
		3	16	0.40	0.12	0.68	0.31
Third	**AY**	**1**	8	0.35	0.04	0.42	0.31
		2	7	0.36	0.05	0.44	0.30
		3	7	0.36	0.05	0.43	0.30
	HO	**1**	1	0.85	-	0.85	0.85
		2	2	0.66	0.49	1.01	0.32
		3	1	0.29	-	0.29	0.29
	JE	**1**	18	0.43	0.09	0.64	0.34
		2	16	0.42	0.09	0.56	0.30
		3	15	0.40	0.11	0.68	0.30

Breeds: Ayrshire (AY), Holstein (HO) and Jersey (JE). Traits: milk (MY), fat (FY) and protein (PY) yield, and somatic cells score (SCS). Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

A lower number of windows with relatively higher proportion of the explained genetic variance was found in the Holstein breed compared to the Ayrshire and Jersey breeds, for MY, FY and PY. This fact suggests that there might be some major genes being differentially expressed in the Holstein breed compared to the others. Pryce et al. (2010), validating SNP and haplotypes associated with milk production traits in Holstein-Friesian and Jersey sires, also found that the average variation explained by SNP was higher in Holstein than in Jersey. However, the authors suggested that their findings are likely a consequence of the population used to select the SNPs, which was mostly composed by Holstein animals.

Table 5. Number of important genomic windows (N), mean, standard deviation (SD), maximum (Max) and minimum (Min) proportion of genetic variance explained (%) for somatic cell score in each lactation stage, parity and breed.

Parity	Breed	Stage	N	Mean	SD	Max	Min
First	**AY**	**1**	11	0.38	0.05	0.47	0.32
		2	11	0.37	0.09	0.55	0.30
		3	9	0.40	0.10	0.62	0.33
	HO	**1**	3	0.38	0.06	0.45	0.32
		2	2	0.37	0.09	0.44	0.31
		3	4	0.34	0.04	0.40	0.31
	JE	**1**	30	0.41	0.10	0.67	0.30
		2	28	0.43	0.09	0.64	0.31
		3	30	0.41	0.09	0.70	0.30
Second	**AY**	**1**	12	0.36	0.07	0.57	0.30
		2	13	0.38	0.08	0.57	0.30
		3	10	0.39	0.07	0.55	0.30
	HO	**1**	5	0.39	0.08	0.52	0.31
		2	6	0.39	0.08	0.51	0.31
		3	6	0.39	0.09	0.51	0.31
	JE	**1**	29	0.41	0.11	0.72	0.30
		2	28	0.42	0.10	0.72	0.31
		3	26	0.40	0.10	0.62	0.30
Third	**AY**	**1**	10	0.36	0.07	0.54	0.31
		2	12	0.37	0.05	0.44	0.31
		3	8	0.41	0.06	0.50	0.32
	HO	**1**	2	0.43	0.12	0.51	0.34
		2	2	0.37	0.06	0.41	0.33
		3	2	0.36	0.07	0.41	0.31
	JE	**1**	25	0.42	0.09	0.63	0.31
		2	25	0.43	0.10	0.71	0.31
		3	28	0.41	0.12	0.85	0.30

Breeds: Ayrshire (AY), Holstein (HO) and Jersey (JE). Traits: milk (MY), fat (FY) and protein (PY) yield, and somatic cells score (SCS). Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

Similar number of important windows were found between different lactation stages of the same parity and trait, for all breeds (Tables 2 to 5). However, for the majority of traits and breeds, a lower number of important windows were found for the third parity compared to the first and second parity of the same trait. The lower number of important windows found in the third parity suggests that the heritability reduction that occurs for MY and milk composition traits from the first to the third parity is also related to the reduction in their genetic control, not only to the greater influence of environmental effects. Average heritabilities of MY, FY, PY and SCS along the lactation curve for the first three parities of Canadian Ayrshire, Holstein and Jersey are detailed in Interbull (2018a,b). For all breeds and parities, windows with low proportion of the explained genetic variance were found especially for SCS, which might be related to the more quantitative nature of this trait compared to MY, FY and PY.

Identification of candidate genes

The total number of genes identified in the important non-overlapping windows for each lactation stage, trait, parity and breed, before and after the prioritization analyses, are shown in Table S5 (Supplementary Material). In general, the prioritization analyses performed using MeSH ORA reduced the number of genes considerably, which facilitated the interpretation of the results. After the prioritization analyses, the statistically significant MeSH ORA terms were associated to 18, 23 and 39 different genes grouped in four categories (i.e., “Anatomy”, “Chemical and Drugs”, “Diseases”, and “Phenomena and Process”), for the Ayrshire, Holstein and Jersey breeds, respectively. The statistically significant ($p\text{-value} < 0.01$) MeSH ORA terms and the genes associated to them are shown in Tables 6 to 9, for MY, FY, PY and SCS.

In general, different candidate genes were found among traits and breeds. However, relatively similar genes were found across lactation stages and parities for the same trait (Tables 6 to 9). Genomic regions associated with the analyzed traits were located on 12, 8 and 15 different chromosomes for the Ayrshire, Holstein and Jerseys breeds, respectively. This result is in agreement with Buitenhuis et al. (2014), who also reported candidate genes spread in a larger number of chromosomes for Jersey compared to Holstein for milk-fat composition in Danish animals. No studies comparing candidate gene results for the Ayrshire with the Holstein and Jersey breeds were found in the literature.

Table 6. Statistically significant (p-value < 0.01) MeSH terms associated to the selected genes for three categories (C1 - Anatomy, C2 - Chemical and Drugs, and C3 - Phenomena and Process) for milk yield in each lactation stage, parity and breed (Ayrshire, AY; Holstein, HO; and Jersey, JE).

Parity	Breed	Stage	(Category) MeSH term	Gene
First	**AY**	**1**	**(C1)** Lactation	MTR
		2	**(C2)** Integrin alphaXbeta2, Protein S	C4BPA, C4BPB
		3	**(C2)** Integrin alphaXbeta2, Protein S	C4BPA, C4BPB
	HO	**1**	**(C1)** Golgi Apparatus, **(C2)** Adenosine Diphosphate, Amyloid, Bicuculline, Glucagon, Lactalbumin, Luminescent Proteins, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen[1], Prekallikrein, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, **(C3)** Lactation, Hydrodynamics	BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		2	**(C1)** Golgi Apparatus, **(C2)** Adenosine Diphosphate, Amyloid, Bicuculline, Glucagon, Lactalbumin, Luminescent Proteins, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen[1], Prekallikrein, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, **(C3)** Lactation, Hydrodynamics	BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		3	**(C2)** Diacylglycerol O-Acyltransferase, **(C3)** Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
	JE	**1**	**(C2)** Adenosine Diphosphate	TTL
		2	**(C2)** Adenosine Diphosphate	TTL
		3	**(C2)** Adenosine Diphosphate	TTL
Second	**AY**	**1**	**-**	-
		2	**(C2)** Integrin alphaXbeta2, Protein S	C4BPA, C4BPB
		3	**(C3)** Hydrodynamics	SDC1
	HO	**1**	**(C1)** Golgi Apparatus, **(C2)** Adenosine Diphosphate, Amyloid, Bicuculline, Glucagon, Lactalbumin, Luminescent Proteins, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen[1], Prekallikrein, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, **(C3)** Lactation, Hydrodynamics	ARFRP1, BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		2	**(C2)** Diacylglycerol O-Acyltransferase, **(C3)** Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		3	**(C2)** Diacylglycerol O-Acyltransferase, **(C3)** Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
	JE	**1**	**(C3)** Adenosine Diphosphate, Lactation, Luminescent Proteins	ARPC2, GPBAR1, HSP90B1, MYO10, SLC11A1, SLC25A4, TCF7L2
		2	**(C2)** Adenosine Diphosphate	TTL
		3	**(C2)** Adenosine Diphosphate	TTL
Third	**AY**	**1**	**-**	-
		2	**(C2)** Integrin alphaXbeta2, Protein S	C4BPA, C4BPB
		3	**(C3)** Hydrodynamics	SDC1
	HO	**1**	**(C1)** Golgi Apparatus, **(C2)** Amyloid, Bicuculline, Glucagon, Lactalbumin, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen[1], Prekallikrein, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, **(C3)** Lactation, Hydrodynamics	BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		2	**(C2)** Diacylglycerol O-Acyltransferase, **(C3)** Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		3	**(C2)** Diacylglycerol O-Acyltransferase, **(C3)** Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
	JE	**1**	**-**	-
		2	**(C2)** Adenosine Diphosphate	TTL
		3	**(C2)** Adenosine Diphosphate	TTL

Table 7. Statistically significant (p-value < 0.01) MeSH terms associated to the selected genes for two categories (C1 - Anatomy and C2 - Chemical and Drugs) for fat yield in each lactation stage, parity and breed (Ayrshire, AY; Holstein, HO; and Jersey, JE).

Parity	Breed	Stage	(Category) MeSH term	Gene
First	AY	1	**(C1)** Kidney Glomerulus, **(C2)** Follicle Stimulating Hormone, Nucleoplasmins	BMP2, ICAM1, PIN1
		2	**(C1)** Kidney Glomerulus, **(C2)** Follicle Stimulating Hormone, Nucleoplasmins	BMP2, ICAM1, PIN1
		3	**(C1)** Kidney Glomerulus, **(C2)** Follicle Stimulating Hormone, Myosin Light Chains	ICAM1, PIN1, ROCK2
	HO	1	**(C2)** Diacylglycerol O-Acyltransferase	CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, RECQL4, TONSL, VPS28
		2	**(C2)** Diacylglycerol O-Acyltransferase	BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, MGP, RECQL4, TONSL, VPS28
		3	**(C2)** Diacylglycerol O-Acyltransferase	BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, RECQL4, TONSL, VPS28
	JE	1	**(C2)** Calcitriol	ANXA2
		2	**(C2)** Calcitriol	ANXA2
		3	**(C2)** Calcitriol, Follicle Stimulating Hormone	ANXA2, CYP19A1
Second	AY	1	**(C1)** Kidney Glomerulus, **(C2)** Follicle Stimulating Hormone, Nucleoplasmins	BMP2, ICAM1, PIN1
		2	**(C1)** Kidney Glomerulus, **(C2)** Follicle Stimulating Hormone, Myosin Light Chains, Nucleoplasmins	BMP2, HIF1A, ICAM1, PIN1, ROCK2
		3	-	-
	HO	1	**(C2)** Diacylglycerol O-Acyltransferase	BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		2	**(C2)** Calcitriol, Diacylglycerol O-Acyltransferase	BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MGP, RECQL4, TONSL, VPS28
		3	**(C2)** Diacylglycerol O-Acyltransferase	BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
	JE	1	**(C2)** Calcitriol, Follicle Stimulating Hormone, Glucuronosyltransferase, Myosin Light Chains, Interferon Regulatory Factor 3 and 7, Diacylglycerol O-Acyltransferase	ANXA2, BOP1, DGAT1, FAS, HSP90B1, IFIT3, IFIT5, STAB2
		2	**(C1)** Cumulus Cells, **(C2)** Calcitriol, Metalloendopeptidases, Glucuronosyltransferase	ANXA2, TMSB4
		3	**(C1)** Cumulus Cells, **(C2)** Calcitriol, Metalloendopeptidases, Glucuronosyltransferase	ANXA2, TMSB4
Third	AY	1	**(C1)** Kidney Glomerulus, **(C2)** Follicle Stimulating Hormone, Metalloendopeptidases, Protease Inhibitors, Tissue Inhibitor of Metalloproteinase-1, Matrix Metalloproteinase, Nucleoplasmins	BMP2, HIF1A, ICAM1, PIN1, TIMP3
		2	**(C1)** Kidney Glomerulus, **(C2)** Follicle Stimulating Hormone, Nucleoplasmins	BMP2, ICAM1, KPNA7, PIN1
		3	Nucleoplasmins	KPNA7
	HO	1	**(C2)** Diacylglycerol O-Acyltransferase	BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MGP, RECQL4, TONSL, VPS28
		2	**(C2)** Calcitriol, Diacylglycerol O-Acyltransferase	BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MGP, RECQL4, TONSL, VPS28
		3	**(C2)** Diacylglycerol O-Acyltransferase	BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
	JE	1	**(C2)** Calcitriol	ANXA2
		2	**(C2)** Calcitriol	ANXA2
		3	**(C1)** Cumulus Cells, **(C2)** Follicle Stimulating Hormone, Glucuronosyltransferase, Myosin Light Chains, Diacylglycerol O-Acyltransferase	BOP1, CYP26A1, DGAT1, FASLG, HSP90B1, STAB2

Table 8. Statistically significant (p-value < 0.01) MeSH terms associated to the selected genes for three categories (C1 - Anatomy, C2 - Chemical and Drugs, and C3 - Phenomena and Process) for protein yield in each lactation stage, parity and breed (Ayrshire, AY; Holstein, HO; and Jersey, JE).

Parity	**Breed**	**Stage**	**(Category) MeSH term**	**Gene**
First	**AY**	**1**	**(C1)** Colostrum, Kidney Glomerulus, **(C3)** Lactation	BAX, FCGRT, ICAM1
		2	**(C1)** Kidney Glomerulus, **(C3)** Lactation	ICAM1, MTR
		3	**(C1)** Kidney Glomerulus	ICAM1
	HO	**1**	**(C2)** Growth Hormone, Diacylglycerol O-Acyltransferase, **(C3)** Energy Metabolism, Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		2	**(C1)** Colostrum, **(C2)** Aldehyde Reductase, Amyloid, Bicuculline, Cortisone Reductase, Glucagon, Lactalbumin, Pancreatic Hormones, Pancreatin, Pituitary Hormones, Plasminogen, Prekallikrein, Prolactin, Growth Hormone, Thyroid Hormones, Receptor Dopamine, Receptor Serotonin, Diacylglycerol O-Acyltransferase, Clusterin, **(C3)** Energy Metabolism, Lactation	AKR1C4, BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		3	**(C2)** Aldehyde Reductase, Cortisone Reductase, Growth Hormone, Diacylglycerol O-Acyltransferase, **(C3)** Energy Metabolism, Lactation	AKR1C4, BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
	JE	**1**	**-**	-
		2	**-**	-
		3	**-**	-
Second	**AY**	**1**	**(C1)** Kidney Glomerulus, **(C2)** Prolactin, Growth Hormone, **(C3)** Lactation	ADRB2, CSH2, ICAM1
		2	-	-
		3	Lactation	PIK3R3
	HO	**1**	**(C2)** Growth Hormone, Diacylglycerol O-Acyltransferase, **(C3)** Energy Metabolism, Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		2	**(C2)** Growth Hormone, Diacylglycerol O-Acyltransferase, **(C3)** Energy Metabolism, Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		3	**(C2)** Aldehyde Reductase, Cortisone Reductase	AKR1C4
	JE	**1**	**-**	-
		2	**(C3)** Energy Metabolism, Lactation	COX4I1, HSP90B1
		3	**-**	-
Third	**AY**	**1**	**(C1)** Kidney Glomerulus, **(C2)** Prolactin, Growth Hormone, **(C3)** Energy Metabolism, Lactation	ADRB2, CSH2, COX4I1, ICAM1
		2	**-**	-
		3	**-**	-
	HO	**1**	**(C2)** Growth Hormone, Diacylglycerol O-Acyltransferase, **(C3)** Energy Metabolism, Lactation	BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		2	**(C2)** Aldehyde Reductase, Cortisone Reductase, Growth Hormone, Diacylglycerol O-Acyltransferase, **(C3)** Energy Metabolism, Lactation	AKR1C4, BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL, VPS28
		3	**(C2)** Aldehyde Reductase, Cortisone Reductase	AKR1C4
	JE	**1**	**-**	-
		2	**-**	-
		3	**(C1)** Colostrum, Kidney Glomerulus, **(C3)** Energy Metabolism, Lactation	COL4A3, COX4I1, CYP26A1, HSP90B1

Table 9. Statistically significant (p-value < 0.01) MeSH terms associated to the selected genes for three categories (C1 - Anatomy, C2 - Chemical and Drugs, C3 – Diseases, and C4 - Phenomena and Process) for somatic cell score in each lactation stage, parity and breed (Ayrshire, AY; Holstein, HO; and Jersey, JE).

Parity	Breed	Stage	(Category) MeSH term	Gene
First	**AY**	**1**	**(C3)** Lysosomal Storage C3	FUCA1
		2	**(C3)** Lysosomal Storage C3	FUCA1
		3	-	-
	HO	**1**	**(C2)** Receptor Epidermal Growth Factor	SYNJ1
		2	-	-
		3	**(C1)** Golgi Apparatus, **(C2)** Large-Conductance Calcium-Activated Potassium Channels, Phosphothreonine, Quercetin, Receptor Epidermal Growth Factor, Calcium Channels, Adiponectin, Amphiregulin, **(C4)** Energy Metabolism, Hydrodynamics	NOS3, SYNJ1
	JE	**1**	**(C2)** Caseins, Receptor Epidermal Growth Factor, Calcium Channels, Serine Proteinase Inhibitors, Amyloid beta-Peptides, Diacylglycerol O-Acyltransferase, **(C3)** Lysosomal Storage C3	CACNA1A, CTSA, CTSD, MMP9, PTK2, SERPINB6, SERPINB9, TUBB2B
		2	**(C2)** Amyloid beta-Peptides, Calcium Channels, Caseins, Diacylglycerol O-Acyltransferase, Large-Conductance Calcium-Activated Potassium Channels, Receptor Epidermal Growth Factor, Receptor Somatotropin, Serine Proteinase Inhibitors, **(C3)** Lysosomal Storage C3	CACNA1A , CTSA, CTSD, IGF2, MMP9, PTI, PTK2
		3	**(C2)** Amyloid beta-Peptides, Calcium Channels, Caseins, Diacylglycerol O-Acyltransferase, Large-Conductance Calcium-Activated Potassium Channels, Receptor Epidermal Growth Factor, Receptors Somatotropin, Serine Proteinase Inhibitors, **(C3)** Lysosomal Storage C3	CACNA1A, CTSA, CTSD, IGF2, MMP9, PTI, PTK2
Second	**AY**	**1**	**(C3)** Lysosomal Storage C3	FUCA1
		2	-	-
		3	**(C2)** Amphiregulin, Receptor Epidermal Growth Factor	NRG1
	HO	**1**	**(C2)** Adiponectin, Calcium Channels, Diacylglycerol O-Acyltransferase, Receptor Epidermal Growth Factor, Receptor Somatotropin, **(C4)** Energy Metabolism	ANGPTL4, BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MCOLN1, RECQL4, RETN, SYNJ1, TONSL, VPS28
		2	**(C1)** Golgi Apparatus, **(C2)** Adiponectin, Amyloid, Amyloid beta-Peptides, Bicuculline, Calcium Channels, Caseins, Clusterin, Diacylglycerol O-Acyltransferase, Glucagon, Lactalbumin, Pancreatic Hormones, Pancreatin, Pancreatic Elastase, Pituitary Hormones, Plasminogen, Prekallikrein, Quercetin, Receptor Somatotropin, Receptor Dopamine, Receptor Serotonin, Thyroid Hormones, **(C4)** Energy Metabolism, Hydrodynamics	ANGPTL4, BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MCOLN1, RECQL4, RETN, STATH, TONSL,VPS28
		3	**(C1)** Golgi Apparatus, **(C2)** Adiponectin, Amyloid, Amyloid beta-Peptides, Bicuculline, Caseins, Calcium Channels, Clusterin, Diacylglycerol O-Acyltransferase, Glucagon, Lactalbumin, Pancreatic Hormones, Pancreatin, Pancreatic Elastase, Pituitary Hormones, Plasminogen, Prekallikrein, Quercetin Receptor Somatotropin, Receptor Dopamine, Receptor Serotonin, Thyroid Hormones, **(C4)** Energy Metabolism, Hydrodynamics	ANGPTL4, BOP1, CPSF1, CSN1S1, CSN1S2, CSN3, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, MCOLN1, RECQL4, RETN, STATH, TONSL,VPS28
	JE	**1**	**(C1)** Golgi Apparatus, **(C2)** Amyloid beta-Peptides, Calcium Channels, Caseins, Diacylglycerol O-Acyltransferase, Large-Conductance Calcium-Activated Potassium Channels, Receptor Epidermal Growth Factor, Receptor Somatotropin, Serine Proteinase Inhibitors, Receptor Antigen T-Cell alpha-beta, **(C4)** Genes MHC Class II	ARCN1, BOLA-DMA, BRD2, CD3E, CD3G, CTSD, IGF2, ITPR3, PLCB1, PTK2, PTI, PSMB9
		2	**(C2)** Amyloid beta-Peptides, Caseins, Calcium Channels, Diacylglycerol O-Acyltransferase, Large-Conductance Calcium-Activated Potassium Channels, Receptor Epidermal Growth Factor, Receptor Somatotropin, Serine Proteinase Inhibitors	CACNA1A, CTSD, IGF2, PLCB1, PTK2, PTI
		3	**(C2)** Amyloid beta-Peptides, Calcium Channels, Caseins, Diacylglycerol O-Acyltransferase, Receptor Epidermal Growth Factor, Receptor Somatotropin	CACNA1A, CTSD, IGF2, PTK2

Third	**AY**	**1**	-	-
		2	-	-
		3	**(C2)** Amphiregulin, Receptor, Epidermal Growth Factor	NRG1
	HO	**1**	**(C2)** Large-Conductance Calcium-Activated Potassium Channels	KCNU1
		2	**(C2)** Large-Conductance Calcium-Activated Potassium Channels	KCNU1
		3	**(C2)** Large-Conductance Calcium-Activated Potassium Channels	KCNU1
	JE	**1**	**(C1)** Golgi Apparatus, **(C2)** Amyloid beta-Peptides, Caseins, Calcium Channels, Diacylglycerol O-Acyltransferase, Receptor Epidermal Growth Factor, Receptor Somatotropin, Receptors Antigen T-Cell alpha-beta, **(C4)** Genes MHC Class II	ARCN1, CACNA1A, CD3E,CD3G, CTSD, BOLA-DMA, BRD2, IGF2, ITPR3, PSMB9, PTK2
		2	**(C1)** Golgi Apparatus, **(C2)** Amyloid beta-Peptides, Caseins, Calcium Channels, Diacylglycerol O-Acyltransferase, Receptor Epidermal Growth Factor Receptor Somatotropin, Receptor Antigen T-Cell alpha-beta	ARCN1, CACNA1A, CD3E,CD3G, CTSD, IGF2, PTK2
		3	**(C2)** Amyloid beta-Peptides, Caseins, Calcium Channels, Diacylglycerol O-Acyltransferase, Phosphothreonine, Receptor Epidermal Growth Factor, Receptor Somatotropin	CACNA1A, CTSD, IGF2, MAPT, PTK2

Candidate genes for MY. Overall, no similar candidate genes were found between the three breeds for MY. Genomic regions associated with MY were located on BTA11 (SDC1), BTA16 (C4BPA and C4BPB), and BTA28 (MTR) for the Ayrshire; BTA6 (CSN1S1, CSN1S2 and CSN3), BTA13 (ARFRP1), and BTA14 (BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL and VPS28) for the Holstein; and BTA2 (ARPC2, GPBAR1 and SLC11A1), BTA5 (HSP90B1), BTA11 (TTL), BTA20 (MYO10), BTA26 (TCF7L2), and BTA27 (SLC25A4) for the Jersey breed.

Our results suggested that the C4BPA *(C4BP α*-chain) and C4BPB (*C4BP β*-chain) genes are related to MY in the middle of lactation for the Ayrshire breed. The C4BPA and C4BPB genes are associated with component 4 binding protein alpha and beta, respectively (Wells et al., 2013). In addition, the SDC1 (syndecan 1) gene appear to be an important candidate gene related to MY for this breed, especially in the end of lactation. The SDC1 gene is a type 1 transmembrane heparan sulfate proteoglycans (HSPG) that interact with adhesion molecules, cytokines, coagulation factors and cell-extracellular matrix (ECM) proteins. The SDC1 gene was related to bovine placentomes during gestation, and Hambruch et al. (2017) showed that the expression of SDC1 increases in the end of gestation (from 250 to 270 gestation days), which may explain the association of SDC1 with MY specially at the end of the lactation.

Many of the candidate genes detected for MY in the Holstein breed corroborated with the literature (e.g., Lafanechère et al., 1998a; Nayeri et al., 2016). In general, several candidate genes appeared consistently in all lactation stages and parities, including CYHR1 (cysteine and histidine rich 1), DGAT1 (diacylglycerol O-acyltransferase 1), EEF1D (eukaryotic translation elongation factor 1 delta), FBXL6 (F-box and leucine rich repeat protein 6) and FOXH1 (forkhead box H1). This indicates that those genes are likely affecting physiological processes of milk synthesis along the entire lactation period. However, interestingly, the CSN1S1 (casein alpha s1), CSN1S2 (casein alpha s2) and CSN3 (casein kappa) genes appeared as candidate genes only in the beginning of lactation (i.e., in the first lactation stage), for all three parities. This fact suggest that those genes are related mainly to the beginning of MY, probably due to higher amount of milk yielded in this stage. The fact that the DGAT1 gene was found consistently in all lactation stages of the Holstein breed may be related to the definition of lactation stages used in this study (e.g., the first lactation stage was defined from 5 to 95 DIM). Therefore, probably the average of QTL effect estimated for this lactation stage was high, which did not allow our results corroborate with Strucken et al. (2011), who reported that known effects of the DGAT1 gene are mainly expressed after 40th lactation day.

For the Jersey breed, the only candidate gene found consistently in more than one lactation stage and parity was the TTL (tubulin tyrosine ligase) gene. Studies have shown that reducing the expression of TTL increases steady state levels of Glutubulin and promotes sarcoma growth (Lafanechère et al., 1998b), which may impact MY. In humans, suppressing the TTL expression is related to breast tumors and neuroblastomas of poor prognosis (Lafanechère et al., 1998b).

Candidate genes for FY. A lower number of candidate genes was found for FY in Ayrshire compared to the Holstein and Jersey breeds. Genomic regions associated with FY were located on BTA7 (ICAM1 and PIN1), BTA10 (HIF1A), BTA11 (ROCK2), BTA13 (BMP2) and BTA25 (KPNA7) for the Ayrshire; BTA5 (MGP) and BTA14 (BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, RECQL4, TONSL and VPS28) for the Holstein; and BTA1 (TMSB4), BTA5 (HSP90B1 and STAB2), BTA10 (ANXA2 and CYP19A1), BTA14 (BOP1 and DGAT1), BTA16 (FASLG) and BTA26 (CYP26A1, FAS, IFIT3 and IFIT5) for the Jersey breed.

In general, the ICAM1 (intercellular adhesion molecule 1) and PIN1 (peptidylprolyl cis/trans isomerase, NIMA-interacting 1) genes appeared as important genes in the expression of FY for Ayrshire during all lactation periods. Rezamand and McGuire (2011) showed that trans fatty acids (i.e., unsaturated fatty acids containing one or more double bonds in the trans configuration) upregulate expression of ICAM1. The partial hydrogenation of unsaturated fatty acids can result in production of elaidic acid to produce more solid fat in the milk (Rezamand and McGuire, 2011). On the other hand, the transcription of other lipolysis genes such as PLIN1, increases following an increase in the enzymatic capacity for continued supply of fatty acids to other organs and rebuilding adipose stores (McNamara, 1988). Different gene expression was suggested for BMP2 (bone morphogenetic protein 2), ROCK2 (Rho associated coiled-coil containing protein kinase 2), and KPNA7 (karyopherin subunit alpha 7). In general, the BMP2, ROCK2 and KPNA7 were candidate genes for the beginning-middle, middle-end, and end of lactation, respectively. The BMP2 was shown to induce commitment of pluripotent stem cells (C3H10T1/2) into adipocytes (Huang et al., 2009), such as occurs with BMP4 (Bowers et al., 2006).

For the Holstein breed, several candidate genes seem to influence the expression of FY during all lactation periods, including BOP1 (block of proliferation 1), CPSF1 (cleavage and polyadenylation specific factor 1), CYHR1 (cysteine and histidine rich 1), DGAT1 (diacylglycerol O-acyltransferase 1), FBXL6 (F-box and leucine rich repeat protein 6), FOXH1 (forkhead box H1), LRRC24 (leucine rich repeat containing 24), RECQL4 (RecQ like helicase 4), TONSL (tonsoku like, DNA repair protein) and VPS28 (ESCRT-I subunit) genes. These genes have also been reported in previous studies (e.g., Lafanechère et al., 1998a; Nayeri et al., 2016). The MGP (matrix Gla protein) gene was associated with FY during the early lactation stage and was previously reported to be associated with milk traits (Ron et al., 2007; Nayeri and Stothard, 2016).

Similarly to Holstein, the DGAT1 (diacylglycerol O-acyltransferase 1) and BOP1 (block of proliferation 1) were identified as important genes along the lactation for FY in Jersey. The presence of DGAT1 only for the FY trait in Jersey may be related to the higher effect of the K232A DGAT1 polymorphism on this trait compared to MY and PY, as showed by Spelman et al. (2002) in New Zealand Jersey cattle. However, specifically for Jersey animals, the ANXA2 (annexin A2) gene also appeared to be associated with FY during all lactation stages. There are evidences that the CYP19A1 (cytochrome P450, family 19, subfamily A, polypeptide 1) and CYP26A1 (cytochrome P450, family 26, subfamily A, polypeptide 1) genes are related to FY in the end of lactation in Jersey cattle. Contradictorily to our findings, Nayeri and Stothard (2016) reported the enzyme related to the CYP19A1 influencing physiological functions in early lactation.

Candidate genes for PY. When compared to the Holstein breed, a reduced number of candidate genes was found for PY in the Ayrshire and Jersey breeds. Genomic regions associated with PY were located on BTA3 (PIK3R3), BTA7 (ADRB2 and ICAM1), BTA18 (BAX, COX4I1 and FCGRT), BTA23 (CSH2) and BTA28 (MTR) for the Ayrshire; BTA6 (CSN1S1, CSN1S2 and CSN3), BTA13 (AKR1C4) and BTA14 (BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL and VPS28) for the Holstein; and BTA2 (COL4A3), BTA5 (HSP90B1), BTA18 (COX4I1), and BTA26 (CYP26A1) for the Jersey breed.

Among the genes identified, the ICAM1 (intercellular adhesion molecule 1) gene appeared to play a major role in the expression of PY along the lactation in the Ayrshire breed, while ADRB2 (adrenoceptor beta 2) and CSH2 (chorionic somatomammotropin hormone 2) are specially related to the beginning of lactation. The ADRB2 has been reported to be associated with milk urea nitrogen in the Brown Swiss breed (Cecchinato et al., 2014). The authors suggested that the effect of the ADRB2 gene may be related to the involvement of β-adrenergic receptors in lipolysis and the regulation of muscle growth to the detriment of fat deposition, since milk urea is synthesized as a consequence of an imbalance between dietary nitrogen and energy in the rumen.

Similarly to MY and FY, the number of candidate genes that appear as important along the entire lactation curve in the Holstein breed is large. Well-studied genes as BOP1 (block of proliferation 1), CPSF1 (cleavage and polyadenylation specific factor 1), CYHR1 (cysteine and histidine rich 1), DGAT1 (diacylglycerol O-acyltransferase 1), EEF1D (eukaryotic translation elongation factor 1 delta), FBXL6 (F-box and leucine rich repeat protein 6), FOXH1 (forkhead box H1), LRRC24 (leucine rich repeat containing 24), RECQL4 (RecQ like helicase 4), TONSL (tonsoku like, DNA repair protein) and VPS28 (ESCRT-I subunit) appeared to be influencing milk traits along all lactation curve. However, AKR1C4 (aldo-keto reductase family 1, member C4) gene seems to affect specifically the end of lactation. In humans, AKR1C4 is involved in the metabolism of potent signaling molecules such as bile acids (Jin and Penning, 2007).

In Jersey, the genes identified were COX4I1 (cytochrome c oxidase subunit 4I1); and HSP90B1 (heat shock protein 90 beta family member 1), which has been associated with heat tolerance in cattle (e.g., Sengar et al., 2018).

Candidate genes for SCS. Genomic regions associated with SCS were located on BTA2 (FUCA1) and BTA27 (NRG1) in Ayrshire; BTA1 (SYNJ1), BTA4 (NOS3), BTA7 (ANGPTL4, MCOLN1 and RETN), BTA14 (BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, RECQL4, TONSL and VPS28) and BTA27 (KCNU1) in Holstein; and BTA7 (CACNA1A), BTA13 (CTSA, MMP9, PLCB1 and PTI), BTA14 (PTK2), BTA15 (ARCN1, CD3E, CD3G), BTA23 (BOLA-DMA, BRD2, ITPR3, PSMB9, SERPINB6, SERPINB9 and TUBB2B) and BTA29 (CTSD and IGF2) in the Jersey breed.

For the Ayrshire breed, FUCA1 (alpha-L-fucosidase 1) and NRG1 (neuregulin 1) were identified as candidate genes. Recently, Song et al., (2016) showed that NRG1 was strongly correlated with the progression of Staphylococcus aureus subclinical mastitis, which likely impacts SCS.

In Holstein, most of the genes for SCS were identified for the second parity. Among the genes, CPSF1 (cleavage and polyadenylation specific factor 1) and DGAT1 (diacylglycerol O-acyltransferase 1) appeared to be influencing somatic cells score. These findings are in agreement with Kaupe et al. (2007) and Cochran et al. (2013).

In Jersey, CACNA1A (calcium voltage-gated channel subunit alpha1), CTSD (cathepsin D), PTK2 (protein tyrosine kinase 2) and IGF2 (insulin like growth factor 2) seemed to affect the traits expression throughout the entire lactation. However, the BOLA-DMA (major histocompatibility complex, class II, DM alpha), ARCN1 (archain 1), CD3E (CD3e molecule), CD3G (CD3g molecule;), ITPR3 (inositol 1,4,5-trisphosphate receptor type 3), PSMB9 (proteasome subunit beta 9) and BRD2 (bromodomain containing 2) seem to be more related to physiological processes at the beginning of lactation. Wang et al. (2013) reported that PTK2 variants are associated with milk production traits, and the authors suggested that it could be considered as convincing genetic markers for individual selection in future cattle breeding program. The PTK2 gene encodes a cytoplasmic non-receptor protein tyrosine kinase (Kanner et al., 1990), and it is related to several signal transduction pathways, such as cell motility and microtubule stability (Harte et al., 1996; Ezratty et al., 2005). In addition, the PTK2 has been found to be important for survival, proliferation, and differentiation of mammary epithelial cells in mice (Palmer et al., 2006). The CD3E and CD3G are part of the CD3 complex, that plays an important role in signal transduction in the T cell activation pathway (MacHugh et al., 1998).

Comparison between breeds, statistical challenges and opportunities

In general, different set of candidate genes were found for the different breeds in this study. Especially for the Holstein breed, many of the candidate genes identified corroborated with previous reports (e.g., Nayeri et al., 2016). However, well-known genes with major effect on milk production traits (e.g., DGAT1, Grisart, 2002) had different expression between lactation stages, traits and parities of different breeds. For instance, the DGAT1 explained high proportion of variation only in the Holstein traits, and FY in Jersey. In addition, there was no evidence of substantial effect of DGAT1 on MY and milk composition traits of Ayrshire breed, and on MY, PY and SCS of Jersey breed, which may explain the lower proportion of genetic variance explained by important windows in these breeds for these traits when compared to Holsteins. However, disagreeing with our results, Spelman et al. (2002) suggested that the DGAT1 mutation has great effect on milk production traits in all three breeds.

Several factors may influence the discrepancy between the set of candidate genes identified between our and previous studies reported in the literature. Among them, it is the fact that we have used “partial accumulated yield” to represent the traits in different lactation stages, while the majority of studies reported in the literature have been using the 305-days accumulated yield to analyze those traits (e.g., Spelman et al., 2002; Raven et al., 2014; Nayeri et al., 2016b). This difference in the phenotypes used in the GWAS probably influences the set of candidate genes identified due to the non-constant effects of QTL over time for those traits, as suggested by the changes in additive genetic variance and heritability over lactation (e.g., Miglior et al., 2009; Oliveira et al., 2017; Sasaki et al., 2017). Thus, our study allowed the identification of specific genes or genetic loci that were more pronounced during a specific lactation stage, not with a high average QTL effect through lactation. On the other hand, the difference between candidate genes identified for the different breeds analyzed in this study may be related to the differences in accuracy of phenotypes used, different genetic background of these breeds, and differences in sample sizes.

In this study we estimated the SNP effects back-solving the DGV estimated using deregressed EBV as response variable. The use of deregressed EBV is important to avoid the double-shrinkage of breeding values and the double-counting of information in the analysis (Garrick et al., 2009). In summary, the deregressed EBV used here were calculated as function of EBV and reliability of EBV for each DIM of each animal, in a way that animals with lower reliability of EBV had deregressed EBV farthest from the original EBV. Comparing the three breeds studied here, the increase in EBV variability was higher for the Ayrshire and Jersey breeds, because the threshold of minimum reliability of EBV to include animals in the analyses was lower for those breeds (the threshold was equal to 0.65 for Holstein and 0.50 for Ayrshire and Jersey breeds; Table 1). Thus, the Ayrshire and Jersey breeds had less accurate response variable when compared to the Holstein breed, which may influence in the detection of QTL (Beecher et al., 2010). Lower threshold was assumed for Ayrshire and Jersey compared to the Holstein breed due to the smaller number of genotyped animals available.

Another fact that may influence the set of candidate genes is the different genetic background between the breeds, as these breeds have been genetically separated for many generations (Kantanen, 2000) and have undergone strong artificial selection (Buitenhuis et al., 2016). Thus, the different selection processes that those breeds were subjected may have fixed different important alleles in each population, and it may reflect in different LD structure between breeds, as showed in this study. In addition, differences in the heritabilities could potentially explain the difference in the number of QTL detected for the different breeds (Buitenhuis et al., 2016).

Differences in sample sizes, i.e., differences in the number of genotyped animals, can also help to explain the difference between the candidate genes identified for the different breeds in this study. The number of genotyped animals is related to the statistical power, and it is a big issue in any GWAS (Meredith et al., 2013). In our study, the power of detecting QTL in the Ayrshire and Jersey breeds was lower than in the Holstein, which may result in a higher number of false-negative findings for those breeds. This may be specifically true for the SCS trait in the Jersey breed, where the number of genotyped animals was small, and the number of candidate genes was big. However, the number of genotyped animals is not under control of the researcher, and sometimes it cannot be increased. In this case, further studies using alternative methods should be used to validate previous results.

Alternative methodologies such as the use of a meta-analysis to combine information from independent studies (Evangelou and Ioannidis, 2013), the use of single-step GWAS considering different weights for different SNPs (Wang et al., 2014), and/or the joint consideration of a set of correlated traits (Pei et al., 2009) can be used in posterior studies to validate the findings from this study. The reason we merely adopted the single-trait analyses herein is because the model used by CDN to estimate the EBV was already a multiple-trait model (Interbull, 2018ab), i.e., the EBVs already contained the information of the other traits. In addition, we have opted for using the multiple-step (i.e., for estimating and deregressing the EBV to be used as pseudo-phenotype to estimate the DGV) instead of the single-step method because the single-step method is not currently used for official genomic evaluations in Canada (Interbull, 2017).

8.5. Conclusions

Genomic regions associated with MY, FY, PY and SCS in the first three parities were identified for different lactation stages in Ayrshire, Holstein and Jersey breeds. Evidence of differential set of candidate genes exists across breeds, traits, parities and lactation stages. As the impact of SNP effect changes over time, the investigation at certain lactation stages seems to contribute to detect quantitative trait genes with overall small effect and great effect at specific stages. Further studies are need to validate our findings.

8.6. Acknowledgements

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES). The authors gratefully acknowledge the financial support to this research from the Food from Thought research program, Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative.

8.7. References

Adams, W. T., and T. R. Skopek. 1987. Statistical test for the comparison of samples from mutational spectra. J. Mol. Biol. 194:391–396. doi:10.1016/0022-2836(87)90669-3.

Aguilar, I., S. Tsuruta, and A. Legarra. 2014. PREGSF90-POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide Association with Ungenotyped Individuals in BLUPF90 Programs. Page in 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada.

Akers, R. M. 2017. A 100-Year Review: Mammary development and lactation. J. Dairy Sci.. doi:10.3168/jds.2017-12983.

Almeida, A. M., and P. D. Eckersall. 2018. Proteomics and Mammary Gland Research in Dairy Species. Springer International Publishing, Cham.

Bauman, D. E., I. H. Mather, R. J. Wall, and A. L. Lock. 2006. Major Advances Associated with the Biosynthesis of Milk. J. Dairy Sci. 89:1235–1243. doi:10.3168/jds.S0022-0302(06)72192-0.

Baumgard, L. H., R. J. Collier, and D. E. Bauman. 2017. A 100-Year Review: Regulation of nutrient partitioning to support lactation. J. Dairy Sci. 100:10353–10366. doi:10.3168/jds.2017-13242.

Beecher, C., M. Daly, S. Childs, D. P. Berry, D.A. Magee, T. V. McCarthy, and L. Giblin. 2010. Polymorphisms in bovine immune genes and their associations with somatic cell count and milk production in dairy cattle. BMC Genet. 11:99. doi:10.1186/1471-2156-11-99.

Bionaz, M., and J. J. Loor. 2008a. Gene networks driving bovine milk fat synthesis during the lactation cycle. doi:10.1186/1471-2164-9-366.

Bionaz, M., and J. J. Loor. 2008b. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 Are the Most Abundant Isoforms in Bovine Mammary Tissue and Their Expression Is Affected by Stage of Lactation. J. Nutr. 138:1019–1024. doi:10.1093/jn/138.6.1019.

Bowers, R. R., J. W. Kim, T. C. Otto, and M. D. Lane. 2006. Stable stem cell commitment to the adipocyte lineage by inhibition of DNA methylation: Role of the BMP-4 gene. Proc. Natl. Acad. Sci. 103:13022–13027. doi:10.1073/pnas.0605789103.

Brito, L. F., F. G. Silva, H. R. Oliveira, N. Souza, G. Caetano, E. V. Costa, G. R. O. Menezes, A. L. P. Melo, M. T. Rodrigues, and R. A. Torres. 2017. Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. Can. J. Anim. Sci. CJAS-2017-0019. doi:10.1139/CJAS-2017-0019.

Buitenhuis, B., L. L. Janss, N. A. Poulsen, L. B. Larsen, M. K. Larsen, and P. Sørensen. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15:1112. doi:10.1186/1471-2164-15-1112.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen. 2016. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC Genet. 17:114. doi:10.1186/s12863-016-0421-2.

Cecchinato, A., C. Ribeca, S. Chessa, C. Cipolat-Gotet, F. Maretto, J. Casellas, and G. Bittante. 2014. Candidate gene association analysis for milk yield, composition, urea nitrogen and somatic cell scores in Brown Swiss cows. animal 8:1062–1070. doi:10.1017/S1751731114001098.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7. doi:10.1186/s13742-015-0047-8.

Cochran, S. D., J. B. Cole, D. J. Null, and P. J. Hansen. 2013. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genet. 14:49. doi:10.1186/1471-2156-14-49.

Evangelou, E., and J. P. A. Ioannidis. 2013. Meta-analysis methods for genome-wide association studies and beyond. Nat. Rev. Genet. 14:379–389. doi:10.1038/nrg3472.

Ezratty, E. J., M. A. Partridge, and G. G. Gundersen. 2005. Microtubule-induced focal adhesion disassembly is mediated by dynamin and focal adhesion kinase. Nat. Cell Biol. 7:581–590. doi:10.1038/ncb1262.

Fragomeni, B. de O., I. Misztal, D. L. Lourenco, I. Aguilar, R. Okimoto, and W. M. Muir. 2014. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. Front. Genet. 5:332. doi:10.3389/fgene.2014.00332.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55. doi:10.1186/1297-9686-41-55.

Grisart, B. 2002. Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition. Genome Res. 12:222–231. doi:10.1101/gr.224202.

Hambruch, N., S. Kumstel, J. D. Haeger, and C. Pfarrer. 2017. Bovine placentomal heparanase and syndecan expression is related to placental maturation. Placenta 57:42–51. doi:10.1016/j.placenta.2017.06.006.

Harte, M. T., J. D. Hildebrand, M. R. Burnham, A. H. Bouton, and J. T. Parsons. 1996. p130 Cas , a Substrate Associated with v-Src and v-Crk, Localizes to Focal Adhesions and Binds to Focal Adhesion Kinase. J. Biol. Chem. 271:13649–13655. doi:10.1074/jbc.271.23.13649.

Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226–231. doi:10.1007/BF01245622.

Huang, H., T. J. Song, X. Li, L. Hu, Q. He, M. Liu, M. D. Lane, and Q. Q. Tang. 2009. BMP signaling pathway is required for commitment of C3H10T1/2 pluripotent stem cells to the adipocyte lineage. Proc. Natl. Acad. Sci. 106:12670–12675. doi:10.1073/pnas.0906266106.

Interbull. 2017. Description of national genomic evaluation systems. Canada. Accessed May. 02, 2018. http://www.interbull.org/ib/nationalgenoforms.

Interbull. 2018a. Description of national genetic evaluation system and trend validation for production traits. Canada. Accessed May. 02, 2018. http://www.interbull.org/ib/geforms.

Interbull. 2018b. Description of national genetic evaluation system and trend validation for milk somatic cell and clinical mastitis. Canada. Accessed May. 02, 2018. http://www.interbull.org/ib/geforms

Jin, Y., and T. M. Penning. 2007. Aldo-Keto Reductases and Bioactivation/Detoxication. Annu. Rev. Pharmacol. Toxicol. 47:263–292. doi:10.1146/annurev.pharmtox.47.120505.105337.

Kanner, S. B., A. B. Reynolds, R. R. Vines, and J. T. Parsons. 1990. Monoclonal antibodies to individual tyrosine-phosphorylated protein substrates of oncogene-encoded tyrosine kinases.. Proc. Natl. Acad. Sci. 87:3328–3332. doi:10.1073/pnas.87.9.3328.

Kantanen, J. 2000. Genetic diversity and population structure of 20 north European cattle breeds. J. Hered. 91:446–457. doi:10.1093/jhered/91.6.446.

Kaupe, B., H. Brandt, E. M. Prinzenberg, and G. Erhardt. 2007. Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. J. Anim. Sci. 85:11–21. doi:10.2527/jas.2005-753.

Kinsella, R. J., A. Kahari, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database 2011:bar030-bar030. doi:10.1093/database/bar030.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories.. Genetics 124:979–93.

Lafanechère, L., C. Courtay-Cahen, T. Kawakami, M. Jacrot, M. Rüdiger, J. Wehland, D. Job, and R. L. Margolis. 1998a. Suppression of tubulin tyrosine ligase during tumor growth. J. Cell Sci. 111 (Pt 2:171–81. doi:10.1186/1471-2164-15-1105.

Lafanechère, L., C. Courtay-Cahen, T. Kawakami, M. Jacrot, M. Rüdiger, J. Wehland, D. Job, and R.L. Margolis. 1998b. Suppression of tubulin tyrosine ligase during tumor growth. J. Cell Sci. 111 (Pt 2:171–81.

Larmer, S., M. Sargolzaei, and F. Schenkel. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. J. Dairy Sci. 97:3128–3141. doi:10.3168/jds.2013-6826.

Lucey, J. A., D. Otter, and D. S. Horne. 2017. A 100-Year Review: Progress on the chemistry of milk and its components. J. Dairy Sci.. doi:10.3168/jds.2017-13250.

MacHugh, N. D., J. K. Mburu, M. J. Hamilton, and W. C. Davis. 1998. Characterisation of a monoclonal antibody recognising the CD3epsilon chain of the bovine T cell receptor complex.. Vet. Immunol. Immunopathol. 61:25–35.

McNamara, J. P. 1988. Regulation of Bovine Adipose Tissue Metabolism During Lactation. 4. Dose-Responsiveness to Epinephrine as Altered by Stage of Lactation. J. Dairy Sci. 71:643–649. doi:10.3168/jds.S0022-0302(88)79602-2.

Meredith, B. K., D. P. Berry, F. Kearney, E. K. Finlay, A. G. Fahey, D. G. Bradley, and D. J. Lynn. 2013. A genome-wide association study for somatic cell score using the Illumina high-density bovine beadchip identifies several novel QTL potentially related to mastitis susceptibility. Front. Genet. 4:229. doi:10.3389/fgene.2013.00229.

Miglior, F., W. Gong, Y. Wang, G. J. Kistemaker, A. Sewalem, and J. Jamrozik. 2009. Short communication: Genetic parameters of production traits in Chinese Holsteins using a random regression test-day model. J. Dairy Sci.. doi:10.3168/jds.2009-2212.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 AND RELATED PROGRAMS (BGF90). Page in 7th World Congress on Genetics Applied to Livestock Production, A, Montpellier, France.

Morota, G., T. M. Beissinger, and F. Peñagaricano. 2016. MeSH-Informed Enrichment Analysis and MeSH-Guided Semantic Similarity Among Functional Terms and Gene Products in Chicken. G3&#58; Genes|Genomes|Genetics 6:2447–2453. doi:10.1534/g3.116.031096.

Morota, G., F. Peñagaricano, J. L. Petersen, D. C. Ciobanu, K. Tsuyuzaki, and I. Nikaido. 2015. An application of MeSH enrichment analysis in livestock. Anim. Genet. 46:381–387. doi:10.1111/age.12307.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F. Schenkel, S. S. Moore, and P. Stothard. 2016a. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 17:75. doi:10.1186/s12863-016-0386-1.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F. Schenkel, S. S. Moore, and P. Stothard. 2016b. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 17:75. doi:10.1186/s12863-016-0386-1.

Nayeri, S., and P. Stothard. 2016. Tissues, Metabolic Pathways and Genes of Key Importance in Lactating Dairy Cattle. Springer Sci. Rev. 4:49–77. doi:10.1007/s40362-016-0040-3.

Nelson, S. J., M. Schopen, A. G. Savage, J. L. Schulman, and N. Arluk. 2004. The MeSH translation maintenance system: structure, interface design, and implementation.. Stud. Health Technol. Inform. 107:67–9.

Nyman, A. K., T. Ekman, U. Emanuelson, A. H. Gustafsson, K. Holtenius, K. P. Waller, and C. H. Sandgren. 2007. Risk factors associated with the incidence of veterinary-treated clinical mastitis in Swedish dairy herds with a high milk yield and a low prevalence of subclinical mastitis. Prev. Vet. Med. 78:142–160. doi:10.1016/j.prevetmed.2006.10.002.

Oliveira, C. S. F., H. Hogeveen, A. M. Botelho, P. V Maia, S. G. Coelho, and J. P. A. Haddad. 2015. Cow-specific risk factors for clinical mastitis in Brazilian dairy cattle. Prev. Vet. Med. 121:297–305. doi:10.1016/j.prevetmed.2015.08.001.

Oliveira, H. R., F. F. e Silva, M.V.G. B. Silva, O. H. G. B. D. Siqueira, M. A. Machado, J. C. C. Panetto, L. S. Glória, and L. F. Brito. 2017. Bayesian Models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. Livest. Sci. 201:78–84. doi:10.1016/j.livsci.2017.05.007.

Oliveira, H. R., L. F. Brito, F. F. Silva, D. A. L. Lourenco, J. Jamrozik, and F. S. Schenkel. 2018. Genomic prediction of lactation curves for milk, fat, protein and somatic cell score in Holstein cattle. J. Dairy Sci. 102:1-12.

Osorio, J. S., J. Lohakare, and M. Bionaz. 2016. Biosynthesis of milk fat, protein, and lactose: roles of transcriptional and posttranscriptional regulation. Physiol. Genomics. doi:10.1152/physiolgenomics.00016.2015.

Palmer, C. A., M. C. Neville, S. M. Anderson, and J. L. McManaman. 2006. Analysis of lactation defects in transgenic mice. J. Mammary Gland Biol. Neoplasia. doi:10.1007/s10911-006-9023-3.

Pei, Y. F., L. Zhang, J. Liu, and H. W. Deng. 2009. Multivariate Association Test Using Haplotype Trend Regression. Ann. Hum. Genet. 73:456–464. doi:10.1111/j.1469-1809.2009.00527.x.

Pryce, J. E., S. Bolormaa, A. J. Chamberlain, P. J. Bowman, K. Savin, M. E. Goddard, and B. J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93:3331–3345. doi:10.3168/jds.2009-2893.

R Development Core Team, R. 2011. R: A Language and Environment for Statistical Computing.

Raven, L. A., B. G. Cocks, and B. J. Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:62. doi:10.1186/1471-2164-15-62.

Rezamand, P., and M. A. McGuire. 2011. Short communication: Effects of trans fatty acids on markers of inflammation in bovine mammary epithelial cells. J. Dairy Sci. 94:316–320. doi:10.3168/jds.2010-3289.

Ron, M., G. Israeli, E. Seroussi, J. I. Weller, J. P. Gregg, M. Shani, and J. F. Medrano. 2007. Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. BMC Genomics 8:183. doi:10.1186/1471-2164-8-183.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2011. FImpute - An efficient imputation algorithm for dairy cattle populations. J. Dairy Sci. 94:421.

Sasaki, O., M. Aihara, A. Nishiura, and H. Takeda. 2017. Genetic correlations between the cumulative pseudo-survival rate, milk yield, and somatic cell score during lactation in Holstein cattle in Japan using a random regression model. J. Dairy Sci. 100:7282–7294. doi:10.3168/jds.2016-12311.

Schaeffer, L. R., J. Jamrozik, G. J. Kistemaker, and B. J. Van Doormaal. 2000. Experience with a Test-Day Model.

Sengar, G. S., R. Deb, U. Singh, T. V. Raja, R. Kant, B. Sajjanar, R. Alex, R. R. Alyethodi, A. Kumar, S. Kumar, R. Singh, S.J. Jakhesara, and C.G. Joshi. 2018. Differential expression of microRNAs associated with thermal stress in Frieswal (Bos taurus x Bos indicus) crossbred dairy cattle. Cell Stress Chaperones 23:155–170. doi:10.1007/s12192-017-0833-6.

Silva, R. M. O., N. B. Stafuzza, B. O. Fragomeni, G. M. F. Camargo, T. M. Ceacero, J. N. S. G. Cyrillo, F. Baldi, A. A. Boligon, M. E. Z. Mercadante, D. L. Lourenco, I. Misztal, and L. G.

Albuquerque. 2017. Genome-Wide Association Study for Carcass Traits in an Experimental Nelore Cattle Population. PLoS One 12:e0169860. doi:10.1371/journal.pone.0169860.

Soares, A. C. C., S. E. F. Guimarães, M. J. Kelly, M. R. S. Fortes, F. F. e Silva, L. L. Verardo, R. Mota, and S. Moore. 2017. Multiple-trait genomewide mapping and gene network analysis for scrotal circumference growth curves in Brahman cattle1. J. Anim. Sci. 95:3331–3345. doi:10.2527/jas.2017.1409.

Song, M., Y. He, H. Zhou, Y. Zhang, X. Li, and Y. Yu. 2016. Combined analysis of DNA methylome and transcriptome reveal novel candidate genes with susceptibility to bovine Staphylococcus aureus subclinical mastitis. Sci. Rep. 6:29390. doi:10.1038/srep29390.

Spelman, R. J., C. A. Ford, P. McElhinney, G.C. Gregory, and R.G. Snell. 2002. Characterization of the DGAT1 Gene in the New Zealand Dairy Population. J. Dairy Sci. 85:3514–3517. doi:10.3168/jds.S0022-0302(02)74440-8.

Strabel, T., J. Szyda, E. Ptak, and J. Jamrozik. 2005. Comparison of Random Regression Test-Day Models for Polish Black and White Cattle. J. Dairy Sci. 88:3688–3699. doi:10.3168/jds.S0022-0302(05)73055-1.

Strucken, E. M., R. H. Bortfeldt, D. J. de Koning, and G. A. Brockmann. 2012a. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375–382. doi:10.1111/j.1365-2052.2011.02278.x.

Strucken, E. M., R. H. Bortfeldt, J. Tetens, G. Thaller, and G. A. Brockmann. 2012b. Genetic effects and correlations between production and fertility traits and their dependency on the lactation-stage in Holstein Friesians. BMC Genet. 13:108. doi:10.1186/1471-2156-13-108.

Strucken, E. M., D. J. de Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442–449. doi:10.3168/jds.2009-2932.

Tsuyuzaki, K., G. Morota, M. Ishii, T. Nakazato, S. Miyazaki, and I. Nikaido. 2015. MeSH ORA framework: R/Bioconductor packages to support MeSH over-representation analysis. BMC Bioinformatics 16:45. doi:10.1186/s12859-015-0453-z.

VanRaden, P. M. 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24. doi:10.3168/jds.2008-1514.

Wang, H., L. Jiang, X. Liu, J. Yang, J. Wei, J. Xu, Q. Zhang, and J. F. Liu. 2013. A Post-GWAS Replication Study Confirming the PTK2 Gene Associated with Milk Production Traits in Chinese Holstein. PLoS One 8:e83625. doi:10.1371/journal.pone.0083625.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, R. L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W.M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front. Genet. 5:134. doi:10.3389/fgene.2014.00134.

Wells, B., S. T. G. Burgess, and A. J. Nisbet. 2013. Characterization of the ovine complement 4 binding protein-beta (C4BPB) chain as a serum biomarker for enhanced diagnosis of sheep scab. Mol. Cell. Probes 27:158–163. doi:10.1016/j.mcp.2013.03.003.

Wiggans, G. R., T. S. Sonstegard, P. M. VanRaden, L. K. Matukumalli, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, and C. P. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92:3431–3436. doi:10.3168/JDS.2008-1758.

Zavadilová, L., J. Jamrozik, and L.R. Schaeffer. 2005. Genetic parameters for test-day model with random regressions for production traits of Czech Holstein cattle. Czech J. Anim. Sci. 50:142–154.

8.8. Supplementary material

Table S1. Location of important genomic windows (Chromosome: Position Initial: Position Final) and proportion of genetic variance explained (Variance, %) for milk yield in each lactation stage, parity and breed (Ayrshire, AY; Holstein, HO; and Jersey, JE).

Parity	Breed	Stage	Chr:PosIn:PosFinal (Variance)
First	**AY**	**1**	1:64891636:65624310 (0.37); 5:49727902:51276764 (0.52); 7:14768778:16371434 (0.38); 9:103270667:104094973 (0.45); 13:76443044:77161984 (0.51); 18:25624516:26527686 (0.47); 20:39576637:40384206 (0.40); 28:8758466:9858113 (0.31)
		2	1:64766940:65466612 (0.39); 5:49727902:51276764 (0.44); 11:86569656:87455875 (0.41); 13:76443044:77161984 (0.31); 16:4185662:5253637 (0.36); 18:25624516:26527686 (0.40); 19:59276357:60165029 (0.32); 20:37553988:38761711 (0.35); 20:39509383:40301530 (0.57)
		3	1:62607089:63372417 (0.31); 1:64766940:65466612 (0.33); 5:49727902:51276764 (0.31); 8:99874540:100967241 (0.33); 11:86569656:87455875 (0.44); 12:60377153:61608282 (0.35); 16:4185662:5253637 (0.33); 16:6259815:8923491 (0.35); 18:25624516:26527686 (0.30); 19:59276357:60165029 (0.31); 20:37553988:38761711 (0.34); 20:39509383:40301530 (0.41)
	HO	**1**	5:40868010:42285835 (0.45); 6:86819633:88069548 (0.43); 14:1489496:2524432 (4.86); 14:2553525:3297177 (0.42); 29:37749427:40281016 (0.34)
		2	5:40868010:42285835 (0.44); 6:86819633:88069548 (0.33); 14:1489496:2524432 (8.74); 14:2553525:3297177 (0.67); 29:37749427:40281016 (0.42)
		3	5:40868010:42285835 (0.37); 14:1489496:2524432 (7.07); 14:2553525:3297177 (0.53); 20:30398960:32293167 (0.31); 29:37749427:40281016 (0.39)
	JE	**1**	1:28978223:29763188 (0.31); 2:18834716:20023792 (0.32); 2:105602018:106952479 (0.39); 5:34989903:37273786 (0.47); 5:66976106:68753154 (0.76); 8:73267165:74336311 (0.35); 10:31735444:32930297 (0.50); 10:63343168:64905140 (0.38); 11:20119555:21123539 (0.52); 11:21496394:22465305 (0.32); 11:45814583:46819460 (0.35); 13:40777908:41833607 (0.36); 17:20940958:22584146 (0.32); 18:4820101:6246055 (0.32); 18:16594926:17343527 (0.34); 23:13714810:14756959 (0.34); 26:32792279:33778915 (0.37); 28:39950373:41154672 (0.35); 29:46703510:47455052 (0.32)
		2	2:105602018:106952479 (0.39); 5:66976106:68753154 (0.69); 8:73267165:74336311 (0.37); 10:63343168:64905140 (0.37); 11:20119555:21123539 (0.44); 11:21461502:22436382 (0.34); 11:45814583:46819460 (0.35); 13:72431970:73369210 (0.35); 18:52264917:53894172 (0.33); 20:33731804:34777599 (0.47); 26:32792279:33778915 (0.52); 28:39950373:41154672 (0.3); 29:46703510:47455052 (0.34)
		3	2:64485917:65574548 (0.31); 5:66976106:68753154 (0.60); 8:73267165:74336311 (0.32); 10:20229329:21131661 (0.32); 10:101521757:102446402 (0.33); 11:20032403:20965840 (0.41); 11:21461502:22436382 (0.32); 11:45814583:46819460 (0.34); 13:72431970:73369210 (0.41); 18:52264917:53894172 (0.43); 20:33731804:34777599 (0.62); 26:32821171:33836268 (0.57); 29:46703510:47455052 (0.31)
Second	**AY**	**1**	1:64864266:65580210 (0.33); 1:119446658:120572284 (0.31); 5:49727902:51276764 (0.34); 5:113137320:114499520 (0.33); 8:61241061:62040920 (0.35); 9:103270667:104094973 (0.31); 11:86569656:87455875 (0.38); 13:76443044:77161984 (0.35); 18:25624516:26527686 (0.45); 19:59276357:60165029 (0.31); 20:39509383:40301530 (0.49)

		2	1:64766940:65466612 (0.36); 5:49727902:51276764 (0.32); 11:86569656:87455875 (0.49); 12:60377153:61608282 (0.39); 16:4185662:5253637 (0.39); 16:12592465:14071161 (0.34); 19:59276357:60165029 (0.40); 20:39509383:40301530 (0.47); 22:26355258:27427804 (0.33)
		3	5:112374372:113499150 (0.30); 8:99874540:100967241 (0.31); 11:77735570:78708174 (0.38); 11:86569656:87455875 (0.33); 12:60377153:61608282 (0.43); 16:11943700:13695211 (0.31); 16:23328862:24948459 (0.40); 22:26657157:27820332 (0.32); 25:36514994:37693781 (0.31)
	HO	**1**	5:41143539:42680103 (0.47); 6:86819633:88069548 (0.49); 13:54397045:55642499 (0.33); 14:1489496:2524432 (7.05); 14:2553525:3297177 (0.70)
		2	5:41143539:42680103 (0.45); 14:1489496:2524432 (8.80); 14:2553525:3297177 (0.82); 29:37749427:40281016 (0.37)
		3	14:1489496:2524432 (2.63)
	JE	**1**	1:29095768:30414871 (0.30); 2:105707203:107370183 (0.43); 3:89347495:89942617 (0.30); 5:34989903:37273786 (0.35); 5:66976106:68753154 (0.63); 8:73267165:74336311 (0.33); 8:86560410:87664192 (0.30); 10:31735444:32930297 (0.41); 10:63343168:64905140 (0.52); 10:65877655:67118053 (0.30); 11:20119555:21123539 (0.65); 11:21496394:22465305 (0.34); 13:40722564:41749885 (0.38); 18:1094150:2047277 (0.35); 18:4820101:6246055 (0.36); 18:16594926:17343527 (0.34); 20:29998448:31752546 (0.32); 20:55992767:56920873 (0.33); 22:52187729:53254546 (0.37); 23:13714810:14756959 (0.35); 26:32792279:33778915 (0.38); 27:13660191:14591550 (0.31); 28:39950373:41154672 (0.39); 29:41512425:42651294 (0.30)
		2	2:105707203:107370183 (0.32); 5:66976106:68753154 (0.60); 8:73267165:74336311 (0.37); 10:63343168:64905140 (0.36); 10:101521757:102446402 (0.35); 11:20032403:20965840 (0.49); 11:21461502:22436382 (0.34); 11:45814583:46819460 (0.34); 20:29998448:31752546 (0.31); 20:33731804:34777599 (0.41); 26:32821171:33836268 (0.57); 28:39950373:41154672 (0.30)
		3	3:118167062:119113936 (0.32); 5:66976106:68753154 (0.53); 7:37028304:38123606 (0.34); 8:73267165:74336311 (0.35); 10:101521757:102446402 (0.36); 11:20032403:20965840 (0.35); 11:45814583:46819460 (0.38); 13:72431970:73369210 (0.33); 15:61241418:62283471 (0.31); 15:79785494:81736304 (0.33); 16:61381572:62948497 (0.31); 18:52264917:53894172 (0.42); 20:33731804:34777599 (0.45); 26:12364119:13697600 (0.34); 26:32851706:33885893 (0.65); 28:29939938:31919826 (0.32)
Third	**AY**	**1**	1:64891636:65624310 (0.38); 1:119446658:120572284 (0.33); 2:56382133:57516075 (0.32); 5:49727902:51276764 (0.33); 5:113137320:114499520 (0.31); 7:14768778:16371434 (0.32); 11:86569656:87455875 (0.34); 13:76443044:77161984 (0.31); 18:25624516:26527686 (0.43); 19:59276357:60165029 (0.34); 20:39509383:40301530 (0.52)
		2	1:64766940:65466612 (0.37); 5:49727902:51276764 (0.30); 5:113137320:114499520 (0.31); 11:86569656:87455875 (0.51); 12:60377153:61608282 (0.33); 16:4185662:5253637 (0.40); 16:12592465:14071161 (0.32); 19:59276357:60165029 (0.37); 20:39509383:40301530 (0.42); 22:26657157:27820332 (0.32)
		3	5:112374372:113499150 (0.31); 8:99874540:100967241 (0.32); 11:77735570:78708174 (0.42); 11:86569656:87455875 (0.35); 12:60377153:61608282 (0.44); 16:23471049:25000153 (0.43); 25:36709894:37927752 (0.34); 29:3259869:4377323 (0.3)
	HO	**1**	5:41143539:42680103 (0.39); 6:86232932:87365840 (0.42); 14:1489496:2524432 (5.87); 14:2553525:3297177 (0.59)
		2	5:40868010:42285835 (0.38); 14:1489496:2524432 (8.38); 14:2553525:3297177 (0.86); 20:30398960:32293167 (0.31); 29:37749427:40281016 (0.37)
		3	14:1489496:2524432 (2.33)
	JE	**1**	2:105707203:107370183 (0.41); 3:89347495:89942617 (0.40); 5:34989903:37273786 (0.50); 5:66976106:68753154 (0.71); 8:73267165:74336311 (0.33); 10:31735444:32930297 (0.44); 10:63343168:64905140 (0.58); 11:20119555:21123539 (0.52); 11:46143526:47183635 (0.32); 18:1094150:2047277 (0.33); 18:4820101:6246055 (0.38); 18:16594926:17343527 (0.42); 20:55992767:56920873 (0.30); 22:52187729:53254546 (0.30); 23:13714810:14756959 (0.38); 26:32792279:33778915 (0.32); 27:13660191:14591550 (0.38); 28:39950373:41154672 (0.41)
		2	1:103434094:104652883 (0.35); 2:64598754:65746051 (0.34); 2:105707203:107370183 (0.40); 5:66976106:68753154 (0.58); 8:73267165:74336311 (0.33); 10:63343168:64905140 (0.45); 10:101521757:102446402 (0.32); 11:20032403:20965840 (0.49); 11:21461502:22436382 (0.31); 11:45814583:46819460 (0.32); 13:72431970:73369210 (0.33); 17:10330387:11419944 (0.34); 20:29998448:31752546 (0.33); 20:33773531:34981347 (0.38); 26:32792279:33778915 (0.53); 28:39950373:41154672 (0.32)

		3	2:64485917:65574548 (0.34); 3:118167062:119113936 (0.32); 5:66976106:68753154 (0.56); 7:37028304:38123606 (0.35); 10:63343168:64905140 (0.32); 10:101521757:102446402 (0.32); 11:20032403:20965840 (0.37); 11:45814583:46819460 (0.33); 13:72431970:73369210 (0.36); 17:10330387:11419944 (0.36); 18:52264917:53894172 (0.39); 20:33731804:34777599 (0.42); 26:32851706:33885893 (0.67)

Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

Table S2. Location of important genomic windows (Chromosome: Position Initial: Position Final) and proportion of genetic variance explained (Variance, %) for fat yield in each lactation stage, parity and breed (Ayrshire, AY; Holstein, HO; and Jersey, JE).

Parity	**Breed**	**Stage**	**Chr:PosIn:PosFinal (Variance)**
First	**AY**	**1**	1:64891636:65624310 (0.35); 4:83555683:84339298 (0.32); 6:31067604:32293744 (0.39); 7:14768778:16371434 (0.44); 13:48711228:50640029 (0.35); 13:54397045:55642499 (0.31); 15:53072568:54497454 (0.34); 20:13731671:15651653 (0.30); 20:57080050:57975104 (0.39); 25:1737669:2299373 (0.38); 28:8758466:9858113 (0.32)
		2	1:64891636:65624310 (0.35); 6:30990436:32128224 (0.34); 7:14768778:16371434 (0.46); 11:86655500:87498559 (0.31); 12:60781993:62028176 (0.35); 13:48711228:50640029 (0.36); 15:53072568:54497454 (0.32); 15:67486632:68557111 (0.30); 21:38049838:39184077 (0.30); 25:1737669:2299373 (0.44); 28:8758466:9858113 (0.31)
		3	1:64864266:65580210 (0.30); 7:14768778:16371434 (0.37); 8:99937935:101012371 (0.38); 10:101724167:102446402 (0.32); 11:86569656:87455875 (0.35); 12:60781993:62028176 (0.42); 21:38049838:39184077 (0.33); 25:1737669:2299373 (0.39)
	HO	**1**	5:41476949:42743482 (0.43); 14:1489496:2524432 (7.46); 14:2754909:3706475 (0.66); 14:7566056:8241174 (0.34)
		2	5:91471989:92844631 (0.36); 14:1489496:2524432 (10.15); 14:2754909:3706475 (1.01); 14:3765019:4468478 (0.40); 4:7646485:8385937 (0.49)
		3	5:91471989:92844631 (0.39); 5:94216856:95743746 (0.34); 5:99569438:101518663 (0.31); 14:1489496:2524432 (8.33); 14:2754909:3706475 (0.96); 14:3765019:4468478 (0.41); 14:7646485:8385937 (0.48)
	JE	**1**	3:5940341:7147596 (0.32); 5:34896625:36987357 (0.61); 5:66976106:68753154 (0.56); 7:2323494:2991449 (0.38); 7:87556048:88540604 (0.42); 10:31735444:32930297 (0.88); 10:49132397:50213424 (0.43); 11:46145227:47210051 (0.41); 14:1801116:2909929 (0.50); 15:79404776:81595546 (0.31); 19:48536450:49512814 (0.31); 20:22093001:23347580 (0.62); 24:39629608:41097326 (0.31)
		2	2:64300614:65301812 (0.33); 5:66976106:68753154 (0.61); 7:2323494:2991449 (0.40); 7:87556048:88540604 (0.33); 8:99527628:100671975 (0.34); 10:31735444:32930297 (0.54); 10:49097110:50185230 (0.37); 10:71694404:73585685 (0.34); 11:17484533:18480826 (0.33); 11:46145227:47210051 (0.44); 14:1801116:2909929 (0.36); 15:79404776:81595546 (0.35); 20:22093001:23347580 (0.39); 24:39629608:41097326 (0.34); 26:11528933:13229500 (0.32)
		3	1:5034441:6024691 (0.33); 2:64375198:65415110 (0.40); 5:66976106:68753154 (0.58); 7:2323494:2991449 (0.41); 8:99267011:100537840 (0.37); 10:31735444:32930297 (0.35); 10:49097110:50185230 (0.32); 10:57832918:59384121 (0.31); 11:17686944:18635979 (0.36); 11:46145227:47210051 (0.46); 15:79531511:81619357 (0.4); 18:52264917:53894172 (0.31); 24:39694262:41187936 (0.31); 26:11528933:13229500 (0.44); 26:32821171:33836268 (0.39)
Second	**AY**	**1**	1:64891636:65624310 (0.36); 6:31067604:32293744 (0.33); 7:14768778:16371434 (0.36); 13:48711228:50640029 (0.43); 13:54397045:55642499 (0.33); 15:53072568:54497454 (0.44); 20:57041322:57916773 (0.34); 25:1737669:2299373 (0.43)
		2	7:14768778:16371434 (0.30); 10:73417064:74273641 (0.30); 11:86569656:87455875 (0.35); 12:60781993:62028176 (0.45); 13:48711228:50640029 (0.33); 15:53072568:54497454 (0.33); 21:38049838:39184077 (0.33); 25:1737669:2299373 (0.52)
		3	8:99937935:101012371 (0.35); 11:77735570:78708174 (0.31); 11:80839258:82409810 (0.35); 12:60568356:61667023 (0.47); 16:23328862:24948459 (0.39); 21:38049838:39184077 (0.33); 25:1737669:2299373 (0.40)
	HO	**1**	14:1489496:2524432 (7.61); 14:2754909:3706475 (0.65)
		2	5:94216856:95743746 (0.33); 14:1489496:2524432 (8.69); 14:2553525:3297177 (0.65)
		3	14:1489496:2524432 (6.19); 14:2553525:3297177 (0.49)

	JE	1	1:3116101:4052161 (0.30); 3:5940341:7147596 (0.32); 5:34896625:36987357 (0.40); 5:66577741:68548817 (0.41); 7:2323494:2991449 (0.32); 7:87556048:88540604 (0.34); 10:31735444:32930297 (0.63); 10:49058682:49981642 (0.40); 14:1801116:2909929 (0.68); 15:79404776:81595546 (0.33); 19:45109206:45976816 (0.33); 19:48536450:49512814 (0.34); 20:22093001:23347580 (0.44); 24:33094977:34013385 (0.31); 24:39629608:41097326 (0.30); 26:10397362:11528933 (0.30)
		2	1:49393074:51378532 (0.32); 2:64375198:65415110 (0.34); 3:6126856:7371235 (0.30); 4:26186690:27339204 (0.34); 5:66976106:68753154 (0.53); 6:115707898:117148273 (0.33); 7:2323494:2991449 (0.37); 8:99527628:100671975 (0.34); 10:31735444:32930297 (0.37); 10:49058682:49981642 (0.38); 10:71694404:73585685 (0.31); 10:81135781:82444410 (0.32); 11:46145227:47210051 (0.38); 14:1801116:2909929 (0.47); 15:79404776:81595546 (0.41); 17:60248818:61485853 (0.31); 19:48405523:49473960 (0.32); 24:40114436:41540027 (0.36); 26:11528933:13229500 (0.39); 26:32644710:33556275 (0.33)
		3	1:5034441:6024691 (0.36); 1:49393074:51378532 (0.33); 2:64375198:65415110 (0.35); 3:1916325:3231882 (0.32); 5:66976106:68753154 (0.43); 6:115707898:117148273 (0.31); 7:2323494:2991449 (0.35); 10:49058682:49981642 (0.30); 11:46145227:47210051 (0.42); 14:1801116:2909929 (0.30); 15:29233683:30409179 (0.39); 15:79531511:81619357 (0.45); 24:40114436:41540027 (0.36); 26:11528933:13229500 (0.54); 26:13601655:14676194 (0.34); 26:32644710:33556275 (0.48)
Third	**AY**	**1**	1:64891636:65624310 (0.37); 5:71144630:72142911 (0.31); 7:14768778:16371434 (0.36); 10:73466092:74310583 (0.31); 13:48711228:50640029 (0.45); 13:54397045:55642499 (0.31); 15:53072568:54497454 (0.43); 20:57041322:57916773 (0.39); 25:1737669:2299373 (0.48)
		2	7:14768778:16371434 (0.32); 11:86655500:87498559 (0.33); 12:60781993:62028176 (0.40); 13:48711228:50640029 (0.34); 15:53072568:54497454 (0.35); 25:1737669:2299373 (0.58)
		3	8:99937935:101012371 (0.35); 11:77735570:78708174 (0.33); 11:80839258:82409810 (0.34); 12:60568356:61667023 (0.45); 16:23328862:24948459 (0.41); 25:1737669:2299373 (0.40); 25:36709894:37927752 (0.31); 26:36056808:37048074 (0.30)
	HO	**1**	14:1489496:2524432 (10.03); 14:2553525:3297177 (0.67)
		2	5:94216856:95743746 (0.33); 14:1489496:2524432 (10.44); 14:2553525:3297177 (0.67)
		3	14:1489496:2524432 (6.41); 14:2553525:3297177 (0.45)
	JE	**1**	1:3116101:4052161 (0.36); 2:105707203:107370183 (0.30); 3:6126856:7371235 (0.32); 5:34896625:36987357 (0.45); 5:66976106:68753154 (0.49); 7:2323494:2991449 (0.33); 7:87556048:88540604 (0.34); 10:31735444:32930297 (0.73); 10:49132397:50213424 (0.35); 11:46145227:47210051 (0.34); 14:1801116:2909929 (0.69); 15:79404776:81595546 (0.33); 16:34656334:35788829 (0.31); 17:60289891:61522805 (0.30); 19:45109206:45976816 (0.36); 19:48368103:49375655 (0.37); 20:22093001:23347580 (0.31); 23:13668169:14658449 (0.35); 26:10310793:11135194 (0.33)
		2	2:64375198:65415110 (0.32); 3:6126856:7371235 (0.34); 4:26186690:27339204 (0.32); 5:66976106:68753154 (0.51); 6:115707898:117148273 (0.35); 7:2323494:2991449 (0.35); 7:87556048:88540604 (0.32); 8:59464494:60665941 (0.30); 8:99267011:100537840 (0.31); 10:31735444:32930297 (0.44); 10:49058682:49981642 (0.36); 10:71694404:73585685 (0.31); 10:81246897:82479371 (0.32); 11:46145227:47210051 (0.39); 12:13903112:15124411 (0.31); 14:1801116:2909929 (0.52); 15:79404776:81595546 (0.41); 17:60248818:61485853 (0.35); 19:48405523:49473960 (0.37); 24:39629608:41097326 (0.35); 26:11528933:13229500 (0.38)
		3	1:49393074:51378532 (0.30); 2:64375198:65415110 (0.35); 3:1916325:3231882 (0.36); 4:25990048:26995149 (0.36); 5:66976106:68753154 (0.43); 6:115707898:117148273 (0.34); 7:2323494:2991449 (0.34); 8:59464494:60665941 (0.36); 11:46145227:47210051 (0.39); 12:13903112:15124411 (0.30); 14:1801116:2909929 (0.31); 15:79404776:81595546 (0.39); 16:40846801:42730174 (0.31); 17:60248818:61485853 (0.36); 18:11777709:12947633 (0.31); 24:40114436:41540027 (0.35); 26:11528933:13229500 (0.50); 26:13630418:14681845 (0.32); 26:32915658:33918405 (0.45)

Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

Table S3. Location of important genomic windows (Chromosome: Position Initial: Position Final) and proportion of genetic variance explained (Variance, %) for protein yield in each lactation stage, parity and breed (Ayrshire, AY; Holstein, HO; and Jersey, JE).

Parity	Breed	Stage	Chr:PosIn:PosFinal (Variance)

First	**AY**	**1**	5:89793252:91118692 (0.39); 6:62408025:63532785 (0.34); 7:14768778:16371434 (0.49); 8:99937935:101012371 (0.35); 9:31971498:32723535 (0.32); 13:76443044:77161984 (0.45); 15:53072568:54497454 (0.38); 18:25709879:26630875 (0.37); 18:55892476:57018636 (0.31)
		2	6:62408025:63532785 (0.36); 7:14768778:16371434 (0.45); 8:99937935:101012371 (0.43); 9:31971498:32723535 (0.36); 11:86569656:87455875 (0.44); 13:76443044:77161984 (0.35); 15:53072568:54497454 (0.34); 18:25709879:26630875 (0.38); 28:8837691:10044965 (0.30)
		3	1:62607089:63372417 (0.31); 7:14768778:16371434 (0.36); 8:99874540:100967241 (0.51); 9:31971498:32723535 (0.32); 11:86569656:87455875 (0.42); 12:60568356:61667023 (0.32); 13:76443044:77161984 (0.33); 16:12592465:14071161 (0.30); 22:22558590:23914818 (0.31); 22:26355258:27427804 (0.32)
	HO	**1**	5:40868010:42285835 (0.37); 5:70804748:71884473 (0.32); 14:1489496:2524432 (1.46)
		2	5:41476949:42743482 (0.39); 6:86819633:88069548 (0.33); 13:43319313:44750541 (0.34); 14:1489496:2524432 (1.94)
		3	5:41476949:42743482 (0.36); 13:43319313:44750541 (0.33); 14:1489496:2524432 (1.52); 19:57235274:58091668 (0.31)
	JE	**1**	2:64598754:65746051 (0.30); 2:105707203:107370183 (0.46); 3:118167062:119113936 (0.44); 5:34896625:36987357 (0.48); 5:66940077:68666925 (0.53); 10:31735444:32930297 (0.51); 11:33642979:34971474 (0.49); 11:45814583:46819460 (0.53); 18:11777709:12947633 (0.35); 18:16594926:17343527 (0.48); 19:48536450:49512814 (0.33); 20:22093001:23347580 (0.37); 23:13267795:14495329 (0.33); 26:11528933:13229500 (0.31); 26:30006329:30858995 (0.35); 26:32792279:33778915 (0.33); 28:39950373:41154672 (0.43); 29:41133252:42342570 (0.34)
		2	2:64485917:65574548 (0.36); 2:105707203:107370183 (0.40); 3:37651069:39654642 (0.34); 3:118167062:119113936 (0.42); 5:34896625:36987357 (0.30); 5:66976106:68753154 (0.57); 5:71045870:72264603 (0.32); 8:73267165:74336311 (0.31); 10:31735444:32930297 (0.36); 10:58028735:59557358 (0.31); 11:17686944:18635979 (0.33); 11:33642979:34971474 (0.47); 11:45814583:46819460 (0.58); 15:79531511:81619357 (0.31); 18:16402220:17234805 (0.43); 26:11528933:13229500 (0.38); 26:30006329:30858995 (0.33); 26:32792279:33778915 (0.51); 28:39950373:41154672 (0.40); 29:46703510:47455052 (0.32)
		3	1:55982457:57660748 (0.31); 2:64485917:65574548 (0.36); 3:37651069:39654642 (0.41); 3:118293854:119150674 (0.46); 5:66976106:68753154 (0.49); 5:71045870:72264603 (0.32); 8:73267165:74336311 (0.33); 10:20229329:21131661 (0.34); 10:58028735:59557358 (0.33); 11:17686944:18635979 (0.36); 11:33642979:34971474 (0.36); 11:45814583:46819460 (0.56); 13:72350125:73329460 (0.34); 15:79923112:81818289 (0.38); 18:16402220:17234805 (0.33); 18:52264917:53894172 (0.33); 20:33731804:34777599 (0.31); 26:11528933:13229500 (0.48); 26:30006329:30858995 (0.32); 26:32821171:33836268 (0.58); 29:18120550:19928179 (0.35); 29:46703510:47455052 (0.31)
Second	**AY**	**1**	7:14768778:16371434 (0.31); 7:61404990:62508803 (0.34); 8:99874540:100967241 (0.34); 9:31971498:32723535 (0.35); 11:86569656:87455875 (0.32); 13:48711228:50640029 (0.30); 13:76443044:77161984 (0.30); 15:53072568:54497454 (0.41); 18:25709879:26630875 (0.30); 23:34379695:35744708 (0.34)
		2	5:112374372:113499150 (0.33); 8:99937935:101012371 (0.41); 9:31971498:32723535 (0.32); 11:86569656:87455875 (0.47); 13:48711228:50640029 (0.30); 16:12592465:14071161 (0.38); 22:22558590:23914818 (0.31); 22:26657157:27820332 (0.37)
		3	3:99853456:100683723 (0.33); 5:112374372:113499150 (0.31); 8:99874540:100967241 (0.43); 11:80839258:82409810 (0.32); 12:60568356:61667023 (0.42); 16:11943700:13695211 (0.34); 16:23328862:24948459 (0.41); 22:26657157:27820332 (0.32); 23:1674622:3167662 (0.31); 25:36709894:37927752 (0.32)
	HO	**1**	5:41476949:42743482 (0.36); 5:70804748:71884473 (0.34); 14:1489496:2524432 (1.80)
		2	5:41476949:42743482 (0.37); 14:1489496:2524432 (1.51)
		3	13:43319313:44750541 (0.31)
	JE	**1**	2:64598754:65746051 (0.38); 2:105707203:107370183 (0.48); 3:118167062:119113936 (0.39); 5:34896625:36987357 (0.34); 5:66940077:68666925 (0.52); 9:65204298:66292441 (0.33); 10:31622218:32771627 (0.37); 10:58257801:59927870 (0.32); 11:33642979:34971474 (0.54); 11:45814583:46819460 (0.45); 15:79531511:81619357 (0.32); 18:11774889:12784439 (0.37); 18:16594926:17343527 (0.51); 23:13122318:14424516 (0.33); 26:11528933:13229500 (0.4); 26:30006329:30858995 (0.34); 26:32644710:33556275 (0.41); 27:18740236:19698845 (0.41); 28:39950373:41154672 (0.43); 29:41133252:42342570 (0.37)

		2	2:64598754:65746051 (0.36); 2:105602018:106952479 (0.31); 3:118293854:119150674 (0.41); 5:66940077:68666925 (0.52); 8:73267165:74336311 (0.31); 11:33642979:34971474 (0.51); 11:45814583:46819460 (0.57); 15:79531511:81619357 (0.38); 18:11777709:12947633 (0.31); 18:16402220:17234805 (0.41); 26:11528933:13229500 (0.52); 26:32644710:33556275 (0.59); 28:39950373:41154672 (0.38)
		3	1:5034441:6024691 (0.34); 2:64485917:65574548 (0.31); 3:37561427:39599273 (0.32); 3:118293854:119150674 (0.49); 5:66976106:68753154 (0.40); 5:71045870:72264603 (0.33); 8:73267165:74336311 (0.32); 11:33642979:34971474 (0.34); 11:45814583:46819460 (0.53); 13:71766574:72735116 (0.33); 15:79971651:82061751 (0.48); 18:52264917:53894172 (0.31); 26:11528933:13229500 (0.62); 26:13300472:14350397 (0.31); 26:32644710:33556275 (0.68); 29:18120550:19928179 (0.36)
Third	**AY**	**1**	7:14768778:16371434 (0.31); 7:61404990:62508803 (0.33); 8:99937935:101012371 (0.31); 9:31971498:32723535 (0.38); 13:48711228:50640029 (0.34); 15:53072568:54497454 (0.42); 18:11245392:12316620 (0.33); 23:34379695:35744708 (0.34)
		2	5:112374372:113499150 (0.34); 8:99937935:101012371 (0.41); 11:86569656:87455875 (0.44); 15:53072568:54497454 (0.31); 16:12592465:14071161 (0.36); 22:22558590:23914818 (0.3); 22:26657157:27820332 (0.33)
		3	5:112374372:113499150 (0.32); 8:99937935:101012371 (0.43); 11:86569656:87455875 (0.30); 12:60377153:61608282 (0.40); 16:11943700:13695211 (0.33); 16:23328862:24948459 (0.42); 25:36709894:37927752 (0.35)
	HO	**1**	14:1489496:2524432 (0.85)
		2	13:43319313:44750541 (0.32); 14:1489496:2524432 (1.01)
		3	13:43319313:44750541 (0.29)
	JE	**1**	2:64485917:65574548 (0.37); 2:105707203:107370183 (0.46); 3:118167062:119113936 (0.42); 5:34896625:36987357 (0.44); 5:66940077:68666925 (0.64); 10:31622218:32771627 (0.37); 10:58257801:59927870 (0.34); 11:33642979:34971474 (0.45); 11:46145227:47210051 (0.53); 15:79531511:81619357 (0.35); 18:11774889:12784439 (0.39); 18:16736436:17540948 (0.62); 23:13668169:14658449 (0.43); 26:11528933:13229500 (0.38); 26:30006329:30858995 (0.36); 26:32644710:33556275 (0.42); 27:18740236:19698845 (0.37); 28:39950373:41154672 (0.41)
		2	2:64598754:65746051 (0.40); 2:105707203:107370183 (0.40); 2:114742983:116142693 (0.33); 3:118293854:119150674 (0.43); 5:66976106:68753154 (0.53); 10:31622218:32771627 (0.30); 11:17686944:18635979 (0.30); 11:33642979:34971474 (0.43); 11:45814583:46819460 (0.54); 15:79531511:81619357 (0.35); 18:11777709:12947633 (0.38); 18:16402220:17234805 (0.50); 26:11528933:13229500 (0.50); 26:32644710:33556275 (0.56); 28:39950373:41154672 (0.38); 29:32070762:33157623 (0.30)
		3	2:64485917:65574548 (0.35); 2:114742983:116142693 (0.33); 3:118293854:119150674 (0.47); 5:66976106:68753154 (0.43); 5:71045870:72264603 (0.30); 11:33642979:34971474 (0.32); 11:46145227:47210051 (0.47); 13:71766574:72735116 (0.32); 15:79923112:81818289 (0.42); 18:11777709:12947633 (0.34); 18:16402220:17234805 (0.31); 26:11528933:13229500 (0.56); 26:13630418:14681845 (0.31); 26:32644710:33556275 (0.68); 29:18120550:19928179 (0.36)

Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

Table S4. Location of important genomic windows (Chromosome: Position Initial: Position Final) and proportion of genetic variance explained (Variance, %) for somatic cell score in each lactation stage, parity and breed.

Parity	Breed	Stage	Chr:PosIn:PosFinal (Variance)
First	**AY**	**1**	1:31840031:32700496 (0.42); 2:85427167:86868918 (0.47); 2:128889674:129882579 (0.41); 3:4522681:5541010 (0.32); 3:60072921:60892186 (0.32); 4:55313701:56479030 (0.34); 5:118605844:119261609 (0.39); 9:64207421:65175762 (0.41); 20:50935019:51789073 (0.34); 22:1644612:2857608 (0.35); 27:8541405:9762847 (0.40)
		2	1:31659179:32598751 (0.50); 2:85485238:86959111 (0.47); 2:128889674:129882579 (0.30); 4:55313701:56479030 (0.30); 5:118605844:119261609 (0.37); 9:3524770:5021874 (0.32); 9:21652331:22652013 (0.31); 14:74814889:76217573 (0.32); 20:50935019:51789073 (0.32); 22:1644612:2857608 (0.32); 27:8541405:9762847 (0.55)

		3	1:31659179:32598751 (0.48); 2:85485238:86959111 (0.40); 5:118605844:119261609 (0.42); 8:30088133:31123907 (0.34); 9:3645428:5045079 (0.37); 9:21652331:22652013 (0.33); 10:85480303:86242499 (0.33); 14:74814889:76217573 (0.35); 27:8541405:9762847 (0.62)
	HO	**1**	1:1288510:2291153 (0.37); 10:69721690:70663937 (0.32); 20:36311419:37237425 (0.45)
		2	14:16245416:17483216 (0.31); 20:36311419:37237425 (0.44)
		3	1:1288510:2291153 (0.32); 4:113506092:114640077 (0.31); 10:69721690:70663937 (0.31); 20:36311419:37237425 (0.40)
	JE	**1**	1:103116743:104280525 (0.49); 2:38748216:39775205 (0.38); 2:40686604:42008005 (0.35); 3:42954965:44028581 (0.34); 3:69785571:70885343 (0.67); 3:117985787:118934229 (0.36); 5:33078266:34101843 (0.42); 7:13468083:14678931 (0.37); 8:90197837:91125290 (0.35); 10:14554414:15528411 (0.41); 10:31622218:32771627 (0.31); 10:98588493:99662832 (0.42); 11:91578841:92718161 (0.36); 13:25873125:27430120 (0.37); 13:75018157:75928133 (0.48); 13:80980304:81743652 (0.35); 14:2985755:4043743 (0.58); 14:39735994:40812291 (0.30); 14:44254668:45945108 (0.31); 14:82053449:83131368 (0.36); 14:83419417:85006003 (0.49); 15:79404776:81595546 (0.34); 18:6858136:7840467 (0.38); 19:58687075:59902586 (0.38); 21:20728681:21532974 (0.60); 23:50009589:50971212 (0.32); 27:13660191:14591550 (0.41); 28:33729338:34530363 (0.64); 29:46703510:47455052 (0.30); 29:50202589:51502869 (0.55)
		2	2:38748216:39775205 (0.33); 2:40881032:42035814 (0.45); 3:42954965:44028581 (0.45); 3:69785571:70885343 (0.55); 3:117985787:118934229 (0.60); 4:56761233:57951154 (0.31); 4:119407450:120227878 (0.36); 5:33078266:34101843 (0.64); 6:97174252:98054956 (0.33); 7:13468083:14678931 (0.42); 8:90249304:91186071 (0.48); 9:75597064:76928003 (0.31); 10:14554414:15528411 (0.32); 10:31622218:32771627 (0.31); 10:98588493:99662832 (0.47); 11:92086008:93241685 (0.46); 13:25985907:27623653 (0.46); 13:74109992:75122713 (0.50); 13:75326854:76103249 (0.41); 13:80980304:81743652 (0.44); 14:2985755:4043743 (0.55); 14:82053449:83131368 (0.35); 14:83419417:85006003 (0.48); 15:83970308:85088227 (0.35); 18:16839215:17988032 (0.39); 21:21065240:21859000 (0.41); 27:14173400:15101058 (0.36); 29:50046572:51481890 (0.51)
		3	1:124043522:125641702 (0.32); 2:38939453:40155558 (0.31); 2:40881032:42035814 (0.47); 3:42954965:44028581 (0.41); 3:69785571:70885343 (0.41); 3:117985787:118934229 (0.48); 4:56761233:57951154 (0.32); 4:119407450:120227878 (0.34); 5:33013488:33981125 (0.70); 6:97174252:98054956 (0.32); 7:13468083:14678931 (0.48); 8:90249304:91186071 (0.50); 9:75597064:76928003 (0.33); 10:31503971:32677973 (0.36); 10:98588493:99662832 (0.44); 11:92179277:93554112 (0.42); 13:25985907:27623653 (0.44); 13:40722564:41749885 (0.39); 13:74109992:75122713 (0.50); 13:75326854:76103249 (0.33); 13:80980304:81743652 (0.45); 14:2985755:4043743 (0.54); 14:43899216:45003977 (0.30); 14:82053449:83131368 (0.35); 14:83419417:85006003 (0.47); 15:83970308:85088227 (0.34); 18:16839215:17988032 (0.42); 21:21065240:21859000 (0.40); 29:46676339:47422720 (0.31); 29:50046572:51481890 (0.56)
Second	**AY**	**1**	1:31659179:32598751 (0.37); 2:85427167:86868918 (0.57); 2:128767807:129845311 (0.30); 3:4522681:5541010 (0.41); 3:60072921:60892186 (0.33); 4:55313701:56479030 (0.35); 5:6357367:7022191 (0.35); 5:118605844:119261609 (0.33); 10:85480303:86242499 (0.33); 19:61205726:62144933 (0.31); 22:1644612:2857608 (0.33); 27:8541405:9762847 (0.37)
		2	1:31659179:32598751 (0.42); 2:85524425:87004473 (0.35); 5:118605844:119261609 (0.39); 8:30088133:31123907 (0.38); 9:3645428:5045079 (0.42); 10:85480303:86242499 (0.46); 13:10048229:10926940 (0.31); 13:48711228:50640029 (0.33); 16:29890723:30859872 (0.30); 19:61362411:62177834 (0.31); 20:48165391:49339693 (0.34); 20:70433323:71433871 (0.32); 27:8541405:9762847 (0.57)
		3	4:115747463:116666797 (0.33); 5:118605844:119261609 (0.34); 9:3971122:5188669 (0.43); 9:8668733:10134275 (0.39); 10:85480303:86242499 (0.38); 13:10048229:10926940 (0.30); 20:48165391:49339693 (0.55); 27:8453977:9599780 (0.47); 27:26713017:27987991 (0.38); 28:45240243:46102047 (0.32)
	HO	**1**	1:1336351:2313042 (0.31); 7:17528449:18379589 (0.35); 14:1489496:2524432 (0.52); 14:16387114:17544926 (0.41); 20:36311419:37237425 (0.37)
		2	6:86819633:88069548 (0.48); 7:17528449:18379589 (0.35); 10:69474936:70296220 (0.37); 14:1489496:2524432 (0.51); 14:16387114:17544926 (0.32); 20:36336225:37264477 (0.31)
		3	6:86819633:88069548 (0.50); 6:92690455:93551941 (0.32); 7:17528449:18379589 (0.32); 10:69474936:70296220 (0.40); 14:1489496:2524432 (0.51); 18:8037156:8765998 (0.31)
	JE	**1**	2:38748216:39775205 (0.41); 2:40881032:42035814 (0.34); 3:42954965:44028581 (0.61); 3:66702642:67594920 (0.30); 3:69716313:70641452 (0.72); 4:103852324:104841985 (0.40); 5:33078266:34101843 (0.42); 6:97174252:98054956 (0.36); 8:90197837:91125290 (0.30); 9:75597064:76928003 (0.33);

			10:12992411:13684162 (0.34); 10:14554414:15528411 (0.36); 10:31503971:32677973 (0.38); 11:92059274:93193697 (0.47); 12:15430493:16437175 (0.30); 13:1278678:2217706 (0.47); 13:74109992:75122713 (0.44); 14:2985755:4043743 (0.51); 14:44254668:45945108 (0.32); 14:83419417:85006003 (0.33); 15:15162470:16273879 (0.30); 15:29113608:30122484 (0.33); 15:83970308:85088227 (0.64); 18:16869103:18072990 (0.52); 20:37468100:38920878 (0.41); 23:6999212:7942859 (0.35); 27:13487036:14526116 (0.39); 28:33729338:34530363 (0.33); 29:50046572:51481890 (0.46)
		2	2:38939453:40155558 (0.40); 2:40881032:42035814 (0.35); 2:114133415:114742983 (0.46); 3:43157454:44364191 (0.56); 3:69716313:70641452 (0.53); 3:118167062:119113936 (0.33); 4:56761233:57951154 (0.37); 5:33078266:34101843 (0.48); 6:97174252:98054956 (0.33); 7:13468083:14678931 (0.34); 7:63935103:64870335 (0.31); 8:90249304:91186071 (0.39); 9:75597064:76928003 (0.35); 10:31364551:32389035 (0.56); 10:98588493:99662832 (0.41); 11:92059274:93193697 (0.54); 13:1278678:2217706 (0.41); 13:40722564:41749885 (0.41); 13:74109992:75122713 (0.42); 14:2985755:4043743 (0.41); 14:43899216:45003977 (0.35); 14:83419417:85006003 (0.33); 15:83970308:85088227 (0.72); 16:34606531:35765411 (0.32); 18:16806560:17659879 (0.55); 20:37468100:38920878 (0.32); 28:23349468:24678582 (0.35); 29:50046572:51481890 (0.58)
		3	1:124043522:125641702 (0.31); 1:128092950:129007020 (0.37); 2:38939453:40155558 (0.32); 2:114133415:114742983 (0.52); 3:43157454:44364191 (0.47); 3:69716313:70641452 (0.34); 4:56761233:57951154 (0.35); 4:119407450:120227878 (0.34); 5:33013488:33981125 (0.62); 6:97174252:98054956 (0.35); 7:13468083:14678931 (0.37); 8:31721931:32749041 (0.31); 8:90197837:91125290 (0.32); 9:75597064:76928003 (0.35); 9:103153156:103947174 (0.36); 10:31364551:32389035 (0.54); 10:98588493:99662832 (0.33); 11:92059274:93193697 (0.48); 13:40666278:41702235 (0.56); 14:2985755:4043743 (0.38); 15:83970308:85088227 (0.55); 16:34606531:35765411 (0.31); 18:16869103:18072990 (0.46); 28:23349468:24678582 (0.30); 29:46703510:47455052 (0.37); 29:50046572:51481890 (0.53)
Third	**AY**	**1**	1:31659179:32598751 (0.31); 2:85427167:86868918 (0.54); 3:4522681:5541010 (0.42); 4:55313701:56479030 (0.40); 5:6266261:6942103 (0.37); 8:30088133:31123907 (0.31); 10:85480303:86242499 (0.32); 19:61205726:62144933 (0.31); 20:4145679:4907906 (0.33); 22:1644612:2857608 (0.31)
		2	1:31659179:32598751 (0.31); 2:85585849:87133202 (0.39); 5:118605844:119261609 (0.4); 8:30088133:31123907 (0.37); 9:3645428:5045079 (0.39); 10:85480303:86242499 (0.44); 14:24407125:25857110 (0.34); 16:29890723:30859872 (0.31); 19:61362411:62177834 (0.37); 20:47919708:49053595 (0.31); 27:8541405:9762847 (0.42); 28:45262079:46146720 (0.34)
		3	4:115747463:116666797 (0.37); 5:118605844:119261609 (0.35); 9:3971122:5188669 (0.47); 9:8668733:10134275 (0.45); 10:85480303:86242499 (0.32); 20:48165391:49339693 (0.50); 27:8453977:9599780 (0.41); 27:26713017:27987991 (0.38)
	HO	**1**	5:106558376:107426006 (0.34); 27:31695928:32513708 (0.51)
		2	5:106558376:107426006 (0.33); 27:31647685:32488843 (0.41)
		3	5:106558376:107426006 (0.31); 27:31647685:32488843 (0.41)
	JE	**1**	2:18834716:20023792 (0.31); 3:42954965:44028581 (0.63); 3:69785571:70885343 (0.59); 3:118167062:119113936 (0.31); 4:103852324:104841985 (0.41); 4:119407450:120227878 (0.39); 5:33078266:34101843 (0.56); 7:13468083:14678931 (0.38); 8:90249304:91186071 (0.36); 9:75597064:76928003 (0.44); 10:14554414:15528411 (0.43); 11:38760385:40290718 (0.34); 11:92059274:93193697 (0.44); 13:25985907:27623653 (0.37); 14:2985755:4043743 (0.60); 14:83419417:85006003 (0.43); 15:29162110:30223756 (0.42); 15:83970308:85088227 (0.41); 16:31605270:32653232 (0.31); 18:16968932:18093362 (0.42); 19:58687075:59902586 (0.38); 23:6999212:7942859 (0.37); 27:13487036:14526116 (0.38); 28:33729338:34530363 (0.39); 29:50046572:51481890 (0.53)
		2	2:114133415:114742983 (0.33); 3:42954965:44028581 (0.59); 3:69716313:70641452 (0.47); 3:117985787:118934229 (0.37); 4:103852324:104841985 (0.33); 4:119407450:120227878 (0.48); 5:33078266:34101843 (0.71); 7:13468083:14678931 (0.43); 7:63935103:64870335 (0.35); 8:90249304:91186071 (0.38); 9:75597064:76928003 (0.42); 10:14554414:15528411 (0.38); 10:31622218:32771627 (0.37); 10:98588493:99662832 (0.38); 11:66744002:67678534 (0.33); 11:92059274:93193697 (0.41); 13:25985907:27623653 (0.39); 14:2985755:4043743 (0.61); 14:43899216:45003977 (0.31); 14:83419417:85006003 (0.40); 15:29162110:30223756 (0.50); 15:83970308:85088227 (0.51); 18:16968932:18093362 (0.45); 27:13487036:14526116 (0.32); 29:50046572:51481890 (0.56)
		3	1:124043522:125641702 (0.36); 2:114133415:114742983 (0.34); 3:43157454:44364191 (0.58); 3:69716313:70641452 (0.33); 3:117985787:118934229 (0.34); 4:56761233:57951154 (0.36);

	4:119407450:120227878 (0.52); 5:33013488:33981125 (0.85); 7:13468083:14678931 (0.42); 8:90249304:91186071 (0.35); 9:75597064:76928003 (0.35); 9:103127530:103847209 (0.43); 10:31503971:32677973 (0.39); 10:98588493:99662832 (0.50); 11:44061322:45096739 (0.30); 11:92059274:93193697 (0.35); 13:25985907:27623653 (0.35); 13:40722564:41749885 (0.48); 14:2985755:4043743 (0.53); 14:43899216:45003977 (0.36); 14:63353353:64151328 (0.31); 14:83419417:85006003 (0.36); 15:83970308:85088227 (0.32); 18:16968932:18093362 (0.33); 19:46478368:47725481 (0.33); 25:11620439:12445653 (0.39); 29:46676339:47422720 (0.33); 29:50046572:51481890 (0.53)

Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

Table S5. Total number of genes identified in the important non-overlapping windows, before (and after) the prioritization analyses, for each trait in the different parities, breeds and lactation stages.

Parity	Breed	Stage	MY	FY	PY	SCS
First	**AY**	**1**	69 (1)	123 (3)	111 (3)	45 (1)
		2	53 (2)	99 (3)	78 (2)	43 (1)
		3	58 (2)	65 (3)	53 (1)	31 (0)
	HO	**1**	78 (14)	55 (8)	49 (11)	16 (1)
		2	78 (14)	59 (10)	72 (15)	9 (0)
		3	71 (11)	95 (9)	65 (12)	37 (2)
	JE	**1**	16 (1)	8 (1)	7 (0)	205 (8)
		2	13 (1)	18 (1)	20 (0)	184 (7)
		3	31 (1)	26 (2)	48 (0)	187 (7)
Second	**AY**	**1**	70 (0)	109 (3)	106 (3)	52 (1)
		2	42 (2)	84 (5)	59 (0)	60 (0)
		3	79 (1)	35 (0)	79 (1)	27 (1)
	HO	**1**	91 (15)	46 (10)	50 (11)	73 (14)
		2	63 (11)	59 (11)	43 (11)	85 (17)
		3	37 (11)	47 (10)	13 (1)	87 (17)
	JE	**1**	130 (7)	12 (8)	19 (0)	184 (12)
		2	12 (1)	26 (2)	82 (2)	170 (6)
		3	8 (1)	160 (2)	8 (0)	139 (4)
Third	**AY**	**1**	91 (0)	115 (5)	91 (4)	44 (0)
		2	63 (2)	79 (4)	52 (0)	42 (0)
		3	67 (1)	59 (1)	66 (0)	20 (1)
	HO	**1**	59 (14)	47 (11)	37 (11)	8 (1)
		2	71 (11)	59 (11)	50 (12)	8 (1)
		3	37 (11)	47 (10)	13 (1)	8 (1)
	JE	**1**	12 (0)	151 (1)	17 (0)	177 (11)
		2	13 (1)	28 (1)	8 (0)	182 (7)
		3	10 (1)	11 (6)	96 (4)	162 (5)

Breeds: Ayrshire (AY), Holstein (HO) and Jersey (JE). Traits: milk (MY), fat (FY) and protein (PY) yield, and somatic cells score (SCS). Lactation stages were defined as: 1) from 5 to 95, 2) from 95 to 215, and 3) from 215 to 305 days in milk.

CHAPTER 9:

Single-step genome wide association analysis for longitudinal traits of Canadian Ayrshire, Holstein and Jersey dairy cattle

Hinayah Rojas de Oliveira[1,2], Daniela Andressa Lino Lourenco[3], Yutaka Matsuda[3], Ignacy Misztal[3], Shogo Tsuruta[3], Janusz Jamrozik[1,4], Luiz Fernando Brito[1,5], Fabyano Fonseca e Silva[2], John Cant[1], Flávio Schramm Schenkel[1]

[1] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, Canada

[2] Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[3] Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, United States of America

[4] Canadian Dairy Network, Guelph, Ontario, Canada

[5] Department of Animal Sciences, Purdue University, West Lafayette, Indiana, 47907, United States of America

9.1. Abstract

Estimating single nucleotide polymorphism (**SNP**) effects over time is essential to identify and validate candidate genes (or Quantitative Trait Loci) associated with time-dependent expression of economically important traits and to better understand the underlying mechanisms of lactation biology. Therefore, this study aimed to estimate time-dependent effects of SNP and identifying candidate genes associated with milk (**MY**), fat (**FY**) and protein (**PY**) yields, and somatic cell score (**SCS**) in the first three lactations of Canadian Ayrshire, Holstein and Jersey breeds, as well as suggest the potential pattern of gene expression over time. Breeding values for random regression coefficients of each animal were estimated using the single-step Genomic Best Linear Unbiased Prediction, based on two random regression models: one considering MY, FY and PY in the first three lactations; and another considering SCS in the first three lactations. Thereafter, the genomic breeding values for the additive genomic random regression coefficients were used to estimate the SNP solutions for the random regression coefficients, which were used to estimate the SNP effects over days 5 to 305 in lactation. The top 1% SNPs that showed high magnitude of SNP effect in at least one day in lactation were selected as relevant SNPs for further analyses of candidate genes, and clustered according to the trajectory of their SNP effects over time. The majority of SNPs selected for MY, FY and PY increased the magnitude of their effects over time, for all breeds. On the other hand, for SCS, the majority of selected SNP decreased the magnitude of their effects over time, especially for the Holstein and Jersey breeds. In general, a different set of candidate genes were identified for each breed, and similar genes were found across different lactations for the same trait of a same breed. For some of the candidate genes, the suggested pattern of gene expression changed among lactations. Among the lactations, candidate genes (and their suggested expression over time) identified for the second and third lactations were more similar with each other than with the first lactation. Well-known candidate genes with major effects on milk production traits presented different suggested patterns of gene expression across breeds, traits and lactations in which they were identified. The candidate genes identified in this study can be used as target genes in transcriptomic studies.

Keywords: fat, milk, protein, somatic cell, test-day

9.2. Introduction

Traits measured repeatedly during an individual's lifetime or physiological cycle are termed longitudinal traits. The majority of economically important longitudinal traits in dairy cattle are lactation-related variables, such as milk, fat and protein yields and somatic cell score (Miglior et al., 2017). The variability observed among the repeated records for those longitudinal traits is due to changes in the physiological state of the animal over the lactation period. Such changes in the trait expression over the lactation period as a consequence of physiological state of the animal are likely due to differential gene expression or unstable effects of Quantitative Trait Loci (**QTL**) over time, as suggested by the changes in additive genetic variance and heritability over lactation (e.g., Miglior et al., 2009; Oliveira et al., 2017; Sasaki et al., 2017). Furthermore, gene expression studies have indicated that the amount of ribonucleic acid (**RNA**) synthesised often varies across lactations (Bionaz and Loor, 2008), and that specific genes or QTLs can be differently expressed depending on the lactation stage (Strucken et al., 2011). Nonetheless, various studies on longitudinal traits have combined repeated records for each animal into a single estimate (e.g., average over all time points or accumulated yield), which are then used as response variables for genomic prediction of breeding values (e.g., Ding et al., 2013; Li et al., 2014; Jenko et al., 2017) and genome-wide association studies (**GWAS;** e.g., Buitenhuis et al., 2014; Raven et al., 2014; Nayeri et al., 2016).

Using a single estimate derived from repeated records when performing GWAS may compromise the understanding of the genetic architecture and underlying biological mechanisms of longitudinal traits. This approach does not enable unraveling the genomic regions and candidate genes with differential effects along the lactation curve. For instance, a certain QTL might have a highly significant effect in a specific lactation stage (e.g., lactation peak), but when averaging its effect across the whole lactation, it might be substantially reduced, which would consequently hamper the identification of this QTL, if combining repeated records. Therefore, using statistical models that consider all the variability from repeated records can increase the statistical power to detect time-dependent QTLs and consequently, candidate genes affecting traits at specific stages of the lactation.

In this context, Das et al. (2011) showed that is possible to use random regression models (**RRMs**) to exploit specific covariance functions to describe changes in the single nucleotide polymorphism (**SNP**) effects over time. By analyzing SNP associations over time, QTLs with greater effects at specific times are more likely to be identified, which may contribute to detect the missing genetic variance that explains the observed phenotypic variation in longitudinal traits (Strucken et al., 2015). Currently, this approach is the most sophisticated and flexible tool to perform GWAS for longitudinal traits. However, despite the SNP availability and several GWAS reports for milk-related traits in dairy cattle, the use of RRMs to estimate SNP effects over time remains underexplored, especially when considering different dairy cattle breeds.

Estimating SNP effects over time for various breeds is essential to identify and validate candidate genes (or QTLs) associated with time-dependent expression of economically important traits across breeds and better understand the underlying mechanisms of lactation biology. Thus, identifying these genomic regions may contribute to fastener genetic progress for longitudinal traits, by developing customized SNP chip panels based on functional markers. Therefore, this study aimed to: 1) estimate time-dependent effects of SNP associated with milk (**MY**), fat (**FY**) and protein (**PY**) yields, and somatic cell score (**SCS**) in the first three lactations of Canadian Ayrshire, Holstein and Jersey breeds; and 2) identify candidate genes associated with the analyzed traits, across lactations and breeds, and suggest their potential pattern of expression.

9.3. Material and methods

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

Phenotypic, Genotypic and Pedigree Data

The Canadian Dairy Network (**CDN**; Guelph, ON, Canada) provided the December 2012 pedigree, genotypes, and the MY, FY, PY and SCS test-day records (phenotypes) from the first three lactations of Ayrshire, Holstein, and Jersey breeds. Phenotypic quality control was performed to remove outlier test-day records for each trait (μ ± 3 SD, within contemporary group, which were defined by the combination of herd and test-day), and contemporary groups that were formed by less than 3 animals. Details on the criteria and number of test-day records removed in each step can be found in Chapter 7.

Bulls and cows were genotyped or imputed by CDN to the Illumina BovineSNP50K BeadChip (Illumina, San Diego, CA, USA), using the FImpute software (Sargolzaei et al., 2011). Details about the genotype imputation process performed by CDN is presented in Larmer et al. (2014). After the imputation process, SNPs with Mendelian conflicts, call rate less than 0.95, minor allele frequency (**MAF**) less than 0.01 and with a difference between observed and expected heterozygosity frequency higher than 0.15 (Wiggans et al., 2009) were removed using the preGSf90 software (Aguilar et al., 2014). The total number of animals, test-day records, genotypes, and SNPs that remained in the analyses for each breed after the quality control are shown in Table 1.

Table 1. Number of animals included in the pedigree file, number of genotyped animals (bulls), number of test-day records, and number of single nucleotide polymorphisms (SNP) used in the analyses for Ayrshire, Holstein and Jersey breeds.

[1]Trait		Ayrshire	Holstein	Jersey
MY, FY and PY	**Pedigree**	204,429	5,085,542	157,718
	Phenotypes	2,143,941	59,724,786	1,353,185
	Genotypes	1,827 (634)	20,663 (8,517)	1,427 (849)
	[2]SNP	38,096	40,658	34,500
SCS	**Pedigree**	195,099	4,983,668	154,123
	Phenotypes	1,856,634	54,532,924	1,211,483
	Genotypes	1,827 (634)	20,663 (8,517)	1,427 (849)
	[2]SNP	38,096	40,658	34,500

[1]Milk (MY), fat (FY), and protein (PY) yields in the 1st, 2nd, and 3rd lactations were analyzed separately from somatic cell score (SCS) in the 1st, 2nd, and 3rd lactations. [2]SNP were spread on the 29 bovine autosomes.

Statistical Analyses

Prediction of Breeding Values. Breeding values for random regression coefficients of each animal were estimated using the single-step Genomic Best Linear Unbiased Prediction (**ssGBLUP**; Misztal et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010), based on RRMs. Two multiple-trait RRMs were used: one considering MY, FY and PY in the first three lactations; and another considering SCS in the first three lactations. Both RRMs included the herd-test day as fixed effect, a fixed regression for age-parity-season of calving, and random regressions for herd-year of calving, animal additive genetic and permanent environmental effects, as routinely used in official genetic evaluations performed by CDN in Canada (Interbull, 2018). All regressions were modeled using the fourth order Legendre orthogonal polynomials (Kirkpatrick et al., 1990) on days in milk (**DIM**) ranged from 5 to 305 days.

In ssGBLUP analyses, the pedigree-based relationship matrix (**A**) traditionally used in the mixed model equations is replaced by the **H** matrix, which combines the pedigree and genomic information. In summary, the **H** matrix can be described as (Aguilar et al., 2010; Christensen and Lund, 2010):

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

where $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix (calculated using the first method presented in VanRaden, 2008), $\mathbf{A}^{-1}$ is the inverse of the pedigree-based relationship matrix, and $\mathbf{A}_{22}^{-1}$ is the inverse of the section of **A** related to the genotyped animals. The τ and ω parameters used in this study assumed the optimal values reported in Chapter 7 for these populations, i.e., $\tau = 2.0$ and $\omega = 0.6$ for the Ayrshire, $\tau = 1.5$ and $\omega = 0.6$ for the Holstein, and $\tau = 1.5$ and $\omega = 0.9$ for the Jersey breed, for the RRMs for MY, FY and PY; and $\tau = 1.0$ and $\omega = 0.9$ for the Ayrshire, $\tau = 2.0$ and $\omega = 0.6$ for the Holstein, and $\tau = 1.0$ and $\omega = 0.6$ for the Jersey breed, for the RRM for SCS. Solutions of the mixed model equations were obtained using the BLUP90IOD software (Misztal et al., 2002; Misztal et al., 2014). Prediction reliability and bias of breeding values for each trait, lactation and breed, as well as complete description of the statistical models used in this study are detailed in Chapter 7.

Derivation of SNP Effects from Genomic Breeding Values. The SNP solutions for the random regression coefficients were calculated using the postGSf90 software (Aguilar et al., 2014). In summary, the postGSf90 software back-solves the Genomic Estimated Breeding Values (**GEBV**; i.e., breeding values for genotyped animals) for random regression coefficients. In a simple way, the back-solving process can be briefly described, for each trait, lactation and breed, as follow (Wang et al., 2014):

$$\hat{\mathbf{u}}_c = \mathbf{M}'[\mathbf{M}\mathbf{M}']^{-1}\mathbf{G\hat{E}BV}_c,$$

where $\hat{\mathbf{u}}_c$ is the vector of estimated SNP solutions for the cth random regression coefficient; $\mathbf{M}$ is the matrix that contains the centered genotypes (i.e., -1, 0 and 1 representing AA, Aa and aa, respectively), and $\mathbf{G\hat{E}BV}_c$ is the vector of GEBV for the cth random regression coefficient estimated by the ssGBLUP, which contains the cth random regression coefficient from all the genotyped animals.

Thereafter, the SNP solutions for all random regression coefficients (c = 1, 2,…,5) of a same SNP k were combined into a vector $\left(\hat{\mathbf{u}}_k = \left[\hat{u}_{k_1} \quad \hat{u}_{k_2} \quad \hat{u}_{k_3} \quad \hat{u}_{k_4} \quad \hat{u}_{k_5}\right]'\right)$, and used to estimate the SNP effects for all DIM (from 5 to 305 days) as:

$$\mathbf{S\hat{N}P}_k = \mathbf{T}\hat{\mathbf{u}}_k,$$

where $\mathbf{S\hat{N}P}_k$ is the vector that contains the SNP effects estimated for every DIM of the kth SNP, $\mathbf{T}$ is a matrix of covariates for each DIM, associated with the Legendre orthogonal polynomials, and $\hat{\mathbf{u}}_k$ is the vector of SNP solutions for all random regression coefficients related to the kth SNP.

Selection of Relevant SNPs and Cluster Analyses. SNPs were selected to be further investigated based on the magnitude of their effects. In this context, the SNPs were ranked according to the magnitude of their daily effects (considering all DIM in the same rank), for each trait, lactation and breed. The top 1% SNPs that showed high magnitude of SNP effect in at least one DIM were selected as relevant SNPs to be further investigated. The strategy for selecting the top 1% SNPs has been implemented in several GWAS (Silva et al., 2014; Abdalla et al., 2016; Mastrangelo et al., 2018), and seems to be enough to show changes in SNP effects over time between breeds.

To facilitate the visualization of results and posterior discussion, the SNP selected as relevant SNP were clustered according to the trajectory of their SNP effects over time. Therefore, an algorithm was created in R (R Development Core Team, 2016) to assign each SNP to one of five pre-established clusters (**C**; defined in preliminary analysis based on graphical inspection and biological knowledge of the traits; results not shown): C1) constant SNP effect over time; C2) SNP effect increases over time; C3) SNP effect decreases over time; C4) SNP effect increases in the middle of lactation; and, C5) SNP effect decreases in the middle of lactation. In brief, the created algorithm estimates the average of the daily SNP effect for three different lactation stages: 1) from 5 to 95 DIM; 2) from 96 to 215 DIM; and 3) from 216 to 305 DIM; as defined in Chapter 8, and it assigns the SNP to their most likely cluster based on the pattern of those averages. Averages that did not differ in more than 10% from one to another lactation stage were considered as constant over time.

In order to describe the differences between clusters, the average of the coefficient of variation (**CV**) was estimated for each cluster. The CV of the kth SNP was calculated as follows:

$$\mathrm{CV}_k(\%) = \frac{s_k}{\bar{X}_k} \times 100,$$

where s_k is the standard deviation for the kth SNP over time, and $\bar{X}_k$ is the kth SNP mean.

Functional Analyses

Gene Mapping, Overrepresentation Analysis and Candidate Genes. Positional candidate genes were mapped using the Biomart tool (Kinsella et al., 2011) embedded in the Ensembl Genes database version 94 (http://useast.ensembl.org/index.html), using the bovine genome UMD 3.1 as the reference genome. Genes located within ± 250 kb from the location of the selected SNP were used in the overrepresentation analysis (**ORA**).

The ORA was performed using the Medical Subject Headings (**MeSH**) vocabulary to retrieve annotations that appear more frequently in a selected gene group relative to their occurrence by chance in a set of reference genes (Nelson et al., 2004). The statistical significance (P-value) of this overrepresentation analysis was assessed by a hypergeometric test (Adams and Skopek, 1987):

$$p = \sum_{i=k}^{\min(M,n)} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}},$$

where N is the total number of genes that were analyzed for each trait (reference genes), M is the total number of selected genes, n is the total number of genes in the MeSH term under study, k is the number of selected genes that belong to the MeSH term under study, and $\binom{y}{x} = \frac{y!}{x!}(y-x)!$ is the binomial coefficient. In order to increase the power of test and to reduce the number of false positive results, all lactations from a same trait and breed were analyzed together in the ORA. The ORA was performed using the MeSH ORA tool (Tsuyuzaki et al., 2015), available in the BIOCONDUCTOR package of the R software (R Development Core Team, 2011).

Genes retrieved from the significant (P-value < 0.01) MeSH terms related to milk production traits were considered as candidate genes for the traits, lactations and breeds that they were uncovered. To understand the biological processes related to the studied traits and to define the most likely candidate genes, complete gene functions were obtained from the National Center for Biotechnology Information database (**NCBI**, www.ncbi.nlm.nih.gov/gene/) and Animal QTL Database (**QTLdb**, www.animalgenome.org/cgi-bin/QTLdb/BT/index).

9.4. Results and discussion

Selection of the Most Relevant SNPs and Cluster Analysis

A total of 381, 407 and 345 SNPs were selected as the most relevant (top 1%) for each trait/lactation of the Ayrshire, Holstein and Jersey breeds, respectively. The number of selected SNP, their average SNP effect, and average of CV estimated for each cluster, for all traits, lactations and breeds are shown in Table 2.

Table 2. Number of selected single nucleotide polymorphisms (N), average of their effect (in grams for MY, FY and PY), and the coefficient of variation in percentage (CV) estimated in each cluster, for all traits, and lactations of Ayrshire, Holstein and Jersey breeds.

[1]Trait	Cluster	Ayrshire			Holstein			Jersey		
		N	Average	CV	N	Average	CV	N	Average	CV
MY1	1	13	11.55	8.92	38	35.88	7.49	35	7.23	8.29
	2	338	10.95	21.71	311	40.34	21.32	299	6.94	18.56
	3	10	10.72	12.48	20	33.39	13.67	4	6.76	11.88
	4	20	11.01	11.31	38	35.37	18.25	7	7.23	12.70
MY2	1	5	14.57	9.08	8	46.17	5.04	10	7.59	8.76
	2	336	12.22	28.69	299	31.09	44.71	290	6.90	25.54
	3	17	12.07	18.71	35	31.67	45.51	14	7.50	15.06
	4	23	13.26	14.09	65	58.27	23.81	31	8.23	21.47
MY3	1	1	13.41	8.99	5	41.17	7.27	7	7.86	8.34
	2	354	12.35	30.97	317	32.21	37.25	272	7.13	24.06
	3	4	13.17	17.69	24	31.26	40.37	21	7.52	14.57
	4	22	13.39	15.13	61	56.30	28.00	45	8.08	21.45
FY1	1	26	0.41	4.48	118	1.21	9.84	40	0.36	4.79
	2	341	0.35	23.96	185	1.51	19.70	295	0.28	19.59
	3	14	0.29	22.99	102	0.86	26.50	10	0.33	14.70
	4	0	-	-	2	1.46	17.77	0	-	-
FY2	1	2	0.59	5.26	49	2.84	6.30	22	0.49	4.84
	2	370	0.42	37.28	291	1.36	31.92	303	0.32	30.86
	3	9	0.36	47.58	61	1.05	54.88	20	0.42	16.71
	4	0	-	-	6	1.95	16.35	0	-	-
FY3	1	2	0.61	7.43	43	3.30	8.23	26	0.49	5.46
	2	370	0.43	39.44	317	1.39	32.43	298	0.33	30.88
	3	9	0.25	30.01	36	1.30	38.30	21	0.44	19.54
	4	0	-	-	11	1.76	18.27	0	-	-
PY1	1	2	0.38	6.42	21	0.91	5.45	2	0.23	7.21
	2	379	0.29	27.30	371	0.85	20.79	342	0.21	25.19
	3	0	-	-	3	0.88	16.63	0	-	-
	4	0	-	-	12	0.94	17.70	1	0.23	8.62
PY2	1	2	0.46	5.68	8	1.13	7.17	0	-	-
	2	373	0.35	36.21	368	0.87	39.74	345	0.24	31.22
	3	4	0.43	11.08	25	0.95	66.49	0	-	-
	4	2	0.43	10.15	6	1.24	20.45	0	-	-
PY3	1	1	0.46	8.03	4	1.18	8.08	0	-	-
	2	377	0.36	39.11	382	0.90	39.19	345	0.24	30.33
	3	1	0.46	9.99	10	0.67	95.17	0	-	-
	4	2	0.44	12.26	11	1.23	35.43	0	-	-
SCS1	1	159	1.39	5.70	129	6.76	6.08	126	0.91	5.87
	2	180	1.34	11.56	13	6.22	13.24	43	0.86	13.20
	3	42	1.27	10.84	265	5.54	13.82	176	0.69	22.70
	4	0	-	-	0	-	-	0	-	-
SCS2	1	25	1.56	6.98	79	8.36	6.68	66	0.96	6.67
	2	270	1.37	25.83	49	7.87	12.73	40	0.94	13.40
	3	69	1.36	22.41	219	7.31	14.88	238	0.78	21.87
	4	17	1.53	11.57	60	8.06	10.06	1	0.95	9.46
SCS3	1	18	1.57	5.11	47	9.81	6.48	28	1.03	6.66
	2	255	1.30	29.89	16	8.98	11.96	16	0.99	44.21
	3	101	1.28	27.41	297	8.08	17.25	293	0.83	25.45
	4	7	1.47	10.79	47	8.35	11.62	8	1.06	10.04

[1]Trait: milk (MY1, MY2, MY3), fat (FY1, FY2, FY3) and protein (PY1, PY2, PY3) yields, and somatic cell score (SCS1, SCS2, SCS3), in the first, second, and third lactations, respectively.

In general, the majority of SNPs selected for MY, FY and PY increased the magnitude of their effects over time (i.e., most SNP were clustered in C2), for all breeds. On the other hand, for SCS, the majority of selected SNP decreased the magnitude of their effects over time (i.e., most SNP were clustered in C3), especially for the Holstein and Jersey breeds, which might coincide with a higher prevalence of mastitis in the early lactation for those breeds (Dego and Tareke, 2003). No SNP whose magnitude of effect decreases in the middle of lactation (C5) were found, for none of the traits, lactations and breeds. In addition, no SNP whose magnitude of effect increases in the middle of lactation (C4) were found for FY in the Ayrshire and Jersey breeds; and for SCS in the first lactation of all breeds (Table 2).

With exception of SCS in the first lactation of all breeds, and FY in the first lactation of the Holstein breed, only a few SNPs were found to have approximately constant effect over DIM (Table 2). These results, combined to the different trajectory of SNP effects over time showed by the majority of SNP, suggest differential gene expression over time, which is in agreement with Strucken et al. (2011). For all traits, the average SNP effect was higher for Holstein, when compared to the Ayrshire and Jersey breeds, which is likely due to the higher average MY in Holstein (i.e., scale effect). These findings are in agreement with Chapter 8, that reported genomic windows with higher proportion of the explained genetic variance in the Holstein breed compared to the Ayrshire and Jersey breeds, and with Pryce et al. (2010), who also found that the average variation explained by the significant SNPs was higher in Holstein compared to Jersey. Considering the average SNP effects for all lactations, a single SNP changed the daily MY, FY, PY and SCS in approximately 12.4 g, 0.4 g, 0.4 g and 1.4, respectively, in the Ayrshire breed; 39.4 g, 1.7 g, 1.0 g, and 7.8, respectively, in the Holstein breed; and 7.4 g, 0.4 g, 0.2 g and 0.9, respectively, in the Jersey breed. For instance, this means that the QTL associated to an average SNP would be responsible for about 0.50 kg (i.e., 1.7 g × 301) of FY in one lactation of the Holstein breed, for example. This is in agreement with Strucken et al. (2012), when studying German Holstein Friesian cows, that concluded that one marker could change the overall FY in approximately 0.46 kg.

Especially for MY in the Holstein breed, the average SNP effect was higher in C2 for the first lactation, and in C4 for the second and third lactations, when compared to the other clusters (Table 2). This finding suggests that there are some specific genomic regions with large effect in MY that have increasing effect over DIM in the first lactation, and higher effects in the middle of lactation in the second and third lactations. Thus, genomic regions with large effect in MY seem to have higher persistency in their SNP effect in the first lactation when compared to the second and third one, which could explain the lower lactation peak and higher lactation persistency found in the first lactation of the Holstein breed when compared to later lactations (Stanton et al., 1992; Márquez et al., 2016). Similar results were found for FY in the Holstein breed, whose higher average SNP effect was observed in C1 when compared to the other SNP clusters, especially for the second and third lactations (Table 2). Hence, it may indicate that genomic regions with large effect in FY have approximately constant effect over DIM, which may be related to the high influence of environmental factors in this trait (Bauman et al., 2006; Bionaz and Loor, 2008b). Similar average SNP effects were estimated among different clusters for all other traits and breeds. As expected, higher CV were estimated in C2 and C3 when compared to the other clusters (Table 2). The trajectory of SNP effects over DIM, estimated for the top 10 SNPs of each cluster, for each trait and lactation, are shown in the Supplementary Figures S1, S2 and S3, for the Ayrshire, Jersey and Holstein breeds, respectively.

Candidate Genes

As expected due to the number of SNP selected for each breed, the number of candidate genes identified was, in general, higher for the for Holstein breed, followed by the Ayrshire and Jersey breeds, respectively. With exception of SCS for the Holstein breed, the highest number of candidate genes was found in C2, for all traits and breeds, which suggests that the majority of genes related to milk related traits have their expressions increased over DIM. The number of candidate genes found for each cluster after the overparameterization analysis performed for each trait and breed, are shown in the Supplementary Table S1.

Candidate genes mapped through Ensembl for the SNPs selected in each cluster, for MY, FY, PY and SCS, are shown in Tables 3, 4, 5 and 6, respectively. In general, the same candidate gene was not found in more than one cluster, which suggests that SNPs associated with the same candidate gene have similar SNP effect pattern over time, and that the method used to cluster the SNPs was efficient. In accordance with Chapter 8, similar genes were found across different lactations for the same trait of a same breed (Tables 3 to 6). For some of the candidate genes, the suggested pattern of gene expression changed among lactations (i.e., some candidate genes were found in different clusters in different lactations).

Table 3. Chromosome information (BTA, inside brackets) and candidate gene symbols mapped through Ensembl for single nucleotide polymorphisms selected in each cluster, for milk yield in the first three lactations of Ayrshire, Holstein and Jersey breeds.

Breed	Lactation	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Ayrshire	**First**	-	(BTA2) COL4A3; (BTA4) IGFBP3; (BTA5) ACO2; (BTA7) ANGPTL4; (BTA13) AVP; (BTA19) FASN; (BTA20) CCL28, PRLR; (BTA23) BTN1A1	-	(BTA6) CSN1S1, CSN1S2, CSN3, STATH
	Second	-	(BTA2) MAP2; (BTA4) IGFBP3; (BTA5) ACO2; (BTA7) ANGPTL4; (BTA14) DGAT1; (BTA17) IGLL1, SLC5A1; (BTA19) FASN; (BTA20) CCL28, PRLR; (BTA23) BTN1A1, (BTA25) KPNA7	(BTA6) CSN1S1	-
	Third	-	(BTA2) MAP2; (BTA4) IGFBP3; (BTA5) ACO2; (BTA7) ANGPTL4; (BTA14) BOP1, DGAT1, FBXL6; (BTA17) IGLL1, SLC5A1; (BTA19) FASN; (BTA20) CCL28, PRLR; (BTA23) BTN1A1; (BTA25) KPNA7	-	-
Holstein	**First**	(BTA2) COL4A3	(BTA5) ACO2; (BTA14) BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, PTK2, RECQL4, TONSL, TRAPPC9, VPS28; (BTA17) IGLL1, MAPK1, SLC5A1; (BTA21) SERPINA3-1, SERPINA5; (BTA23) BoLA; (BTA27) SLC25A4	(BTA3) GNG12; (BTA5) TNFRSF1A	(BTA5) PPARA
	Second	(BTA2) COL4A3; (BTA17) IGLL1, SLC5A1	(BTA5) ACO2; (BTA14) XKR4; (BTA17) MAPK1; (BTA19) COL1A1; (BTA21) SERPINA3-1, SERPINA5; (BTA23) BoLA; (BTA26) DNTT; (BTA27) SLC25A4	(BTA3) GNG12; (BTA5) TNFRSF1A	(BTA5) PPARA; (BTA14) BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, PTK2, RECQL4, TONSL, TRAPPC9, VPS28
	Third	(BTA2) COL4A3; (BTA17) IGLL1, SLC5A1	(BTA1) APP; (BTA5) ACO2; (BTA17) MAPK1; (BTA19) COL1A1; (BTA23) BoLA; (BTA26) DNTT; (BTA27) SLC25A4	(BTA3) GNG12; (BTA5) TNFRSF1A	(BTA5) PPARA; (BTA14) BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, PTK2, RECQL4, TONSL, TRAPPC9, VPS28
Jersey	**First**	(BTA2) COL4A3; (BTA24) NPC1; (BTA25) ACTB	(BTA3) GNG12; (BTA6) SLC34A2; (BTA8) TLR4; (BTA10) SLC7A7; (BTA19) STAT5A; (BTA21) IGF1R; (BTA25) KPNA7, SLC5A2; (BTA26) ACSL5; (BTA27) SLC25A4	-	(BTA19) SLC25A11
	Second	-	(BTA3) GNG12; (BTA6) SLC34A2; (BTA19) STAT5A; (BTA21) IGF1R; (BTA25) KPNA7, SLC5A2; (BTA26) ACSL5; (BTA27) IKBKB, NRG1; (BTA28) ANXA11	(BTA25) ACTB	(BTA10) SLC7A7; (BTA19) SLC25A11; (BTA24) NPC1
	Third	(BTA24) NPC1	(BTA6) SLC34A2; (BTA19) STAT5A; (BTA21) IGF1R; (BTA25) KPNA7, SLC5A2; (BTA27) IKBKB, NRG1; (BTA28) ANXA11	-	(BTA10) SLC7A7; (BTA25) ACTB; (BTA26) ACSL5

Clusters were defined as: Cluster 1) constant SNP effect over time; Cluster 2) SNP effect increases over time; Cluster 3) SNP effect decreases over time; Cluster 4) SNP effect increases in the middle of lactation; and, Cluster 5) SNP effect decreases in the middle of lactation. No candidate genes were found in Cluster 5.

Table 4. Chromosome information (BTA, inside brackets) and candidate gene symbols mapped through Ensembl for single nucleotide polymorphisms selected in each cluster, for fat yield in the first three lactations of Ayrshire, Holstein and Jersey breeds.

Breed	Lactation	Cluster 1	Cluster 2	Cluster 3
Ayrshire	**First**	(BTA6) CSN1S1	(BTA3) ACP6; (BTA4) IGFBP3; (BTA5) ACO2; (BTA7) ANGPTL4; (BTA11) SLC2A8; (BTA14) PTK2, TRAPPC9; (BTA17) SLC5A1; (BTA19) DCXR, FASN, PYCR1; (BTA23) BTN1A1	-
	Second	-	(BTA3) ACP6; (BTA4) IGFBP3; (BTA5) MYBPC1; (BTA7) ANGPTL4; (BTA11) SLC2A8; (BTA17) SLC5A1; (BTA18) SLC12A3; (BTA19) DCXR, FASN, PYCR1; (BTA25) KPNA7	-
	Third	-	(BTA3) ACP6; (BTA4) IGFBP3; (BTA5) MYBPC1; (BTA7) ANGPTL4; (BTA17) SLC5A1; (BTA18) SLC12A3; (BTA19) DCXR, FASN, PYCR1; (BTA24) NPC1; (BTA25) KPNA7	(BTA6) SNCA; (BTA17) MAPK1
Holstein	**First**	(BTA14) BOP1, DGAT1, FBXL6, OPLAH, PTK2, TRAPPC9	(BTA1) AGPAT3, APP; (BTA5) ACO2, ADIPOR2; (BTA13) PLCB4; (BTA14) CPSF1, CYHR1, FOXH1, LRRC24, RECQL4, SLC39A4, TONSL, VPS28; (BTA23) AGPAT1, BMP5, BoLA, CYP21; (BTA26) DNTT, FAS, IFIT3, IFIT5	(BTA14) EEF1D
	Second	(BTA14) BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, OPLAH, PTK2, RECQL4, TONSL, TRAPPC9, VPS28	(BTA1) AGPAT3, APP; (BTA5) ACO2, ADIPOR2; (BTA19) COL1A1; (BTA23) AGPAT1, BMP5, BoLA, CYP21; (BTA26) ACSL5, DNTT, FAS, IFIT3, IFIT5	(BTA14) EEF1D
	Third	(BTA14) BOP1, CPSF1, CYHR1, DGAT1, FBXL6, FOXH1, LRRC24, OPLAH, PTK2, RECQL4, TONSL, TRAPPC9, VPS28	(BTA1) AGPAT3, APP; (BTA5) ACO2; (BTA13) PLCB4; (BTA19) COL1A1; (BTA23) MAPK14; (BTA26) DNTT, FAS, IFIT3, IFIT5	(BTA5) ADIPOR2; (BTA14) EEF1D
Jersey	**First**	(BTA14) BOP1, DGAT1	(BTA6) SLC34A2; (BTA10) SLC7A7; (BTA19) SLC25A11; (BTA25) KPNA7; (BTA27) IKBKB	-
	Second	(BTA14) BOP1, DGAT1	(BTA2) IGFBP2; (BTA6) SLC34A2; (BTA18) FTO; MMP2; (BTA19) SLC25A11; (BTA27) IKBKB	-
	Third	(BTA14) BOP1, DGAT1	(BTA2) IGFBP2; (BTA6) SLC34A2; (BTA7) SLC26A2; (BTA18) FTO; (BTA19) SLC25A11; (BTA27) IKBKB	-

Clusters were defined as: Cluster 1) constant single nucleotide polymorphism (SNP) effect over time; Cluster 2) SNP effect increases over time; Cluster 3) SNP effect decreases over time; Cluster 4) SNP effect increases in the middle of lactation; and, Cluster 5) SNP effect decreases in the middle of lactation. No candidate genes were found in Clusters 4 and 5.

Table 5. Chromosome information (BTA, inside brackets) and candidate gene symbols mapped through Ensembl for single nucleotide polymorphisms selected in each cluster, for protein yield in the first three lactations of Ayrshire, Holstein and Jersey breeds.

Breed	Lactation	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Ayrshire	First	-	(BTA2) COL4A3; (BTA4) IGFBP3; (BTA5) ACO2; (BTA6) CSN1S1; (BTA7) ANGPTL4; (BTA13) MMP9; (BTA17) IGLL1, SLC5A1; (BTA19) EIF1, FASN; (BTA21) PDIA3	-	-
	Second	-	(BTA3) ACP6; (BTA4) IGFBP3; (BTA5) ACO2; (BTA7) ANGPTL4; (BTA17) IGLL1, SLC5A1; (BTA18) SLC12A3; (BTA19) EIF1, FASN; (BTA20) RPL37; (BTA21) PDIA3	(BTA6) CSN1S1	-
	Third	-	(BTA3) ACP6; (BTA4) IGFBP3; (BTA5) ACO2; (BTA7) ANGPTL4; (BTA17) IGLL1, MAPK1, SLC5A1; (BTA18) SLC12A3; (BTA19) EIF1, FASN; (BTA20) RPL37; (BTA21) PDIA3; (BTA24) NPC1	-	-
Holstein	First	(BTA3) GNG12	(BTA1) APP; (BTA5) ACO2; (BTA13) PLCB4; (BTA14) BOP1, DGAT1, EEF1D, FBXL6, PTK2; (BTA23) AGPAT1, BMP5, BoLA, CYP21	-	-
	Second	-	(BTA1) APP; (BTA5) ACO2; (BTA13) PLCB4; (BTA23) AGPAT1, BMP5, BoLA, CYP21; (BTA25) SLC5A2; (BTA26) FAS, IFIT3, IFIT5	-	-
	Third	-	(BTA1) APP; (BTA5) ACO2; (BTA13) PLCB4; (BTA17) SCARB1; (BTA23) BoLA, MAPK14; (BTA26) FAS, IFIT3, IFIT5	-	-
Jersey	First	(BTA6) CSN1S1, CSN1S2, CSN3, STATH; (BTA20) PRLR	(BTA6) SLC34A2; (BTA8) TLR4; (BTA10) SLC7A7; (BTA21) PDIA3; (BTA24) NPC1; (BTA25) ACTB; (BTA27) SLC25A4	-	(BTA20) RPL37
	Second	-	(BTA6) SLC34A2; (BTA8) TLR4; (BTA18) MMP2; (BTA21) PDIA3; (BTA24) NPC1; (BTA27) IKBKB, NRG1	-	-
	Third	-	(BTA2) COL4A3; (BTA6) SLC34A2; (BTA8) TLR4; (BTA21) PDIA3; (BTA24) NPC1; (BTA27) IKBKB, NRG1	-	-

Clusters were defined as: Cluster 1) constant single nucleotide polymorphism (SNP) effect over time; Cluster 2) SNP effect increases over time; Cluster 3) SNP effect decreases over time; Cluster 4) SNP effect increases in the middle of lactation; and, Cluster 5) SNP effect decreases in the middle of lactation. No candidate genes were found in Cluster 5.

Table 6. Chromosome information (BTA, inside brackets) and candidate gene symbols mapped through Ensembl for single nucleotide polymorphisms selected in each cluster, for somatic cell score in the first three lactations of Ayrshire, Holstein and Jersey breeds.

Breed	Lactation	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Ayrshire	**First**	(BTA3) ACP6; (BTA9) TNFAIP3; (BTA10) CAPN3; (BTA14) EEF1D, TRAPPC9	(BTA19) MAPT; (BTA23) TUBB2B; (BTA25) SLC5A2	(BTA25) ACTB	-
	Second	-	(BTA5) APOBEC3H, TNFRSF1A; (BTA6) SNCA; (BTA14) EEF1D; (BTA17) MAPK1; (BTA23) TUBB2B; (BTA25) SLC5A2	(BTA3) ACP6; (BTA9) TNFAIP3; (BTA10) CAPN3; (BTA14) TRAPPC9; (BTA19) MAPT; (BTA25) ACTB	-
	Third	-	(BTA5) APOBEC3H, TNFRSF1A; (BTA6) SNCA; (BTA14) EEF1D; (BTA17) MAPK1; (BTA23) TUBB2B; (BTA25) SLC5A2	(BTA3) ACP6; (BTA9) TNFAIP3; (BTA10) CAPN3; (BTA14) TRAPPC9; (BTA19) MAPT; (BTA25) ACTB	
Holstein	**First**	(BTA5) TNFRSF1A; (BTA6) ABCG2, FAM13A, SNCA; (BTA7) ANGPTL4; (BTA13) PTI; (BTA14) TRAPPC9; (BTA27) GPAT4, HTRA4	(BTA19) SOCS3	(BTA2) IL22RA1; (BTA7) SLC26A2; (BTA14) BOP1, CPSF1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL; (BTA16) C4BPA, C4BPB; (BTA20) PRLR	-
	Second	(BTA6) ABCG2, FAM13A; (BTA27) HTRA4	(BTA14) TRAPPC9	(BTA5) TNFRSF1A; (BTA6) SNCA; (BTA13) PTI; (BTA14) BOP1, CPSF1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, RECQL4, TONSL; (BTA16) C4BPA, C4BPB; (BTA27) NRG1	-
	Third	(BTA14) TRAPPC9	-	(BTA2) IL22RA1; (BTA5) TNFRSF1A; (BTA6) ABCG2, FAM13A, SNCA; (BTA14) BOP1, DGAT1, EEF1D, FBXL6; (BTA16) C4BPA, C4BPB; (BTA17) SLC5A1; (BTA27) HTRA4	-
Jersey	**First**	(BTA27) SLC25A4	(BTA2) COL4A3; (BTA6) SLC34A2; (BTA14) TRAPPC9; (BTA20) PRLR	(BTA14) PTK2	-
	Second	(BTA27) SLC25A4	(BTA6) SLC34A2; (BTA20) PRLR	(BTA2) COL4A3; (BTA5) SLC38A4; (BTA14) TRAPPC9; (BTA27) SLC25A4	(BTA7) SLC26A2
	Third	(BTA27) SLC25A4	(BTA6) SLC34A2; (BTA20) PRLR	(BTA2) COL4A3; (BTA7) SLC26A2; (BTA14) PTK2, TRAPPC9; (BTA27) SLC25A4	(BTA6) ABCG2

Clusters were defined as: Cluster 1) constant single nucleotide polymorphism (SNP) effect over time; Cluster 2) SNP effect increases over time; Cluster 3) SNP effect decreases over time; Cluster 4) SNP effect increases in the middle of lactation; and, Cluster 5) SNP effect decreases in the middle of lactation. No candidate genes were found in Cluster 5.

Among the lactations, candidate genes (and their suggested expression over time) identified for the second and third lactations were more similar with each other than with the first lactation. This may be related to the fact that the activity level of different genes that contribute to milk-related traits in different lactations depends on the metabolic changes happening in the individual (Strucken et al., 2011), and in the first lactation cows are still growing, i.e., the mammary gland and the digestive tract are under development (Taylor et al., 2003). Similar to the findings reported by Chapter 8, a different set of candidate genes were identified for each breed. This reinforces the hypothesis that the underlying lactation biology and genes associated with milk related traits differ across breeds or populations. Only some of the selected candidate genes will be further discussed in this paper.

Candidate Genes for MY

No functional candidate genes with constant suggested effect over DIM (i.e., candidate genes associated with SNP clustered in C1) were found for MY in the Ayrshire breed. Genes such as IGFBP3, ACO2, ANGPTL4, FASN, CCL28, PRLR, and BTN1A1 had increased effect over time for all three lactations. For instance, BTN1A1 has been linked to the secretion of milk lipid droplets because it is highly expressed in the mammary epithelium during lactation (Jeong et al., 2009). The CCL28 gene has been reported to play a role in lactation persistence in Canadian Holstein cattle (Do et al., 2017), which may explain its increased effect over time suggested in our study for the Ayrshire breed. As discussed by Tong et al. (2018), milking frequency affects lactation persistency, the expression of the PRL receptor (PRLR), and the milk response to prolactin inhibition (Bernier-Dodier et al., 2011; Lacasse et al., 2011). Therefore, it is possible that some factors modulating milk production are affecting prolactin secretion or the responsiveness of the mammary gland to this hormone (Tong et al., 2018).

The genes MAP2, DGAT1, IGLL1, and SLC5A1 had increasing effect over DIM (C2), and were suggested as candidate genes only for the second and third lactations of the Ayrshire breed. These genes are well known to affect MY and other milk-related traits (Huang et al., 2012; Nilsen et al., 2009; Prinzenberg et al., 2003). Cerri et al. (2012) investigating the effects of lactation and pregnancy on endometrial gene expression on 17th day of the estrous cycle and pregnancy, reported that IGLL1 was up-regulated by lactation. The casein group (i.e., CSN1S1, CSN1S2 and CSN3 genes) was suggested as candidate with increased effect in the middle of lactation, only for the first lactation. Interestingly, the CSN1S1 was suggested to have a decreasing effect over DIM in the second lactation. Strucken et al. (2012) reported that the largest effects for the casein group genes occurred at the beginning of the lactation, but they did not found statistical significance for MY. In addition, Strucken et al. (2012), studying how the influence of genomic regions changes during the first 60 DIM in German Holstein Friesian cattle, showed that the DGAT1 gene had increasing effects for MY.

Some genes were identified to be affecting all production traits (i.e., MY, FY and PY) in all lactations of Ayrshire, with differential expression patterns, such as ANGPTL4, which is known to play important role in lipid metabolism. Wang et al. (2018) investigated variations of serum ANGPTL4 in periparturient and early lactation dairy cows and reported that serum ANGPTL4 concentrations peaked at parturition, and declined rapidly over the following 2 weeks, indicating that energy balance changes that were associated with parturition contributed significantly to this effect. Koltes and Spurlock (2012) reported that the abundance of ANGPTL4 mRNA expression increased 2- to 6-fold with declining energy balance. According to the authors, the consistency and magnitude of the increase in ANGPTL4 mRNA across multiple models of altered energy balance identifies it as an adipokine that is uniquely responsive to changes in energy balance in the lactating dairy cow. In our study, depending on the trait, this gene was clustered in different groups (e.g., C1 or C2).

For the Holstein breed, our results suggested that the COL4A3 gene has constant effect over DIM, for all three lactations. In addition, the IGLL1 and SLC5A1 genes also showed constant effect over DIM, but only for the second and third lactation. The GNG12 and TNFRSF1A were candidate genes with suggested decreasing effect over DIM, for all three lactations. Interestingly, genes in the BTA14 very well-known for their effects in MY, such as BOP1, CPSF1, CYHR1, DGAT1, EEF1D, FBXL6, FOXH1, LRRC24, PTK2, RECQL4, TONSL, TRAPPC9 and VPS28, had increasing effect over DIM in the first lactation, and higher effect in the middle of lactation for the second and third lactations. This suggests that the pattern of expression for this genes also changes across lactations, even though they are still important in all lactations. The gene EEF1D has been previously associated with fat percentage (Jiang et al., 2010), which is a trait dependent on the level of MY and FY across lactation. The importance of these gene for various milk traits and breeds highlights its importance to lactation biology. In addition, the different pattern of gene expression suggested for genes in BTA14 may explain the higher average SNP effect in C2 for the first lactation, and in C4 for the second and third lactations. Differences in the conclusions presented by Strucken et al. (2012) for the DGAT1 gene are probably due to the short period of time studied by the refereed authors, and to the use of MY averaged over the first three lactations. Thus, the authors could not notice the suggested decrease in this gene expression after the middle of lactation, for the second and third lactations.

Wang et al. (2015a) investigated the expression pattern of the clock circadian regulator network and selected metabolic genes simultaneously in mammary gland, liver, and subcutaneous adipose tissue to better understand the transition from late gestation to early lactation. The referred authors reported that the expression of PPARA was lower at -10 days, but increased after parturition. However, Schaff et al. (2013) reported that PPARA expression continuously decreased after parturition, while for Holstein we observed an effect increase in the middle of lactation (C4).

For the Jersey breed, the SLC34A2, STAT5A, IGF1R, KPNA7 and SLC5A2 genes showed increasing effects over DIM in all three lactations (C2). The STAT5A gene, a member of the JAK-STAT signaling pathway has been associated with proliferation, secretory differentiation, and survival of mammary epithelial cells and milk synthesis (Ao et al., 2015; Cui et al., 2004; Yamaji et al., 2009; Furth et al., 2011; Liu et al., 2012). Additionally, STAT5A can mediate a signal-transduction pathway to regulate milk protein synthesis, mammary gland epithelial cell differentiation in pregnancy, to maintain survival of breast epithelial cells (Faraci-Orf et al., 2006; Buser et al., 2007) and mastitis susceptibility (Usman et al., 2014). Interestingly, the pattern of gene expression suggested for the ACTB gene changed completely among lactations, as it had suggested constant effect over time in the first lactation (C1), decreasing effect in the second lactation (C3), and higher effect in the middle of lactation, in the third lactation (C4). The IKBKB, NRG1, and ANXA11 genes were suggested as candidate genes with increasing effect over time, for the second and third lactations. In agreement with Oliveira et al. (mtGWAS), no potential effect of DGAT1 gene was found for MY in the Jersey breed.

Candidate Genes for FY

The CSN1S1 and PTK2 genes were suggested as candidate genes only for the first lactation of the Ayrshire breed, and they had suggested constant (C1) and increasing (C2) effect over DIM, respectively. The ACP6, IGFBP3, ANGPTL4, SLC5A1, DCXR, FASN, and PYCR1 genes had increasing effects over DIM (C2), for all three lactations. Bionaz and Loor (2008) studied mammary tissue mRNA expression across changes in physiological state, including different lactation stages, and they reported that the FASN gene is associated with fatty acid synthesis and desaturation. Moreover, the referred authors reported that the peak expression of the FASN gene was around 60 DIM.

For the Holstein breed, the BOP1, DGAT1, FBXL6, OPLAH, PTK2, and TRAPPC9 genes had constant effect over DIM (C1) for all three lactations. However, as suggested by the average SNP effect estimated in the clusters, some important genes from BTA14 had different expression pattern in the first lactation when compared to the second and third ones. Those genes were CPSF1, CYHR1, FOXH1, LRRC24, RECQL4, SLC39A4, TONSL and VPS28, that had increasing effect over DIM in the first lactation, and constant effect in the second and third lactations. For instance, CPSF1 and CYHR1 have been reported to be associated with milk fat components (Buitenhuis et al., 2014; Jiang et al., 2014; Nayeri et al., 2016). FOXH1 has also been linked to milk related traits (Nayeri et al., 2016). The ADIPOR2 gene showed increasing effect in the first and second lactation, and decreasing effect over DIM in the third lactation. Weber et al. (2016) reported that the ADIPOR2 gene expression changed from late gestation to early lactation. Ohtani et al. (2011) investigated the gene expression of ADIPOR2 in various bovine tissues and mammary glands among different lactation stages, and the effects of lactogenic hormones (insulin, dexamethasone and prolactin) and growth hormone (**GH**) in cultured bovine mammary epithelial cells (**BMEC**). They reported that in BMEC, lactogenic hormones and GH up-regulated ADIPOR2 mRNA expression. Loor et al. (2006) studied temporal gene expression profiles during the dry period and early lactation in liver of Holstein cows. They identified 85 genes with expression patterns that were affected by level of energy intake prepartum over time. Restricted energy intake prepartum resulted in more pronounced upregulation of genes with key functions in hepatic fatty acid oxidation that were also identified in our study (ADIPOR2) and ad libitum feeding upregulated a number of genes associated with liver triacylglycerol synthesis (DGAT1) and proinflammatory cytokines (TNFAIP3). However, in our study, TNFAIP3 was associated with SCS. Other very well-known candidate genes were suggested to have the same expression pattern over DIM in the all three lactations, such as AGPAT3, APP, ACO2, DNTT, FAS, IFIT3 and IFIT5, which had increasing effect over DIM, and EEF1D, that had suggested decreasing effect over DIM for all three lactations.

Similar to what found for Holstein, the BOP1 and DGAT1 genes were suggested to have constant effect over DIM for all three lactations in the Jersey breed. Those results corroborate with Oliveira et al. (mtGWAS), that also reported the BOP1 and DGAT1 genes as important genes along the lactation for FY in Jersey. The SLC34A2, SLC25A11 and IKBKB genes were found to have increasing effects over DIM in all three lactations. On the other hand, the IGFBP2 and FTO genes were suggested to have increasing effects over DIM only for the second and third lactation. Zielke et al. (2013) investigated the biological role of FTO (Fat Mass and Obesity associated gene) on milk composition in German Holstein cattle. They reported that higher frequency SNP alleles and haplotypes within the FTO gene increased FY and PY by up to 2.8 and 2.2 kg per lactation, respectively. According to the authors, the pleiotropic effects observed suggest a biological function on the usage of energy resources and the control of energy balance rather than directly affecting fat and protein synthesis. Clempson et al. (2012) reported that polymorphisms in IGFBP2 were associated with the circulating patterns of glucose, insulin, and BHB, particularly around second calving. Interestingly, for the Jersey breed, we observed that this gene was identified in second and third parities for FY. They also reported associations between the IGFBP2 SNPs and 305-day yield in the first lactation, as well as trends of association with DIM in first and second lactations in UK dairy cows (Clempson et al., 2012).

Candidate Genes for PY

The majority of candidate genes found for PY were suggested to have an increasing effect over DIM. Specifically for the Ayrshire breed, the IGFBP3, ACO2, ANGPTL4, IGLL1, SLC5A1, EIF1, FASN, and PDIA3 genes were suggested as candidate genes for all three lactations. Interestingly, it was suggested that the CSN1S1 gene is a candidate gene with decreasing effect over DIM for the second lactation. Strucken et al. (2012) found statistical significance in the decreasing gene expression of CSN1S1 over time for PY, when studying the first 60 DIM of German Holstein Friesian cattle.

For the Holstein breed, the GNG12 gene was suggested as a candidate gene with constant effect over DIM, for the first lactation. The APP, ACO2, PLCB4, and BoLA genes were suggested to have an increasing effect over DIM for all three lactations. The bovine major histocompatibility complex (BoLA) encode proteins involved in the presentation of extracellular antigens to T CD4+ cells (Ellis et al., 1999), and is well known to have a significant role in the development of immunological response and disease resistance (Ellis and Codner, 2012; Stear et al, 2001). Various studies have associated the BoLA genes with clinical mastitis, SCS, and antibody-mediated and cell-mediated immune responses in Holstein cattle (Sharif et al., 1998; Rupp et al., 2007; Pashmi et al., 2009; Wu et al., 2010; Yoshida et al., 2012; Thompson-Crispi et al., 2014). Interestingly, we identified its association with MY, PY and FY, but not with SCS. Pokorska et al. (2018) also identified a higher MY for a specific allele in Holstein cattle. The BOP1, DGAT1, EEF1D, FBXL6, and PTK2 genes were suggested as candidate genes only for the first lactation. In addition, it is important to notice that the AGPAT1, BMP5, and CYP21 genes were suggested as candidate genes for the first and second lactations, and that the FAS, IFIT3 and IFIT5 were suggested as candidate genes for the second and third lactations; all of them with increasing effect over DIM.

For the Jersey breed, the casein group (i.e., CSN1S1, CSN1S2 and CSN3 genes), STATH, and PRLR genes were identified as having a constant effect over DIM for the first lactation. The SLC34A2, TLR4, PDIA3, and NPC1 were suggested as candidate genes with increasing effect over DIM, for all three lactations. The gene TLR4, which was identified to be important for MY and PY, had been previously associated with lactation persistency in Canadian Holstein bulls (Sharma et al., 2006), as well as MY, FY, and fat and protein percentage and other milk composition traits (Viale et al., 2017; Zhou et al., 2017). The IKBKB and NRG1 genes were found to have an increasing effect over DIM only for the second and third lactations. Interestingly, the RPL37 gene was suggested to have a higher effect in the middle of lactation, in the first lactation. The gene MAPK1, which was identified for all four traits in this study, regulates milk protein synthesis (Lu et al., 2013).

Candidate Genes for SCS

For Ayrshire, the TUBB2B and SLC5A2 were suggested as candidate genes with increasing effect over DIM, for all lactations, while the ACTB gene was suggested as candidate gene with decreasing effect over DIM, for all lactations. Interestingly, the ACP6, TNFAIP3, CAPN3, and TRAPPC9 genes were suggested as candidate genes with constant effect in the first lactation and with decreasing effect in the second and third lactations. In the same way, the EEF1D gene was also suggested as candidate genes with constant effect in the first lactation, while it seems to increase its effect in the second and third lactations.

For the Holstein breed, genes such as BOP1, DGAT1, EEF1D, FBXL6, C4BPA and C4BPB were suggested as candidate genes with higher effect in the middle of lactation, for all three lactations. However, while the TNFRSF1A and SNCA genes also seems to have a higher effect in the middle of lactation for the second and third lactations, they had suggested constant effect in the first lactation. The ABCG2, FAM13A and HTRA4 genes were suggested as candidate genes with constant effect in the first and second lactations, and with higher effect in the middle of lactation for the third lactation. The ATP-binding cassette sub-family G member 2 (ABCG2) gene has been associated with various milk-related traits in Holstein and other dairy breeds, including milk production (Cohen-Zinder et al., 2005; Pausch et al., 2017), coagulation properties (Cecchinato et al., 2015) and fatty acid composition (Li et al., 2016). In addition, various signature of selection studies have identified this gene as playing an important role in milk production traits (e.g., Rothammer et al., 2013; Zhao et al., 2015). Li et al. (2016) also reported association between FASN, ABCG2 and IGF1 on milk fatty acid composition. Wei et al. (2012) concluded that ABCG2 plays an important role in mammary epithelial cell proliferation and that functional polymorphisms in this gene may influence the cellular compartment of the mammary gland and potentially milk production. As there is an inferred positive correlation between mammary cell numbers and milk production, mammary progenitor cells may be important in renewing damaged mammary cells during lactation (Capuco et al., 2005), which could also explain its differential association with SCS during the lactation. Mani et al. (2009) investigated the differential expression and localization of lipid transporters in the bovine mammary gland during the pregnancy-lactation cycle and reported that ABCG2 mRNA abundance was significantly increased after parturition and declined from the lactating to the nonlactating state in the bovine mammary gland. In accordance with this hypothesis, for the Holstein breed, ABCG2 was identified in C1, however, for Jersey, it was clustered in C4 (higher effect in the middle of lactation). Do et al. (2017) suggested the association of this gene with lactation persistence. The NRG1 gene has been strongly correlated with the progression of Staphylococcus aureus subclinical mastitis (Song et al., 2016). The SOCS3 gene, identified for SCS in our study, was previously suggested to have a primary role in apoptosis and tissue remodeling during involution through programmed cell death and tissue remodeling, initiated after the termination of lactation (Sutherland et al., 2007). Arun et al. (2015) reported that the levels of SOCS3 expression increased from pregnancy to involution.

For the Jersey breed, the SLC25A4 gene was suggested to have constant effect, and the SLC34A2 and PRLR genes were suggested to have increasing effects over DIM, for all three lactations. Prolactin receptor (PRLR), known to be associated with production traits and SCS (Meredith et al., 2012), was found to be downregulated in bovine mammary epithelial cells infected with Staphylococcus aureus (Lara-Zarate et al., 2011), as well as in bovine hepatic tissue following intra-mammary injection of Escherichia coli lipopolysaccharide to simulate mastitis infection (Jiang et al., 2008). Prolactin was found to be significantly increased in udder quarters with high SCS and chronic mastitis (Boutet et al., 2007). In addition, PRLR has been suggested to be associated with mammary gland epithelium development and mammary gland alveolus development and milk production traits (Marete et al., 2018; Viale et al., 2017; Viitala et al., 2006; Lu et al., 2011). The COL4A3 and TRAPPC9 genes were suggested as candidate genes with increasing effect over DIM in the first lactation, and decreasing effects over DIM in the second and third lactations. Wang et al. (2015) identified that TRAPPC9 was associated with mastitis susceptibility in Chinese Holstein cattle.

9.5. Conclusion

Different pattern of SNP effects over time were estimated for MY, FY, PY and SCS in the first three lactations of the Canadian Ayrshire, Holstein and Jersey breeds. Moreover, the potential pattern of gene expression was suggested for several candidate genes associated with the traits, lactations and breeds analyzed in this study. Selected genes previously reported to be associated with milk related traits were also discussed. The candidate genes identified in this study can be used as target genes in transcriptomic studies, which are necessary to validate our findings.

9.6. Acknowledgements

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES). The authors gratefully acknowledge Dr. Ignacio Aguilar for sharing the PCG algorithm for random regression models. In addition, the authors acknowledge the financial support for this research provided by the Food From Thought research program, Agriculture and Agri-Food Canada, and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative.

9.7. References

Abdalla, E.A., F. Peñagaricano, T.M. Byrem, K.A. Weigel, and G.J.M. Rosa. 2016. Genome-wide association mapping and pathway analysis of leukosis incidence in a US Holstein cattle population. Anim. Genet. 47:395–407. doi:10.1111/age.12438.

Adams, W.T., and T.R. Skopek. 1987. Statistical test for the comparison of samples from mutational spectra. J. Mol. Biol. 194:391–396. doi:10.1016/0022-2836(87)90669-3.

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.

Aguilar, I., S. Tsuruta, and A. Legarra. 2014. PREGSF90-POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 Programs. In: Proceedings of the 10th World Congress of Genetics Applied to Livestock Production (WCGALP), Vancouver, Canada.

Ao, J., Wei, C., Si, Y., Luo, C., Lv, W., Lin, Y., Cui, Y. and Gao, X., 2015. Tudor-SN regulates milk synthesis and proliferation of bovine mammary epithelial cells. Inter. J. Mol. Sci., 16(12):29936-29947.

Arun, S.J., Thomson, P.C., Sheehy, P.A., Khatkar, M.S., Raadsma, H.W. and Williamson, P., 2015. Targeted analysis reveals an important role of JAK-STAT-SOCS genes for milk production traits in Australian dairy cattle. Front. Gen. 6:342.

Bauman, D.E., I.H. Mather, R.J. Wall, and A.L. Lock. 2006. Major advances associated with the biosynthesis of milk. J. Dairy Sci. 89:1235–1243. doi:10.3168/jds.S0022-0302(06)72192-0.

Bernier-Dodier, P., C. L. Girard, B. G. Talbot, and P. Lacasse. 2011. Effect of dry period management on mammary gland function and its endocrine regulation in dairy cows. J. Dairy Sci. 94:4922–4936.

Bionaz, M., and J. J. Loor. 2008a. *ACSL1*, *AGPAT6*, *FABP3*, *LPIN1*, and *SLC27A6* are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J. Nutr. 138:1019–1024. doi:10.1093/jn/138.6.1019.

Bionaz, M., and J.J. Loor. 2008b. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9:366. doi:10.1186/1471-2164-9-366.

Boutet P., Sulon J., Closset R., Detilleux J., Beckers J.F., et al. (2007) Prolactin-Induced Activation of Nu- clear Factor κ B in Bovine Mammary Epithelial Cells: Role in Chronic Mastitis. J Dairy Sci. 90: 155–164. PMID: 17183084

Buitenhuis, B., L.L. Janss, N.A. Poulsen, L.B. Larsen, M.K. Larsen, and P. Sørensen. 2014. Genome-wide association and biological pathway analysis for milk-fat composition in Danish Holstein and Danish Jersey cattle. BMC Genomics 15:1112. doi:10.1186/1471-2164-15-1112.

Buser, A. C., E. K. Gass-Handel, S. L. Wyszomierski, W. Doppler, S. A. Leonhardt, J. Schaak, J. M. Rosen, H. Watkin, S. M. Anderson, and D. P. Edwards. 2007. Progesterone receptor repression of pro- lactin/signal transducer and activator of transcription 5-mediated transcription of the beta-casein gene in mammary epithelial cells. Mol. Endocrinol. 21:106–125.

Capuco A.V., Ellis S. Bovine mammary progenitor cells: current concepts and future directions. 2005. J Mamm. Gland Biol. Neop. 10:5–15.

Cecchinato, A., S. Chessa, C. Ribeca, C. Cipolat-Gotet, T. Bobbo, J. Casellas, and G. Bittante. 2015. Genetic variation and effects of candidate-gene polymorphisms on coagulation properties, curd firmness modeling and acidity in milk from Brown Swiss cows. Animal. 9(7):1104-1112.

Cerri, R.L.A., Thompson, I.M., Kim, I.H., Ealy, A.D., Hansen, P.J., Staples, C.R., Li, J.L., Santos, J.E.P. and Thatcher, W.W., 2012. Effects of lactation and pregnancy on gene expression of endometrium of Holstein cows at day 17 of the estrous cycle or pregnancy. J. Dairy Sci., 95(10):5657-5675.

Christensen, O.F., and M.S. Lund. 2010. Genomic relationship matrix when some animals are not genotyped Genomic prediction models. Genet. Sel. Evol. 42.

Clempson, A.M., Pollott, G.E., Brickell, J.S. and Wathes, D.C., 2012. Associations between bovine

IGFBP2 polymorphisms with fertility, milk production, and metabolic status in UK dairy cows. Animal Biotech., 23(2):101-113.

Cohen-Zinder, Miri, Eyal Seroussi, Denis M. Larkin, Juan J. Loor, Annelie Everts-van der Wind, Jun-Heon Lee, James K. Drackley et al. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. 15(7):936-944.

Cui, Y., G. Riedlinger, K. Miyoshi, W. Tang, C. Li, C. X. Deng, G. W. Robinson, and L. Hennighausen. 2004. Inactivation of Stat5 in mouse mammary epithelium during pregnancy reveals distinct functions in cell proliferation, survival, and differentiation. Mol. Cell. Biol. 24:8037–8047.

Das, K., J. Li, Z. Wang, C. Tong, G. Fu, Y. Li, M. Xu, K. Ahn, D. Mauger, R. Li, and R. Wu. 2011. A dynamic model for genome-wide association studies. Hum. Genet. 129:629–639. doi:10.1007/s00439-011-0960-6.

Dego, O. K., and F. Tareke. 2003. Bovine mastitis in selected areas of southern Ethiopia. Trop. Anim. Health Pro. 35:197-205.

Ding, X., Z. Zhang, X. Li, S. Wang, X. Wu, D. Sun, Y. Yu, J. Liu, Y. Wang, Y. Zhang, S. Zhang, Y. Zhang, and Q. Zhang. 2013. Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. J. Dairy Sci. doi:10.3168/jds.2012-6194.

Do, D.N., Bissonnette, N., Lacasse, P., Miglior, F., Sargolzaei, M., Zhao, X. and Ibeagha-Awemu, E.M., 2017. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. J. Dairy Sci., 100(3):1955-1970.

Ellis S.A., Holmes E.C., Staines K.A., Smith K.B., Stear M.J., McKeever D.J., MacHugh N.D., Morrison W.I. (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. Immunogenetics 50:319–328. https://doi.org/10.1007/s002510050 608.

Ellis S.A., Codner G. 2012. The impact of MHC diversity on cattle T cell responses. Vet. Immunol. Immunopathol. 2012, 148:74–77.

Faraci-Orf, E., C. Mcfadden, and W. F. Vogel. 2006. DDR1 signaling is essential to sustain stat5 function during lactogenesis. J. Cell. Biochem. 97:109–121.

Furth, P.A., R.E. Nakles, S. Millman, E.S. Diaz-Cruz, and M. C. Ca- brera. 2011. Signal transducer and activator of transcription 5 as a key signalling pathway in normal mammary gland developmental biology and breast cancer. Breast Cancer Res. 13:220.

Huang, W., Peñagaricano, F., Ahmad, K.R., Lucey, J.A., Weigel, K.A. and Khatib, H., 2012. Association between milk protein gene variants and protein composition traits in dairy cattle. J. Dairy Sci., 95(1):440-449.

Jenko, J., G.R. Wiggans, T.A. Cooper, S.A.E. Eaglen, W.G. de. L. Luff, M. Bichard, R. Pong-Wong, and J.A. Woolliams. 2017. Cow genotyping strategies for genomic selection in a small dairy cattle population. J. Dairy Sci. 100:439–452. doi:10.3168/jds.2016-11479.

Jeong, J., Rao, A.U., Xu, J., Ogg, S.L., Hathout, Y., Fenselau, C. and Mather, I.H., 2009. The PRY/SPRY/B30. 2 domain of butyrophilin 1A1 (BTN1A1) binds to xanthine oxidoreductase: implications for the function of BTN1A1 in the mammary gland and other tissues. J. Biol. Chem. pp.jbc-M109.

Jiang L., Sørensen P., Røntved C., Vels L., Ingvartsen K.L. 2008. Gene expression profiling of liver from dairy cows treated intra-mammary with lipopolysaccharide. BMC Genomics, 9: 443. doi: 10.1186/ 1471-2164-9-443 PMID: 18816405

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y. and Zhang, Q., 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PloS one, 5(10), p.e13661.

Jiang, L., X. Liu, J. Yang, H. Wang, J. Jiang, L. Liu, S. He, X. Ding, J. Liu, and Q. Zhang. 2014. Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. BMC Genomics 15:1105 https://doi.org/10.1186/1471-2164-15-1105.

Kinsella, R.J., A. Kahari, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database 2011:bar030-bar030. doi:10.1093/database/bar030.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979–93.

Larmer, S., M. Sargolzaei, and F. Schenkel. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. J. Dairy Sci. 97:3128–3141. doi:10.3168/jds.2013-6826.

Lacasse, P., V. Lollivier, R. M. Bruckmaier, Y. R. Boisclair, G. F. Wagner, and M. Boutinaud. 2011. Effect of the prolactin-re- lease inhibitor quinagolide on lactating dairy cows. J. Dairy Sci. 94:1302–1309.

Lara-Zárate L., López-Meza J.E., Ochoa-Zarzosa A. 2011. Staphylococcus aureus inhibits nuclear factor kappa B activation mediated by prolactin in bovine mammary epithelial cells. Microb. Pathog. 51: 313– 318. doi: 10.1016/j.micpath.2011.07.010 PMID: 21843629

Li, X., S. Wang, J. Huang, L. Li, Q. Zhang, and X. Ding. 2014. Improving the accuracy of genomic prediction in Chinese Holstein cattle by using one-step blending. Genet. Sel. Evol. doi:10.1186/s12711-014-0066-4.

Li, C., Sun, D., Zhang, S., Yang, S., Alim, M.A., Zhang, Q., Li, Y. and Liu, L., 2016. Genetic effects of FASN, PPARGC1A, ABCG2 and IGF1 revealing the association with milk fatty acids in a Chinese Holstein cattle population based on a post genome-wide association study. BMC Genetics, 17(1):110.

Liu, X.F., Li, M., Li, Q.Z., Lu, L.M., Tong, H.L. and Gao, X.J., 2012. Stat5a increases lactation of dairy cow mammary gland epithelial cells cultured in vitro. In Vitro Cellular & Developmental Biology-Animal, 48(9), pp.554-561.

Loor, J.J., Dann, H.M., Guretzky, N.A.J., Everts, R.E., Oliveira, R., Green, C.A., Litherland, N.B., Rodriguez-Zas, S.L., Lewin, H.A. and Drackley, J.K., 2006. Plane of nutrition prepartum alters hepatic gene expression and function in dairy cows as assessed by longitudinal transcript and metabolic profiling. Physiol. Gen., 27(1):29-41.

Lü, A., Hu, X., Chen, H., Dong, Y., Zhang, Y. and Wang, X., 2011. Novel SNPs of the bovine PRLR gene associated with milk production traits. Biochem. Gen., 49(3-4):177-189.

Lu, L.M., Li, Q.Z., Huang, J.G. and Gao, X.J., 2012. Proteomic and functional analyses reveal MAPK1 regulates milk protein synthesis. Molecules, 18(1):263-275.

Mani, O., Sorensen, M.T., Sejrsen, K., Bruckmaier, R.M. and Albrecht, C., 2009. Differential expression and localization of lipid transporters in the bovine mammary gland during the pregnancy-lactation cycle. J. Dairy Sci., 92(8):3744-3756.

Márquez, G.C., Y. Zare, K.L. Stephan, and K. Olson. 2016. 0367 Genetic evaluation of mastitis, metritis, and ketosis in Holstein cattle using producer recorded data. J. Anim. Sci. doi:10.2527/jam2016-0367.

Mastrangelo, S., M.T. Sardina, M. Tolone, R. Di Gerlando, A.M. Sutera, L. Fontanesi, and B. Portolano. 2018. Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds. Animal 1–9. doi:10.1017/S1751731118000629.

Meredith B.K., Kearney F.J., Finlay E.K., Bradley D.G., Fahey A.G., et al. 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genetics 13: 21. doi: 10.1186/1471-2156-13-21 PMID: 22449276

Miglior, F., A. Fleming, F. Malchiodi, L.F. Brito, P. Martin, and C.F. Baes. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J. Dairy Sci. 100:10251–10271. doi:10.3168/jds.2017-12968.

Miglior, F., W. Gong, Y. Wang, G.J. Kistemaker, A. Sewalem, and J. Jamrozik. 2009. Short communication: Genetic parameters of production traits in Chinese Holsteins using a random regression test-day model. J. Dairy Sci. doi:10.3168/jds.2009-2212.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. doi:10.3168/jds.2009-2064.

Nayeri, S., M. Sargolzaei, M.K. Abo-Ismail, N. May, S.P. Miller, F. Schenkel, S.S. Moore, and P. Stothard. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genetics. 17:75. doi:10.1186/s12863-016-0386-1.

Nelson, S.J., M. Schopen, A.G. Savage, J.-L. Schulman, and N. Arluk. 2004. The MeSH translation maintenance system: structure, interface design, and implementation. Stud. Health Technol. Inform. 107:67–9.

Nilsen, H., Olsen, H.G., Hayes, B., Sehested, E., Svendsen, M., Nome, T., Meuwissen, T. and Lien, S., 2009. Casein haplotypes and their association with milk production traits in Norwegian Red cattle. Gen. Sel. Evol., 41(1):24.

Ohtani, Y., Yonezawa, T., Song, S.H., Takahashi, T., Ardiyanti, A., Sato, K., Hagino, A., ROH, S.G. and Katoh, K., 2011. Gene expression and hormonal regulation of adiponectin and its receptors in bovine mammary gland and mammary epithelial cells. Animal Sci. J., 82(1):99-106.

Oliveira, H.R., F.F. e Silva, M.V.G. Barbosa da Silva, O.H.G. Barbosa Dias de Siqueira, M.A. Machado, J.C. do Carmo Panetto, L.S. Glória, and L.F. Brito. 2017. Bayesian models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. Livest. Sci. 201:78–84. doi:10.1016/j.livsci.2017.05.007.

Pashmi M., Qanbari S., Ghorashi S.A., Sharifi A.R., Simianer H. 2009. Analysis of relationship between bovine lymphocyte anti- gen DRB3.2 alleles, somatic cell count and milk traits in Iranian Holstein population. J. Anim. Breed. Genet. 126:296–303. https:// doi.org/10.1111/j.1439-0388.2008.00783.x

Pausch, H., Emmerling, R., Gredler-Grandl, B., Fries, R., Daetwyler, H.D. and Goddard, M.E.,

2017. Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. BMC Genomics, 18(1):853.

Pokorska, J., Kułaj, D., Dusza, M., Ochrem, A. and Makulska, J., 2018. The influence of BoLA-DRB3 alleles on incidence of clinical mastitis, cystic ovary disease and milk traits in Holstein Friesian cattle. Mol. Biol. Rep. 45(5):917-923.

Prinzenberg, E.M., Weimann, C., Brandt, H., Bennewitz, J., Kalm, E., Schwerin, M. and Erhardt, G., 2003. Polymorphism of the bovine CSN1S1 promoter: linkage mapping, intragenic haplotypes, and effects on milk production traits. J. Dairy Sci., 86(8):2696-2705.

Pryce, J.E., S. Bolormaa, A.J. Chamberlain, P.J. Bowman, K. Savin, M.E. Goddard, and B.J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93:3331–3345. doi:10.3168/jds.2009-2893.

R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. Vienna Austria. doi:10.1038/sj.hdy.6800737.

R Development Core Team, R. 2011. R: A Language and Environment for Statistical Computing.

Raven, L.-A., B.G. Cocks, and B.J. Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:62. doi:10.1186/1471-2164-15-62.

Rothammer, S., Seichter, D., Förster, M. and Medugorac, I., 2013. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. BMC Genomics, 14(1):908.

Rupp R., Hernandez A., Mallard B.A. 2007. Association of bovine leukocyte antigen (BoLA) DRB3.2 with immune response, mastitis, and production and type traits in Canadian Holsteins. J. Dairy Sci. 90:1029–1038. https://doi.org/10.3168/jds.S0022 -0302(07)71589-8

Sargolzaei, M., J.P. Chesnais, and F.S. Schenkel. 2011. FImpute - An efficient imputation algorithm for dairy cattle populations. J. Dairy Sci. 94:421.

Sasaki, O., M. Aihara, A. Nishiura, and H. Takeda. 2017. Genetic correlations between the cumulative pseudo-survival rate, milk yield, and somatic cell score during lactation in Holstein cattle in Japan using a random regression model. J. Dairy Sci. 100:7282–7294. doi:10.3168/jds.2016-12311.

Schäff, C., Börner, S., Hacke, S., Kautzsch, U., Sauerwein, H., Spachmann, S.K., Schweigel-Röntgen, M., Hammon, H.M. and Kuhla, B., 2013. Increased muscle fatty acid oxidation in dairy cows with intensive body fat mobilization during early lactation. J. Dairy Sci, 96(10), pp.6449-6460.

Sharif S., Mallard B.A., Wilkie B.N., Sargeant J.M., Scott H.M., Dekkers J.C., Leslie K.E. 1998 Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle. Anim. Genet. 29:185–193. https://doi.org/10.104 6/j.1365-2052.1998.00318.x

Sharma, B. S., I. Leyva, F. Schenkel, and N. Karrow. 2006. Association of toll-like receptor 4 polymorphisms with somatic cell score and lactation persistency in Holstein bulls. J. Dairy Sci. 89:3626–3635.

Silva, F.F., H.A. Mulder, E.F. Knol, M.S. Lopes, S.E.F. Guimarães, P.S. Lopes, P.K. Mathur, J.M.S. Viana, and J.W.M. Bastiaansen. 2014. Sire evaluation for total number born in pigs using a genomic reaction norms approach. J. Anim. Sci. 92:3825–3834. doi:10.2527/jas.2013-6486.

Song, M., He, Y., Zhou, H., Zhang, Y., Li, X. and Yu, Y., 2016. Combined analysis of DNA methylome and transcriptome reveal novel candidate genes with susceptibility to bovine Staphylococcus aureus subclinical mastitis. Scient. Rep., 6:29390.

Stanton, T.L., L.R. Jones, R.W. Everett, and S.D. Kachman. 1992. Estimating milk, fat, and protein lactation curves with a Test Day model. J. Dairy Sci. 75:1691–1700. doi:10.3168/jds.S0022-0302(92)77926-0.

Stear M.J., Bishop S.C., Mallard B.A., Raadsma H. 2001. The sustainability, feasibility and desirability of breeding livestock for disease resistance. Res. Vet. Sci. 71:1–7.

Strucken, E.M., R.H. Bortfeldt, D.J. de Koning, and G.A. Brockmann. 2012. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375–382. doi:10.1111/j.1365-2052.2011.02278.x.

Strucken, E.M., D.J. de Koning, S.A. Rahmatalla, and G.A. Brockmann. 2011. Lactation curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442–449. doi:10.3168/jds.2009-2932.

Strucken, E.M., Y.C.S.M. Laurenson, and G.A. Brockmann. 2015. Go with the flow - biology and genetics of the lactation cycle. Front. Genet. 6:118. doi:10.3389/fgene.2015.00118.

Sutherland, K. D., Lindeman, G. J., and Visvader, J. E. 2007. Knocking off SOCS genes in the mammary gland. Cell Cycle 6, 799–803. doi: 10.4161/cc.6.7.4037

Taylor, V.J., D.E. Beever, M.J. Bryant, and D.C. Wathes. 2003. Metabolic profiles and progesterone cycles in first lactation dairy cows. Theriogenology. doi:10.1016/S0093-691X(02)01225-6.

Thompson-Crispi, K.A., Sargolzaei, M., Ventura, R., Abo-Ismail, M., Miglior, F., Schenkel, F. and Mallard, B.A., 2014. A genome-wide association study of immune response traits in Canadian Holstein cattle. BMC Genomics, 15(1), p.559.

Tong, J.J., Thompson, I.M., Zhao, X. and Lacasse, P., 2018. Effect of 17β-estradiol on milk production, hormone secretion, and mammary gland gene expression in dairy cows. J. Dairy Sci., 101(3):2588-2601.

Tsuyuzaki, K., G. Morota, M. Ishii, T. Nakazato, S. Miyazaki, and I. Nikaido. 2015. MeSH ORA framework: R/Bioconductor packages to support MeSH over-representation analysis. BMC Bioinformatics 16:45. doi:10.1186/s12859-015-0453-z.

Usman, T., Yu, Y., Liu, C., Wang, X., Zhang, Q. and Wang, Y., 2014. Genetic effects of single nucleotide polymorphisms in JAK2 and STAT5A genes on susceptibility of Chinese Holsteins to mastitis. Mol. Biol. Rep., 41(12):8293-8301.

Viale, E., Tiezzi, F., Maretto, F., De Marchi, M., Penasa, M. and Cassandro, M., 2017. Association of candidate gene polymorphisms with milk technological traits, yield, composition, and somatic cell score in Italian Holstein-Friesian sires. J. Dairy Sci., 100(9):7271-7281.

Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Maki-Tanila A, et al. 2006. The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. Genetics. 173: 2151–2164.

https://doi.org/10.1534/genetics.105.046730 PMID: 16751675

Wang, H., I. Misztal, I. Aguilar, A. Legarra, R.L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken, and W.M. Muir. 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front. Genet. 5:134. doi:10.3389/fgene.2014.00134.

Wang, M., Zhou, Z., Khan, M.J., Gao, J. and Loor, J.J., 2015a. Clock circadian regulator (CLOCK) gene network expression patterns in bovine adipose, liver, and mammary gland at 3 time points during the transition from pregnancy into lactation. J. Dairy Sci., 98(7), pp.4601-4612.

Wang, J., Zhu, X., She, G., Kong, Y., Guo, Y., Wang, Z., Liu, G. and Zhao, B., 2018. Serum hepatokines in dairy cows: periparturient variation and changes in energy-related metabolic disorders. BMC Vet. Res., 14(1):236.

Wang, X., Ma, P., Liu, J., Zhang, Q., Zhang, Y., Ding, X., Jiang, L., Wang, Y., Zhang, Y., Sun, D. and Zhang, S., 2015b. Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. BMC Genetics, 16(1):111.

Weber, M., Locher, L., Huber, K., Rehage, J., Tienken, R., Meyer, U., Dänicke, S., Webb, L., Sauerwein, H. and Mielenz, M., 2016. Longitudinal changes in adipose tissue of dairy cows from late pregnancy to lactation. Part 2: The SIRT-PPARGC1A axis and its relationship with the adiponectin system. J. Dairy Sci., 99(2):1560-1570.

Wei, J., Geale, P.F., Sheehy, P.A. and Williamson, P., 2012. The impact of ABCG2 on bovine mammary epithelial cell proliferation. Animal Biotech. 23(3):221-224.

Wiggans, G.R., T.S. Sonstegard, P.M. VanRaden, L.K. Matukumalli, R.D. Schnabel, J.F. Taylor, F.S. Schenkel, and C.P. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92:3431–3436. doi:10.3168/JDS.2008-1758.

Wu X.X., Yang Z.P., Wang X.L., Mao Y.J., Li S.C., Shi X.K., Chen Y. 2010. Restriction fragment length polymorphism in the exon 2 of the BoLA-DRB3 gene in Chinese Holstein of the south China. J. Biomed. Sci. Eng. 3:221–225. https://doi.org/10.4236/jbise .2010.32030

Yamaji, D., R. Na, Y. Feuermann, S. Pechhold, W. Chen, G. W. Robinson, and L. Hennighausen. 2009. Development of mammary luminal progenitor cells is controlled by the transcription factor STAT5A. Genes Dev. 23:2382–2387.

Yoshida T., Furuta H., Kondo Y., Mukoyama H. 2012. Association of BoLA-DRB3 alleles with mastitis resistance and susceptibility in Japanese Holstein cows. Anim. Sci. J. 83:359–366. https://doi. org/10.1111/j.1740-0929.2011.00972.x

Zhao, F., McParland, S., Kearney, F., Du, L. and Berry, D.P., 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Gen. Sel. Evol., 47(1):49.

Zhou, H., Cheng, L., Gong, H., Byun, S.O., Edwards, G.R. and Hickford, J.G., 2017. Variation in the Toll-like Receptor 4 (TLR4) gene affects milk traits in dairy cows. J. Dairy Res., 84(4), pp.426-429.

Zielke, L.G., Bortfeldt, R.H., Reissmann, M., Tetens, J., Thaller, G. and Brockmann, G.A., 2013. Impact of variation at the FTO locus on milk fat yield in Holstein dairy cattle. PLoS One,

8(5).

9.8. Supplementary material

Table S1. Number of candidate genes found in each cluster (C1, C2, C3, and C4), for all traits and lactations of Ayrshire, Holstein and Jersey breeds.

[1]Trait	Ayrshire					Holstein					Jersey				
	C1	**C2**	**C3**	**C4**	**Total**	**C1**	**C2**	**C3**	**C4**	**Total**	**C1**	**C2**	**C3**	**C4**	**Total**
MY1	0	9	0	4	**13**	1	21	2	1	**25**	3	10	0	1	**14**
MY2	0	12	1	0	**13**	3	9	2	14	**28**	0	10	1	3	**14**
MY3	0	14	0	0	**14**	3	7	2	14	**26**	1	8	0	3	**12**
FY1	1	12	0	0	**13**	6	21	1	0	**28**	2	5	0	0	**7**
FY2	0	11	0	0	**11**	13	14	1	0	**28**	2	6	0	0	**8**
FY3	0	11	2	0	**13**	13	10	2	0	**25**	2	6	0	0	**8**
PY1	0	11	0	0	**11**	1	12	0	0	**13**	5	7	0	1	**13**
PY2	0	11	1	0	**12**	0	11	0	0	**11**	0	7	0	0	**7**
PY3	0	13	0	0	**13**	0	9	0	0	**9**	0	7	0	0	**7**
SCS1	5	3	1	0	**9**	9	1	14	0	**24**	1	4	1	0	**6**
SCS2	0	7	6	0	**13**	3	1	15	0	**19**	1	2	4	1	**8**
SCS3	0	7	6	0	**13**	1	0	13	0	**14**	1	2	5	1	**9**
Mean	**0.5**	**10.1**	**1.4**	**0.3**	**12.3**	**4.4**	**9. 7**	**4.3**	**2.4**	**20.8**	**1.5**	**6.2**	**0.9**	**0.8**	**9.4**

[1]Trait: milk (MY1, MY2, MY3), fat (FY1, FY2, FY3) and protein (PY1, PY2, PY3) yields, and somatic cell score (SCS1, SCS2, SCS3), in the first, second, and third lactations, respectively. Clusters were defined as: Cluster 1) constant single nucleotide polymorphism (SNP) effect over time; Cluster 2) SNP effect increases over time; Cluster 3) SNP effect decreases over time; Cluster 4) SNP effect increases in the middle of lactation; and, Cluster 5) SNP effect decreases in the middle of lactation. No candidate genes were found in Cluster 5.

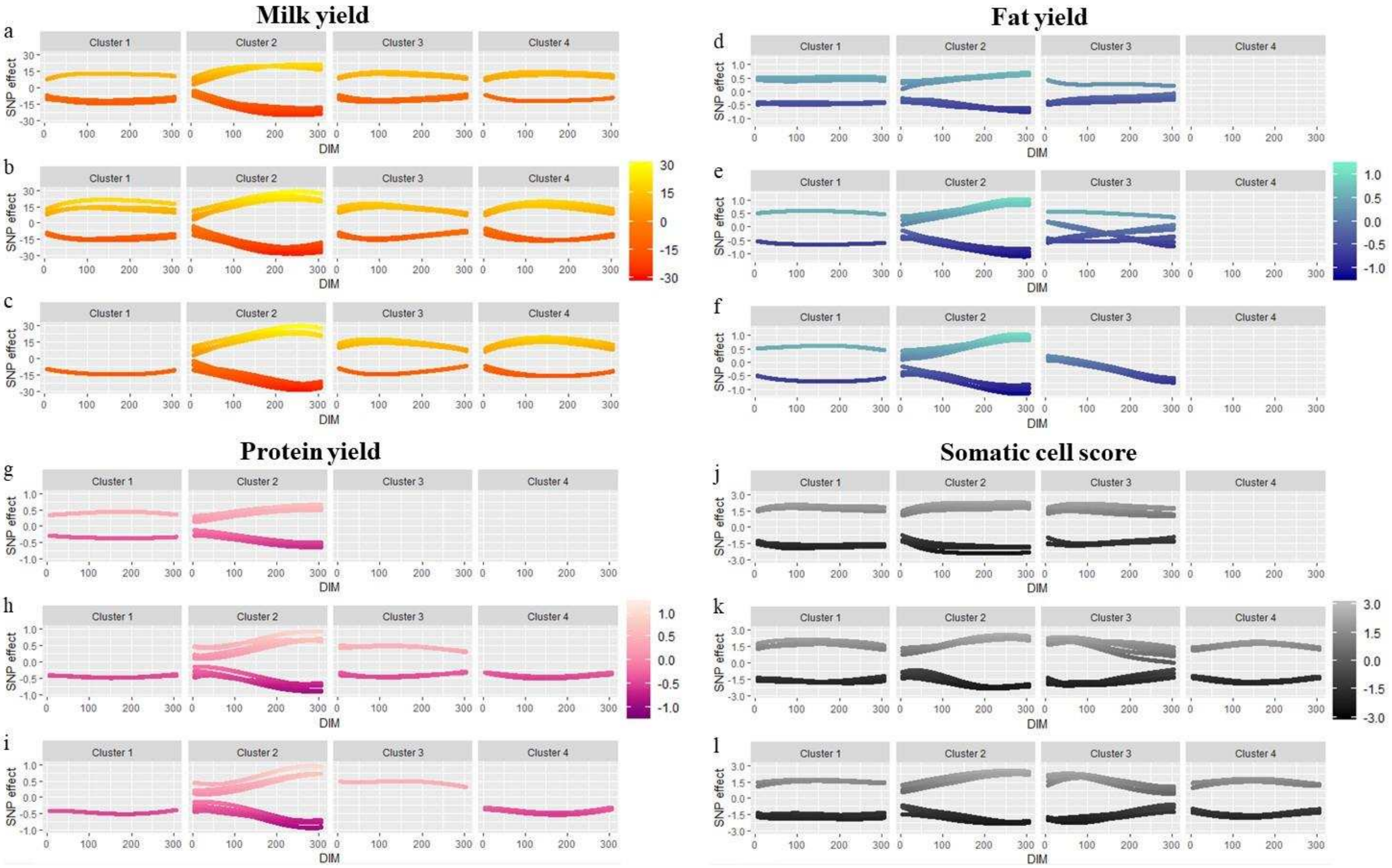

Figure S1. Single nucleotide polymorphism (SNP) effects pattern of the top 10 SNPs of each cluster over days in milk for milk (a, b, c), fat (d, e, f), and protein (g, h, i) yields, and somatic cell score (j, k, l) from the first, second and third lactation of the Ayrshire breed, respectively.

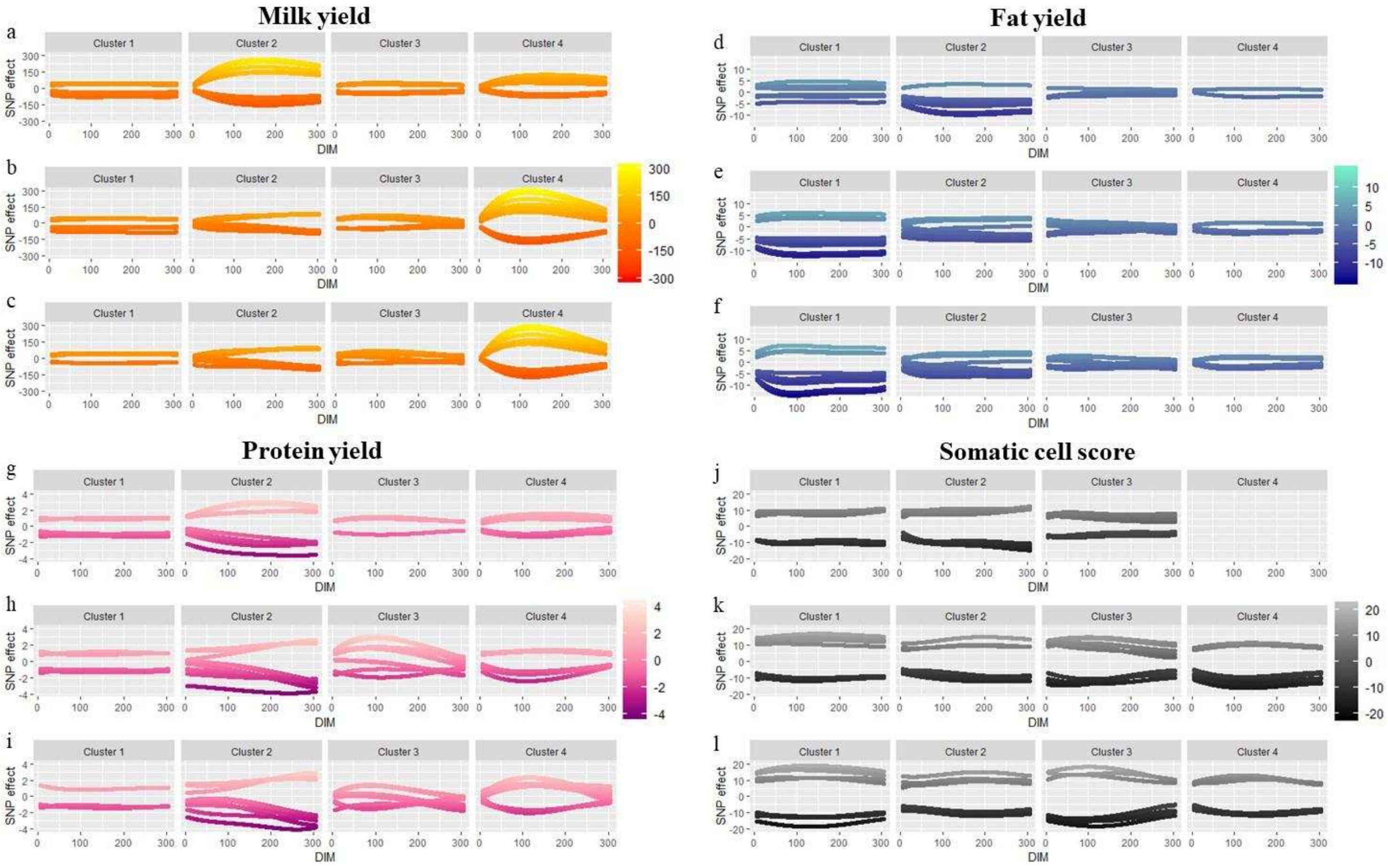

Figure S2. Single nucleotide polymorphism (SNP) effects pattern of the top 10 SNPs of each cluster over days in milk for milk (a, b, c), fat (d, e, f), and protein (g, h, i) yields, and somatic cell score (j, k, l) from the first, second and third lactation of the Holstein breed, respectively.

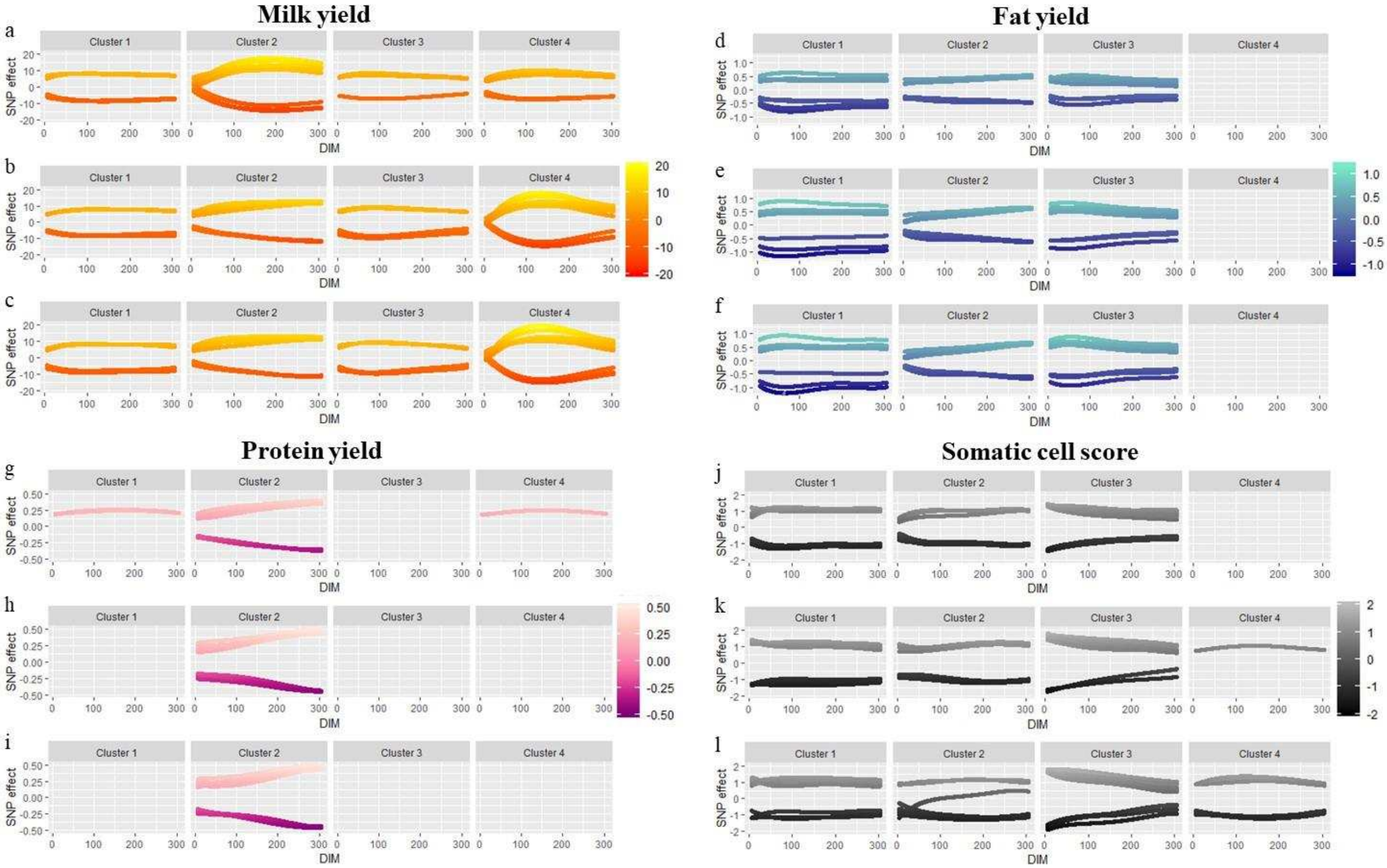

Figure S3. Single nucleotide polymorphism (SNP) effects pattern of the top 10 SNPs of each cluster over days in milk for milk (a, b, c), fat (d, e, f), and protein (g, h, i) yields, and somatic cell score (j, k, l) from the first, second and third lactation of the Jersey breed, respectively.

CHAPTER 10: General conclusions

Genetic or genomic predictions for various points in time can be simultaneously done using random regression models, which enable selection of breeding animals based on the complete pattern of the production curve. Combining different functions to model the fixed and random effects in a multiple-trait random regression model seems a feasible alternative for genetic modeling of lactation curves in dairy cattle (based on the goodness-of-fit of model, breeding values and variance component estimates), as shown here for Gyr cattle. In addition, the self-organizing map was used to define the classes of residual variances for all lactations underling the same multiple-trait random regression model.

Deregressed longitudinal estimated breeding values obtained using well established methods of deregression for non-longitudinal traits can be used for genomic prediction of longitudinal traits. For all longitudinal traits analyzed in this thesis, the VR and WG deregression methods yielded similar results, being, however, more accurate and less biased than the GR method. Based on our findings, removing the parent average and the genotyped daughters' average from the deregressed estimated breeding values can increase the reliability of genomic predictions. The new deregression method presented in this thesis showed to be a better alternative to generate deregressed estimated breeding values to be used in multiple-step genomic evaluations, increasing prediction reliability.

Regarding to the multiple-step genomic best linear unbiased prediction, reliability of genomic breeding values estimated using the random regression model was in general lower than the reliability from the accumulated 305-d model. Nevertheless, in general random regression model provided less biased genomic estimated breeding values compared to the accumulated 305-d model. More importantly, the use of random regression model in both steps of two-step genomic prediction produced fairly accurate genomic estimated breeding values over the entire lactation curve for milk, fat and protein yields, and somatic cell score. Thus, selecting young animals based on the pattern of genomically predicted lactation curves seems possible in Holstein cattle. The use of single-step genomic best linear unbiased predictor to predict genomic breeding values for longitudinal traits based on random regression models increased reliability and reduced bias of breeding values compared to traditional parentage average, in the Canadian Ayrshire, Holstein, and Jersey breeds. Moreover, similar genomic predictions in terms of daily validation reliability and bias were obtained when including all available genotypes or only genotypes from animals with accurate estimated breeding values in the analyses.

Different genomic regions associated with milk, fat and protein yields, and somatic cell score were identified for different lactation stages in Ayrshire, Holstein and Jersey breeds. Evidence of different set of candidate genes was found across lactation stages of the three breeds. For all breeds, the pattern of the effect of several SNPs related to the analyzed longitudinal traits changed over time. In addition, putative chromosomal regions and prospective candidate genes with different potential pattern of expression over time were identified for the longitudinal traits and dairy cattle breeds analyzed.

APPENDIX:

Additional research work and other academic experiences

1. Introduction

The main objective of this section is to briefly present the additional research work that I could be involved, as well as the other academic experiences that I had during my PhD program. All the research work and activities showed here were developed during my PhD program, which lasted from August/2015 to December/2018.

2. Academic experiences

2.1. Internships

a. Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada. **Supervisors:** Dr. Flavio Schramm Schenkel and Dr. Luiz Fernando Brito.
b. Department of Animal and Dairy Science, University of Georgia, Athens, Georgia, United States of America. **Supervisor:** Dr. Daniela Andressa Lino Lourenco.
c. Brazilian Association of Zebu Breeders (ABCZ), Uberaba, Minas Gerais, Brazil. **Supervisor:** Dr. Henrique Torres Ventura.

2.2. Participation in conferences and meetings

a. **2018:** Annual Meeting of the European Federation of Animal Science, Dubrovnik, Croatia.
b. **2018:** World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand.
c. **2018:** Dairy Cattle Breeding and Genetics Committee Meeting, Guelph, Canada.
d. **2018:** CCSI Genetics Committee Meeting, Guelph, Canada.
e. **2017:** VIII International Symposium in Genetics and Animal Breeding, Vicosa, Brazil.
f. **2017:** ADSA Annual Meeting, The American Dairy Science Association, Pittsburgh, United States of America.
g. **2017:** Dairy Cattle Breeding and Genetics Committee Meeting, Guelph, Canada.

2.3. Short courses

a. **2018:** Multivariate Statistics in Animal Breeding Using R, Guelph, Canada. Instructor: Dr. Nicolo Macciotta.

b. **2018:** Design of Breeding Programs with Genomic Selection. Guelph, Canada. Instructor: Dr. Jack Dekkers and Dr. Julius van der Werf.
c. **2017:** Advanced topics in quantitative genetics, genomics and breeding, University of Guelph, Guelph, Ontario, Canada. Instructors: Dr. Christian Maltecca and Dr. Francesco Tiezzi.
d. **2017:** Mixed Models, ADSA, Pittsburgh, United States of America. Instructors: Dr. Nora Bello, Dr. Nick Keuler and Dr. Kevin McCarter.
e. **2017:** Writing Tune-up: English Grammar, University of Guelph, Guelph, Ontario, Canada.
f. **2015:** How to build a genomic-enhanced selection program, Federal University of Vicosa, Vicosa, Brazil. Instructor: Egbert Knol.

2.4. Invited presentations and courses

a. **2018:** "Application of ssGBLUP using random regression models in the Ayrshire and Jersey breeds". Talk at the Annual Meeting of the European Federation of Animal Science, Dubrovnik, Croatia (speaker).
b. **2018: "**A genomic approach of lactation curves". Invited seminar, Federal University of Vicosa, Brazil (speaker).
c. **2017:** "Lactation curves: A genetic and genomic approach". Invited seminar at CGIL, University of Guelph, Canada (speaker).
d. **2016:** "Genomic selection: Theory and new perspectives". Invited seminar at III Genetics and Breeding Workshop, Vicosa, Brazil (speaker).
e. **2016:** "Accessing breeding value accuracies from Bayesian random regression models". Invited seminar at GDMA, Federal University of Vicosa, Brazil (speaker).
f. **2016:** "Understanding bulls catalogues". Invited seminar at Gezebu, Federal University of Vicosa, Brazil (speaker).
g. **2016:** "Dairy cattle course". Farmer's Week, extension event at Federal University of Vicosa, Brazil (main speaker).
h. **2016:** "Beef cattle course". Farmer's Week, extension event at Federal University of Vicosa, Brazil (main speaker).
i. **2016:** "Statistical Analysis using SAS". Course at Federal University of Vicosa, Brazil (teaching assistant).

2.5. Awards

a. **2018:** Annual Meeting of the European Federation of Animal Science (EAAP) 2018 Travel Award.

b. **2017:** Emerging Leaders in the Americas Program (ELAP) 2017 Travel Award.

3. Additional research work

3.1. Peer reviewed papers: first author

a. **Oliveira**, **H. R.**, L. F. Brito, J. Jamrozik, F. F. Silva, F. S. Schenkel. 2017. T48 Genomic prediction of lactation curves for milk, fat, protein, and somatic cell score in Canadian Jersey cattle. Journal of Dairy Science, 100 (Supplement 2).

b. **Oliveira, H. R.**, H. T. Ventura, E. V. Costa, M. A. Pereira, R. Veroneze, M. S. Duarte, O. H. G. B. D. Siqueira, F. F. Silva. 2017. Meta-Analysis of genetic-parameter estimates for reproduction, growth and carcass traits in Nellore cattle by using a random-effects model. Animal Production Science, 58:1575–1583. DOI:10.1071/AN16712.

3.2. Peer reviewed papers: co-author

a. Cruz, V. A. R., L. F. Brito, F. S. Schenkel, M. Jafarikia, **H. R. Oliveira**, Z. Feng. Strategies for within-litter selection of piglets using ultra-low density SNP panels. Livestock Science (submitted).

b. Paiva, J. T., M. D. V. Resende, R. T. Resende, **H. R. Oliveira**, H. T. Silva, G. C. Caetano, A. A. Calderano, P. S. Lopes, J. M. S. Viana, F. F. Silva. 2018. A note on transgenerational epigenetics affecting egg quality traits in meat-type quail. British Poultry Science, 21:1-5. DOI: 10.1080/00071668.2018.1514582.

c. Barili, L. D., N. M. D. Vale, J. E. D. S. Carneiro, **H. R. Oliveira,** R. P. Vianello, P. A. M. R. Valdisser, M. Nascimento. 2018. Genome prediction accuracy of common bean via Bayesian models. Ciência Rural, 48(8). DOI: 10.1590/0103-8478cr20170497.

d. Paiva, J. T., M. D. V. Resende, R. T. Resende, **H. R. Oliveira**, H. T. Silva, G. C. Caetano, P. S. Lopes, F. F. Silva. 2018. Transgenerational epigenetic variance for body weight in meat quails. Journal of Animal Breeding and Genetics, 135(3). DOI:10.1111/jbg.12329.

e. Stachowicz, K., L. F. Brito, **H. R. Oliveira**, S. P. Miller, F. S. Schenkel. 2018. Assessing genetic diversity of various Canadian sheep breeds through pedigree analyses. Canadian Journal of Animal Science. DOI:10.1139/CJAS-2017-0187.

f. Brito, L. F., F. G. Silva, **H. R. Oliveira**, N. O. Souza, G. C. Caetano, E. V. Costa, G. R. O. Menezes, A. L. P. Melo, M. T. Rodrigues, R. A. Torres. 2017. Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. Canadian Journal of Animal Science, 98(1):73–83. DOI:10.1139/CJAS-2017-0019.
g. Oliveira, L. T., C. M. Bonafé, F. F. Silva, H. T. Ventura, **H. R. Oliveira**, G. R. O. Menezes, M. D. V. Resende, J. M. S. Viana. 2017. Bayesian random regression threshold models for genetic evaluation of pregnancy probability in Red Sindhi heifers. Livestock Science, 202:166–170. DOI:10.1016/j.livsci.2017.06.005
h. Caetano, G. C., R. R. M., D. A. Silva, **H. R. Oliveira**, J. M. S. Viana, O. H. G. B. D. Siqueira, P. H. F. Freitas, F. F. Silva. 2017. Bayesian estimation of genetic parameters for individual feed conversion and body weight gain in meat quail. Livestock Science, 200:76-79. DOI:10.1016/j.livsci.2017.04.011
i. Siqueira, O. H. G. B. D., R. R. Mota, **H. R. Oliveira**, D. A. S. Duarte, L. S. Glória, M. T. Rodrigues, F. F. Silva. 2017. Genetic evaluation of lactation persistency and total milk yield in dairy goats. Livestock Research for Rural Development, 29(7).

3.3. Conferences' publications: first author

a. **Oliveira, H. R.**, L. F. Brito, D. A. L. Lourenco, Y. Matsuda, I. Misztal, S. Tsuruta, J. Jamrozik, F. F. Silva, F. S. Schenkel. 2018. Application of ssGBLUP using random regression models in the Ayrshire and Jersey breeds. In: Annual Meeting of the European Federation of Animal Science, Dubrovnik, Croatia.
b. **Oliveira, H. R.**, L. F. Brito, D. A. L. Lourenco, Y. Matsuda, I. Misztal, S. Tsuruta, J. Jamrozik, F. F. Silva, F. S. Schenkel. 2018. Effect of including only genotype of animals with accurate proofs in ssGBLUP using random regression. In: Annual Meeting of the European Federation of Animal Science, Dubrovnik, Croatia.
c. **Oliveira, H. R.**, F. F. Silva, L. F. Brito, J. Jamrozik, D. A. L. Lourenco, F. S. Schenkel. 2018. Genome-wide association study for milk, fat and protein yields in different lactation stages in Canadian Holstein and Jersey cattle. In: World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand.

d. **Oliveira, H. R.**, F. F. Silva, L. F. Brito, J. Jamrozik, F. S. Schenkel. 2017. Deregression methods for genomic selection. In: VIII International Symposium in Genetics and Animal Breeding, Vicosa, Brazil.

3.4. Conferences' publications: co-author

a. Brito L. F., F. S. Schenkel, **H. R. Oliveira,** A. Canovas, F. Miglior. 2018. Meta-analysis of heritability estimates for methane emission indicator traits in cattle and sheep. In: World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand.
b. Brito L. F., A. R. Guarini, **H. R. Oliveira**, Z. Karimi, K. Alves, F. Miglior, F. S. Schenkel. 2018. Maximizing rates of genetic progress in Canadian dairy cattle by using alternate genomic methodologies. In: Western Canadian Dairy Seminar - Advances in Dairy Technology, Alberta, Canada.
c. Fernandes, C.G.; J.T. Paiva; **H. R**. **Oliveira**; G.C. Caetano; R. Veroneze; F. F. Silva. 2018. Genetic parameters for egg number and eggshell weight in meet quails. In: XI Semana da Zootecnia, Rio Pomba, Brazil.
d. Paiva, J. T.; R. T. Resende; **H. R. Oliveira**; G. C. Caetano, G. C.; F. F. Silva; M. D. V. Resende. 2017. Transgenerational epigenetic heritability for body weight and mean egg weight in meat-type quails. In: XII Brazilian Symposium of Animal Breeding, Ribeirao Preto, Brazil.
e. Silva, H. T.; G. C. Caetano; D. A. Silva; J. T. Paiva; **H. R. Oliveira**; O. H. G. B. D. Siqueira. 2017. Genetic parameter estimates for body weight in meet quails using Bayesian inference. In: XII Brazilian Symposium of Animal Breeding, Ribeirao Preto, Brazil.
f. Freitas, P. H. F.; F. F.Silva; J. T. Paiva; H. T. Silva; **H. R. Oliveira**; M. T. Rodrigues. 2017. Mutiple-trait random regression models combining different functions via Bayesian analysis in dairy goats. In: Simpósio de Integracão Acadêmica, Vicosa, Brazil.
g. Freitas, P. H. F.; O. H. G. B. D. Siqueira; J. T. Paiva; **H. R. Oliveira**; H. T. Silva; F. F. Silva. 2016. Bayesian random regression models to describe milk, fat and protein yield in goats. In: VII International Symposium in Genetics and Animal Breeding, Vicosa, Brazil.
h. Freitas, P. H. F.; J. T. Paiva; O. H. G. B. D. Siqueira; **H. R. Oliveira**; L. P. Oliveira; F. F. Silva. 2016. Genetic parameters estimates in dairy goats using Bayesian random regression models. In: VII International Symposium in Genetics and Animal Breeding, Vicosa, Brazil.

i. Pereira, C. R. M.; F. F. Silva; **H. R. Oliveira**; P. H. F. Freitas; L. P. Oliveira; M. T. Rodrigues. 2016. Bayesian inference of genetic parameters for prolificity in dairy goats. In: Simpósio de Integracão Acadêmica, Vicosa, Brazil.
j. Silva, H. T.; F. F. Silva; V. S. Junqueira; **H. R. Oliveira**; N. O. Souza; M. T. Rodrigues. 2015. Bayesian comparison of random regression models for genetic evaluation of body weight in dairy goats. In: Simpósio de Integracão Acadêmica, Vicosa, Brazil.
k. Pereira, C. R. M.; F. F. Silva; J. T. Paiva; V. S. T. Corrêa; D. A. Silva; **H. R. Oliveira**. 2015. Influence of protein levels on carcass yield in meet quail. In: Simpósio de Integracão Acadêmica, Vicosa, Brazil.

3.5. Book chapters: co-author

a. Junqueira, V. S., N. O. Souza, D. B. D. Marques, F. A. Reimann, G. C. Caetano, **H. R. Oliveira**, L. L. Cardoso, L. L. Verardo, L. P. Silva, R. Veroneze, F. F. Silva. 2018. Chapter: Genomic Selection Applied to Animal Breeding. In: Genomic Selection Applied to Animal and Plant Breeding, 1st ed. Vicosa, 2018, v. 1. (printing).
b. Lopes, P. S., **H. R. Oliveira**, V. S. Junqueira, R. Veroneze, D. A. Silva, F. F. Silva. 2016. Chapter: Biometrics applied to animal breeding and its challenges. In: Biometric Challenges in Genetic Improvement, 1st ed. Vicosa, 2016, v. 1. 166p.