GENOMIC STUDIES OF LACTATION CURVE PARAMETERS IN HOLSTEIN CATTLE BASED ON DIFFERENT MODELS AND MILKING SYSTEMS DATA

PATRICK ROMUALD FOTSO KENMOGNE

2025

UNIVERSIDADE ESTADUAL DO SUDOESTE DA BAHIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

GENOMIC STUDIES OF LACTATION CURVE PARAMETERS IN HOLSTEIN CATTLE BASED ON DIFFERENT MODELS AND MILKING SYSTEMS DATA

Autor: Patrick Romuald Fotso Kenmogne
Orientador: Prof. DSc. Paulo Luiz Souza Carneiro

ITAPETINGA
BAHIA – BRASIL
Setembro de 2025

PATRICK ROMUALD FOTSO KENMOGNE

GENOMIC STUDIES OF LACTATION CURVE PARAMETERS IN HOLSTEIN CATTLE BASED ON DIFFERENT MODELS AND MILKING SYSTEMS DATA

Tese apresentada, como parte das exigências para obtenção do título de DOUTOR EM ZOOTECNIA, no Programa de Pós-Graduação em Zootecnia da Universidade Estadual do Sudoeste da Bahia.

Orientador: Prof. DSc. Paulo Luiz Souza Carneiro
Co-orientadores: Prof. PhD. Luiz Fernando Brito
Prof. DSc. Delvan Alves da Silva

ITAPETINGA
BAHIA – BRASIL
Setembro de 2025

636.214 Kenmogne, Patrick Romuald Fotso.
K43g Genomic studies of lactation curve parameters in Holstein cattle based on different models and milking systems data. / Patrick Romuald Fotso Kenmogne. – Itapetinga-BA: Universidade Estadual do Sudoeste da Bahia, 2025.
92 p.

Tese apresentada ao Programa de Pós-Graduação em Zootecnia da Universidade Estadual do Sudoeste da Bahia como requisito parcial para obtenção do título de Doutor em Zootecnia. Sob a orientação do Prof. DSc. Paulo Luiz Souza Carneiro; coorientadores Prof. PhD Luiz Fernando Brito; Prof. DSc. Delvan Alves da Silva.

1. Vacas leiteiras – Persistência na lactação. 2. Gado Holandês – Produção de leite. 3. Gado Holandês – Seleção genômica – Lactações estendidas. 4. Bovinos – Raça Holandesa – Sistema de ordenha. I. Universidade Estadual do Sudoeste da Bahia - Programa de Pós-Graduação em Zootecnia, C*ampus* de Itapetinga. II. Carneiro, Paulo Luiz Souza. III. Brito, Luiz Fernando. IV. Silva, Delvan Alves da. V. Título,

CDD(21): 636.214

Catalogação na Fonte:
Cláudia Aparecida de Souza – CRB 1014-5ª Região
Bibliotecária – UESB – Campus de Itapetinga-BA

Índice Sistemático para desdobramentos por Assunto:

1. Vacas leiteiras
2. Gado Holandês
3. Seleção genômica
4. Produção de leite
5. Persistência na lactação
6. Sistemas de ordenha

DECLARAÇÃO DE APROVAÇÃO

Título: "Genomic studies of lactation curve parameters in Holstein cattle based on different models and milking systems data"

Autor: Patrick Romuald Fotso Kenmogne
Orientador: Prof. DSc. Paulo Luiz Sousa Carneiro
Coorientadores: Prof. PhD. Luiz Fernando Brito
Prof. DSc. Delvan Alves da Silva

Aprovado como parte das exigências para obtenção do Título de DOUTOR EM ZOOTECNIA, ÁREA DE CONCENTRAÇÃO PRODUÇÃO DE RUMINANTES, pela Banca Examinadora:

gov.br Documento assinado digitalmente
PAULO LUIZ SOUZA CARNEIRO
Data: 20/09/2025 17:06:38-0300
Verifique em https://validar.iti.gov.br

Prof. DSc. Paulo Luiz Sousa Carneiro (Orientador) – UESB

DocuSigned by:
Luiz Fernando Brito 9/23/2025
FCACE46F89DA4C2...

Prof. PhD. Luiz Fernando Brito (Coorientador) - Purdue University, EUA

gov.br Documento assinado digitalmente
FABIANA MARTINS COSTA MAIA
Data: 22/09/2025 13:21:26-0300
Verifique em https://validar.iti.gov.br

Profa. DSc. Fabiana Martins Costa Maia - Universidade Tecnológica Federal do Paraná

gov.br Documento assinado digitalmente
LUIS FERNANDO BATISTA PINTO
Data: 22/09/2025 10:12:22-0300
Verifique em https://validar.iti.gov.br

Prof. DSc. Luís Fernando Batista Pinto - Universidade Federal da Bahia

gov.br Documento assinado digitalmente
JAIME ARAUJO COBUCI
Data: 22/09/2025 08:51:12-0300
Verifique em https://validar.iti.gov.br

Prof. DSc. Jaime Araújo Cobuci - Universidade Federal de Rio Grande do Sul

Data de realização: 19 de setembro de 2025.

“If you can not measure it, you can not improve it.”

Lord Kelvin

Ao Povo Brasileiro,
DEDICO.

AGRADECIMENTOS

À Deus, que permitiu que tudo pudesse ser realizado.

À Universidade Estadual do Sudoeste da Bahia (UESB), pela oportunidade que me foi dada de crescimento pessoal e intelectual, me possibilitando desenvolver este trabalho, que privilégio!

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

Ao Ministro do Emprego e Formação Profissional (MINEFOP) pela autorização de estudos concedida.

A Universidade Federal de Viçosa (UFV), pelo período de mobilidade acadêmica.

Aos meus orientadores, professores Paulo Carneiro, Luiz Brito e Delvan A. Silva, pela orientação, confiança, paciência, ensinamentos e por me permitir “abrir asas e voar”. Obrigado pelos exemplos e dedicação. Eternamente grato.

Ao grupo de estudos em Genômica Aplicada a Conservação e Melhoramento (GACOM) da UESB de Jequié, pela amizade, oportunidade de convivência, aprendizado e prontidão.

Ao Grupo de Discussão em Melhoramento Animal (GDMA) da UFV, por me receber, e pelos conhecimentos compartilhados.

Ao Brito’s Lab da Purdue University, por infraestrutura para análises e todo apoio.

Aos meus irmãos, irmãs, primos, primas e aos amigos por me darem o apoio, estímulo e me ajudaram chegar até aqui.

Ao Doctor Kenette Fru Mbangari pelo apoio e incentivo.

A todos que direta e indiretamente colaboraram para a realização desse trabralho, que só foi possível graças a cada um de vocês. Obrigado, Thank you, Gracias, ..., Merci.

BIOGRAFIA

PATRICK ROMUALD FOTSO KENMOGNE – nascido em 19 de abril de 1990, na cidade de Mbo-Bafoussam, Camarões, filho de Dominique Fotso (*in memoriam*) e Elisabeth Guiadem (*in memoriam*).

Iniciou em outubro de 2009 o curso de Agronomia e Ciência Agrícola na Universisté de Dschang em Camarões, obtendo os títulos de Bachelor of Science in Agriculture (setembro de 2012), Master of Engineering Degree with a major in Animal Production (janeiro de 2015) e Master of Sciences in Biotechnology and Animal Production (dezembro 2016).

Em abril de 2015, ingressou no Higher Technical Teachers' Training College (HTTTC), Kumba na University of Buea em Camarões, obtendo o título de Technical Teachers' Postgraduate Certificate of Education in Agriculture (janeiro de 2017).

Em junho de 2017, ingressou no Programa de Pós-graduação em Zootecnia, em nível de Mestrado, na área de concentração de Produção Animal, na Universidade Tecnológica Federal do Paraná, campus de Dois Vizinhos, concluindo em julho de 2018.

Trabalhou como resource person no Ministry of Agriculture and Rural Development (MINADER), professor no Ministry of Secondary Education (MINESEC), e professor visitante nas universidades de Dschang, Buea e Douala em Camarões.

Entre março de 2021 e março de 2022, foi chefe do Serviço de Formação, Orientação Profissional e Avaliações no Ministry of Employment and Vocational Training (MINEFOP), onde assumiu várias responsabilidades.

Em abril de 2022, iniciou o curso de Pós-Graduação em Zootecnia, em nível de Doutorado, área de concentração Produção de Ruminantes, na Universidade Estadual do Sudoeste da Bahia (UESB) na cidade de Itapetinga - BA, realizando estudos na linha de pesquisa em Genética e Melhoramento Animal. Realizou entre agosto de 2023 e agosto de 2024, uma mobilidade acadêmica na Universidade Federal de Viçosa. Em 29 de julho de 2024, submeteu-se a exame geral de qualificação. Em 19 de setembro de 2025, submeteu-se à banca de defesa da presente Tese de Doutorado para obtenção do título de DOUTOR EM ZOOTECNIA.

SUMÁRIO

Página

DEDICO ii
AGRADECIMENTOS iii
BIOGRAFIA iv
SUMÁRIO v
LISTA DE FIGURAS vii
LISTA DE TABELAS viii
RESUMO ix
ABSTRACT xii
I – REFERENCIAL TEÓRICO 1
1.1 Introduction 1
1.2 Milk production 3
1.2.1 Lactation curve 3
1.2.2 Lactation persistency 4
1.2.3 Physiology of lactation 6
1.3 Model and functions used to describe lactation curve in dairy farming 7
1.3.1 Random regression models 7
1.3.2 Wood, Wilmink, and Ali-Schaeffer functions 8
1.3.3 Legendre orthogonal polynomials 8
1.3.4 B-splines functions 10
1.4 Genomic background of lactation 10
1.5 Genotype-by-milking systems 11
1.6 References 13
II – OBJETIVOS 19
2.1 General objective 19
2.2 Specific objectives 19
III – CAPÍTULO I 20
ABSTRACT 21

INTRODUCTION 22
MATERIALS AND METHODS 23
RESULTS 33
DISCUSSION 38
CONCLUSIONS 44
NOTES 44
REFERENCES 45
SUPPLEMENTARY MATERIAL 51
IV – CAPÍTULO II 57
ABSTRACT 58
INTRODUCTION 59
MATERIALS AND METHODS 60
RESULTS 64
DISCUSSION 66
CONCLUSIONS 70
ACKNOWLEDGMENTS 70
TABLES AND FIGURE 71
REFERENCES 77
SUPPLEMENTAL MATERIAL 83
V – CAPÍTULO III 86
Abstract 86
Introduction 87
Materials & methods 87
Results 89
Discussion 90
Conclusions 91
Acknowledgments 91
References 91
VI – CONCLUSÕES FINAIS 92

LISTA DE FIGURAS

Página

I – REFERENCIAL TEÓRICO

Figure 1. World milk production by region 3

Figure 2. Standard shape of the lactation curve for Holstein cattle 4

III – CAPÍTULO I

Figure 1. Phenotypic average estimated in the first (MY1), second (MY2), and third (MY3) lactation for daily milk yield (DMY) in American Holstein cattle according to DIM, obtained based on random regression models using the LEG4 model with HE residual variance. 36

Figure 2. Additive genetic variance, permanent environmental variance, residual variance, and phenotypic variance estimated in the first (MY1), second (MY2), and third (MY3) lactation for daily milk yield (DMY) in American Holstein cattle according to DIM based on random regression models using the LEG4 model with HE residual variance. 37

IV – CAPÍTULO II

Figure 1. Heritability estimated using the Wood (WD), Wilmink (WL), and Ali-Schaeffer (AS) models. (a) represents heifers and (b) represents cows. 73

V – CAPÍTULO III

Figure 1. Phenotypic average estimated for daily milk yield (DMY). (a) represents heifers and (b) represents cows. 89

LISTA DE TABELAS

Página

I – REFERENCIAL TEÓRICO

Table 1. Measures of milk yield persistency (PSi)1 for ***i*** = 1 to 9..... 5

III – CAPÍTULO I

Table 1. Summary statistics for number of cows, number of records, average records per cow, number of contemporary groups (CG), average daily milk yield, minimum and maximum daily milk yield, average age at calving, minimum and maximum age at calving, and milking systems in American Holstein cattle[1] 24

Table 2. Measures of milk yield persistency (PS_i) for ***i*** = 1 to 10..... 30

Table 3. Number of estimated parameters (P) and deviance information criterion (DIC) values for the models evaluated fitting homogenous (HO) or heterogeneous (HE) residual variance for the lactation curves among the 3 lactations in American Holstein cattle. 35

Table 4. Estimates of genetic correlation between daily milk yield (DMY) across DIM in three lactations (MY1, MY2, and MY3) in American Holstein cattle based on random regression models (RRM) using the LEG4 model with HE residual variance. 39

Table 5. Heritability (diagonal) and genetic (above diagonal) correlation estimates for different lactation persistency measures (PSi) and 305-d MY (MY305) within lactations (MY1, MY2, and MY3) in American Holstein cattle based on random regression models (RRM) using LEG4 model with HE residual variance. 40

IV – CAPÍTULO II

Table 1. Descriptive statistics for number of animals, number of records, average records per animal, CG number, average daily milk yield, minimum and maximum

daily milk yield, average age at calving, minimum and maximum age at calving, and milking systems in American Holstein cattle. 71

Table 2. Mean ± standard error (SE), minimum, and maximum values for each parameter (P) from the Wood, Wilmink, and Ali-Schaeffer models for daily milk yield (DMY) in American Holstein cattle. 72

Table 3. Summary of the most significant SNP associated with each lactation curve parameters in American Holstein. The complete list is presented in the Supplemental Tables S1 and S2. 74

Table 4A. Summary of the most significant terms in the functional annotation analysis for group of genes identified in American Holstein heifers. The full list is presented in the Supplemental Tables S3, S4, and S5. 75

Table 4B. Summary of the most significant terms in the functional annotation analysis for group of genes identified in American Holstein cows. The full list is presented in the Supplemental Tables S6, S7, and S8. 76

V – CAPÍTULO III

Table 1. Posterior means (SD in parentheses) of variance components, heritability, repeatability, and genetic correlation (milking robots x milking parlors) for daily milk yield (DMY) in American Holstein cattle. 90

RESUMO

FOTSO KENMOGNE, Patrick Romuald. **Estudos genômicos dos parâmetros da curva de lactação em bovinos da raça Holandesa com base em diferentes modelos e sistemas de ordenha**. Itapetinga, BA: UESB, 2025. 92p. Tese. (Doutorado em Zootecnia, Área de Concentração em Produção de Ruminantes).*

A persistência na lactação pode ser definida como a capacidade de uma vaca manter a produção de leite em um nível elevado depois de atingir o pico de produção. Essa característica é de interesse particular na pecuária leiteira, pois representa um fator econômico importante, impactando a fertilidade, a saúde e os custos com alimentação. A seleção genômica para maior persistência na lactação pode, indiretamente, aumentar a produção total de leite e reduzir o risco de doenças metabólicas, infecciosas ou problemas de fertilidade. No entanto, até o momento, a persistência na lactação tem sido definida apenas com o uso de modelos genéticos tradicionais (pedigree-BLUP). Nesse contexto, o objetivo foi otimizar a persistência na lactação em vacas leiteiras de alta produção, o que permitiria uma transição da lactação tradicional para a lactação estendida, com impacto significativo no bem-estar animal, produtividade e sustentabilidade da indústria leiteira. Primeiramente, foram estimados parâmetros genéticos, usando dados genômicos, para a produção diária de leite e persistência na lactação para as primeiras três lactações em gado Holandês dos EUA. A variância residual heterogênea com dez (10) classes foram definidas. Vinte e seis (26) modelos foram comparados para cada lactação, e os modelos foram selecionados com base na qualidade de ajuste e parcimônia, estimativas dos componentes de variância, herdabilidades, correlações, credibilidade e interpretação dos resultados, bem como na demanda e tempo computacional. Levando em consideração tais critérios, os polinômios ortogonais de Legendre de ordem quarto foram usados. As estimativas de herdabilidade, ao longo dos dias em lactação, variaram de 0,12 a 0,22 para primeira lactação; de 0,11 a 0,19 para a segunda lactação; e de 0,03 a 0,15 para a terceira lactação. Para a persistência na lactação, as estimativas de herdabilidade variaram de 0,05 a 0,18 para primeira lactação; de 0,06 a 0,20 para segunda lactação; e de 0,05 a 0,16 para terceira lactação. No geral, os valores das estimativas de herdabilidade variaram de baixa

* Orientador: Paulo Luiz Souza Carneiro, DSc. UESB e Co-orientadores: Luiz Fernando Brito, PhD. Purdue University e Delvan Alves da Silva, DSc. UFV.

à moderada magnitude. Correlações genéticas nulas a moderadas foram observadas entre a persistência na lactação e a produção de leite em 305 dias, indicando uma associação fraca (que não deve ser negligenciada) entre persistência da produção de leite e produção de leite total em 305 dias, o que é desejável. Dez medidas de persistência na lactação foram avaliadas usando valores genéticos genômicos. A terceira medida de persistência (calculada subtraindo-se a área sob a curva da lactação no terço final da área sob a curva no terço inicial da lactação) é recomendada como critério de seleção, pois essa medida apresentou herdabilidade moderada (0,18 para primeira lactação, 0,20 para segunda lactação e 0,15 para terceira lactação) e correlação genética baixa a moderada (0,38 para primeira lactação, 0,23 para segunda lactação e 0,20 para terceira lactação) com a produção de leite em 305 dias. O segundo estudo teve como objetivo realizar estudos de associação genômica ampla (GWAS) e análises de enriquecimento funcional para os parâmetros da curva de lactação. Os efeitos dos SNPs foram estimados para os coeficientes de regressão aleatória utilizando o método GBLUP. Os marcadores significativos foram identificados com base na correção de Bonferroni estrita e na correção de Bonferroni modificada, esta última baseada no número de segmentos cromossômicos independentes. Foram reportados os genes e QTL, localizados até 100 kb em torno dos SNPs significativos. As análises de enriquecimento funcional dos genes identificados para cada parâmetro da curva de lactação foram realizadas. Em novilhas, foram identificados 81, 128 e 196 SNPs significativos para os parâmetros dos modelos de Wood (WD), Wilmink (WL) e Ali-Schaeffer (AS), respectivamente; enquanto em vacas, foram identificados 120, 125 e 128 SNPs significativos para os mesmos parâmetros. Os SNPs significativos estavam localizados nos cromossomos BTA1, BTA2, BTA3, BTA5, BTA6, BTA7, BTA8, BTA9, BTA11, BTA13, BTA14, BTA15, BTA17, BTA18, BTA19, BTA20, BTA23 e BTA29. Uma região genômica (BTA14: 1.801.116 pb) foi comum a todas as funções paramétricas (WD, WL e AS). As regiões genômicas localizadas nos cromossomos BTA15 e BTA19 foram exclusivas dos parâmetros WL; enquanto aquelas localizadas nos cromossomos BTA3 e BTA20 foram exclusivas dos parâmetros AS. Dez genes candidatos (*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1* e *TSNARE1*) foram comuns a todas as funções paramétricas. Nenhum *Gene Ontology* significativo foi encontrado para o parâmetro *c* (WD) em novilhas, nem para os parâmetros *d* e *f* (AS) em vacas. O terceiro estudo teve como objetivo avaliar a interação genótipo x tipo de ordenha (robôs x salas tipo carrossel) para a produção diária de leite. Assim, foram estimados componentes de variância e

* Orientador: Paulo Luiz Souza Carneiro, DSc. UESB e Co-orientadores: Luiz Fernando Brito, PhD. Purdue University e Delvan Alves da Silva, DSc. UFV.

parâmetros genéticos, usando dados genômicas, para a produção diária de leite em sistemas de ordenha automática (robôs) e salas de ordenha (tipo carrossel), utilizando um modelo bicaracterísticas. As correlações genômicas entre os tipos de ordenha (robôs x salas tipo carrossel) foram altas (> 0,80), indicando ausência de respostas genéticas diferenciais entre os sistemas de ordenha. No entanto, as diferenças nas herdabilidades entre ordenha por robôs e ordenha em salas tipo carrossel foram consideráveis. Este estudo apresentou parâmetros genéticos, usando dados genômicas, para a produção diária de leite e dez medidas de persistência na lactação, bem como a arquitetura genômica dos parâmetros da curva de lactação e a interação genótipo x tipo de ordenha (robôs x salas tipo carrossel) para a produção diária de leite em gado Holandês dos EUA. Os resultados podem possibilitar a implementação bem-sucedida da seleção genômica para maior persistência na lactação e, assim, aumentar a lucratividade da indústria leiteira por meio de lactações estendidas.

Palavras-chave: interação genótipo x tipo de ordenha, polinômios ortogonais de Legendre, produção diária de leite, funções paramétricas, coeficientes de regressão aleatória

* Orientador: Paulo Luiz Souza Carneiro, DSc. UESB e Co-orientadores: Luiz Fernando Brito, PhD. Purdue University e Delvan Alves da Silva, DSc. UFV.

ABSTRACT

FOTSO KENMOGNE, Patrick Romuald. **Genomic studies of lactation curve parameters in Holstein cattle based on different models and milking systems data**. Itapetinga, BA: UESB, 2025. 92p. Thesis. (Doctorate in Animal Science, Area of Concentration in Ruminant Production).*

Lactation persistency can be defined as the ability of a cow to maintain milk production at a high level after reaching the milk production peak. This trait is of particular interest in dairy farming because it is an important economic factor, impacting fertility, health, and feed costs. Genomic selection for increased lactation persistency could indirectly raise total milk yield and reduce the risk of metabolic and infectious diseases or fertility issues. However, lactation persistency has so far been defined only using traditional genetic models (pedigree-BLUP). In this context, the objective was to optimize lactation persistency in high-producing dairy cows, which would enable to shift from traditional lactation to extended lactation and highly impact animal welfare, and sustainability of the dairy industry. First, a genomic-based genetic parameters for daily milk yield and lactation persistency were estimated for the first 3 lactations in American Holstein cattle. In this study, 26 models were compared for each lactation and models were selected based on goodness of fit and parsimony, estimates of variance components, heritabilities, correlations, credibility and interpretation of the results, and computational demand and time. Consequently, the model using fourth-order Legendre orthogonal polynomials and 10 classes of heterogeneous residual variance was considered the optimal model. The daily heritability estimates across days in milk ranged from 0.12 to 0.22 for milk yield in parity 1 (MY1), 0.11 to 0.19 for milk yield in parity 2 (MY2), and from 0.03 to 0.15 for milk yield in parity 3 (MY3). For lactation persistency, the heritability estimates ranged from 0.05 to 0.18 for MY1, from 0.06 to 0.20 for MY2, and from 0.05 to 0.16 for MY3. Overall, the heritability estimate values were of low to moderate magnitude. In general, null to moderate genetic correlations were observed between lactation persistency and 305-day milk yield, indicating weak association between lactation persistency and milk yield, which is desirable (but should not be neglected). Ten measures of lactation

* Orientador: Paulo Luiz Souza Carneiro, DSc. UESB e Co-orientadores: Luiz Fernando Brito, PhD. Purdue University e Delvan Alves da Silva, DSc. UFV.

persistency were evaluated using genomic breeding values. The third lactation persistency measurement (derived by subtracting the area under the curve of lactation in the final third from the area under the curve in the initial third of lactation) is recommended as the selection criterion, as these measures showed a moderate heritability (0.18 for MY1, 0.20 for MY2, and 0.15 for MY3) and a low to moderate genetic correlation (0.38 for MY1, 0.23 for MY2, and 0.20 for MY3) with 305-day milk yield. The second study aimed to perform genome-wide association studies and functional enrichment analyses for lactation curve parameters. The SNP effects were estimated for the random regression coefficients using GBLUP method. Significant markers were identified using the strict Bonferroni correction and the modified Bonferroni correction, based on the number of independent chromosomal segments. Genes and QTL located up to 100 kb upstream or downstream of the significant SNP were reported, and functional enrichment analyses of the candidate genes identified for each lactation curve parameters were performed. In heifers, 81, 128, and 196 significant SNP were identified for the Wood (WD), Wilmink (WL), and Ali-Schaeffer (AS) parameters, respectively; whereas in cows, 120, 125, and 128 significant SNP were identified for the same parameters. The significant SNP were located on chromosomes BTA1, BTA2, BTA3, BTA5, BTA6, BTA7, BTA8, BTA9, BTA11, BTA13, BTA14, BTA15, BTA17, BTA18, BTA19, BTA20, BTA23, and BTA29. One genomic region (BTA14: 1,801,116 bp) was common to all the parametric functions (WD, WL, and AS). The genomic regions located on BTA15 and BTA19 were unique to the WL parameters; whereas those located on BTA3 and BTA20 were unique to the AS parameters. Ten candidate genes (*LY6D*, *LYNX1*, *LYPD2*, *SLURP1*, *THEM6*, *PSCA*, *JRK*, *ARC*, *ADGRB1*, and *TSNARE1*) were common to all the parametric functions. No significant gene ontology terms were found for the parameter *c* (WD), for heifers and for the parameters *d* and *f* (AS), for cows. The third study aimed to evaluate genotype-by-milking type interaction (milking robots x milking parlors) for daily milk yield. We estimated genomic-based variance components and genetic parameters for daily milk yield (milking robots x milking parlors) based on a bi-trait genomic model. The genomic-based genetic correlations between milking types (milking robots x milking parlors) were high (> 0.80), indicating no potential differential genetic responses across milking systems. However, the differences in heritabilities between milking robots and milking parlors were considerable. This study reported genomic-based genetic parameters for daily milk yield and various lactation persistency traits, as well as the genomic background of lactation curve parameters, and the genotype-

* Orientador: Paulo Luiz Souza Carneiro, DSc. UESB e Co-orientadores: Luiz Fernando Brito, PhD. Purdue University e Delvan Alves da Silva, DSc. UFV.

by-milking type interaction (milking robots x milking parlors) for daily milk yield in American Holstein cattle. These findings could enable the successful implementation of genomic selection for greater lactation persistency and, therefore, enhance the profitability of the dairy industry through extended lactations.

Key words: daily milk yield, genotype-by-milking type interaction, Legendre orthogonal polynomials, parametric functions, random regression coefficients

---

***** Orientador: Paulo Luiz Souza Carneiro, DSc. UESB e Co-orientadores: Luiz Fernando Brito, PhD. Purdue University e Delvan Alves da Silva, DSc. UFV.

I – REFERENCIAL TEÓRICO

1.1 Introduction

The world population growth is greater than the increase in agricultural production and food demand has been projected to increase by 50% by 2050 (FAO, 2019). This imbalance between agricultural production and that of population growth is responsible for the deficit in food products, especially animal protein. This protein deficit, which is becoming common in some parts of the world, is the root cause of malnutrition. In the livestock sector, animal products supply human diets with 39% of protein and 18% of energy requirements (FAO, 2018), making cattle production the first source of milk in the world. Its nutritional benefits play an important role in human nutrition (Wu et al., 2023).

In fact, milk is one of the most widely produced beverages in the developing world and a valuable source of essential nutrients, including protein, calcium, and vitamins (FAO, 2018). Milk and milk products are consumed in various forms, including fresh, pasteurized, and fermented products such as cheese, yogurt, and butter. Milk production and consumption are crucial aspects of the global food system, and their impact is felt across economic, social, and environmental spheres. Therefore, genetic and genomic improvement of cattle populations is a promising strategy to address the challenges of the global milk supply.

In addition to genetic and genomic improvement, dairy producers often make decisions to maximize total milk yield per cow throughout extended lactations (Lehmann et al., 2016). Extended lactations in dairy production breaks with the tradition of aiming for one-year calving intervals by deliberately delaying the start of the next reproductive cycle. In most dairy systems, one-year calving intervals have been the norm since the beginning of artificial selection and the concept of extended lactation emerged in an attempt to alleviate fertility issues and problems with drying off high-yielding cows (Sehested et al., 2019) and also to take advantage of precision management, precision health and care management, and wide adoption of reproductive biotechnologies as sperm sexing, artificial insemination, and embryo transfer (Brito et al., 2021).

Therefore, the full capacity of the animals is well exploited as they are not dried off at a time when they are still producing large amounts of milk (Knight, 2005). Switching a herd to extended lactation management decrease the number of health risks periods associated with calving (Ingvartsen, 2006), reduces the number of calving per year, and this alters herd composition, feed use and production (Lehmann et al., 2016). Delaying insemination after the peak yield would increase pregnancy rates (Schindler et al., 1991) as a result of sufficient time to recover from the effects of early lactation stress and calving (Borman et al., 2004), decrease in number of inseminations per pregnancy and the frequency of anestrus treatment (Larsson and Berglund, 2000). The health risks such as mastitis, lameness, and metabolic stress at peak lactation, from which 60% occur around the calving period (Ingvartsen, 2006), will be reduced when delaying rebreeding. On the other hand, a random selection of cows results in an average loss of milk, whereas selecting the most suitable cows allows milk production to be maintained for a longer duration throughout the lactation period. This persistency in milk production is the key to increase milk yield and ultimately making extended lactations profitable (Sorensen et al., 2008).

Lactation persistency, which is known as the number of days during which the level of constant yield is maintained (Grossman et al., 1999), is of particular interest in dairy farming. This is because it is an important economic trait in dairy cattle, impacting fertility, health, and feed costs (Dekkers et al., 1998). Furthermore, the shape of the lactation curve, which is determined by genetic and environmental factors, also plays a crucial role in understanding lactation persistency. Several studies (Bertilsson et al., 1997; Kolver et al., 2007; Lehmann et al., 2016; Pedrosa et al., 2021; Pedrosa et al., 2023) show that there is a considerable variation in individual milk production during lactation. Abdelsayed et al. (2015) noted this as well, and further argued for using both phenotypic and genomic markers to manage high-persistent cows. Based on this, it is possible to know already in early life which of these cows will end up being among the higher or the lower yielding cows during the coming lactation. Cows with more persistent lactations tend to be more productive and profitable than average lactation cows (Cole and Null, 2009). Therefore, persistency is essential if farmers are to strive for an extended lactation where cows are milked for more than 305 days.

To our best knowledge, little is known about lactation persistency using both phenotypic and genomic information, longitudinal daily milk yield, and factors that can predict cows' ability to maintain a high level of milk production.

1.2 Milk production

According to FAO (2023) the world milk production reached around 937.3 million tonnes in 2022, increased by 0.6 percent and Asia is the world's largest milk-producing region (Figure 1). The dairy sector is one of the most important in the agroindustry in the world. This sector involves the production and processing of milk, whether cow, goat, buffalo, sheep or camel. Milk can be consumed and processed into different types of derivatives, such as: cheese, butter, cream, yogurt, among others. Dairy products have several nutrients, playing a fundamental role in food security, nutrition, and well-being.

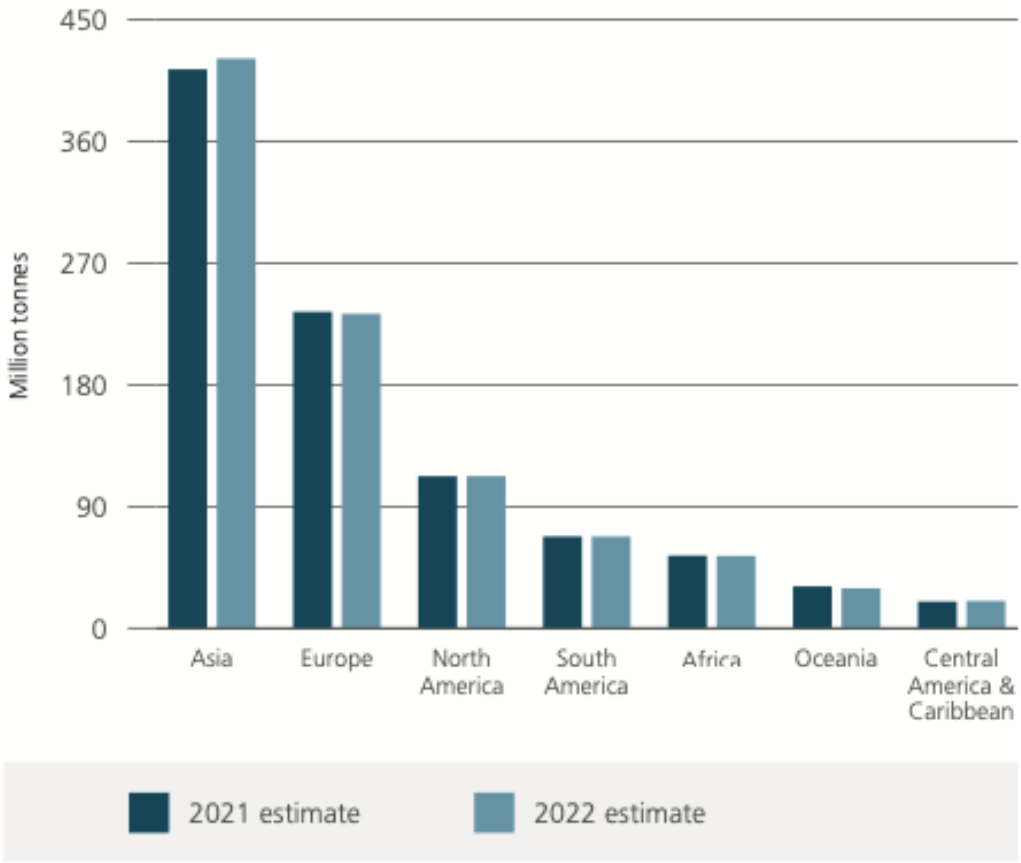

Figure 1. World milk production by region

Source: FAO, 2023.

Milk production as a function of the lactation time is represented by a lactation curve.

1.2.1 Lactation curve

According to Nicolò et al. (2011) lactation curve is the mathematical representation of milk production during the lactation period. At cow level, lactation

curve modelling is of help for monitoring individual yields for diet planning, early detection of diseases, and for selecting animals. The standard shape of the lactation curve (Figure 2) is characterized by a first ascending phase from parturition till a maximum, the lactation peak, followed by a second declining slope that ends with the dry off of the animal.

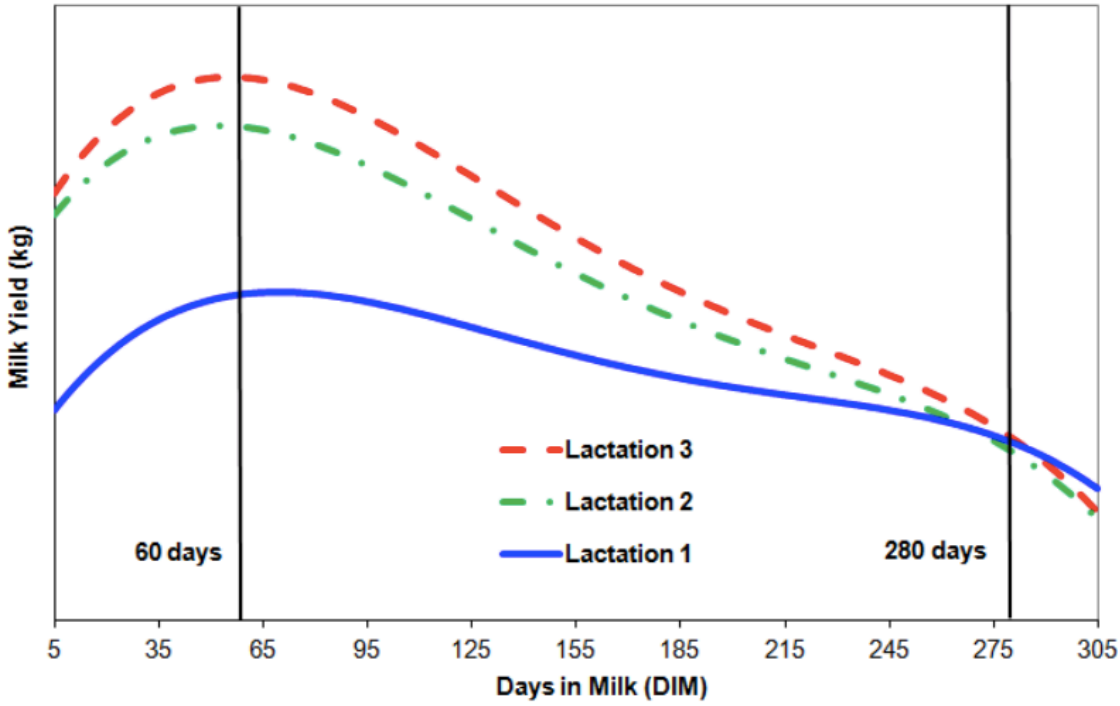

Figure 2. Standard shape of the lactation curve for Holstein cattle

Source: CDN, 2017.

Most of the mathematical functions proposed to fit lactation patterns in dairy cattle are mainly aimed at describing the phenomenon. The mathematical tool used in this approach is represented by an analytical function of time:

$$y_t = (f_t)\,, [1]$$

where y_t is the daily milk production recorded at time t. The Wood incomplete gamma function (Wood, 1967) is probably the most popular empirical model of the lactation curve. Basic traits of the curve are represented by the rate of increase of milk production in the first phase, time at peak occurrence, peak production, the rate of decline of milk yield in the second phase, and lactation persistency.

1.2.2 Lactation persistency

Persistency of yields during the lactation is an important lactation curve parameter. It can be described as the ability of the cow to maintain production following peak yield, and have an economic interest because of its impact on fertility, health, and

feed costs (Dekkers et al., 1998; Cobuci et al., 2003). Dekkers et al. (1998) found that the economic value of persistency nearly tripled when the average calving interval was 13 months, and consideration of health and reproductive costs would further increase this value. Additionally, persistency had a substantial impact on optimal insemination decisions. Definition of persistency can be extended to fat and protein yields, and can be measured in many ways (Table 1). In Canada the measure of persistency **(PS_i)** is essentially the breeding value for milk yield on day 280 minus the breeding value for milk yield on day 60 (PS_1).

Table 1. Measures of milk yield persistency (PSi)1 for $\boldsymbol{i}$ = 1 to 9

Measures*	**Authors**
$\boldsymbol{PS_1 = (EBV_{280} - EBV_{60})}$	Jamrozik et al. (1997)
$\boldsymbol{PS_2 = \left(\sum_{t=106}^{205} EVB_t - \sum_{t=6}^{105} EVB_t\right)}$	Jakobsen et al. (2002)
$\boldsymbol{PS_3 = \left(\sum_{t=206}^{305} EVB_t - \sum_{t=6}^{105} EVB_t\right)}$	Jakobsen et al. (2002)
$\boldsymbol{PS_4 = \sum_{t=61}^{280} (EBV_t - EBV_{60})}$	Jamrozik et al. (1997)
$\boldsymbol{PS_5 = \sum_{t=60}^{279} (EBV_t - EBV_{280})}$	Jakobsen et al. (2002)
$\boldsymbol{PS_6 = (EBV_{290} - EBV_{90})}$	Cobuci et al. (2004)
$\boldsymbol{PS_7 = \left(\frac{1}{51}\sum_{t=255}^{305} EVB_t - \frac{1}{21}\sum_{t=50}^{70} EVB_t\right)}$	Kistemaker (2003)
$\boldsymbol{PS_8 = \left(\sum_{t=101}^{300} EVB_t - (300 - 100)\, EBV_{100}\right)}$	Pösö (2003) in Kistemaker (2003)
$\boldsymbol{PS_9 = \left(\sum_{t=61}^{305} EVB_t - (305 - 60)\, EBV_{60}\right)}$	de Ross et al. (2001)

[1]Smaller values of PS_1, PS_2, PS_3, PS_4 and PS_6 indicates a high persistency of lactation, while a large values of PS_5 indicates a high level of persistency.

The problem in measuring persistency is to express the shape of the lactation curve by a single term. Many attempts have been made to find a scale along which this can be done, and there are two types of measures in the literature (Jamrozik et al., 1997; Jakobsen et al., 2002). The first type uses ratios between different parts of the lactation, and the second uses more or less complicated mathematical functions (Wood, 1967; Wilmink, 1987; Ali and Schaeffer, 1987).

Kistemaker (2003) compared three measures of persistency (PS_7, PS_8, and PS_9) and conclude that the performance of PS_8 and PS_7 was very similar with slightly better properties for PS_7 in most cases.

Cobuci et al. (2004), Cobuci and Costa (2012) working with Holstein cows in Brazil on persistency of lactation using random regression models, obtained heritability values for persistency ranging from 0.04 to 0.43. The authors explain that choosing animals based on their ability to persist, with the goal of altering the lactation curve, can be effective.

In Holstein dairy cattle, Jakobsen et al. (2002) and Cobuci et al. (2004) reported genetic correlation values for persistency measures (PS_1 to PS_5) within the range of -0.91 to 0.94. The authors concluded that genetic correlations between individual persistency measures, as well as between persistency measures and production, are highly dependent on how persistency is defined. This suggests that the specific method used to quantify persistency can significantly impact the observed genetic correlations.

Many factors may influence the total milk yield of a single lactation, but the general shape of the curve, remains substantially unchanged (Wood, 1967). The declining production after peak yield is a result of a changing imbalance between cell proliferation and cell apoptosis (Stefanon et al., 2002). High lactation persistency is associated with a slow rate of decline in production, whereas low lactation persistency is associated with a rapid rate of decline.

1.2.3 Physiology of lactation

At the level of the mammary gland, milk production capacity is determined by the number of mammary epithelial cells and their metabolic activity, with the number of cells being considered the primary determining factor (Capuco and Choudhary, 2020). The net loss of mammary epithelial cells due to cell death that occurs throughout a cow's lactation is believed to be the primary reason for the decline in milk production. By 305 days in milk (**DIM**), the net number of mammary epithelial cells is approximately 50% of the

level present in early lactation. Although there are half the number of cells in the gland in late lactation, cell renewal continually occurs throughout lactation. By the end of lactation, most cells in the gland were found to be newly formed after calving (Capuco and Choudhary, 2020). The limitation in forming new cells is believed to be governed by the number of mammary stem and progenitor cells and their proliferation capacity. Thus, identifying traits that reflect the factors governing mammary stem and progenitor cell numbers and renewal will enable their selection. Identifying cows with traits related to high lactation persistency will likely contribute to the continued growth in production efficiency of the modern dairy cow, which currently shows no clear limits to her production capacity (Pulina et al., 2020). The use of model and functions for dairy cows is important, since it allows the study of the genetic relations between milk yield and its components during lactation.

1.3 Model and functions used to describe lactation curve in dairy farming

1.3.1 Random regression models

Random regression model (**RRM**), can be used for traits like daily milk yields, which is expressed repeatedly. In the RRM, a different regression coefficient for each animal is estimated. The RRM provide robust genetic modelling of time-specific effects for longitudinal traits, with key advantages compared with other statistical models, such as: estimation of genetic parameters, including genetic correlation among traits, for the time-series within the range of data collection (Schaeffer, 2004); more precise use of data across time; and no need to assume that all longitudinal records collected throughout lactation represent the same trait. The model equation can be defined as:

$$\mathbf{y}_{ijk} = \boldsymbol{HYS}_{\boldsymbol{k}} + \sum_{\mathbf{m=1}}^{\mathbf{m}} \boldsymbol{\beta}_{\boldsymbol{m}}(\boldsymbol{t}_{\boldsymbol{ij}}) + \sum_{\mathbf{m=1}}^{\mathbf{m}} \boldsymbol{AG}_{\boldsymbol{im}}\, \emptyset_{\boldsymbol{m}\,(\boldsymbol{t}_{\boldsymbol{ij}})} + \sum_{\mathbf{m=1}}^{\mathbf{m}} \boldsymbol{PE}_{\boldsymbol{m}}\, \emptyset_{\boldsymbol{m}\,(\boldsymbol{t}_{\boldsymbol{ij}})} + \mathbf{e}_{\boldsymbol{ijk}}\,, [2]$$

where $\boldsymbol{y}$ is the predicted trait record of $\boldsymbol{i^{th}}$ animal on DIM $\boldsymbol{j}$, $\boldsymbol{HYS}$ is the fixed effects of herd-year-season; $\boldsymbol{\beta}$, $\boldsymbol{AG}$, and $\boldsymbol{PE}$ are the $\boldsymbol{m^{th}}$ random regression coefficients describing the fixed, additive genetic, and the permanent environment effects for the $\boldsymbol{i^{th}}$ animal, respectively; $\boldsymbol{t_{ij}}$ is the time of data collection (DIM $\boldsymbol{j}$); $\boldsymbol{\emptyset_m(t_{ij})}$ is a regression function according to DIM $\boldsymbol{j}$ using a Legendre Polynomial or B-Splines and $\boldsymbol{e_{ijk}}$ is the random residual term.

The application of RRM has become popular since the use of parametric functions to represent the pattern of milk yield during lactation.

1.3.2 Wood, Wilmink, and Ali-Schaeffer functions

The Wood (Wood, 1967), Wilmink (Wilmink, 1987), Ali and Schaeffer (Ali and Schaeffer, 1987) are parametric functions used to describe the shape of the lactation curve. Parametric functions are not widely used for random regression analysis as estimated covariance matrices show very high correlations between random regression coefficients (Schaeffer, 2004).

The equation of Wood can be described as:

$$\mathbf{y_t = at^b exp^{ct}}\text{ , [3]}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are parameters; $\boldsymbol{Y_t}$ is the recorded DMY at $\mathbf{t}$ DIM; $\boldsymbol{a}$ is a parameter associated with peak yield; $\boldsymbol{b}$ represents the increasing slope; and $\boldsymbol{c}$ represents the decreasing slope. The equation of Wilmink can be described as:

$$\mathbf{y_t = a + b\ exp^{-0.05t} + ct}\text{ , [4]}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are parameters; $\boldsymbol{Y_t}$ is the recorded DMY at $\boldsymbol{t}$ DIM; $\boldsymbol{a}$ is a factor associated with level of production; $\boldsymbol{b}$ is a factor associated with production increase towards peak yield; and $\boldsymbol{c}$ is a factor associated with production decrease after peak yield, and the factor -0,05 with the moment of peak yield, i.e., about 50 days postpartum.

The equation of Ali and Schaeffer can be described as:

$$\mathbf{y_t = a + b\theta_t + c\theta_t^2 + d\delta_t + g\delta_t^2}\text{ , [5]}$$

where $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$, and $\boldsymbol{g}$ are the parameters that define the scale or shape of the lactation curve; $\boldsymbol{Y_t}$ is the recorded DMY at $\boldsymbol{t}$ DIM; $\boldsymbol{a}$ is associated with peak yield; $\boldsymbol{b}$ and $\boldsymbol{c}$ are associated with the decreasing slope; $\boldsymbol{d}$ and $\boldsymbol{g}$ are associated with the increasing slope of the curve; $\boldsymbol{\theta_t = t\ /\ 305}$; and $\boldsymbol{\delta_t = ln\ (305\ /t)}$.

Based on the fact that polynomial regression is appropriate and flexible to adjust linear functions of the averages, the non-parametric functions, as the Legendre orthogonal polynomials (**LEG**), are used in the processes of estimating covariance functions.

1.3.3 Legendre orthogonal polynomials

Legendre orthogonal polynomials are non-parametric functions obtained by transforming the independent variable. In the case of milk yield, the transformation of the lactation days ($\boldsymbol{a_i}$) can be derived through the equation shown in Kirkpatrick et al. (1990):

$$\mathbf{a_i^* = \frac{2(a_i - a_{min})}{a_{max} - a_{min}} - 1}\text{, [6]}$$

where a_{max} and a_{min} are the higher and smaller days in lactation that can be found in the data set, respectively.

According to Kirkpatrick et al. (1990) the j^{th} normalized LEG, Φ_k, can be defined as follow:

$$\Phi_k(a_i^*) = \frac{1}{2^j}\sqrt{\frac{2k+1}{2}}\sum_{m=0}^{[k/2]}(-1)^m\binom{k}{m}\binom{2k-2m}{k}(a_i^*)^{k-2m}, [7]$$

where a_i^* is the established day for the range in which the polynomial k is defined as [-1 to +1].

The use of LEG to fit random curves is widespread due to their ability to describe the variation along lactation, better convergence and being normalised (Takman and Akbas, 2007).

One of the biggest advantages of the LEG is that, since they are orthogonal, they reduce the problems of multicollinearity of the data, which means the columns of the matrix referring to the days are independent, preventing the matrix to become non-defined positive (Oliveira et al., 2016). When both the additive genetic (AG) and permanent environmental (PE) components are modelled by LEG coefficients during time, the estimate of variance components and prediction of breeding value become more accurate (Pool et al., 2000). The minimum order of fit required for modelling observed covariance should be determined in order to obtain reliable estimate of the dispersion parameters. Moreover, the order of LEG is important in random regression because of its ability to influence genetic parameter estimation. Several studies indicated that LEG of orders four and five are the most appropriate to fit RRM for milk yield and persistency of lactation, among them Pool and Meuwissen (2000); Cobuci and Costa (2012); and Pereira et al. (2012).

However, it is possible for the polynomial interpolation to present oscillations in the extremes of the curve of the interpolated interval that tend to the infinite. This effect is common in analyses that use the LEG and is known as Runge's phenomenon; that is, a problem of oscillation at the edges of an interval that occurs when using polynomial interpolation with polynomials of high degree over a set of interpolation points (Boyd and Ong, 2009). Differently from what happens with LEG, in B-splines the Runge's phenomenon is not observed.

1.3.4 B-splines functions

Splines are generally defined as curves which consist of individual segments which are joined smoothly; the segments are given by polynomials and the points at which they join are referred to as knots.

Rice and Wu (2001) described the use of B-splines to model random effects curves in mixed model analyses, and demonstrated their efficacy in estimating covariance functions.

According to Meyer (2005) the use of splines is interesting when a low polynomial degree does not fit well the data. B-splines yield the same fit as splines based on truncated power functions, but have better numerical properties. Here, the "B" stands for basis. B-splines comprise a set of overlapping, smooth and non-negative functions, which are unimodal and sum to unity for all values of t. Like splines with a truncated power basis, they can have different degree p. Basis functions of degree $p = 0$ have values of unity for all points in a given interval, and zero otherwise.

For the k^{th} interval given by knots T_k and T_{k+1} with $T_k \leq T_{k+1}$,

$$B_{k,0}(t) = f(x) = \begin{cases} 1 & if\ T_k \leq t \leq T_{k+1} \\ 0 & otherwise. \end{cases}, [8]$$

Higher degree basis functions, $B_{k,p}$ for $p > 0$, are then determined from the values of the lower degree basis functions, already evaluated, and the width of the adjoining intervals between knots.

The general relationship is:

$$B_{k,p}(t) = \frac{t - T_k}{T_{k+p} - T_k} B_{k,p-1}(t) + \frac{T_{k+p+1} - t}{T_{k+p+1} - T_{k+1}} B_{k+1,p-1}(t), [9]$$

For each p, there are a limited number of non-zero basis functions of lower order.

In order to better understand the molecular mechanisms underlying the phenotypic expression of lactation persistency, which can be useful for improving the genomic evaluation; the locations of quantitative trait loci (**QTL**) regions can be investigated by integrating ssGWAS (single-step genome-wide association study) and RRM (Oliveira et al., 2019).

1.4 Genomic background of lactation

The ssGWAS approach is an extension of the single-step genomic best linear unbiased prediction (ssGBLUP) to the GWAS framework (Wang et al., 2012), and has

recently been used to investigate the genome-to-phenome relationships in animals (Wang et al., 2014).

For instance, Nayeri et al. (2017) conducted ssGWAS for lactation persistency in the North American Holstein dairy cattle population and reported several significant single nucleotide polymorphisms (SNPs) associated with lactation persistency on specific chromosomes, including BTA6, 13, 20, and 27. These regions contain several candidate genes, including *MYT1, SLC2A4RG,* and *SLC17A9* on BTA13 and *THRB* on BTA27.

Pedrosa et al. (2021) identified 49 SNPs located on 18 chromosomes and significantly associated with 85 genes. The most significant genes (p-value < 10^{-7}) were located on BTA18 and BTA28, represented by *ARHGAP35, NPAS1, TMEM160, ZC3H4, SAE1* (BTA18); and *ZMIZ1* and *PPIF* (BTA28).

It is necessary to emphasize that Nayeri et al. (2017) and Pedrosa et al. (2021) did not investigate the GWAS for lactation curve parameters. The lactation curve parameters have biological interpretations, and therefore, they may be interpreted (Strucken et al., 2012; Atashi et al., 2020; Lázaro et al., 2024).

Some genetic effects may be expressed differently depending on the milking systems (e.g., parlors vs. robots), indicating potential milking systems interaction.

1.5 Genotype-by-milking systems

The genotype-by-environment (**GxE**) interaction occurs when the performance of genotypes, exhibit different response to changes in environmental conditions. When the genotypes show differences in estimated breeding value (**EBV**) across environmental conditions, but without changes in ranking, the environmental effect leads to a scaling effect of EBV. However, when genotypes exhibit different levels of response according to environmental condition, often the re-ranking of genotypes could occur. The scaling (simple interaction), without re-ranking (complex interaction) effect caused by GxE interaction on EBV value show less importance, because the best genotype in determined environment would still in other environments.

According to Kolmodin et al. (2002) the GxE interaction leading to animal re-rank of EBV, of the trait evaluated in different environmental conditions can be considered as different traits and the magnitude of GxE interaction effect could be assessed by the genetic correlation of trait in different environmental conditions like temperature and humidity index (**THI**).

Robertson (1959) suggested that correlation for a trait evaluated in different environments lower than 0.80 would indicate an important biological effect of GxE interaction in a trait and would justify the separate selection scheme for breeding programs. Hence, the selection of animal for determinate trait in a determined environmental condition may not be favorable for offspring raised in a different environment.

To evaluate the GxE interaction effect and to obtain breeding values in different environmental conditions for beef cattle, two models are frequently used: multi-trait and reaction norm (**RN**) models. The multi-trait model is useful when environmental conditions can be described by few categories, e.g., milking systems (parlors or milking robots), barn type (free-stall or tie-stall). Thus, the (co)variance components can be estimated for a trait in each environmental condition and the animal re-ranking across environmental conditions is quantified by the genetic correlation (Robertson, 1959).

The multi-trait models used to assess the GxE interaction show limitations when many environmental conditions are evaluated, leading to problems in convergence of models. Hence, the genetic parameters estimation could be biased. In this context, the RN model may be useful, particularly in situations in which the environmental conditions could be described as continuous variable (Kolmodin et al., 2002; Schaeffer, 2004). The GxE interaction is fitted considering that the animal performance could change continuously across the environment (Schaeffer, 2004).

The RN model assume that phenotypic value is express as a polynomial function associated with the environmental condition, in which the polynomial coefficients show genetic influence (de Jong, 1995). Assuming a linear effect, the animal sensitivity can be expressed in function of the regression slope of the animal performance on known environmental conditions (de Jong, 1995). Hence, genetic variations across environmental conditions resulting from variations on slope coefficient. The animals' response to environmental conditions (**EC**) changes can be describes by RN model as follows:

$$\boldsymbol{y_{ij}} = \boldsymbol{X\beta} + \sum_{\mathrm{f}=\mathbf{0}}^{\mathbf{1}} \boldsymbol{\omega}_{\mathrm{f}}\, \boldsymbol{\phi}_{\mathrm{f}}(\boldsymbol{EC_j}) + \sum_{\mathrm{f}=\mathbf{0}}^{\mathbf{1}} \boldsymbol{\alpha}_{\mathbf{i}\mathrm{f}}\, \boldsymbol{\phi}_{\mathrm{f}}(\boldsymbol{EC_j}) + \boldsymbol{e_{ij}}\,, [10]$$

where, $\boldsymbol{y_{ij}}$ is the vector for the phenotypic information of animal $\boldsymbol{i}$ recorded in the environment $\boldsymbol{j}$; $\boldsymbol{X\beta}$ is the fixed effects, $\boldsymbol{\omega}_{\mathrm{f}}$ are the fixed regression coefficients on $\boldsymbol{\phi}_{\mathrm{f}}$; $\boldsymbol{\phi}_{\mathrm{f}}$ are the Legendre polynomials corresponding to $\boldsymbol{EC}$ level ($\boldsymbol{ECj}$); $\boldsymbol{\alpha}_{\mathbf{jf}}$ are the random

regression for additive effects of intercept and slope for animal ***i*** on the **EC*j***, and $\boldsymbol{e}_{ij}$ is a random residual.

Jeon et al. (2023) evaluated the impact of heat stress on milk production in Korean Holstein cows. The results showed that both milk production and lactation persistency sharply declined after surpassing their respective temperature-humidity index break points.

Chen et al. (2023) investigated the effects of heat stress on the milk yield, milk composition, serum oxidative status, and metabolites in Chinese Holstein cows during mid-lactation. They reported a significantly lower average daily milk yield in the heat stress group than in the non-heat stress group ($P < 0.05$) and concluded that heat stress reduces milk yield and composition of Holstein cows and destroys the oxidation balance.

However, it is important to note that Chen et al. (2023) and Jeon et al. (2023) did not use genomic information. The majority of the studies found only use the pedigree relationship matrix. More recently, the use of RN model with the addition of genomic information for the genetic evaluation of longitudinal phenotypic and environmental data has been recommended.

Understanding the GxE interaction considering all sources of information allows stakeholders to make efficient decisions, to ensure the sustainability and resilience of the livestock industry.

1.6 References

ABDELSAYED, M.; THOMSON, P.C.; RAADSMA, H.W. A review of the genetic and non- genetic factors affecting extended lactation in pasture-based dairy systems. **Animal Production Science**, v.55, p.949-966, 2015.

ALI, T.E.; SCHAEFFER, L.R. Accounting for covariances among test day milk yields in dairy cows. **Canadian Journal of Animal Science**. v.67(3), p.637-644, 1987. https://doi.org/10.4141/cjas87-067

ATASHI, H.; SALAVATI, M.; DE KOSTER, J.; EHRLICH, J.; CROWE, M.; OPSOMER, G.; GplusE CONSORTIUM; HOSTENS, M. Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. **Journal of Animal Breeding and Genetics**., v.137, p.292-304. 2020. https://doi.org/10.1111/jbg.12442

BERTILSSON, J.; BERGLUND, B.; RATNAYAKE, G.; SVENNERSTEN, S. K.; WIKTORSSON, H. Optimising lactation cycles for the high-yielding dairy cow. **Livestock Production Science**, v.50, p.5-13, 1997.

BORMAN, J.M.; MACMILLAN, K.L.; FAHEY, J. The potential for extended lactations in Victorian dairying: a review. **Australian Journal of Experimental Agriculture**, v.44, p.507- 519, 2004.

BOYD, J.P.; ONG, J.R. Exponentially-convergent strategies for defeating the Runge phenomenon for the approximation of non-periodic functions, part two: Single-interval schemes. **Communications in Computational Physics**, v.5, p.484–497, 2009.

BRITO, L.F.; BEDERE, N.; DOUHARD, F.; OLIVEIRA, H.R.; ARNAL, M.; PEÑAGARICANO, F.; SCHINCKEL, A.P.; BAES, C.F.; MIGLIOR, F. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. **Animal**, v.15. 2021. https://doi.org/10.1016/j.animal.2021.100292

CANADIAN DAIRY NETWORK (CDN). Increased profits from higher lactation persistency. 3 pages, 2017.

CAPUCO, A.V.; CHOUDHARY, R.K. Symposium review: Determinants of milk production: Understanding population dynamics in the bovine mammary epithelium. **Journal of Dairy Science**, 103(3), p.2928-2940, 2020. https://doi.org/10.3168/jds.2019-17241

CHEN, X.; LI, Y.Q.; XIAO, J.; DONG, J.N.; ZHAO, W.; HAN, Z.Y.; XIN, L.; QIN, G.X.; WANG, T.; ZHEN, Y.G.; SUN, Z.; ZHANG, X.F. Heat stress affects the milk yield, milk composition, serum oxidative status, and metabolites of Holstein cows during mid-lactation. **Czech Journal of Animal Science,** v.68, p.333–345, 2023. https://doi.org/10.17221/2/2023-CJAS

COBUCI, J.A.; EUCLYDES, R.F.; COSTA, C.N.; LOPES, P.S.; TORRES, R.A.; PEREIRA, C.S. Análises da persistência na lactação de vacas da raça holandesa, usando produção no dia do controle e modelo de regressão aleatória. **Revista Brasileira de Zootecnia**, v.33(3), p.546-554, 2004. https://doi.org/10.1590/S1516-35982004000300004

COBUCI, J.A.; EUCLYDES, R.F.; PEREIRA, C.S.; DE AMEIDA TORRES, R.; COSTA, C.N.; LOPES, P.S. Persistência na lactação - uma revisão. **Archivos Latinoamericanos de Producción Anim**al, 11(3), p.163-173, 2003.

COBUCI, J.A.; COSTA, C.N. Persistency of lactation using random regression models and different fixed regression modeling approaches. **Revista Brasileira Zootecnia**, v.41(9), p.1996-2004, 2012. https://doi.org/10.1590/S1516-35982012000900005

COLE, J.B.; NULL, D.J. Genetic evaluation of lactation persistency for five breeds of dairy cattle. **Journal of Dairy Science**, v.92, n.5, p.2248-2258, 2009.

DE JONG, G. Phenotypic plasticity as a product of selection in a variable environment. **American Naturalist**, v.145, p.493–512, 1995.

DE ROOS, A.P.W.; HARBERS, A.G.F.; DE JONG, G. Random regression test-day model in The Netherlands. **Interbull Bulletin 27**, p.155-158, 2001.

DEKKERS, J.C.M.; TEN HAG, J.H.; WEERSINK, A. Economic aspects of persistency of lactation in dairy cattle. **Livestock Production Science**, v.53, p.237–252, 1998.

FAO (FOOD AND AGRICULTURE ORGANIZATION). Dairy Market Review – Overview of global dairy market and policy developments in 2022. Rome, 2023.

FAO (FOOD AND AGRICULTURE ORGANIZATION). Five practical actions towards low-carbon livestock. Rome, 29 pages, 2019.

FAO (FOOD AND AGRICULTURE ORGANIZATION). Shaping the future of livestock – sustainably, responsibly, efficiently. The 10th Global Forum for Food and Agriculture (GFFA), 17 pages, 2018.

GROSSMAN, M.; HARTZ, S.M.; KOOPS, W.J. Persistency of Lactation Yield: A Novel Approach. **Journal of Dairy Science**, v.82, p. 2192–2197, 1999.

INGVARTSEN, K.L. Feeding and management related diseases in the transition cow physiological adaptations around calving and strategies to reduce feeding related diseases. **Animal Feed Science and Technology**, v.126, p.175-213, 2006.

JAKOBSEN, J.H.; MADSEN, P.; JENSEN, J.; PEDERSEN, J.; CHRISTENSEN, L.G.; SORENSEN, D.A. Genetic parameters for milk production and persistency for Danish Holstein estimated in random regression models using REML. **Journal of Dairy Science**, v.85(6), p.1607-1616, 2002. https://doi.org/10.3168/jds.S0022-0302(02)74231-8

JAMROZIK, J.; SCHAEFFER, L.R; DEKKERS, J.C.M. Genetic evaluation of dairy cattle using test day yields and random regression model. **Journal of Dairy Science**, v.80(6), p.1217-1226, 1997. https://doi.org/10.3168/jds.S0022-0302(97)76050-8

JEON, E.; JANG, S.; YEO, J.-M.; KIM, D.-W.; CHO, K. Impact of climate change and heat stress on milk production in Korean Holstein cows: A Large-Scale Data Analysis. **Animals**, v.13, 2946, 2023. https://doi.org/10.3390/ani13182946

KIRKPATRICK, M.; LOFSVOLD, D.; BULMER, M. Analysis of inheritance, selection and evolution of growth trajectories. **Genetics**, v.124, p.979–993, 1990.

KISTEMAKER, G.J. Comparison of persistency definitions in random regression test day models. **Interbull Bulletin 30**, p.96-98, 2003.

KNIGHT, C.H. 2005. Extended lactation: Turning theory into reality. **Advances in Dairy Technology**, v.17, p.113-123, 2005

KOLMODIN, R.; STRANDBERG, E.; MADSEN, P.; JENSEN, J.; JORJANI, H. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. **Acta Agriculturae Scandinavica**, v.52, p.11–24, 2002.

KOLVER, E.S.; ROCHE, J.R.; BURKE, C.R.; KAY, J.K.; ASPIN, P.W. Extending lactation in pasture based dairy cows: Genotype and diet effect on milk and reproduction. **Journal of Dairy Science**, v.90, p.5518-5530, 2007.

LARSSON, B.; BERGLUND, B. Reproductive performance in cows with extended calving interval. **Reproduction in Domestic Animals**, v.35, p.277-280, 2000.

LÁZARO, S.F.; TONHATI, H.; OLIVEIRA, H.R.; SILVA, A.A.; SCALEZ, D.C.B.; NASCIMENTO, A.V.; SANTOS, D.J.A.; STEFANI, G.; CARVALHO, I.S.;

SANDOVAL, A.F.; BRITO, L.F. Genetic parameters and genome-wide association studies for mozzarella and milk production traits, lactation length, and lactation persistency in Murrah buffaloes. **Journal of Dairy Science**, v.107, p.992-1021, 2024. https://doi.org/10.3168/jds.2023-23284

LEHMANN, J.O.; FADEL, J.G.; MOGENSEN, L.; KRISTENSEN, T.; GAILLARD, C.; KEBREAB, E. Effect of calving interval and parity on milk yield per feeding day in Danish commercial dairy herds. **Journal of Dairy Science,** v.99, p.621-633, 2016.

MEYER, K. Random regression analyses using B-splines to model growth of Australian Angus cattle. **Genetics Selection Evolution**, v.37, p.473–500, 2005. https://doi.org/10.1186/1297-9686-37-6-473

NAYERI, S.; SARGOLZAEI, M.; ABO-ISMAIL, M. K.; MILLER, S.; SCHENKEL, F.; MOORE, S. S.; STOTHARD, P. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. **Journal of Dairy Science**, v.100, p.1246–1258, 2017. https://doi.org/10.3168/jds.2016-11770

NICOLÒ, P.P.M.; CORRADO, D.; SALVATORE P.G.R.; ROBERTO, S. & GIUSEPPE, P. The mathematical description of lactation curves in dairy cattle. **Italian Journal of Animal Science**, 2011. https://doi.org/10.4081/ijas.2011.e51

OLIVEIRA, H.R.; LOURENCO, D.A.L.; MASUDA, Y.; MISZTAL, I.; TSURUTA, S.; JAMROZIK, J.; BRITO, L.F.; SILVA, F.F.; CANT, J.P.; SCHENKEL, F.S. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. **Journal of Dairy Science**, v.102, p.9995–10011, 2019. https://doi.org/10.3168/jds.2019-16821

OLIVEIRA, H.R.; SILVA, F.F.; SIQUEIRA, O.H.G.B.D.; SOUZA, N.O.; JUNQUEIRA, V.S.; RESENDE, M.D.V.; BORQUIS, R.R.A.; RODRIGUES. M.T. Combining different functions to describe milk, fat, and protein. **Journal of Animal Science**, v.94, p.1865–1874, 2016. https://doi.org/10 .2527/jas.2015-0150

PEDROSA, V.B.; SCHENKEL, F.S.; CHEN, S.Y.; OLIVEIRA, H.R.; CASEY, T.M.; MELKA, M.G.; BRITO, L.F. Genome wide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. **Genes**, v.12(11), 1830, 2021. https://doi.org/10.3390/genes12111830

PEDROSA, V.B.; BOERMAN, J.P.; GLORIA, L.S.; CHEN, S.Y.; MONTES, M.E.; DOUCETTE, J.S.; BRITO, L.F. Genomic-based genetic parameters for milkability traits derived from automatic milking systems in North American Holstein cattle. **Journal of Dairy Science,** v.106, p.2613–2629, 2023. https://doi:10.3168/jds.2022-22515

PEREIRA, R.J.; VERNEQUE, R.S.; LOPES, P.S.; SANTANA, J.R.M.L.; LAGROTTA, M.R.; TORRES, R.A.; VERCESI FILHO, A.E.; MACHADO, M.A. Milk yield persistency in Brazilian Gyr cattle based on a random regression model. **Genetics and Molecular Research**, v.11(2), p.1599-1609, 2012.

POOL M.H.; JANSS L.L.G.; MEUWISSEN T.H.E. Genetic parameters of Legendre polynomials for first parity lactation curves. **Journal of Dairy Science**, v.83, p.2640–2649, 2000. https://doi:10.3168/jds.S0022-0302(00)75157-5

POOL, M.H.; MEUWISSEN T.H.E. Reduction of the number of parameters needed for a polynomial random regression test day model. **Livestock Production Science**, v.64, p.133–145, 2000. https://doi.org/10.1016/S0301-6226(99)00166-9

PULINA, G.; TONDO, A.; DANIELI, P.P.; PRIMI, R.; MATTEO CROVETTO, G.; FANTINI, A.; MACCIOTTA, N.P.P.; ATZORI, A.S. How to manage cows yielding 20,000 kg of milk: technical challenges and environmental implications. **Italian Journal of Animal Science**, 19(1), p.865–879, 2020. https://doi.org/10.1080/1828051X.2020.1805370

RICE, J.A.; WU, C.O. Nonparametric mixed effects models for unequally sampled noisy curves. **Biometrics**, v.57, p.253–259, 2001.

ROBERTSON A The sampling variance of the genetic correlation coefficient. **Biometrics**, v.15, p.469–485, 1959.

SCHAEFFER L.R. Application of random regression models in animal breeding. **Livestock Science**, v.86, p.35–45, 2004. https://doi:10.1016/S0301-6226(03) 00151-9

SCHINDLER, H.; EGER, S.; DAVIDSON, M.; OCHOWSKI, D.; SCHERMERHORN, E.C.; FOOTE, R.H. Factors affecting response of groups of dairy cows managed for different calving conception intervals. **Theriogenology,** v.36, p.495-503, 1991.

SEHESTED, J.; GAILLARD, C.; LEHMANN, J.O.; MACIEL, G.M.; VESTERGAARD, M.; WEISBJERG, M.R.; MOGENSEN, L.; LARSEN, L.B.; POULSEN, N.A.; KRISTENSEN, T. Review: extended lactation in dairy cattle. **Animal**, v.13, p65-74, 2019. https://doi:10.1017/S1751731119000806

SORENSEN, A.; MUIR, D.D.; KNIGHT, C.H. Extended lactation in dairy cows: effects of miking frequency, calving season and nutrition on lactation persistency and milk quality. **Journal of Dairy Research**, v.75, p.90-97, 2008.

STEFANON, B.; COLITTI, M.; GABAI, G.; KNIGHT, C.H.; WILDE, C.J. Mammary apoptosis and lactation persistency in dairy animals. **Journal of Dairy Research**, v.69(1), p.37-52, 2002. https://doi.org/10.1017/S0022029901005246

STRUCKEN, E.M.; BORTFELDT, R.H.; DE KONING, D.J.; BROCKMANN, G.A. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. **Animal Genetics**, 43(3), p.375-382, 2012. https://doi.org/10.1111/j.1365-2052.2011.02278.x

TAKMAN C.; AKBAS Y. Estimates of genetic parameters for test-day milk yields of Holstein Friesian herd in turkey with random regression models. **Arch Tierz Dummerstorf**, v.50, p.327–336, 2007.

WANG, H., MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; FERNANDO, R.L.; VITEZICA, Z.; OKIMOTO, R.; WING, T.; HAWKEN, R.; MUIR, W.M. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. **Frontiers in Genetics**, v.5:134, 2014. https://doi.org/10.3389/fgene.2014.00134

WANG, H.; MISZTAL, I.; AGUILAR, I.; LEGARRA, A.; MUIR, W.M. Genome- wide association mapping including phenotypes from relatives without genotypes. **Genetics Research (Cambridge)**, v.94, p.73–83, 2012. https://doi.org/10.1017/S0016672312000274

WILMINK, J.B.M. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. **Livestock Production Science**, v.16, p.335–348, 1987. https://doi.org/10.1016/0301-6226(87)90003-0

WOOD, P.D.P. Algebraic model of the lactation curve in cattle. **Nature**, v.216, p.164–165, 1967. https://doi.org/10.1038/216164a0

WU, X-L.; DING, X.; ZHAO, Y.; MILES, A.M.; BRITO, L.F.; HERINGSTAD, B.; ZHAO, S.; JIANG, Z. Editorial: Lactation genomics and phenomics in farm animals: Where are we at? **Frontiers in Genetics**, 14, 1173595, 2023. https://doi.org/10.3389/fgene.2023.1173595

II – OBJETIVOS

2.1 General objective

The main objective of this thesis was to develop the background research to optimize lactation persistency in high-producing dairy cows through selection, which would enable to shift from traditional lactation to extended lactation and highly impact animal welfare, productivity, and overall sustainability of the dairy industry.

2.2 Specific objectives

The specific objectives were to:

- ✓ estimate genomic-based genetic parameters for daily milk yield and various lactation persistency traits;
- ✓ investigate genomic regions, candidate genes, and metabolic pathways associated with lactation curve parameters; and,
- ✓ evaluate genotype-by-milking type interaction (milking robots x milking parlors) for daily milk yield.

III – CAPÍTULO I

This chapter was published in the Journal of Dairy Science

DOI: https://doi.org/10.3168/jds.2024-25836

Genomic-based genetic parameters for daily milk yield and various lactation persistency traits in American Holstein cattle

Patrick R. Fotso-Kenmogne[1,2], Paulo L. S. Carneiro[3,*], Delvan A. Silva[4], Jaime A. Cobuci[5], Pedro F. C. Aponte[1], Hinayah R. Oliveira[6], and Luiz F. Brito[6,*]

[1]Postgraduate Program in Animal Sciences, State University of Southwestern Bahia, Itapetinga, BA, 45700-000, Brazil

[2] Ministry of Employment and Vocational Training (MINEFOP), Yaoundé, CE, P.O. Box 16273, Cameroon

[3]Department of Biology, State University of Southwestern Bahia, Jequié, BA, 45206-190, Brazil

[4]Department of Animal Sciences, Federal University of Viçosa, Viçosa, MG, 36570-000, Brazil

[5]Department of Animal Science, Federal University of Rio Grande do Sul, Porto Alegre, RS, 91540-000, Brazil

[6]Department of Animal Sciences, Purdue University, West Lafayette, IN, 47907, USA

[*]**Corresponding authors:**

Luiz F. Brito. **E-mail**: britol@purdue.edu **Address**: Department of Animal Sciences, Purdue University, West Lafayette, 47907, IN, USA

Paulo L. S. Carneiro. **E-mail**: plscarneiro@gmail.com **Address:** Department of Biology, State University of Southwest Bahia, Jequié, BA, 45205-490, Brazil

ABSTRACT

Genomic-based genetic parameters for daily milk yield and lactation persistency were estimated for the first three lactations in American Holstein cattle. Data included 5,235,411 daily milk yield records on automatic milking systems and milking parlors from 11,788 genotyped cows that calved from 2012 to 2019. A total of 62,029 SNPs remained after quality control. Single-trait random regression models were used to estimate variance and covariance components based on Bayesian inference. The systematic effects in the model included calving year, calving season, milking system, and calving age as linear and quadratic covariates. Random effects included additive genetic, permanent environment, and residual effects modelled with parametric (i.e., Wood, Wilmink, and Ali and Schaeffer) and non-parametric functions (i.e., Legendre orthogonal polynomials and B-splines). The same number of coefficients was considered for all fixed and random regressions in each analysis. Twenty-six models were compared for each lactation and models were selected based on goodness of fit and parsimony, estimates of variance components, heritabilities, correlations, credibility and interpretation of the results, and computational demand and time. Consequently, the model using fourth-order Legendre orthogonal polynomials and 10 classes of heterogeneous residual variance was considered the optimal model. The daily heritability estimates across days in milk ranged from 0.12 to 0.22 for milk yield in parity 1 (MY1), 0.11 to 0.19 for milk yield in parity 2 (MY2), and from 0.03 to 0.15 for milk yield in parity 3 (MY3). For lactation persistency, the heritability estimates ranged from 0.05 to 0.18 for MY1, from 0.06 to 0.20 for MY2, and from 0.05 to 0.16 for MY3. Overall, the heritability estimate values were of low to moderate magnitude. In general, null to moderate genetic correlations were observed between lactation persistency and 305-day milk yield, indicating weak association between lactation persistency and milk yield, which is desirable (but should not be neglected). Ten measures of lactation persistency were evaluated using genomic breeding values. The third lactation persistency measurement (derived by subtracting the area under the curve of lactation in the final third from the area under the curve in the initial third of lactation) is recommended as the selection criterion, as these measures showed a moderate heritability (0.18 for MY1, 0.20 for MY2, and 0.15 for MY3) and a low to moderate genetic correlation (0.38 for MY1, 0.23 for MY2, and 0.20 for MY3) with 305-day milk yield. In summary, this study provides genetic parameter estimates for longitudinal daily milk yield and various lactation persistency traits in American Holstein cattle. These estimates will be useful to

optimize lactation persistency in high-producing dairy cows through selection as they suggest null or weak association between LP and milk yield, and therefore, enhance the profitability of the dairy industry through extended lactations.

Key words: automatic milking systems, extended lactation, longitudinal data, parametric functions, random regression models

INTRODUCTION

Longitudinal daily milk yield records are collected from milking robots or automatic milking systems (**AMS**), as well as traditional milking parlors, and generate a large amount of daily phenotypic records. These records allow a shift from traditional models such as 305-day MY or total lactation yield, as well as test-day milk yield (**TDMY**) records used in dairy cattle genetic evaluation, to a more precise daily milk yield (**DMY)** records approach. Moreover, daily records can yield more accurate genetic parameters and breeding values (Carlström et al., 2013) and could be used for deriving novel traits such as overall resilience (e.g., Chen et al., 2023; Wang et al., 2022, 2024). Furthermore, DMY records enable the modeling of systematic and genetic effects along the lactation curve.

Genomic selection is of paramount importance to improve the efficiency of milk yield (**MY**) and other relevant traits in dairy cattle (Brito et al., 2021). However, its success is mostly due to the availability of data, and the use of ssGBLUP. Presently, random regression models (**RRM**) are an efficient tool for genetic assessment of longitudinal traits (Oliveira et al., 2016, 2019a), offering significant benefits over alternative statistical models, such as more accurate utilization of data across time; not assuming that all records represent records from the same trait; and prediction of genetic parameters for the time-series fall into the range of data collection (Schaeffer, 2004).

Genetic evaluations of the lactation curve can be used to estimate lactation persistency (**LP**) traits. In brief, LP is the ability of a cow to maintain milk production at a high level after reaching the milk production peak (Sehested et al., 2019) and numerous metrics of LP have been proposed over the years (Jamrozik et al., 1997; Jakobsen et al., 2002; Cole and Null, 2009). Greater LP is usually associated with a lower peak production in early lactation, resulting in reduced physiological stress due to high DMY, and risk of reproductive and metabolic disorders (Wahinya et al., 2020). In addition, genetic selection for increased LP could indirectly increase total milk yield and decrease infectious diseases

(e.g., mastitis) or fertility problems (Walsh et al., 2011), which is crucial to increase the economic profitability of the herd (Dekkers et al., 1998; Pedrosa et al., 2021). Improved LP is also associated with lower feeding and health costs (Dekkers et al., 1998).

The traditional approach to improve LP has primarily relied on phenotypic records and pedigree-based information. However, the advent of genomic data, especially SNP markers, enables better predictions of LP (Nayeri et al., 2017; Pedrosa et al., 2021; Lázaro et al., 2024). Therefore, the simultaneous integration of phenotypes, pedigree, and genomic information in the single-step genomic best linear unbiased prediction approach (**ssGBLUP**; Legarra et al., 2009; Aguilar et al., 2010) could lead to more accurate and less biased GEBV (Guarini et al., 2018; Silva et al., 2023), which are essential for the development or enhancement of selection indices in dairy cattle breeding programs.

Genomic estimates of genetic parameters of LP based on daily milk yield records and using parametric and non-parametric functions and models such as RRM, are rare in the literature (Pedrosa et al., 2023). Lactation persistency has been defined mostly based on traditional genetic models (pedigree-BLUP; Jakobsen et al., 2002; Cobuci et al., 2004; Cole and Null, 2009) using TDMY records and could benefit from larger phenotypic and genomic datasets, as currently available in many dairy cattle breeding programs around the world allowing for more accurate estimates of genetic parameters. Thus, the objectives of this study were to (1) define the best functions (i.e., Wood [**WD**; Wood, 1967], Wilmink [**WL**; Wilmink, 1987], Ali and Schaeffer [**AS**; Ali and Schaeffer, 1987], Legendre orthogonal polynomials [**LEG;** Kirkpatrick et al., 1990], or B-splines [**BSP;** Meyer, 2005]) for genetically evaluating DMY and LP in the first three lactations of Holstein cows using the GBLUP approach based on single-trait RRM; (2) estimate genomic-based genetic parameters for DMY and various LP traits, including heritability and genetic correlations; and (3) define the optimal indicators of LP in American Holstein cattle.

MATERIALS AND METHODS

Ethics Statement

Animal care and use committee approvals were unnecessary because data were obtained from pre-existing databases.

Datasets

The data were obtained from a large commercial dairy cattle farm in North-Central Indiana. The phenotypic and genomic datasets used were previously described by

Chen et al. (2023). In brief, 7,083,609 DMY records were initially collected using Lely Astronaut robots and milking parlors (Afimilk system) for 29,343 lactations of 14,469 Holstein cows born from 2012 to 2019. Data editing was performed to remove records with missing information, biologically impossible values, sparse recording (<30 records in total, or time span <70 d, or average daily interval >2), truncated records, records from the fourth or higher lactations, from cows with first calving before 640 d of age, and within small (<10 cows) contemporary groups (**CG**, defined by concatenating calving year and season [March to May, June to August, September to November, December to February] of calving).

In this study, only DMY records within the range of 5 and 305 DIM were considered, and DMY records deviating ±3.5 phenotypic SD from the mean within the CG were removed from further analyses (Silva et al., 2023; Lázaro et al., 2024). Finally, 5,235,411 records from 21,355 lactations of 11,788 Holstein cows remained for further analyses, including 2,754,840 DMY of 11,139 cows in lactation 1; 1,642,653 DMY of 6,735 cows in lactation 2; and 837,918 DMY of 3,481 cows in lactation 3. The descriptive statistics of the datasets are presented in Table 1.

Table 1. Summary statistics for number of cows, number of records, average records per cow, number of contemporary groups (CG), average daily milk yield, minimum and maximum daily milk yield, average age at calving, minimum and maximum age at calving, and milking systems in American Holstein cattle[1]

Variable	MY1	MY2	MY3
Number of cows	11,139	6,735	3,481
Number of records	2,754,840	1,642,653	837,918
Average (SD) records per cow	247.31 (69.38)	243.90 (68.60)	240.71 (69.12)
CG number	30	25	20
Average (SD) daily milk yield, kg	32.86 (08.11)	39.88 (10.73)	42.25 (12.02)
Min (Max) daily milk yield, kg	0.22 (68.21)	0.21 (82.90)	0.15 (90.45)
Average (SD) age at calving, d	713.5 (40.04)	1,106 (69.01)	1,502 (89.18)
Min (Max) age at calving, d	640 (962)	916 (1,556)	1,256 (1,960)
Milking systems			
Number of records from the Lely Astronaut robots (%)	459,505 (16.68)	485,473 (29.55)	312,201 (37.26)
Number of records from the Afimilk milking parlors (%)	2,295,335 (83.32)	1,157,180 (70.45)	525,717 (62.74)

[1]CG: contemporary group; Max: maximum; Min: minimum; SD: standard deviation; MY1: milk yield in first lactation; MY2: milk yield in second lactation; MY3: milk yield in third lactation.

The genomic dataset contained 72,820 SNPs for 22,090 animals that were imputed from multiple SNPs panels as described by Chen et al. (2023). A quality control (**QC**) was performed to remove SNPs and cows with call rate lower than 0.90, minor allele frequency lower than 0.01, and extreme deviation from the Hardy-Weinberg equilibrium ($P < 10^{-8}$). After the QC, the clean genomic dataset contained 62,029 SNPs and 14,469 animals.

Statistical Models

The analyses were performed using single-trait random regression models. These models included the average curve, systematic effects (CG and milking system – AMS or parlor), calving age as linear and quadratic covariates, additive genetic, permanent environment, and residual effects. We also compared models with heterogeneous (**HE**) and homogeneous (**HO**) residual variances. The HE residual variances were defined based on DIM and grouped into 10 classes as follows: 5 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 (Bohmanova et al., 2008; Oliveira et al., 2019a,b). The random regressions for additive genetic and permanent environmental effects, as well as fixed regression (average curve) were modelled using nonlinear functions (WD, WL, and AS) and linear functions (BSP and LEG), considering the same number of coefficients for fixed and random regressions in each analysis, according to the model. The LEG can be defined as follows (Kirkpatrick et al., 1990):

$$\mathbf{DMY_t} = \sum_{i=0}^{n} \boldsymbol{k_i}\, \boldsymbol{\phi_i}\, (\boldsymbol{x}), \text{ [1]}$$

where x is the standardized time unit, where $x = -1 + 2(t - t_{min})/(t_{max} - t_{min})$, in which t_{min} and t_{max} are the minimum and maximum records of DIM, respectively; k_i are the estimated coefficients of the lactation curve. The $\phi_i\ (x)$ parameter was calculated as follows:

$$\boldsymbol{\phi_i}\ (\boldsymbol{x}) = \sqrt{\frac{\boldsymbol{2n+1}}{\boldsymbol{2}}}\boldsymbol{P_n}\ (\boldsymbol{x}), \text{ [2]}$$

where $P_n(x)$ is a polynomial of degree n, $\phi_i\ (x)$ is the normalized polynomial, and other parameters were as described above. The BSP can be defined as follows (Meyer, 2005):

$$\boldsymbol{B_{ijq}} = (\boldsymbol{t} - \boldsymbol{T_q}/\boldsymbol{T_{q+1}} - \boldsymbol{T_q})\boldsymbol{I_{ij}}, \text{ [3]}$$

where B_{ijq} is the covariate for the qth interval associated with DIM j from cow i; t is the observed day and I_{ij} is a binary indicator variable given by knots T_q and T_{q+1}, with:

$$\boldsymbol{T_q} \leq \boldsymbol{t} < \boldsymbol{T_{q+1}} \text{ [4]}$$

$$\boldsymbol{I}_{ij=}\begin{cases}\mathbf{1}, & \boldsymbol{if\ T_q \leq t \leq T_{q+1}}\\ \mathbf{0}, & \boldsymbol{otherwise}\end{cases} \text{ [5]}$$

The LEG models were described as LEG*i*, in which *i* corresponds to the LEG order ranging from 3 to 6, whereas models based on BSP functions were described as BSP*ij*, where *i* = 1, 2, or 3 represented the linear, quadratic, and cubic order, respectively, and *j* = 4 or 5 represented 4 (located at DIM 5, 105, 205, and 305) or 5 (located at DIM 5, 80, 155, 230, and 305) equidistant knots (3 and 4 segments, respectively). The DMY records from lactations 1, 2, and 3 were separately fitted using the 26 models. The RRM can be described as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e}, \text{ [6]}$$

where $\mathbf{y}$ is the vector of phenotypic DMY records of the animal for each lactation (MY1, MY2, or MY3); $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\mathbf{pe}$, and $\mathbf{e}$ are the vectors of systematic (i.e., fixed regression, CG, milking system, and calving age as linear and quadratic covariates), additive genetic, permanent environmental, and residual effects, respectively; and $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{W}$ are the incidence matrices relating $\boldsymbol{\beta}$, $\boldsymbol{a}$, and $\mathbf{pe}$, to $\mathbf{y}$. The model assumptions are:

$$\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{a}, \mathbf{pe}, \mathbf{G}, \mathbf{P}, \mathbf{R} \sim \mathbf{N}(\mathbf{X\beta} + \mathbf{Za} + \mathbf{Wpe}, \mathbf{R}), \text{ where } \mathbf{R} = \boldsymbol{\sigma}^2_{e_h}\mathbf{I}, \text{ [7]}$$

$$\mathbf{a} \mid \mathbf{G} \sim \mathbf{N}(\mathbf{0}, \mathbf{G}), \text{ where } \mathbf{G} = \mathbf{G_0} \otimes \mathbf{G_{pure}}, \text{ [8]}$$

$$\mathbf{pe} \mid \mathbf{P} \sim \mathbf{N}(\mathbf{0}, \mathbf{P}), \text{ where } \mathbf{P} = \mathbf{P_0} \otimes \mathbf{I_P}, \text{ [9]}$$

where $\mathbf{G_0}$ and $\mathbf{P_0}$ are the additive genetic and permanent environmental covariance matrices between regression coefficients of each function (LEG, BSP, AS, WD, WL), respectively, which were assumed to follow an inverted Wishart distribution; $\mathbf{G_{pure}}$ is the pure $\mathbf{G}$ matrix constructed based on genomic information; $\boldsymbol{\sigma}^2_{e_h}$ is the residual variance and h represents the 10 classes considered; and $\mathbf{I_P}$ and $\mathbf{I}$ are the identity matrices. A uniform distribution was assumed for systematic effects. The LEG was used to model the systematic and random regression coefficients in accordance with Kirkpatrick et al. (1990), in which the third to sixth LEG orders were tested. The linear, quadratic, and cubic BSP functions were defined according to Meyer (2005).

The general model applied for WD, WL, and AS functions can be described as:

$$\mathbf{y}_{ijkl} = \boldsymbol{\mu} + \mathbf{CG}_k + \mathbf{DATS}_l + \sum_{\mathbf{q=0}}^{\mathbf{q}} \mathbf{a}_{iq}\,\boldsymbol{\psi}_{ijq} + \sum_{\mathbf{q=0}}^{\mathbf{q}} \boldsymbol{p}_{iq}\,\boldsymbol{\psi}_{ijq} + \mathbf{e}_{ijkl}, \text{ [10]}$$

where $\mathbf{y}_{ijkl}$ is the vector of the DMY recorded i (i = 1, 2, …, i) of the animal for trait measured on DIM j (j = 5, 6, . . ., 305), with the CG_k (k = 1, 2, …, K), and data source (DATS_l = 1 for Lely Astronaut robots or 2 for Afimilk milking parlors) of DMY; μ is the overall mean and the fixed regression coefficient; a_{iq} and p_{iq} are the random regression

coefficients associated with covariate q for the additive genetic and permanent environmental effects of animal i, respectively; ψ_{ijq} is the qth covariate for DIM j from cow i and e_{ijkl} is the random residual term. The parameters that define the scale or shape of the lactation curve were defined according to WD, WL, and AS. The WD, WL, and AS functions can be described, respectively, as follows:

$\Lambda_{ij0} = \mathbf{1}$, $\Lambda_{ij1} = \mathbf{ln}\, \boldsymbol{t}$, and $\Lambda_{ij2} = (-\boldsymbol{t})$, [11]

$\boldsymbol{\psi_{ij0}} = \mathbf{1}$, $\boldsymbol{\psi_{ij1}} = \mathbf{exp}(-\mathbf{0.05}\boldsymbol{t})$, and $\boldsymbol{\psi_{ij2}} = \boldsymbol{t}$, [12]

$\boldsymbol{\Gamma_{ij0}} = \mathbf{1}$, $\boldsymbol{\Gamma_{ij1}} = \mathbf{t/305}$, $\boldsymbol{\Gamma_{ij2}} = (\boldsymbol{t/305})^{\mathbf{2}}$, $\boldsymbol{\Gamma_{ij3}} = \boldsymbol{ln}(\mathbf{305}/\boldsymbol{t})$, and $\boldsymbol{\Gamma_{ij4}} = \boldsymbol{ln}(\mathbf{305}/\boldsymbol{t})^{\mathbf{2}}$, [13]

where t is the day of lactation.

The heterogeneity of residual variance was assumed as $e_{ijkl} \sim N(0, \sigma^2_{e_h})$, where h represents the 10 classes of residual variance. The analyses were based on GBLUP approach rather than ssGBLUP because no pedigree information was available. The system of mixed model equations was as follows:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}W} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + G^{-1}} & \mathbf{Z'R^{-1}W} \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}Z} & \mathbf{W'R^{-1}W + P^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^{\circ} \\ \hat{\mathbf{a}} \\ \widehat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \\ \mathbf{W'R^{-1}y} \end{bmatrix}, \text{ [14]}$$

where $\mathbf{G^{-1}}$ is the inverse of the genomic relationship matrix, and was computed as in the first method proposed by VanRaden (2008), **R** and **P** are diagonal matrices, and $\mathbf{R^{-1}}$ and $\mathbf{P^{-1}}$ are the inverse, respectively.

The (co)variance components were estimated based on Bayesian inference using the Gibbs sampler algorithm implemented in the GIBBSF90+ software, version 2023 from the BLUPF90 family of programs (Misztal et al., 2018), because these random regression models are computationally hard to deal with, in the variance components estimation stage. A chain length of 500,000 cycles was used with a burn-in of 200,000 cycles and a sampling interval of 10 cycles, which resulted in 30,000 samples for subsequent analyses. The convergence of the MCMC chains was verified by graphical inspection and based on the Geweke's criterion (Geweke, 1992), assessed using the Bayesian Output Analysis R package (Smith, 2007). The models were compared based on the best goodness of fit according to the deviance information criterion (**DIC**; Spiegelhalter et al., 2002). The DIC can be described as follows:

$$\mathbf{DIC} = \mathbf{D}(\bar{\boldsymbol{\theta}}) + \mathbf{2p_D}, \text{ [15]}$$

where $D(\bar{\theta})$ is a point estimate of the deviance obtained by replacing the parameters by their posterior mean estimates in the likelihood function and $2p_D$ is the effective number

of parameters in the model. Models with lower DIC are considered to have better goodness of fit. However, DIC only expresses if one model presented the best fit in relation to other models, but the magnitude of this difference may be subjective. To complement this information, in terms of the superiority of one model over another, the Model Posterior Probabilities (**MPP**) were calculated from DIC as presented by Wilberg and Bence (2008), in which:

$$\boldsymbol{p(M_t|\theta) = \exp\left(-\frac{\Delta_t}{2}\right) / \sum_{t=1}^{2} \exp\left(-\frac{\Delta_t}{2}\right)}, \text{ [16]}$$

where $p(M_t|\theta)$ is the posteriori probability of model t; Δ_t is the DIC difference between model t and the model that presented the smallest DIC value. Models with higher MPP are considered to be better.

Genetic Parameter Estimates

The covariance matrices for additive genetic $(\mathbf{G_1})$ and permanent environmental $(\mathbf{P_1})$ effects along DIM in each lactation were estimated as:

$$\mathbf{G_1 = \Phi\, \widehat{Go}\, \Phi'} \text{ [17]}$$

$$\mathbf{P_1 = \Phi\, \widehat{Po}\, \Phi'} \text{ [18]}$$

where, $\widehat{\mathbf{Go}}$ and $\widehat{\mathbf{Po}}$ were previously defined and $\mathbf{\Phi}$ is a matrix of LEG or BSP coefficients or a matrix of WD, WL or AS parameters.

The posterior marginal distribution samples for heritabilities throughout the lactation curve were obtained by:

$$\mathbf{h_j^2} = \frac{\boldsymbol{\sigma_{a_j}^2}}{\boldsymbol{\sigma_{a_j}^2 + \sigma_{p_j}^2 + \sigma_{e_h}^2}} \text{ [19]}$$

where h_j^2 is the heritability estimate on DIM j, $\sigma_{a_j}^2$ is the additive genetic variance on DIM j, $\sigma_{p_j}^2$ is the permanent environmental variance on DIM j, and $\sigma_{e_h}^2$ is the residual variance, which depends on the residual variance classes h (h = 1 to 10). The posterior marginal distribution samples for correlation throughout the lactation curve were obtained by:

$$\mathbf{r_{(j,j')}} = \frac{\boldsymbol{\sigma_{a_{(j,j')}}}}{\sqrt{\boldsymbol{\sigma_{a_{(j)}}^2\ \sigma_{a_{(j')}}^2}}} \text{ [20]}$$

where $r_{(j,j')}$ is the correlation between DIM j and j', $\sigma_{a_{(j,j')}}$ is the genetic covariance between j and j', $\sigma_{a_{(j)}}^2$ is the additive genetic variance for DIM j and $\sigma_{a_{(j')}}^2$ is the additive genetic variance for DIM j'.

Genomic Estimated Breeding Values

The additive genetic effect of the coefficients estimated were used to derive the GEBV for each DIM. The GEBVs were computed using the GBLUP method (Aguilar et al., 2010). The vectors of the GEBV for all DIM of each animal were obtained as follows:

$$(\mathbf{GEBV_{ij}}) = \mathbf{\Phi}\,\widehat{\boldsymbol{\delta}}_{\mathbf{ij}}\,, \quad [21]$$

where $\widehat{\boldsymbol{\delta}}_{\mathbf{ij}}$ is the vector of the predicted genomic coefficients for each animal i and lactation j; and $\mathbf{\Phi}$ is a matrix of the independent covariates for each DIM associated with the better function (i.e., WD, WL, AS, LEG, or BSP). The $\mathbf{GEBV}$ for accumulated 5 to 305 DIM ($\mathbf{GEBV305}$) were obtained as the sum of all $\mathbf{GEBVs}$ per animal within lactation. The GEBVs are needed for estimating the proposed lactation persistency measures. Additionally, GEBV305 was calculated to examine the correlation with PS_i, and MY_{305} is the adjusted milk production for 305 d.

Lactation Persistency

Different methods of quantifying LP (Table 2) were considered to compare them and define the best definition of this trait according to the design of lactation curves. The first, PS_1, was calculated based on the difference between the GEBV for DIM 280 and 60 (adapted from Jamrozik et al., 1997). The second, PS_2, was obtained based on the difference between the areas under the lactation curve in the middle and initial thirds of lactation (adapted from Jakobsen et al., 2002). The third, PS_3, was derived by subtracting the area under the curve of lactation in the final third from the area under the curve in the initial third of lactation (adapted from Jakobsen et al., 2002). The fourth, PS_4, was calculated by summing GEBV in the period from DIM 61 to 280, as a deviation from DIM 60 (adapted from Jamrozik et al., 1997). The fifth, PS_5, was obtained by summing the contributions of each day of production in the period from DIM 60 to 279, as a deviation from DIM 280 (adapted from Jakobsen et al., 2002). The sixth, PS_6, was derived by calculating the difference between GEBV for DIM 290 and 90 (adapted from Cobuci et al., 2004). The seventh, PS_7, was obtained by the average of GEBV for DMY from DIM 255 to 305 as a deviation from the average of GEBV from DIM 50 to 70 (adapted from Kistemaker, 2003). The eighth, PS_8, was calculated by summing the contribution for each day of production in the period from DIM 101 to 300 as a deviation from DIM 100 (adapted from Pösö, 2003 in Kistemaker, 2003). The ninth, PS_9, was calculated by summing the contribution for each day of production in the period from DIM 61 to 305 as a deviation from DIM 60 (adapted from De Ross et al., 2003). The tenth, PS_{10}, was

adapted from Jamrozik et al. (1997), which takes into account the DIM 280 and the peak yield day of the population sampled. Lower values of PS_1, PS_2, PS_3, PS_4, PS_6, and PS_{10} indicate higher levels of LP, whereas higher values of PS_5, PS_7, PS_8, and PS_9 indicate higher levels of LP.

Table 2. Measures of milk yield persistency (PS_i) for $\boldsymbol{i}$ = 1 to 10

Measures	Authors
$\mathbf{PS_1} = (\boldsymbol{GEBV_{280}} - \boldsymbol{GEBV_{60}})$	Adapted from Jamrozik et al. (1997)
$\mathbf{PS_2} = \left(\sum_{t=106}^{205} \boldsymbol{GEBV_t} - \sum_{t=6}^{105} \boldsymbol{GEBV_t} \right)$	Adapted from Jakobsen et al. (2002)
$\mathbf{PS_3} = \left(\sum_{t=206}^{305} \boldsymbol{GEBV_t} - \sum_{t=6}^{105} \boldsymbol{GEBV_t} \right)$	Adapted from Jakobsen et al. (2002)
$\mathbf{PS_4} = \sum_{t=61}^{280} (\boldsymbol{GEBV_t} - \boldsymbol{GEBV_{60}})$	Adapted from Jamrozik et al. (1997)
$\mathbf{PS_5} = \sum_{t=60}^{279} (\boldsymbol{GEBV_t} - \boldsymbol{GEBV_{280}})$	Adapted from Jakobsen et al. (2002)
$\mathbf{PS_6} = (\mathbf{G}\boldsymbol{EBV_{290}} - \mathbf{G}\boldsymbol{EBV_{90}})$	Adapted from Cobuci et al. (2004)
$\boldsymbol{PS_7} = \left(\frac{1}{51} \sum_{t=255}^{305} \boldsymbol{GEBV_t} - \frac{1}{21} \sum_{t=50}^{70} \boldsymbol{GEBV_t} \right)$	Adapted from Kistemaker (2003)
$\boldsymbol{PS_8} = \left(\sum_{t=101}^{300} \boldsymbol{GEBV_t} - (\mathbf{300} - \mathbf{100})\, \boldsymbol{GEBV_{100}} \right)$	Adapted from Pösö (2003) in Kistemaker (2003)
$\boldsymbol{PS_9} = \left(\sum_{t=61}^{305} \boldsymbol{GEBV_t} - (\mathbf{305} - \mathbf{60})\, \boldsymbol{GEBV_{60}} \right)$	Adapted from de Ross et al. (2001)
$\boldsymbol{PS_{10}} = \sum_{t=\mathbf{pyd}+1}^{280} (\boldsymbol{GEBV_t} - \boldsymbol{GEBV_{\mathbf{pyd}}})$	Adapted from Jamrozik et al. (1997)

GEBV = genomic estimated breeding value; t = day of lactation; pyd = peak yield day.

The genetic variances for PS_i ($\sigma^2_{aPS_i}$) were obtained by:

$$\left(\boldsymbol{\sigma^2_{aPSi}}\right) = \mathbf{f_i} \bigwedge_{\mathbf{A}} \mathbf{f_i'}\text{ , [22]}$$

where $\mathbf{f_i}$ is the vector of covariates corresponding to the function of the measure of LP, and $\bigwedge_{\boldsymbol{A}}$ is the matrix of additive genetic covariances between the independent covariates. Permanent environmental variances for PS_i ($\sigma^2_{pePS_i}$) were obtained in a similar manner by replacing $\bigwedge_{\boldsymbol{A}}$ with $\bigwedge_{\mathbf{P}}$, the matrix of permanent environmental covariances between the independent covariates.

The heritabilities of the LP measures ($h^2_{PS_i}$) were calculated as:

$$\mathbf{h^2_{PS_i}} = \boldsymbol{\sigma^2_{aPS_i}} / \boldsymbol{\sigma^2_{aPS_i}} + \boldsymbol{\sigma^2_{pePS_i}} + \boldsymbol{\sigma^2_{ePS_i}}\text{, [23]}$$

where $\sigma^2_{ePS_i}$ represents the heterogeneous residual variance; $\sigma^2_{ePS_1}$ and $\sigma^2_{ePS_2}$ were obtained as follows:

$$\left(\sigma^2_{ePS_1}\right) = V_e(PS_1) = V_e(MY_{280} - MY_{60})\text{, [24]}$$

$$V_e(PS_1) = V_e(MY_{280}) + V_e(MY_{60}) - 2Cov_e(MY_{280}, MY_{60})\text{. [25]}$$

The residual variances were considered to be independent, the terms $Cov_e(MY_{280}, MY_{60})$ were considered as equal to zero. Thus,

$$V_e(PS_1) = V_e(MY_{280}) + V_e(MY_{60})\text{. [26]}$$

According to the residual variance classes used,

$$V_e(MY_{60}) = \sigma^2_{e_2}\text{ [27]}$$

$$V_e(MY_{280}) = \sigma^2_{e_{10}}\text{ [28]}$$

Therefore,

$$V_e(PS_1) = \sigma^2_{e_2} + \sigma^2_{e_{10}}\text{ , [29]}$$

$$\left(\sigma^2_{ePS_2}\right) = V_e(PS_2) = V_e\left[\left(\sum_{t=106}^{205} MY_t\right) - \left(\sum_{t=6}^{105} MY_t\right)\right]\text{, [30]}$$

$$V_e(PS_2) = V_e\left(\sum_{t=106}^{205} MY_t\right) + \left(\sum_{t=6}^{105} MY_t\right) - 2Cov_e\left[\left(\sum_{t=106}^{205} MY_t\right) - \left(\sum_{t=6}^{105} MY_t\right)\right]\text{, [31]}$$

where t is the day of lactation and MY_t is the milk yield on day t. Because in the present study the residual variances were considered to be independent and t is always different from t', the terms containing Cov_e are equal to zero. Thus,

$$V_e(PS_2) = V_e\left(\sum_{t=106}^{205} MY_t\right) + \left(\sum_{t=6}^{105} MY_t\right)\text{, [32]}$$

$$V_e(PS_2) = [V_e(MY_{106}) + \ldots + V_e(MY_{205})] + [V_e(MY_6) + \ldots + V_e(MY_{105})]\text{. [33]}$$

According to the residual variance classes used,

$$V_e(MY_5) \text{ to } V_e(MY_{35}) = \sigma^2_{e_1}, [34]$$

$$V_e(MY_{36}) \text{ to } V_e(MY_{65}) = \sigma^2_{e_2}, [35]$$

$$V_e(MY_{66}) \text{ to } V_e(MY_{95}) = \sigma^2_{e_3}, [36]$$

$$V_e(MY_{96}) \text{ to } V_e(MY_{125}) = \sigma^2_{e_4}, [37]$$

$$V_e(MY_{126}) \text{ to } V_e(MY_{155}) = \sigma^2_{e_5}, [38]$$

$$V_e(MY_{156}) \text{ to } V_e(MY_{185}) = \sigma^2_{e_6}, [39]$$

$$V_e(MY_{186}) \text{ to } V_e(MY_{215}) = \sigma^2_{e_7}, [40]$$

$$V_e(MY_{216}) \text{ to } V_e(MY_{245}) = \sigma^2_{e_8}, [41]$$

$$V_e(MY_{246}) \text{ to } V_e(MY_{275}) = \sigma^2_{e_9}, [42]$$

$$V_e(MY_{276}) \text{ to } V_e(MY_{305}) = \sigma^2_{e_{10}}. [43]$$

Therefore,

$$V_e(PS_2) = (20\sigma^2_{e_4} + 30\sigma^2_{e_5} + 30\sigma^2_{e_6} + 20\sigma^2_{e_7}) + (30\sigma^2_{e_1} + 30\sigma^2_{e_2} + 30\sigma^2_{e_3} + 10\sigma^2_{e_4}),$$

[44]

$$\left(\sigma^2_{ePS_2}\right) = V_e(PS_2) = 30\sigma^2_{e_1} + 30\sigma^2_{e_2} + 30\sigma^2_{e_3} + 30\sigma^2_{e_4} + 30\sigma^2_{e_5} + 30\sigma^2_{e_6} + 20\sigma^2_{e_7};$$

[45]

$\sigma^2_{ePS_i}$ values were obtained in a similar manner for the other measures of PS_i, based on the residual variance classes used (Menezes et al., 2010).

Additive ($\sigma^2_{a_{305}}$) and permanent environment ($\sigma^2_{pe_{305}}$) variances, and heritability (h^2_{305}) for 305-d MY were obtained as follows:

$$\left(\boldsymbol{\sigma}^2_{\mathbf{a_{305}}}\right) = \mathbf{f_{305}} \Lambda_A \mathbf{f}'_{\mathbf{305}} \text{ and } \left(\boldsymbol{\sigma}^2_{\mathbf{pe_{305}}}\right) = \mathbf{f_{305}} \Lambda_P \mathbf{f}'_{\mathbf{305}} \text{ with } \mathbf{f_{305}} = \sum_{i=5}^{305} \sum_{j=0}^{n} \emptyset_{ij}, [46]$$

$$h^2_{305} = \sigma^2_{a_{305}} / \sigma^2_{a_{305}} + \sigma^2_{pe_{305}} + \sigma^2_{e_{305}}, [47]$$

where $\mathbf{f_{305}}$ is a row vector of order q, with the jth elements equal to the sum of the jth orthogonal polynomial from d 5 to 305, ($\sigma^2_{e_{305}}$) is the heterogeneous residual variance estimated from d 5 to 305 for 305-d MY, and other parameters were as described above.

Genetic correlations between PS_i and PS_i', for $i' \neq i$, and PS_i and MY_{305} were estimated as:

$$\mathbf{r}_{(PS_i, PS_i')} = \frac{\boldsymbol{\sigma}_{a_{(PS_i, PS_i')}}}{\sqrt{\boldsymbol{\sigma}^2_{a_{(PS_i)}} \boldsymbol{\sigma}^2_{a_{(PS_i')}}}} \text{ with } \boldsymbol{\sigma}_{a_{(PS_i, PS_i')}} = \mathbf{f}_{PS_i} \Lambda_A \mathbf{f}'_{PS_i'}, [48]$$

$$\mathbf{r}_{(PS_i, MY_{305})} = \frac{\boldsymbol{\sigma}_{a_{(PS_i, MY_{305})}}}{\sqrt{\boldsymbol{\sigma}^2_{a_{(PS_i)}} \boldsymbol{\sigma}^2_{a_{(MY_{305})}}}} \text{ with } \boldsymbol{\sigma}_{a_{(PS_i, MY_{305})}} = \mathbf{f}_{PS_i} \Lambda_A \mathbf{f}'_{MY_{305}}. [49]$$

where $\sigma^2_{a_{(PS_i)}}$ and $\sigma^2_{a_{(PS_i')}}$ are the genetic variances for LP, $\sigma_{a_{(PS_i,PS_i')}}$ is the genetic covariance between PS_i and PS_i', $\sigma_{a_{(PS_i,MY_{305})}}$ is the genetic covariance between PS_i and MY_{305}, and $\sigma^2_{a_{(MY_{305})}}$ the genetic variance for 305-day MY, and other parameters were as described above.

RESULTS

Descriptive Statistics and Model Comparison

The DMY records were extracted from two milking systems: Lely Astronaut robots (n = 1,257,179; 24.01%) and Afimilk milking parlors (n = 3,978,232; 75.99%). The average records (±SD) of DMY per cow were 247.31 ± 69.38, 243.90 ± 68.60, and 240.71 ± 69.12 for MY1, MY2, and MY3, respectively (Table 1). The average DMY (±SD) for the first, second, and third lactation were 32.86 ± 08.11 kg, 39.88 ± 10.73 kg, and 42.25 ± 12.02 kg, respectively. These cows were assigned into 20 to 30 CG based on the calving year and season.

The convergence was achieved for all models. The number of parameters and the DIC calculated for the comparison between the single-trait RRM for DMY for the lactation curve models based on third, fourth, fifth, and sixth Legendre orthogonal polynomials (LEG3, LEG4, LEG5, and LEG6), linear, quadratic, and cubic B-splines with 4 and 5 knots (BSP14, BSP15, BSP24, BSP25, BSP34, and BSP35), WD, WL, and AS are shown in Table 3. The number of estimated parameters ranged from 21 to 66 when increasing the LEG order from 3 to 6 or the BSP order from 1 to 3. Additionally, when the number of knots for the BSP models was increased from 4 to 5, the number of estimated parameters also increased from 10 to 14. Furthermore, when the assumption of heterogeneous residual variance was taken into account, 9 additional parameters were estimated. In general, more complex models with higher orders of LEG or BSP in all lactations tended to have better DIC (with lower values). The models with HE residual variance assumptions showed lower DIC values (more appropriately fitted) than models with homogeneous residual variance. The computational time (in hours) for the LEG6 model with HE residual variances was 1,776, 1,008, and 720 for MY1, MY2, and MY3, respectively. In comparison, the computing time for the LEG4 model with HE residual variances was 960, 696, and 504 for MY1, MY2, and MY3, respectively. Results from the DIC values and the MPP (Supplementary Table S1) suggested that the LEG6 model with HE residual variances for MY1, MY2, and MY3 provided the best goodness of fit. Nevertheless, this model also presented the greatest degree of complexity and produced

results that deviated from the expected pattern for the genetic correlation estimates for different PS_i and MY_{305}, particularly for LP of MY3 (Supplementary Table S2), with little basis in the literature to explain these findings (i.e., few studies in the literature have examined daily recorded milk production using random regression models while incorporating genomic information; as a result, there are limited references available for comparison). Furthermore, those genetic correlation values of PS_i and MY_{305} are atypical when compared with those found for the other models in this study. Therefore, to balance goodness of fit and parsimony, various factors such as estimates of variance components, heritabilities, correlations, credibility and interpretations of the results, and computational demand and time were considered. Hence, the LEG4 model with HE residual variances was used for the subsequent analyses, as it appeared to be the most appropriate for the studied population. The results obtained with the LEG6 model with HE residual variances are presented in Supplemental Tables S2 and S3 and Supplemental Figures S1, S2, and S3.

Phenotypic Information and Variance Components

Phenotypic averages observed in the first, second, and third lactations according to DIM for DMY are illustrated in Figure 1. The phenotypic curve pattern was the same for the second and third lactations, and the first lactation cows were more persistent. Additionally, the first lactation curve was lower than the second and third lactation curves up to DIM 245. The DMY tended to be higher in the third lactation. The additive genetic variance, permanent environmental variance, residual variance, and phenotypic variance according to DIM for DMY are shown in Figure 2. Higher oscillation in the pattern of the curve was observed for MY3. In general, the additive genetic variances were lower at the beginning of the lactation, with a gradual increase as DIM increased. The pattern of the permanent environmental variance and phenotypic variance were the same. Moreover, these estimates tended to increase from the first to the third lactation. The curves plotted for phenotypic variance and permanent environmental variance presented higher estimates at the extreme DIM, whereas additive genetic variances did not. The residual variances were smaller for MY1 and remained approximately consistent across the 10 classes. However, for MY2 and MY3, the trajectory of residual variances tended to increase at the beginning of lactation, with a peak around DIM 35 to 125, and then gradually decreasing until the end of lactation. Furthermore, as the lactation number increased from MY1 to MY3, the residual variances also increased.

Table 3. Number of estimated parameters (P) and deviance information criterion (DIC) values for the models evaluated fitting homogenous (HO) or heterogeneous (HE) residual variance for the lactation curves among the 3 lactations in American Holstein cattle.

	P		DIC-MY1		DIC-MY2		DIC-MY3	
Models	**HO**	**HE**	**HO**	**HE**	**HO**	**HE**	**HO**	**HE**
LEG3	21	30	16,188,309.30	16,175,300.27	10,489,530.78	10,436,053.00	5,497,177.97	5,459,605.57
LEG4	31	40	16,028,201.35	16,020,613.90	10,392,944.62	10,340,465.13	5,441,415.14	5,406,213.50
LEG5	43	52	15,937,657.03	15,931,379.27	10,340,305.22	10,288,138.10	5,408,270.51	5,374,274.42
LEG6	57	66	15,876,727.27	15,869,904.16	10,298,831.90	10,244,926.47	5,385,753.93	5,351,078.07
BSP14	21	30	16,300,076.34	16,256,158.15	10,551,074.53	10,475,330.51	5,537,328.36	5,483,693.35
BSP15	31	40	16,146,173.20	16,118,303.64	10,459,348.60	10,391,679.29	5,482,449.41	5,435,007.44
BSP24	31	40	16,061,129.05	16,049,208.01	10,412,216.96	10,355,430.53	5,454,250.81	5,414,975.89
BSP25	43	52	15,947,451.28	15,939,231.02	10,351,304.43	10,296,914.90	5,416,494.90	5,380,020.14
BSP34	43	52	15,937,942.60	15,930,800.59	10,343,690.58	10,290,119.00	5,416,311.65	5,381,144.99
BSP35	57	66	15,876,182.07	15,869,974.88	10,301,038.12	10,247,003.63	5,389,119.19	5,352,040.97
WD	13	22	16,240,742.74	16,237,016.54	10,508,542.37	10,471,582.62	5,507,949.92	5,482,903.86
WL	13	22	16,284,179.20	16,279,414.22	10,517,589.40	10,483,940.12	5,510,281.58	5,488,575.54
AS	31	40	16,032,353.23	16,028,899.29	10,383,509.43	10,338,642.32	5,435,165.28	5,405,692.50

MY1: milk yield in first lactation; MY2: milk yield in second lactation; MY3: milk yield in third lactation; LEG*i*: where *i* correspond to Legendre orthogonal polynomials order ranging from 3 to 6; BSP*ij*: where $i = 1, 2, 3$ represents linear, quadratic, and cubic order, respectively, and $j = 4$ or 5 represents 4 (DIM=5, 105, 205, and 305) or 5 (DIM=5, 80, 155, 230, and 305) equidistant segments for B-splines functions; WD: Wood; WL: Wilmink; AS: Ali and Schaeffer; P: number of parameters.

Heritability Estimates and Genetic Correlations

The shape of the heritability curves for MY1, MY2, and MY3 was similar, with lower values around the extremes DIM, increasing smoothly in subsequent days (Figure 3). The heritability curves showed a gradual increase until they reached DIM 280, after which they decreased. The heritability curve for MY2 was reasonably flat until DIM 155. The daily heritability estimates ranged from 0.12 to 0.22 for MY1, 0.11 to 0.19 for MY2, and 0.03 to 0.15 for MY3. Higher heritabilities were estimated around DIM 280 for MY1, MY2, and MY3. Compared with the other lactations, the first lactation had higher heritabilities, across DIM estimates. These estimates tend to decrease from the second to the third lactation.

The genomic-based genetic correlations among DMY are presented in Table 4, showing moderate to lower estimates between DIM 5 and DIM 305. This outcome may be attributed to the intrinsic variability present along DIM, but it could also indicate that different sets of genes are responsible for the phenotypic expression of DMY across

lactation stages. Overall, low correlations were observed between further DIM, but there was an increase in estimates of correlation between the days spanning the lactation period from 65 to 275 d. Genetic correlations between DIM in MY1, MY2, and MY3 decreased gradually as the time interval increased.

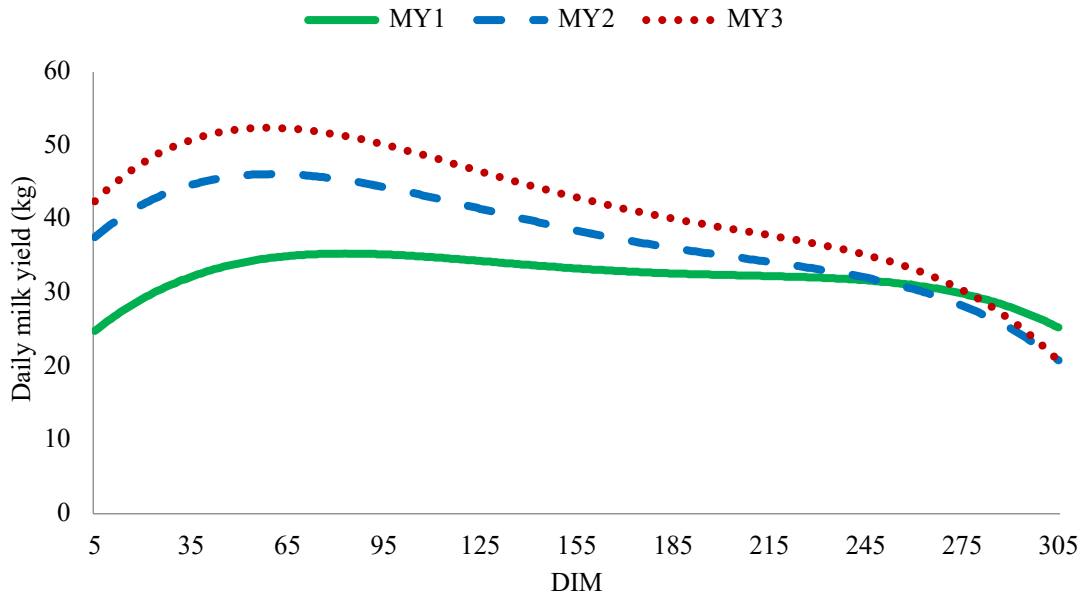

Figure 1. Phenotypic average estimated in the first (MY1), second (MY2), and third (MY3) lactation for daily milk yield (DMY) in American Holstein cattle according to DIM, obtained based on random regression models using the LEG4 model with HE residual variance.

Genetic Parameters for Lactation Persistency

The heritability estimates and the genetic correlation for the 10 measurements of LP (PS_i) and 305-d MY (MY_{305}) are shown in Table 5. PS_1 to PS_{10} represent the first to the tenth measures of LP for MY1, MY2, and MY3, respectively, with PS_{10} being adjusted based on the peak yield day. The heritability estimates for the 10 measures ranged from 0.05 to 0.18 for MY1, from 0.06 to 0.20 for MY2, and from 0.05 to 0.16 for MY3. The heritability estimates for the PS_5 and PS_6 were similar for MY2 and MY3. Overall, the heritability estimate values were of low to moderate magnitude. The highest heritability estimates were observed for PS_3 (MY1 and MY2) and PS_7 (MY3). These measures are indicated as selection criteria due to their greater heritabilities and low genetic correlations with MY_{305}. The heritability estimates for MY_{305} decreased according to parity (0.16 to 0.12).

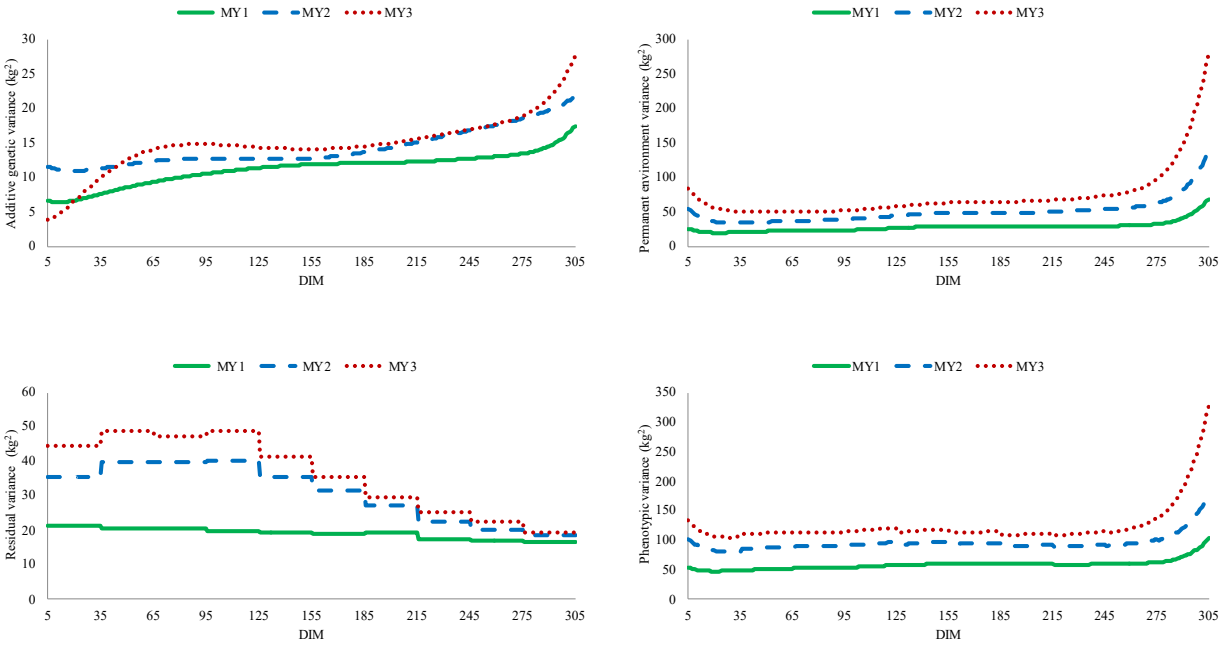

Figure 2. Additive genetic variance, permanent environmental variance, residual variance, and phenotypic variance estimated in the first (MY1), second (MY2), and third (MY3) lactation for daily milk yield (DMY) in American Holstein cattle according to DIM based on random regression models using the LEG4 model with HE residual variance.

The genetic correlations between LP traits (Table 5) were generally high for MY1, MY2, and MY3, indicating a strong genetic association between the measures. The negative genetic correlation between PS_5 is related to the method used to define these lactation persistency measures; conversely, a high positive genetic correlation indicates differences in milk production between 2 similar parts of the lactation. The genetic correlation estimates between PS_i and MY_{305} ranged from low to moderate (-0.10 to 0.46) for MY1, MY2, and MY3. However, PS_2 showed the highest genetic correlation values with MY_{305} over the first 3 lactations (0.46, 0.35, and 0.44 for the first, second, and third lactations, respectively). Low genetic correlations between LP and MY_{305} are preferable, as they suggest null or weak association between LP and milk yield, assuming that the animals with similar MY_{305} may have differences in their LP or the shape of the lactation curve. It is of interest to increase LP without impacting MY_{305}. Therefore, PS_3 is recommended as the selection criterion in the first 3 lactations of American Holstein cattle.

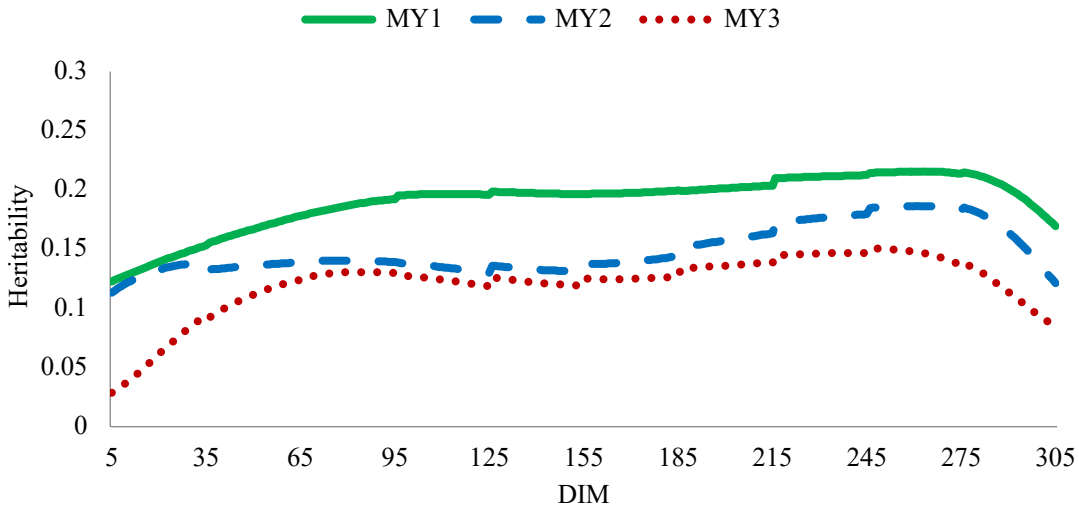

Figure 3. Heritability (h2) estimated in the first (MY1), second (MY2), and third (MY3) lactation for daily milk yield (DMY) in American Holstein cattle according to DIM, obtained based on random regression models using the LEG4 model with HE residual variance.

DISCUSSION

Random regression models have become the most commonly adopted models for routine genetic evaluations of longitudinal traits in dairy populations, as they accurately account for genetic and environmental effects throughout lactation. Daily milk yield records and genomic information from American Holstein cattle were analyzed in this study to estimate genomic-based genetic parameters and establish the optimal indicators of LP in the first 3 lactations. Single-trait RRM were used to estimate (co)variance components for DMY and LP in the first 3 lactations, which provided valuable insights of the trait over time and could enable the successful implementation of genomic selection for more persistent and extended lactations in high-producing dairy cows. Extended lactation strategy holds the potential to improve dairy cow longevity, lifetime efficiency, and farm profitability (van Knegsel et al., 2022).

In this study, the models were compared based on various criteria, including goodness of fit and parsimony, estimates of variance components, heritabilities, genetic correlations, credibility and interpretations of the results, and computational demand and time. Overall, DIC, MPP, and weighted average of heterogeneous residual variance estimates tended to be in favor of more complex models with higher orders of LEG or BSP, indicating better goodness of fit. Breda et al. (2010), Li et al. (2020), and Silva et al. (2023) also reported similar results when comparing the effect of LEG and BSP order

in genetic evaluation of Murrah buffaloes and dairy cattle. However, models with higher orders have more parameters, which could result in atypical results and increased computational demand and time. Consequently, in practice, model selection needs to balance goodness of fit and model complexity.

Table 4. Estimates of genetic correlation between daily milk yield (DMY) across DIM in three lactations (MY1, MY2, and MY3) in American Holstein cattle based on random regression models (RRM) using the LEG4 model with HE residual variance.

		DIM										
MY (kg)	DIM	5	35	65	95	125	155	185	215	245	275	305
MY1	5	-	0.807	0.621	0.509	0.447	0.426	0.436	0.457	0.464	0.431	0.301
	35		-	0.951	0.864	0.776	0.719	0.701	0.705	0.699	0.641	0.433
	65			-	0.972	0.913	0.862	0.832	0.816	0.794	0.741	0.563
	95				-	0.982	0.950	0.920	0.891	0.857	0.814	0.686
	125					-	0.990	0.969	0.936	0.897	0.865	0.773
	155						-	0.992	0.966	0.930	0.904	0.824
	185							-	0.990	0.965	0.940	0.846
	215								-	0.992	0.968	0.848
	245									-	0.986	0.858
	275										-	0.924
	305											-
MY2	5	-	0.872	0.694	0.567	0.483	0.428	0.389	0.351	0.294	0.188	-0.002
	35		-	0.956	0.888	0.820	0.750	0.677	0.602	0.518	0.397	0.195
	65			-	0.981	0.941	0.880	0.804	0.722	0.635	0.519	0.329
	95				-	0.986	0.945	0.882	0.808	0.727	0.620	0.438
	125					-	0.985	0.944	0.886	0.818	0.720	0.543
	155						-	0.986	0.950	0.898	0.812	0.641
	185							-	0.989	0.956	0.885	0.726
	215								-	0.989	0.939	0.801
	245									-	0.979	0.877
	275										-	0.955
	305											-
MY3	5	-	0.746	0.591	0.503	0.464	0.475	0.522	0.555	0.503	0.271	-0.210
	35		-	0.970	0.911	0.837	0.769	0.723	0.680	0.595	0.378	-0.058
	65			-	0.980	0.927	0.858	0.792	0.725	0.625	0.425	0.041
	95				-	0.980	0.932	0.868	0.790	0.681	0.489	0.138
	125					-	0.983	0.936	0.862	0.753	0.570	0.239
	155						-	0.982	0.929	0.833	0.664	0.340
	185							-	0.980	0.913	0.763	0.439
	215								-	0.974	0.860	0.545
	245									-	0.948	0.684
	275										-	0.873
	305											-

Breda et al. (2010), Li et al. (2020), and Silva et al. (2023) also reported similar results when comparing the effect of LEG and BSP order in genetic evaluation of Murrah buffaloes and dairy cattle. However, models with higher orders have more parameters, which could result in atypical results and increased computational demand and time.

Table 5. Heritability (diagonal) and genetic (above diagonal) correlation estimates for different lactation persistency measures (PSi) and 305-d MY (MY305) within lactations (MY1, MY2, and MY3) in American Holstein cattle based on random regression models (RRM) using LEG4 model with HE residual variance.

		Trait										
MY	Trait	PS_1	PS_2	PS_3	PS_4	PS_5	PS_6	PS_7	PS_8	PS_9	PS_{10}	MY_{305}
MY1	PS_1	**0.088**	0.755	0.976	0.912	-0.875	0.967	0.999	0.949	0.939	0.953	0.271
	PS_2		**0.164**	0.850	0.942	-0.362	0.576	0.767	0.628	0.923	0.799	0.459
	PS_3			**0.178**	0.954	-0.780	0.893	0.973	0.913	0.967	0.953	0.381
	PS_4				**0.070**	-0.600	0.789	0.917	0.848	0.997	0.953	0.362
	PS_5					**0.054**	-0.957	-0.867	-0.852	-0.657	-0.735	-0.102
	PS_6						**0.077**	0.966	0.935	0.833	0.883	0.144
	PS_7							**0.163**	0.937	0.945	0.948	0.266
	PS_8								**0.144**	0.869	0.965	0.183
	PS_9									**0.167**	0.960	0.345
	PS_{10}										**0.054**	0.258
	MY_{305}											**0.155**
MY2	PS_1	**0.114**	0.809	0.987	0.945	-0.958	0.984	0.999	0.975	0.968	0.945	0.148
	PS_2		**0.124**	0.891	0.942	-0.619	0.692	0.807	0.763	0.922	0.942	0.353
	PS_3			**0.197**	0.981	-0.904	0.943	0.986	0.957	0.993	0.981	0.235
	PS_4				**0.059**	-0.812	0.873	0.942	0.935	0.997	1	0.251
	PS_5					**0.097**	-0.992	-0.960	-0.922	-0.856	-0.812	-0.045
	PS_6						**0.106**	0.984	0.961	0.909	0.873	0.060
	PS_7							**0.183**	0.971	0.966	0.942	0.139
	PS_8								**0.159**	0.952	0.935	0.139
	PS_9									**0.154**	0.997	0.223
	PS_{10}										**0.059**	0.251
	MY_{305}											**0.146**
MY3	PS_1	**0.108**	0.613	0.985	0.875	-0.945	0.985	0.999	0.961	0.928	0.872	0.042
	PS_2		**0.084**	0.717	0.893	-0.336	0.476	0.617	0.597	0.846	0.899	0.442
	PS_3			**0.149**	0.926	-0.887	0.942	0.982	0.955	0.961	0.924	0.197
	PS_4				**0.048**	-0.668	0.785	0.875	0.891	0.992	0.999	0.228
	PS_5					**0.116**	-0.983	-0.944	-0.873	-0.755	-0.662	0.091
	PS_6						**0.116**	0.985	0.934	0.856	0.780	-0.081
	PS_7							**0.159**	0.954	0.929	0.872	0.024
	PS_8								**0.156**	0.924	0.886	0.049
	PS_9									**0.123**	0.991	0.171
	PS_{10}										**0.048**	0.236
	MY_{305}											**0.124**

Consequently, in practice, model selection needs to balance goodness of fit and model complexity. Li et al. (2020) suggested that higher-order models may not be more optimal than lower-order models. Several studies (e.g., Bohmanova et al., 2008; Li et al., 2020) have shown that LEG performed well in random regression test-day models, but there is no gold-standard method for genomic predictions in livestock reported in the literature on choosing the optimal order of LEG in the model. The model mostly used in countries that employ RRM is the fourth-order Legendre orthogonal polynomials (Padilha et al., 2018).

In general, additive, permanent environmental, residual, and phenotypic variances increased with parity (MY1, MY2, and MY3). This indicates that the influence of genes on DMY varies along the lactation trajectory. The additive variances showed higher estimates at the extremes DIM for MY3, probably due to the reduced number of records, resulting in Runge's effect (Meyer, 2005). The Runge effect, also known as the endpoint effect or end-of-range (Meyer, 2005) refers to the problems frequently observed when using polynomial interpolation, particularly with higher-order Legendre polynomials. These errors are predominantly attributed to oscillations at the extremes of the curve. The additive variances reported in this study differ from those reported by Cobuci et al. (2005), Li et al. (2020), and Wahinya et al. (2020). Those differences may be because most studies used TDMY, which might indirectly introduce bias associated with the nondaily MY data collection system. In this study, we used longitudinal DMY records collected with milking robotic systems (Lely Astronaut robots) and milking parlor (Afimilk milking parlors), allowing for more accurate estimates of genetic parameters.

The permanent environmental variance and phenotypic variance remained constant throughout the lactation period but increased at the end of lactation, particularly for MY3. Li et al. (2020) obtained similar pattern of the variances, working with TDMY using RRM based on first to fifth orders Legendre orthogonal polynomials. The high variance estimates at the end of lactation are likely artifacts of the RRM used, due to the reduced amount of information available. Misztal (2006) attributed this poor fit at the extremes to the Legendre polynomials, suggesting that the Runge effect may be influencing the estimates generated at the extremes of the curve. The permanent environmental variance plays a crucial role in the lactation. Its constancy throughout the lactation period in this study may be attributed to all data being obtained from a single large farm in the United States, which uses AMS and milking parlors and houses over 5,000 lactating cows. Additionally, this farm has partnered with artificial insemination (**AI**) companies, utilizing semen from a diverse range of bulls over time, and owns dams

and close relatives of widely used AI bulls (Pedrosa et al., 2023; Chen et al., 2023). Consequently, it is reasonable to assume that the variance components obtained in this study are a good representation of the American Holstein population parameters. However, it is recommended that future studies use datasets from additional farms.

Heterogeneous residual variances were generally observed throughout the lactation period (Figure 2). The residual variances changed across the 10 classes defined in this study, although the differences were small for MY1. The estimated residual variances increased with the lactation number and were highest during the lactation peak, specifically between DIM 35 to 125, for MY2 and MY3. This contrasts with findings by Bohmanova et al. (2008), Li et al. (2020), and Silva et al. (2023), who reported higher residual variances between DIM 5 to 35. The highest values of residual variances observed during the lactation peak may be explained by significant cow physiological changes, as they are producing maximum milk yield. These physiological changes can result in greater variability in individual cow responses.

Milk yield in the first, second, and third lactations showed fluctuations in heritability patterns across DIM (Figure 3). The heritability estimates found in our study were slightly lower than those reported by Cobuci et al. (2005), Li et al. (2020), and Wahinya et al. (2020). These lower estimates may be related to the differences in data structure, the population sampled, and the relationship matrix used. As already mentioned, this study used DMY records and only genomic information. De los Campos et al. (2015) stated that genomic heritability, which is the proportion of trait variance explained by a linear regression on a set of markers, generally tends to be lower but more accurate compared to pedigree-based estimates.

One of the major advantages of our study is the utilization of daily milk production data, which enabled more accurate RRM analyses. Even though DMY, an economically relevant trait, presents low to moderate heritabilities (0.03 - 0.22), it has sufficient genetic variance to justify selection for this trait. Pedrosa et al. (2023) using the same dataset and third-order Legendre polynomials, reported heritability estimates for DMY ranging from 0.07 to 0.28. This difference can be attributed to the population sampled, as only data collected with the AMS were used. This suggests that the AMS provides more variability, possibly due to differences in milking frequency and highlights the need to explore potential genotype by milking system interactions.

Genetic correlations between DMY were higher and positive between DIM adjacent to each other and then decreased as the distance between DIM increased (Table

4). DMY also had a high genetic correlation around mid-lactation, consistent with findings reported in the literature (Prakash et al., 2017; Pedrosa et al., 2023; Soumri et al., 2023), suggesting a need to investigate genomic regions influencing DMY across DIM. Positive genetic correlations estimated between yields within lactations indicate that genetic gain could be achieved by selecting for yield in early lactation. However, negative correlation estimates between the extreme points of the lactation curve were found in MY2 and MY3. These negative values, which may be explained by the reduction in the number of observations at the end of lactation, were also reported by Breda et al. (2010), Oliveira et al. (2017), and Wahinya et al. (2020).

Persistency of milk yield during lactation is an important parameter of the lactation curve and can be defined as the ability of a cow to maintain a constant level of production throughout lactation. LP was defined in 10 different ways, with heritability values ranging from 0.05 to 0.20. Pereira et al. (2015) working with Gyr cattle, found a low heritability (0.08) for LP. Cobuci et al. (2004) reported heritability estimates for LP ranging from 0.11 to 0.27 for Brazilian Holstein cattle. Additionally, Jakobsen et al. (2002) working with Danish Holstein cattle, reported heritability estimates ranging from 0.09 to 0.24. The differences in heritabilities for LP measures may be attributed to the population sampled, which reflects genetic variability, the methods used to describe LP, and the statistical models used in the analyses. Furthermore, Jakobsen et al. (2002) and Pereira et al. (2012) reported heritability estimates of 0.09 and 0.10, respectively, for PS_4, which are close to the value obtained in this present study (0.07).

In general, 3 factors can affect heritability estimates for LP: the biological efficiency of the LP indicator, the statistical efficiency of the type of measurement of persistency, and the part of lactation used in the calculation of persistency (Madsen, 1975). In our study, the highest heritability estimates were observed for PS_3 and PS_7 over the first 3 lactations (0.18, 0.20, and 0.16 for the first, second, and third lactations, respectively). This is likely due to additive genetic factors, indicating that selecting animals for LP to modify the lactation curve could be effective. Furthermore, a constant MY along DIM will increase total production as well as milk quality and cows' welfare (Padilha et al., 2018).

The genetic correlations between PS_i (Table 5) were generally high, with a few moderate ones, indicating a strong genetic association between the measures. Similar results were reported in studies with dairy cattle (Jakobsen et al., 2002; Cobuci et al., 2004; Pereira et al., 2012) and buffaloes (Lázaro et al., 2024). The high and positive

genetic correlations between PS_i suggest that these measures of LP effectively evaluate the difference in milk production between 2 similar parts of the lactation. Conversely, high negative genetic correlations between PS_i are related to the method used for defining these LP measures.

In general, weak genetic correlations were found between PS_i and MY_{305}. According to Jakobsen et al. (2002) the ideal measure of LP should have high genetic variance, high heritability, and low genetic correlation with MY_{305}. This ensures that selection remains reasonably independent; in other words, when selecting to increase lactation persistence, there should be small indirect responses in milk production. High LP is associated with a slow rate of decline in production, whereas low LP is associated with a rapid rate of decline. Cobuci and Costa (2012) reported genetic correlations between different LP measures and 305-d MY ranging from -0.08 to 0.52 for the same 9 measures (PS_1 to PS_9) using fourth-order Legendre orthogonal polynomials. These estimates are close to those obtained in this study, with slight differences probably due to the population sampled and the method used. Cows with greater LP are more profitable than average cows when yield and LP are genetically correlated (Dekkers et al., 1998). Therefore, selection for both LP and milk yield is possible, enabling genetic gains in milk yield and altering the lactation curve in a desirable direction.

CONCLUSIONS

We defined statistical models and calculated genomic-based genetic parameters for DMY and various indicators of LP. The DIC and MPP criteria indicated that the LEG6 model with HE residual variance best fitted the data in the studied population. However, considering a balance between goodness of fit and parsimony, estimates of variance components, heritabilities, correlations, credibility and interpretations of the results, and practical issues such as computational demand and time, the LEG4 model with HE residual variance appeared to be the most appropriate for practical implementations in the industry. Among the different measures of LP studied, PS_3 is recommended for genetic evaluation of LP in American Holstein cattle, as they showed moderate heritability and a low genetic correlation with 305-d MY.

NOTES

This research was funded by the Agriculture and Food Research Initiative Competitive Grant number 2022-67021-37022 from the USDA National Institute of Food

and Agriculture (Washington, DC). We are very grateful for the support received from the Purdue Agriculture Data Services (West Lafayette, IN), the University of Guelph (Guelph, ON, Canada), and the Federal University of Viçosa (Viçosa, MG, Brazil). The authors would like to thank Homestead Dairy, LLC (Plymouth, IN), and especially Brian and Jill Houin, for allowing our research group to use their farm records. We also thank Lely for giving us access to their system for extracting the datasets. In addition, the first author is thankful to the Minister of Employment and Vocational Training (MINEFOP) of the Republic of Cameroon for granting permission for study leave and to the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES, Brasília, DF, Brazil) for granting financial assistance for doctoral studies.

REFERENCES

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

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. https://doi.org/10.4141/cjas87-067.

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. https://doi:10.3168/jds.2007-0945.

Breda, F. C., L. G. Albuquerque, R. F. Euclydes, A. B. Bignardi, F. Baldi, R. A. Torres, L. Barbosa, and H. Tonhati. 2010. Estimation of genetic parameters for milk yield in Murrah buffaloes by Bayesian inference. J. Dairy Sci. 93:784–791. https://doi.org/10.3168/ jds.2009-2230.

Brito, L.F., N. Bedere, F. Douhard, H.R. Oliveira, M. Arnal, F. Peñagaricano, A.P. Schinckel, C.F. Baes, and F. Miglior. 2021. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. Animal 15:100292. https://doi.org/10.1016/j.animal.2021.100292.

Carlström, C., G. Pettersson, K. Johansson, E. Strandberg, H. Stålhammar, and J. Philipsson. 2013. Feasibility of using automatic milking system data from

commercial herds for genetic analysis of milkability. J. Dairy Sci. 96:5324–5332. https://doi.org/10.3168/ jds.2012-6221.

Chen, S.-Y., J. P. Boerman, L. S. Gloria, V. B. Pedrosa, J. Doucette, and L. F. Brito. 2023. Genomic-based genetic parameters for resilience across lactations in North American Holstein cattle based on variability in daily milk yield records. J. Dairy Sci. 106:4133–4146. https://doi.org/10.3168/jds.2022-22754.

Cobuci, J. A., and C. N. Costa. 2012. Persistency of lactation using random regression models and different fixed regression modeling approaches. R. Bras. Zootec. 41:1996-2004. https://doi.org/10.1590/S1516-35982012000900005.

Cobuci, J. A., R. F. Euclydes, C.N. Costa, P.S. Lopes, R.A. Torres, and C.S. Pereira. 2004. Analysis of persistency in the lactation of Holstein cows using test-day yield and random regression model. R. Bras. Zootec. 33:546-554. https://doi.org/10.1590/S1516-35982004000300004.

Cobuci, J. A., R. F. Euclydes, P.S. Lopes, C.N. Costa, R.A. Torres, and C.S. Pereira. 2005. Estimation of genetic parameters for test-day milk yield in Holstein cows using a random regression model. Genet. Mol. Biol. 28:75-83. https://doi.org/10.1590/S1415-47572005000100013.

Cole, J. B., and D. J. Null. 2009. Genetic evaluation of lactation persistency for five breeds of dairy cattle. J. Dairy Sci. 92:2248-2258. https://doi.org/10.3168/jds.2008-1825.

de los Campos, G., D. Sorensen, D. Gianola. 2015. Genomic Heritability: What Is It? PLoS Genet. 11:e1005048. https://doi.org/10.1371/journal.pgen.1005048.

de Roos, A.P.W., A.G.F. Harbers, and G. de Jong, 2001. Random regression test-day model in The Netherlands. Interbull Bull. 27:155-158.

Dekkers, J. C. M., J. H. Ten Hag, and A. Weersink. 1998. Economic aspects of persistency of lactation in dairy cattle. Livest. Prod. Sci. 53:237-252. https://doi.org/10.1016/S0301-6226(97)00124-3.

Geweke, J. 1992. Comment: Inference and prediction in the presence of uncertainty and determinism. Stat. Sci. 7:94-101. https://doi.org/ 10.1214/ss/1177011446.

Guarini, A. R., D. A. L. Lourenco, L. F. Brito, M. Sargolzaei, C. F. Baes, F. Miglior, I. Misztal, and F. S. Schenkel. 2018. Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. J. Dairy Sci. 101:8076–8086. https://doi.org/10.3168/jds.2017-14193.

Jakobsen, J. H., P. Madsen, J. Jensen, J. Pedersen, L. G. Christensen, and D. A. Sorensen. 2002. Genetic parameters for milk production and persistency for Danish Holstein estimated in random regression models using REML. J. Dairy Sci. 85:1607-1616. https://doi.org/10.3168/jds.S0022-0302(02)74231-8.

Jamrozik, J., L.R. Schaeffer, and J.C.M. Dekkers. 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. J. Dairy Sci. 80:1217-1226. https://doi.org/10.3168/jds.S0022-0302(97)76050-8.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979–993. https://doi.org/10.1093/genetics/124.4.979.

Kistemaker, G.J. 2003. Comparison of persistency definitions in random regression test day models. Interbull Bull. 30:96-98.

Lázaro, S. F., H. Tonhati, H. R. Oliveira, A. A. Silva, D. C. B. Scalez, A. V. Nascimento, D. J. A. Santos, G. Stefani, I. S. Carvalho, A. F. Sandoval, and L. F. Brito. 2024. Genetic parameters and genome-wide association studies for mozzarella and milk production traits, lactation length, and lactation persistency in Murrah buffaloes. J. Dairy Sci. 107:992–1021. https://doi.org/10.3168/jds.2023-23284.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. https://doi.org/10.3168/jds.2009-2061.

Li, J., H. Gao, P. Madsen, R. Li, W. Liu, P. Bao, G. Xue, Y. Gao, X. Di, and G. Su. 2020. Impact of the order of Legendre polynomials in random regression model on genetic evaluation for milk yield in dairy cattle population. Front. Genet. 11:586155. https://doi.org/10.3389/fgene.2020.586155.

Madsen, O. 1975. A comparison of some suggested measures of persistency of milk yield in dairy cows. Anim. Sci. 20:191–197. https://doi.org/10.1017/S0003356100035182.

Menezes, G. R. D. O., R. D. A. Torres, J. L. R. Sarmento, M. T., Rodrigues, A. L. P. D. Melo, F. G. D. Silva, and L. F. Brito. 2010. Evaluation of persistency lactation measures of Saanen goats under random regression model. R. Bras. Zootec. 39:1691-1698. https://doi.org/10.1590/S1516-35982010000800010.

Meyer, K. 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. Genet. Sel. Evol. 37:473. https://doi.org/10.1186/1297-9686-37-6-473.

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

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for BLUPF90 family of programs. University of Georgia, Athens. Accessed Jan. 10, 2024. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all8.pdf/

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, S. Miller, F. Schenkel, S. S. Moore and P. Stothard. 2017. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. J. Dairy Sci. 100:1246–1258. https://doi.org/10.3168/jds.2016-11770.

Oliveira, H. R., L. F. Brito, D. A. L. Lourenco, F. F. Silva, J. Jamrozik, L. R. Schaeffer, and F. S. Schenkel. 2019a. Invited review: Advances and applications of random regression models: From quantitative genetics to genomics. J. Dairy Sci. 102:7664–7683. https://doi.org/10.3168/jds.2019-16265.

Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, and F. S. Schenkel. 2019b. Application of single-step genomic evaluation using multiple-trait random regression test-day models in dairy cattle. J. Dairy Sci. 102:2365–2377. https://doi.org/10.3168/jds.2018-15466.

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. 2017. Bayesian Models combining Legendre and B-spline polynomials for genetic analysis of multiple lactations in Gyr cattle. Livest. Sci. 201:78-84. https://doi.org/10.1016/j.livsci.2017.05.007.

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. J. Anim. Sci. 94:1865–1874. https://doi.org/10 .2527/jas.2015-0150.

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. 59:207-215. https://doi.org/10.1071/AN16835.

Pedrosa, V. B., J. P. Boerman, L. S. Gloria, S. Y. Chen, M. E. Montes, J. S. Doucette, and L. F. Brito. 2023. Genomic-based genetic parameters for milkability traits

derived from automatic milking systems in North American Holstein cattle. J. Dairy Sci. 106:2613-2629. https://doi.org/10.3168/jds.2022-22515.

Pedrosa, V. B., F. S. Schenkel, S. Y. Chen, H. R. Oliveira, T. M. Casey, M. G. Melka, and L. F. Brito 2021. Genome wide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. Genes (Basel), 12:1830. https://doi.org/10.3390/genes12111830.

Pereira, R. J., D. R. Ayres, L. E. Faro, A. E. V. Filho, R. S. Verneque, and L. G. Albuquerque. 2015. A new way to measure milk yield persistency: a genetic point of view with application to Gyr (Bos indicus) cattle. J. Dairy Res. 82:385-390. https://doi:10.1017/S0022029915000448.

Pereira, R. J., R. S. Verneque, P. S. Lopes, M. L. Santana Jr., M. R. Lagrotta, R. A. Torres, A.E. Vercesi Filho, and M.A. Machado. 2012. Milk yield persistency in Brazilian Gyr cattle based on a random regression model. Genet. Mol. Res. 11:1599-1609. http://dx.doi.org/10.4238/2012.June.15.9.

Prakash, V., A. Gupta, M. Singh, G. Ambhore, A. Singh, and R. Gandhi. 2017. Random regression test-day milk yield models as a suitable alternative to the traditional 305-day lactation model for genetic evaluation of Sahiwal cattle. Indian J. Anim. Sci. 87: 340–344. https://doi.org/10.56093/ijans.v87i3.68863.

Schaeffer, L. R. 2004. Application of random regression models in animal breeding. Livest. Prod. Sci. 86:35–45. https://doi.org/10 .1016/S0301-6226(03)00151-9.

Sehested, J., C. Gaillard, J.O. Lehmann, G.M. Maciel, M. Vestergaard, M.R. Weisbjerg, L. Mogensen, L.B. Larsen, N.A. Poulsen, and T. Kristensen. 2019. Review: Extended lactation in dairy cattle. Animal 13:s65-s74. https://doi.org/10.1017/S1751731119000806.

Silva, A.A., L.F. Brito, D.A. Silva, S.F. Lazaro, K.R. Silveira, G. Stefani, and H. Tonhati. 2023. Random regression models using B-splines functions provide more accurate genomic breeding values for milk yield and lactation persistence in Murrah buffaloes. J. Anim. Breed. Genet. 140:167-184. https://doi.org/10.1111/jbg.12746.

Smith, B. J. 2007. boa: An R package for MCMC output convergence assessment and posterior inference. J. Stat. Softw. 21:1–37. https: //doi.org/10.18637/jss.v021.i11.

Soumri, N., M. J. Carabaño, O. González-Recio, and S. Bedhiaf-Romdhani. 2023. Random regression models to estimate genetic parameters for milk yield, fat, and

protein contents in Tunisian Holsteins. J. Anim. Breed. Genet. 140:440-461. https://doi.org/10.1111/jbg.12770.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc. B. J. R. Stat. Soc. Series B Stat. Methodol. 64:583–639. https: //doi.org/10.1111/1467-9868.00353.

van Knegsel, A.T., Burgers, E.E., Ma, J., Goselink, R.M. and A. Kok. 2022. Extending lactation length: Consequences for cow, calf, and farmer. J. Anim. Sci., 100:skac220. https://doi.org/10.1093/jas/skac220.

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

Wahinya, P. K., M. G. Jeyaruban, A. A. Swan, A. R. Gilmour, and T. M. Magothe. 2020. Genetic parameters for test-day milk yield, lactation persistency, and fertility in low-, medium-, and high-production systems in Kenya. J. Dairy Sci. 103:10399–10413. https://doi.org/10.3168/jds.2020-18350.

Walsh, S. W., E. J. Williams, and A. C. O. Evans. 2011. A review of the causes of poor fertility in high milk producing dairy cows. Anim. Reprod. Sci., 123:127-138. https://doi.org/10.1016/j.anireprosci.2010.12.001.

Wang, A., Brito, L.F., Zhang, H., Shi, R., Zhu, L., Liu, D., Guo, G. and Y. Wang. 2022. Exploring milk loss and variability during environmental perturbations across lactation stages as resilience indicators in Holstein cattle. Front. Genet. 13:1031557. https://doi.org/10.3389/fgene.2022.1031557.

Wang, A., Su, G., Brito, L.F., Zhang, H., Shi, R., Liu, D., Guo, G. and Y. Wang. 2024. Investigating the relationship between fluctuations in daily milk yield as resilience indicators and health traits in Holstein cattle. J. Dairy Sci. 107:1535-1548. https://doi.org/10.3168/jds.2023-23495.

Wilberg, M.J., and J.R. Bence. 2008. Performance of deviance information criterion model selection in statistical catch-at-age analysis. Fish. Res. 93:212–221. https://doi.org/10.1016/j.fishres.2008.04.010.

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. https://doi.org/10.1016/0301-6226(87)90003-0.

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

SUPPLEMENTARY MATERIAL

https://doi.org/ 10.6084/m9.figshare.28926179.v1

Table S1. Model posterior probabilities (MPP) estimates for the models evaluated fitting homogenous (HO) or heterogeneous (HE) residual variance for the lactation curves among the 3 lactations in American Holstein cattle

Models	MPP-MY1		MPP-MY2		MPP-MY3	
	HO	HE	HO	HE	HO	HE
LEG3	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
LEG4	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
LEG5	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
LEG6	≈ 0.00	≈ 1.00	≈ 0.00	≈ 1.00	≈ 0.00	≈ 1.00
BSP14	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
BSP15	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
BSP24	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
BSP25	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
BSP34	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
BSP35	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
WD	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
WL	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00
AS	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00	≈ 0.00

MY1: milk yield in first lactation; MY2: milk yield in second lactation; MY3: milk yield in third lactation; LEG*i*: where *i* correspond to the order of the Legendre orthogonal polynomials ranging from 3 to 6; BSP*ij*: where i = 1, 2, 3 represents linear, quadratic, and cubic order, respectively, and j = 4 or 5 represents 4 (DIM = 5, 105, 205, and 305) or 5 (DIM = 5, 80, 155, 230, and 305) equidistant segments for B-splines functions; WD: Wood; WL: Wilmink; AS: Ali and Schaeffer.

Table S2. Heritability (diagonal) and genetic (above diagonal) correlation estimates for different lactation persistency measures (PS_i) and 305-d MY (MY_{305}) within lactations (MY1, MY2, and MY3) in American Holstein cattle based on random regression models (RRM) using the LEG6 model with HE residual variance.

		Trait										
MY	Trait	PS_1	PS_2	PS_3	PS_4	PS_5	PS_6	PS_7	PS_8	PS_9	PS_{10}	MY_{305}
MY1	PS_1	**0.095**	0.744	0.979	0.917	-0.879	0.942	0.999	0.935	0.945	0.928	0.241
	PS_2		**0.168**	0.856	0.935	-0.351	0.504	0.735	0.653	0.911	0.915	0.479
	PS_3			**0.189**	0.965	-0.780	0.865	0.974	0.903	0.979	0.967	0.355
	PS_4				**0.077**	-0.616	0.752	0.913	0.857	0.997	0.998	0.356
	PS_5					**0.059**	-0.963	-0.883	-0.823	0.673	0.641	0.050
	PS_6						**0.073**	0.948	0.938	0.799	0.782	0.090
	PS_7							**0.172**	0.936	0.941	0.924	0.222
	PS_8								**0.128**	0.882	0.889	0.203
	PS_9									**0.176**	0.997	0.331
	PS_{10}										**0.073**	0.339
	MY_{305}											**0.212**
MY2	PS_1	**0.078**	0.646	0.945	0.902	-0.950	0.971	0.996	0.922	0.940	0.820	-0.084
	PS_2		**0.071**	0.807	0.873	-0.404	0.489	0.659	0.605	0.833	0.938	0.120
	PS_3			**0.154**	0.961	-0.820	0.890	0.960	0.937	0.976	0.922	0.010
	PS_4				**0.033**	-0.722	0.800	0.914	0.871	0.995	0.976	-0.072
	PS_5					**0.088**	-0.977	-0.935	-0.849	-0.787	-0.608	0.082
	PS_6						**0.080**	0.971	0.935	0.857	0.684	-0.143
	PS_7							**0.139**	0.941	0.951	0.835	-0.113
	PS_8								**0.101**	0.901	0.764	-0.085
	PS_9									**0.093**	0.957	-0.095
	PS_{10}										**0.042**	-0.005
	MY_{305}											**0.135**
MY3	PS_1	**0.042**	0.104	0.853	0.693	-0.903	0.947	0.982	0.818	0.805	0.590	0.780
	PS_2		**0.045**	0.410	0.721	0.292	-0.080	0.142	0.243	0.616	0.805	-0.022
	PS_3			**0.069**	0.800	-0.647	0.820	0.917	0.878	0.881	0.799	-0.566
	PS_4				**0.017**	-0.317	0.525	0.713	0.749	0.982	0.949	-0.657
	PS_5					**0.071**	-0.934	-0.868	-0.631	-0.475	-0.211	0.635
	PS_6						**0.054**	0.961	0.844	0.671	0.408	-0.723
	PS_7							**0.078**	0.867	0.831	0.635	-0.734
	PS_8								**0.051**	0.825	0.610	-0.808
	PS_9									**0.045**	0.924	-0.704
	PS_{10}										**0.022**	-0.415
	MY_{305}											**0.107**

Table S3. Estimates of genetic correlation between daily milk yield (DMY) across DIM in three lactations (MY1, MY2, and MY3) in American Holstein cattle based on random regression models (RRM) using the LEG6 model with HE residual variance.

		DIM										
MY (kg)	DIM	5	35	65	95	125	155	185	215	245	275	305
MY1	5	-	0.758	0.584	0.476	0.422	0.412	0.426	0.441	0.435	0.418	0.450
	35		-	0.946	0.854	0.785	0.746	0.721	0.690	0.642	0.596	0.571
	65			-	0.972	0.923	0.879	0.842	0.806	0.761	0.714	0.660
	95				-	0.985	0.953	0.921	0.889	0.849	0.805	0.737
	125					-	0.990	0.970	0.944	0.906	0.862	0.791
	155						-	0.994	0.975	0.939	0.898	0.827
	185							-	0.992	0.966	0.929	0.861
	215								-	0.990	0.965	0.900
	245									-	0.991	0.936
	275										-	0.968
	305											-
MY2	5	-	0.733	0.550	0.473	0.442	0.397	0.300	0.191	0.149	0.142	-0.076
	35		-	0.949	0.894	0.876	0.858	0.770	0.621	0.538	0.506	0.234
	65			-	0.986	0.974	0.963	0.899	0.770	0.683	0.624	0.361
	95				-	0.996	0.983	0.917	0.790	0.710	0.652	0.377
	125					-	0.991	0.923	0.795	0.721	0.673	0.384
	155						-	0.964	0.862	0.794	0.739	0.469
	185							-	0.964	0.912	0.835	0.634
	215								-	0.980	0.900	0.780
	245									-	0.963	0.869
	275										-	0.907
	305											-
MY3	5	-	0.588	0.448	0.405	0.366	0.303	0.216	0.133	0.128	0.184	0.030
	35		-	0.959	0.922	0.905	0.885	0.834	0.733	0.708	0.700	0.253
	65			-	0.988	0.963	0.946	0.931	0.877	0.868	0.816	0.400
	95				-	0.988	0.972	0.947	0.880	0.866	0.811	0.369
	125					-	0.992	0.949	0.846	0.806	0.748	0.269
	155						-	0.971	0.867	0.798	0.699	0.256
	185							-	0.955	0.879	0.703	0.392
	215								-	0.959	0.744	0.579
	245									-	0.880	0.680
	275										-	0.684
	305											-

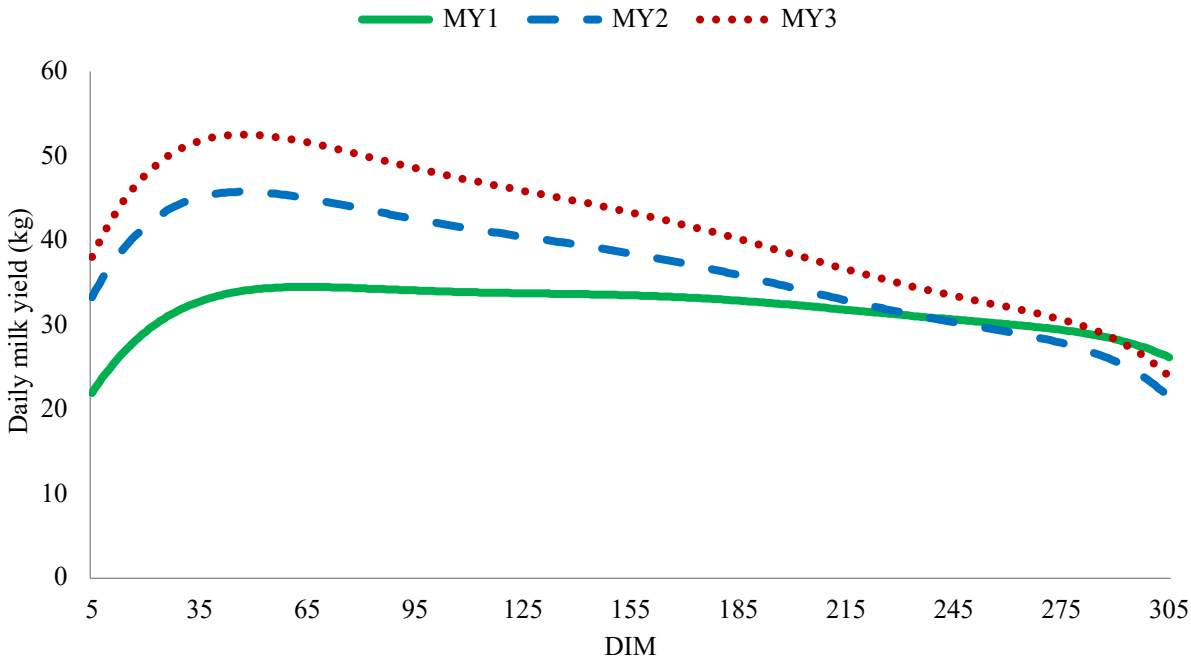

Figure S1. Phenotypic average estimated in the first (MY1), second (MY2), and third (MY3) lactation for daily milk yield (DMY) in American Holstein cattle according to DIM, obtained based on random regression models using the LEG6 model with HE residual variance.

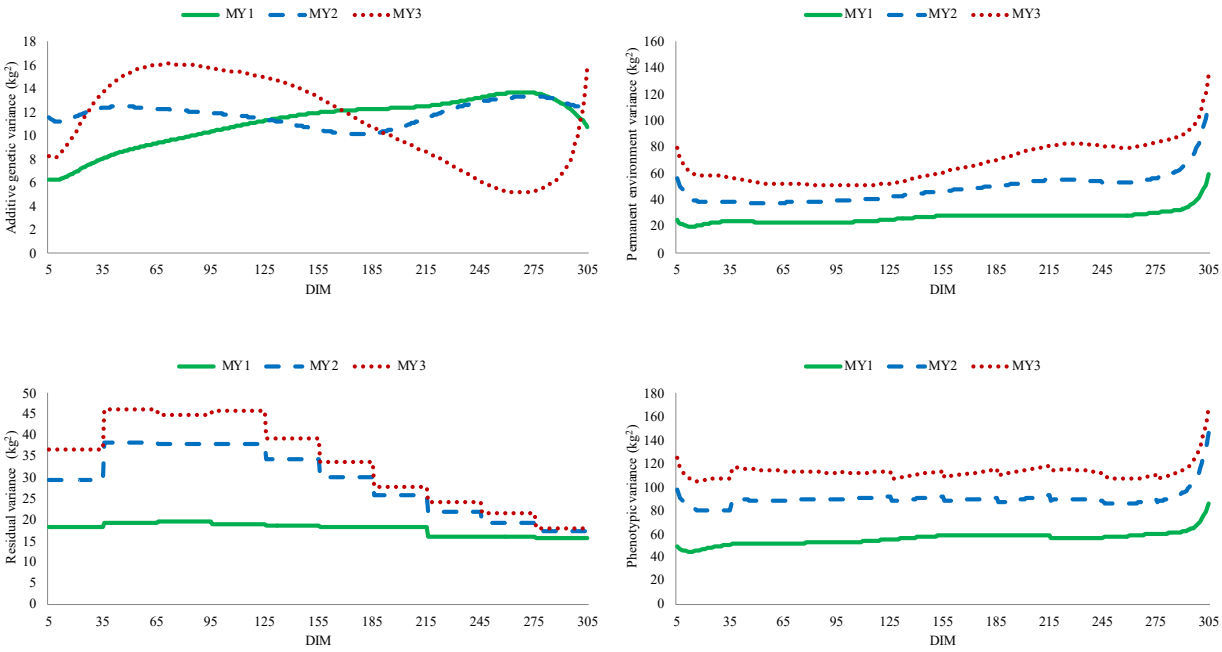

Figure S2. Additive genetic variance, permanent environmental variance, residual variance, and phenotypic variance estimated in the first (MY1), second (MY2), and third (MY3) lactation for daily milk yield (DMY) in American Holstein cattle according to DIM based on random regression models using the LEG6 model with HE residual variance.

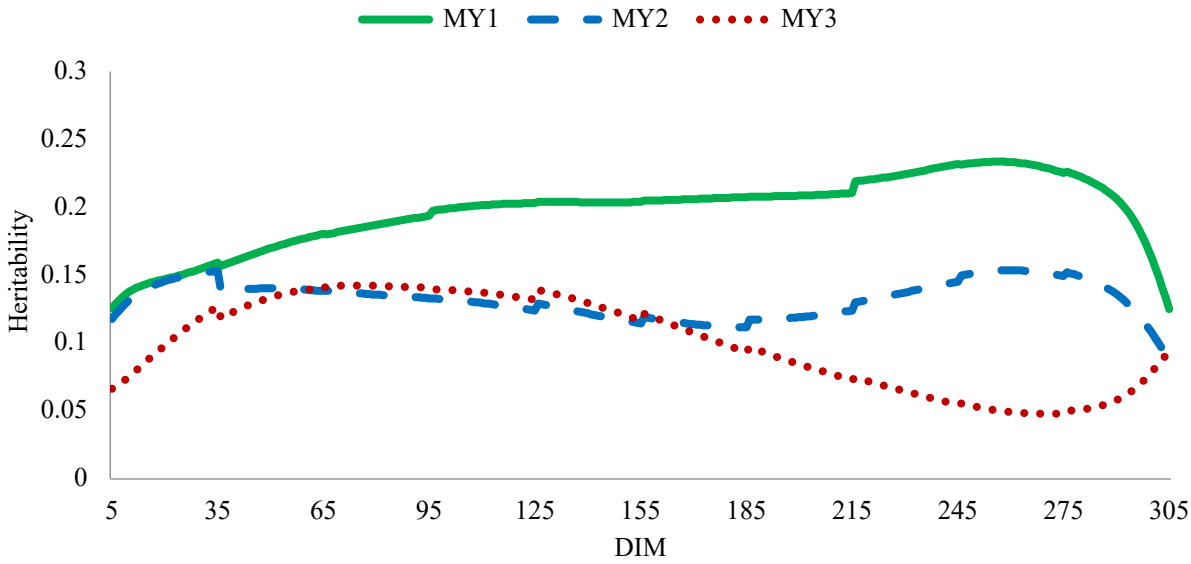

Figure S3. Heritability (h^2) estimated in the first (MY1), second (MY2), and third (MY3) lactation for daily milk yield (DMY) in American Holstein cattle according to DIM, obtained based on random regression models using the LEG6 model with HE residual variance.

IV – CAPÍTULO II

This chapter will be submitted to the Journal of Dairy Science

Genomic background of various lactation curve parameters in American Holstein cattle

Patrick R. Fotso-Kenmogne[1,2], Paulo L. S. Carneiro[3,*], Delvan A. Silva[4], Sirlene F. Lázaro[5], Pedro F. C. Aponte[1], Jaime A. Cobuci[6], Hinayah R. Oliveira[7], and Luiz F. Brito[7,*]

[1] Postgraduate Program in Animal Science, State University of Southwest Bahia, Itapetinga, 45700-000, BA, Brazil

[2] Ministry of Employment and Vocational Training (MINEFOP), Yaoundé, P.O. Box 16273, CE, Cameroon

[3] Department of Biology, State University of Southwest Bahia, Jequié, 45206-190, BA, Brazil

[4] Department of Animal Sciences, Federal University of Viçosa, Viçosa, 36570-000, MG, Brazil

[5] Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, N1G2W1, Canada

[6]Animal Science Department, Federal University of Rio Grande do Sul, Porto Alegre, 91540-000, RS, Brazil

[7] Department of Animal Sciences, Purdue University, West Lafayette, 47907, IN, USA

[*]**Corresponding authors:**

Luiz F. Brito. **E-mail**: britol@purdue.edu **Address**: Department of Animal Sciences, Purdue University, West Lafayette, 47907, IN, USA

Paulo L. S. Carneiro. **E-mail**: plscarneiro@uesb.edu.br **Address:** Department of Biology, State University of Southwest Bahia, Jequié, BA, 45205-490, Brazil

ABSTRACT

Lactation curve parameters (LCP) are essential in the design of genetic improvement programs of dairy cattle due to their relationship with the shape of the lactation curve and their biological interpretations. We investigated genomic regions influencing LCP in Holstein heifers and cows using longitudinal genome-wide association approach. Therefore, the objectives of this study were to perform genome-wide association studies and functional enrichment analyses of various LCP (Wood [WD], Wilmink [WL], and Ali-Schaeffer [AS]) in American Holstein cattle. We used 2,754,840 records for heifers and 1,642,653 records for cows with 11,139 and 6,735 animals, respectively, born between 2012 and 2019. The analysis included 14,464 genotyped animals with 60,277 SNP. The SNP effects were estimated for the random regression coefficients using GBLUP method. Significant markers were identified using the strict Bonferroni correction and the modified Bonferroni correction, based on the number of independent chromosomal segments. Genes and QTL located up to 100 kb upstream or downstream of the significant SNP were reported, and functional enrichment analyses of the candidate genes identified for each LCP were performed. In heifers, 81, 128, and 196 significant SNP were identified for the WD, WL, and AS parameters, respectively; whereas in cows, 120, 125, and 128 significant SNP were identified for the same parameters. The significant SNP were located on chromosomes BTA1, BTA2, BTA3, BTA5, BTA6, BTA7, BTA8, BTA9, BTA11, BTA13, BTA14, BTA15, BTA17, BTA18, BTA19, BTA20, BTA23, and BTA29. One genomic region (BTA14: 1,801,116 bp) was common to all the parametric functions (WD, WL, and AS). The genomic regions located on BTA15 and BTA19 were unique to the WL parameters; whereas those located on BTA3 and BTA20 were unique to the AS parameters. Ten candidate genes (*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1,* and *TSNARE1*) were common to all the parametric functions. No significant gene ontology terms were found for the parameter *c* (WD), for heifers and for the parameters *d* and *f* (AS), for cows. This study unveiled the candidate genes associated with LCP; the identified candidate genes will contribute to a better understanding of the genomic basis of LCP in American Holstein cattle.

Key words: dairy cattle, genes, lactation curve parameters, longitudinal genome-wide association, parametric functions

INTRODUCTION

Milk production is the most intensively selected trait in dairy cattle breeding programs worldwide due to its significant economic impact on the industry (Brito et al., 2021; Pedrosa et al., 2021). In this context, modeling lactation curves is essential, as it enables the characterization of important traits such as lactation peak and persistency, and novel indicators of resilience (Chen et al., 2023; Oliveira et al., 2024). To model the lactation curve, various parametric functions have been developed, including those proposed by Wood (**WD**; Wood, 1967), Wilmink (**WL**; Wilmink, 1987), and Ali-Schaeffer (**AS**; Ali and Schaeffer, 1987). Beyond describing the lactation curve, unraveling the genomic architecture of its parameters (i.e., *a*, *b*, *c* for WD and WL functions, and *a*, *b*, *c*, *d*, *f* for AS function) is paramount, as they have meaningful biological interpretations and can be incorporated into selection indices to change the shape of the lactation curve. Additionally, lactation curve parameters (**LCP**) are heritable, indicating that the average lactation curve of a population can be modify through selective breeding. Furthermore, LCP provides valuable insights for herd management and breeding strategies.

Selecting animals based on their complete lactation curves integrates both the lactation biology and the physiological implications of different curve patterns, contributing to the mitigation of energy deficits during early and peak lactation while improving animal health, welfare, and milk production efficiency (Oliveira et al., 2019b). In line with this selection strategy, several studies have reported the genetic architecture of milk production. For instance, Oliveira et al. (2019c) identified 407 SNP associated with milk yield (**MY**) in Canadian Holstein cattle, while Strucken et al. (2012) investigated the time-dependent genetic effects for milk production traits in German dairy cattle, revealing that numerous genomic regions influence milk production.

Identifying genomic regions and candidate genes associated with LCP is essential to elucidate the biological mechanisms underlying milk production. Genome-wide association studies (**GWAS**) have been widely performed for this purpose, but most focus on milk production traits rather than on lactation curve parameters. For instance, Lázaro et al. (2024) performed a GWAS for mozzarella, milk production traits, lactation length, and lactation persistency in Murrah buffaloes using parametric functions and identified several candidate genes. However, similar studies on the genomic background of LCP in dairy cattle remain limited. Therefore, the objectives of this study were to (1) perform a GWAS for LCP (WD, WL, and AS) aiming to identify key genomic regions and candidate

genes influencing these parameters in American Holstein cattle; and (2) conduct functional genomic analysis to better understand the biological pathways associated.

MATERIALS AND METHODS

Ethics Statement

Animal Care and Use Committee approvals were unnecessary for this study, as data were obtained from pre-existing databases.

Datasets

The data were obtained from a large commercial dairy cattle farm in North-Central Indiana. The phenotype and genomic datasets used were described in our previous study (Fotso-Kenmogne et al., 2025). In brief, 7,083,609 daily milk yield (**DMY**) records were initially collected using Lely Astronaut robots (automatic milking systems; **AMS**) and a milking parlor (Afimilk system) for 29,343 lactations of 14,469 Holstein cows. Data editing was performed to remove records with missing information, biologically-impossible values, sparse recording (<30 records in total, or time span <70 d, or average day interval >2), truncated records, records from the fourth or higher lactations, from cows with first calving before 640 d of age, and within small (<10 cows) contemporary groups (**CG**, defined by concatenating calving year and season [March to May, June to August, September to November, December to February] of calving).

In this study, only DMY records within the range of 5 and 305 DIM were considered, and DMY records deviating ±3.5 phenotypic SD from the mean within the CG were removed from further analyses. Finally, 4,397,493 records from 17,874 lactations of 11,788 animals remained for further analyses, including 2,754,840 DMY of 11,139 heifers; and 1,642,653 DMY of 6,735 cows. The descriptive statistics of the datasets are presented in Table 1.

The genomic dataset contained 72,820 SNP for 22,090 animals that were imputed from multiple SNP panels as described by Chen et al. (2023). A quality control (**QC**) was performed to remove SNP and animals with call rate lower than 0.90, minor allele frequency lower than 0.01, duplicated samples, and extreme deviation from the Hardy-Weinberg equilibrium ($P < 10^{-8}$). After the QC, the genomic dataset contained 60,277 SNP and 14,464 animals.

Statistical Analyses

The analyses were performed using single-trait random regression models (**RRM**). These models included the average curve, systematic effects (CG and milking

system – AMS or parlor), calving age as linear and quadratic covariates, additive genetic, permanent environment, and residual effects. The random regressions for additive genetic and permanent environmental effects, as well as fixed regression (average curve) were modelled using nonlinear functions (WD, WL, and AS). The general model applied for WD, WL, and AS functions can be described as:

$$\mathbf{y_{ijkl}} = \boldsymbol{\mu} + \mathbf{CG_k} + \mathbf{DATS_l} + \sum_{\mathbf{q=0}}^{\mathbf{q}} \mathbf{a_{iq}}\,\boldsymbol{\psi_{ijq}} + \sum_{\mathbf{q=0}}^{\mathbf{q}} \boldsymbol{p_{iq}}\,\boldsymbol{\psi_{ijq}} + \mathbf{e_{ijkl}}\text{ , [1]}$$

where $\mathbf{y_{ijkl}}$ is the vector of the DMY recorded i (i = 1, 2, …, i) of the animal for trait measured on DIM j (j = 5, 6, . . ., 305), with the CG_k (k= 1, 2, …, K), and data source ($\mathbf{DATS}_l$ = 1 for Lely Astronaut robots or 2 for Afimilk milking parlors) of DMY; μ is the overall mean and the fixed regression coefficient; a_{iq} and p_{iq} are the random regression coefficients associated with covariate q for the additive genetic and permanent environmental effects of animal i, respectively; ψ_{ijq} is the qth covariate for DIM j from animal i and e_{ijkl} is the random residual term. The parameters that define the scale or shape of the lactation curve were defined according to WD, WL, and AS. These WD, WL, and AS functions can be described, respectively, as follows:

$$\Lambda_{ij0} = \mathbf{1},\ \Lambda_{ij1} = \mathbf{ln}\ \boldsymbol{t},\ \text{and}\ \Lambda_{ij2} = (-\boldsymbol{t}),\ [2]$$

$$\boldsymbol{\psi_{ij0}} = \mathbf{1},\ \boldsymbol{\psi_{ij1}} = \mathbf{exp}(-\mathbf{0.05}\boldsymbol{t}),\ \text{and}\ \boldsymbol{\psi_{ij2}} = \boldsymbol{t},\ [3]$$

$$\mathbf{\Gamma_{ij0}} = \mathbf{1},\ \mathbf{\Gamma_{ij1}} = \mathbf{t/305},\ \mathbf{\Gamma_{ij2}} = (\boldsymbol{t}/\mathbf{305})^{\mathbf{2}},\ \mathbf{\Gamma_{ij3}} = \mathbf{ln}(\mathbf{305}/\boldsymbol{t}),\ \text{and}\ \mathbf{\Gamma_{ij4}} = \boldsymbol{ln}(\mathbf{305}/\boldsymbol{t})^{\mathbf{2}},\ [4]$$

where t is the day of lactation. The variance structure of RRM was assumed as follows:

$$\mathbf{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{p}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{\Sigma_{G_o}} \otimes \mathbf{G} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma_{P_o}} \otimes \mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \boldsymbol{\sigma}_{\mathbf{e}}^{\mathbf{2}}\,\mathbf{I}\end{bmatrix},\ [5]$$

where $\mathbf{\Sigma_{G_o}}$ and $\mathbf{\Sigma_{P_o}}$ are the additive genetic and permanent environmental covariance matrices between regression coefficients of each function (WD, WL, AS); $\mathbf{G}$ is the genomic-based relationship matrix, and its inverse, $\mathbf{G^{-1}}$ was computed as in the first method proposed by VanRaden (2008); $\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$ is the residual variance, which was assumed to be homogeneous to reduce computational demand and expedite computation speed; $\mathbf{I}$ is an identity matrix, and $\otimes$ denotes the Kronecker product. The analyses were based on GBLUP rather than single-step GBLUP because no pedigree information was available.

The (co)variance components were estimated based on Bayesian inference using the Gibbs sampler algorithm implemented in the GIBBSF90+ software from the BLUPF90 family of programs (Misztal et al., 2018). Uniform a priori distributions were

assumed for the systematic effects, and Gaussian distributions were assumed for the additive genetic and permanent environmental effects; and an inverse Wishart distribution for the variance components. A chain length of 500,000 cycles was used with a burn-in of 200,000 cycles and a sampling interval of 10 cycles, which resulted in 30,000 samples for subsequent analyses. The convergence of the MCMC chains was verified by graphical inspection and based on the Geweke's criterion (Geweke, 1992), assessed using the Bayesian Output Analysis, R package (Smith, 2007).

The genomic heritabilities estimates over time were obtained based on the covariance components estimated for the regression coefficients of each function (WD, WL, AS). The posterior marginal distribution samples for heritabilities throughout the lactation curve were obtained by:

$$\mathbf{h_j^2} = \frac{\boldsymbol{\sigma_{a_j}^2}}{\boldsymbol{\sigma_{a_j}^2 + \sigma_{p_j}^2 + \sigma_e^2}}\ , \quad [6]$$

where h_j^2 is the genomic heritability estimate on DIM j, $\sigma_{a_j}^2$ is the additive genomic variance on DIM j, $\sigma_{p_j}^2$ is the permanent environmental variance on DIM j, and σ_e^2 is the residual variance.

Genomic Estimated Breeding Values, GWAS, and Functional Analyses

The additive genetic effect of the coefficients estimated were used to derive the GEBV for each DIM. The GEBV were computed using the GBLUP method (Aguilar et al., 2010). The vectors of the GEBV for all DIM of each animal were obtained as follows:

$$(\mathbf{GEBV_i}) = \boldsymbol{\Phi}\ \widehat{\boldsymbol{\delta}}_\mathbf{i}, \quad [7]$$

where $\hat{\delta}_i$ is the vector of the predicted genomic coefficients for each animal i; and $\boldsymbol{\Phi}$ is a matrix of the independent covariates for each DIM associated with the WD, WL or AS function.

After calculating the GEBV based on the RRM, the postGSf90 software (Aguilar et al., 2014) was used to back-solve the GEBV predicted from the additive genomic random regression coefficients for the SNP effects. The SNP effects calculated for the regression coefficients were used to estimate the SNP effects over time (from 5 to 305 d in lactation), as detailed in Oliveira et al. (2019c) and Lázaro et al. (2024). The SNP effects can be described as follows (Wang et al., 2012):

$$\widehat{\mathbf{u}}_\mathbf{c} = \mathbf{Z}'[\boldsymbol{ZZ}']^{-\mathbf{1}}\ \hat{\mathbf{a}}_\mathbf{g}, \quad [8]$$

where $\hat{u}_c$ is the vector of estimated SNP solutions for the cth random regression coefficient; $\mathbf{Z}$ is the matrix containing the centered genotypes; and $\hat{\mathbf{a}}_\mathbf{g}$ is a vector of GEBV

for the sth random regression coefficient estimated by the GBLUP analyses, which contains the sth random regression coefficients for all genotyped animals. Subsequently, the SNP solutions for all random regression coefficients (c = 1, 2, 3 for WD and WL functions; or c = 1, 2,…, 5 for AS function) of the same SNP k were combined into a vector ($\widehat{\mathbf{u}}_{\mathbf{k}} = [\widehat{\mathbf{u}}_{\mathbf{k1}}\ \widehat{\mathbf{u}}_{\mathbf{k2}}\ \widehat{\mathbf{u}}_{\mathbf{k3}}\ \widehat{\mathbf{u}}_{\mathbf{k4}}\ \widehat{\mathbf{u}}_{\mathbf{k5}}]'$), and used to estimate the SNP effects for all DIM (from 5 to 305 d), as proposed by Oliveira et al. (2019a):

$$\widehat{\mathbf{SNP}_{\mathbf{k}}} = \mathbf{T}\widehat{\mathbf{u}}_{\mathbf{K}}, \quad [9]$$

where $\widehat{\mathbf{SNP}_{\mathbf{k}}}$ is the vector that contains the SNP effects estimated for every DIM of the kth SNP; **T** is a matrix of covariates for each DIM, associated with the WD, WL, and AS functions; and $\widehat{\mathbf{u}}_{\mathbf{K}}$ is the vector of SNP solutions for all random regression coefficients related to the kth SNP. To identify significant SNP, approximate *P*-values of SNP effects were calculated by:

$$\boldsymbol{P-\mathbf{values}_{\mathbf{ij}}} = \mathbf{2}\left[\mathbf{1}-\boldsymbol{\Phi}\left(\left|\frac{\hat{\mathbf{g}}_{\mathbf{ij}}}{\mathbf{sd}\ (\hat{\mathbf{g}}_{\mathbf{ij}})}\right|\right)\right], \quad [10]$$

where $P-values_{ij}$ is the *P*-value associated with the jth SNP in parameter i and Φ is the cumulative standard normal function (Gualdrón Duarte et al., 2014).

The genomic inflation factor (λ), defined as the median of the observed chi-squared test statistics divided by the expected median of the corresponding chi-squared distribution [$\boldsymbol{\lambda = \mathbf{median}(\chi^2)/0.456}$], [11] was used to assess population stratification.

The SNP effects were ranked according to the significance of association with each parameter and the cut-off for considering a significant association was established using a Bonferroni correction for multiple testing, which was equals to 6.08 [-log10 (0.05/60,277)], where 60,277 represents the total number of SNP. Some disadvantages of this much stricter Bonferroni threshold are that it is too conservative, it does not consider the dependence among the tests due to strong linkage disequilibrium between the SNP and it is known for being less tolerant to type I errors than other adaptations of this method or other methods. Therefore, many biologically important associations might not be identified when considering this strict Bonferroni correction (i.e., higher incidence of type II errors-false negatives). Hence, to avoid excessive false-negative results, a modified Bonferroni correction was also applied by using the number of independent chromosomal segments (**Me**) at the chromosome-wise level (Li et al., 2015) instead of the total number of SNP. The Me is a function of effective population size (Ne) and chromosome length in Morgan (L), and it was calculated as (Goddard et al., 2011): $\mathbf{Me = 2NeL/log(NeL)}$,

[12]. The effective population size (Ne) was assumed to be 66, as it was the most conservative value recently reported for a North American Holstein cattle population (Makanjuola et al., 2020). A SNP effect was considered to be statistically significant if its *P*-value was smaller than the modified Bonferroni genome-wide threshold, i.e., 0.05/Me. The value of Me and the corresponding -log10 of significance threshold used were 2,044.7 and 4.61, respectively. This approach has been used in other studies (Pedrosa et al., 2021; Chen et al., 2024; Sousa Junior et al., 2024).

The BioMart tool from Ensembl was used to seek candidate genes potentially associated with the selected SNP according to the *Bos taurus* ARS-UCD1.3 reference genome (Durinck et al., 2009). Genes and QTL located up to 100 kb upstream or downstream of the significant SNP were kept for further analyses. The QTL database used was the Animal QTLdb Release 56 (Hu et al., 2022). Thereafter, functional enrichment analyses of the candidate genes identified for each LCP were performed using the DAVID platform (Huang et al. 2009; Sherman et al. 2022). Genes and gene ontology (**GO**) terms were considered enriched when the false discovery rate (**FDR**; Benjamini and Hochberg, 1995) was less than 5% ($P < 0.05$).

RESULTS

Descriptive Statistics for each Parameter and Heritabilities

Table 2 presents the mean (±SE), minimum, and maximum values for the parameters *a*, *b*, and *c* from the WD model; parameters *a*, *b*, and *c* from the WL model; and parameters *a*, *b*, *c*, *d*, and *f* from the AS model, for heifers and cows. Heritability estimated are shown in Figure 1. For heifers, the heritability estimates ranged from 0.14 to 0.21 for WD model, 0.14 to 0.20 for WL model, and from 0.10 to 0.21 for AS model; and for cows, the heritability estimates ranged from 0.10 to 0.16 for WD model, 0.11 to 0.16 for WL model, and from 0.11 to 0.16 for AS model. Despite being low to moderate, genomic heritabilities are more accurate compared with pedigree-based heritabilities.

Association Analyses

The Q-Q plots for all LCP are shown in Supplemental File S1 and the λ values ranged from 0.91 to 0.99 (Supplemental File S1: Figures S1 - S6). According to the Q-Q plots and λ values, no signs of uncorrected population stratification was observed. Supplemental File S1 presents the Manhattan plots for all LCP (Supplemental File S1: Figures S7 - S12); and the candidate genes and QTL found within an interval of 100 kb upstream and downstream from the significant SNP are provided in Supplemental Files

S2 (Tables S1 – S2), S3, and S4. Supplemental Files S5 and S6 present the significant SNP. Table 3 shows a summary of the main significant findings of SNP and genes, and the complete results are presented in Supplemental File S2 (Tables S1 and S2). For heifers, a total of 81, 128, and 196 significant SNP were identified for the WD, WL, and AS parameters, respectively (Supplemental File S2: Table S1). For cows, a total of 120, 125, and 128 significant SNP were identified for the WD, WL, and AS parameters, respectively (Supplemental File S2: Table S2). Additionally, BTA14 had the highest number of significant SNP, 358 SNP for heifers and 350 SNP for cows (Supplemental File S2: Tables S1 and S2). Furthermore, when considering a stricter threshold for the Bonferroni correction based on the total number of SNP, significant SNP were mostly found on BTA14, and one genomic region (BTA14: 1,801,116 bp; *PSCA*, *JRK*, *ARC*, and *ADGRB1*) was common to all the parametric functions. In general, most of the QTL are related to milk production traits, but there are some others associated to production, reproduction, and health (Supplemental Files S3 and S4). The summary of the most significant terms in the functional annotation analysis for group of genes identified is presented in Tables 4A and 4B, and the full results are presented in the Supplemental File S2 (Tables S3 - S8); common biological functions were found.

Wood Parameters

A total of 201 SNP distributed across the chromosomes BTA1, BTA2, BTA5, BTA6, BTA7, BTA8, BTA9, BTA13, BTA14, BTA17, BTA18, BTA23, and BTA29 were significantly associated with WD parameters (Supplemental Files S5 and S6). The most statistically significant SNP for WD parameters (Table 3) were located on BTA14 (1,736,599 bp; 1,801,116 bp; and 35,732,773 bp). Sixty-eight and 76 candidate genes were found near the significant SNP for heifers and cows, respectively (Supplemental File S2: Tables S1 and S2). The *ZDHHC14*, *SNX9*, *FBXO5*, *MTRF1L*, and *RGS17* genes located on BTA9 were unique to the WD parameters. Several QTL were previously reported in the genomic regions of the SNP associated with WD parameters (Supplemental Files S3 and S4). Three significant GO terms were associated with WD parameters, and 28 genes associated with WD parameters are involved with these GO terms (Tables 4A and 4B). However, no significant GO terms were found for the parameter *c*, for heifers (i.e., FDR higher than 5%).

Wilmink Parameters

Two hundred fifty-three SNP located across 12 chromosomes (BTA1, BTA2, BTA5, BTA6, BTA7, BTA9, BTA11, BTA14, BTA15, BTA18, BTA19, and BTA23) were

found to be significantly associated with WL parameters (Supplemental Files S5 and S6). The genomic regions located on BTA15 (*FOLR1*, *FOLR2*, *INPPL1*, *PHOX2A*, and *CLPB*), and BTA19 (*NCOR1*, *TTC19*, *ZSWIM7*, *ADORA2B*, and *SPECC1*) were unique to the WL parameters. The most statistically significant SNP for WL parameters (Table 3) were located on BTA6 (88,919,352 bp), and BTA14 (1,736,599 bp; and 1,801,116 bp). The genomic regions around the significant SNP harbor 101 candidate genes for heifers and 81 candidate genes for cows (Supplemental File S2: Tables S1 and S2). Moreover, several QTL were previously reported in the same genomic regions (Supplemental Files S3 and S4). Three significant GO containing 25 genes were identified for WL parameters (Tables 4A and 4B).

Ali and Schaeffer Parameters

A total of 324 SNP were found to be associated with AS parameters (Supplemental Files S5 and S6). These SNP are located on chromosomes BTA1, BTA3, BTA5, BTA6, BTA7, BTA8, BTA9, BTA11, BTA13, BTA14, BTA17, BTA18, BTA20, BTA23, and BTA29. Among them, genomic regions located on BTA3 (*MAB21L3* and *SLC22A15*), and BTA20 (*HCN1*, *U6*, *CARD6*, *SNORD72*, *PRKAA1*, *TTC33*, and *PTGER4*) were unique to the AS parameters. The most statistically significant SNP for AS parameters (Table 3) were located on BTA11 (66,602,594 bp), and BTA14 (1,736,599 bp; and 1,801,116 bp). The genomic regions around these SNP contain 146 and 134 candidate genes for heifers and cows, respectively (Supplemental File S2: Tables S1 and S2). Moreover, these regions contain QTL previously associated with various traits, such as milk yield, somatic cell score, lifetime profit index, lactation persistency, udder depth, and milking speed (Supplemental Files S3 and S4). Three significant GO terms were associated with AS parameters, but for cows, no significant GO terms were found for the parameters *d* and *f* (i.e., FDR higher than 5%).

DISCUSSION

In this study, our objective was to investigate the genomic background of various LCP in American Holstein cattle. We conducted a GWAS to identify candidate markers associated with LCP and performed the functional annotation of candidate genes in order to unveil the candidate genes associated with lactation curve parameters and their biological pathways. The *a* parameter (WD and WL) is an approximation of the initial MY after calving; the *b* parameter (WD and WL) is the increasing slope parameter up to lactation peak yield; the *c* parameter (WD and WL) is the decreasing slope parameter

after the lactation peak yield; the *a* parameter (AS) is associated with lactation peak yield; the *b* and *c* parameters (AS) are associated with the decreasing slope; and the *d* and *f* parameters (AS) are associated with the increasing slope (Strucken et al., 2012; Lázaro et al., 2024).

Wood Parameters. A total of 148 candidate genes were found for heifers and cows (Supplemental File S2: Tables S1 and S2). A larger number of candidate genes was found for the *a* and *b* parameters than for the *c* parameter. The *c* coefficient also had low heritability estimates (Supplemental File S2: Tables S9). Overall, 12 similar candidate genes (*GML, CYP11B1, LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1,* and *TSNARE1*) were identified for all parameters (*a*, *b*, and *c*). Eight candidate genes (*SLC9A1, TENT5B, TRNP1, KDF1, NUDC, ZDHHC14, SNX9*, and *TRPA1*) were specific to the parameter *c* in heifers, suggesting their potential role in the biological processes underlying this parameter. In the most significant regions, BTA14 houses the genes *PSCA* (prostate stem cell antigen), *JRK* (Jrk helix-turn-helix protein), *ARC* (Activity regulated cytoskeleton associated protein), *ADGRB1* (Adhesion G protein-coupled receptor B1), and *TRPA1* (Transient receptor potential cation channel subfamily A member 1). The *PSCA* gene is a member of the LY-6 family of surface proteins, regulating the major histocompatibility complex and has been reported to be associated with clinical mastitis in first parity US Holstein cows (Tiezzi et al., 2015). Additionally, the *ARC* gene was also reported for both clinical mastitis-related traits and somatic cell (Narayana et al., 2023). The *JRK* gene has been associated with 305-d MY and peak yield in multiparous cows (Atashi et al., 2020). The *c* parameter represents the decreasing slope after the lactation peak yield and can be considered as a measure for lactation persistency. Furthermore, the *ADGRB1* gene is associated with milk lactose (Costa et al., 2019), while *TRPA1* is associated with growth regulator (Aponte et al., 2024). Wu et al. (2019) reported that *TRPA1* is related with growth traits in 3 Chinese cattle breeds (Jiaxian red cattle, Qinchuan cattle, and Luxi cattle). On BTA2, the genes *KDF1* (Keratinocyte differentiation factor 1) and *NUDC* (Nuclear distribution C, dynein complex regulator) were identified as an essential regulator of epidermis formation and udder morphology, respectively (Miles et al., 2021). Various QTL (Supplemental Files S3 and S4) were previously reported to be located on the genomic regions around the significant SNP. Milking speed was reported among the QTL. Indeed, we used longitudinal DMY records collected with milking robotic systems (Lely Astronaut robots) and milking parlor (Afimilk milking parlors). It is unknown if there are genotype-by-milking type

interaction. The productive performance, longevity, and welfare of high-producing dairy cows are highly dependent on the interaction between the environments in which they are raised and their genetic background. Therefore, there is a need to explore potential genotype-by-milking type interaction.

Wilmink Parameters. One hundred four candidate genes were found for heifers and 84 for cows (Supplemental File S2: Tables S1 and S2). The *a* parameter had the highest number of candidate genes and high heritability estimates (Supplemental File S2: Tables S9). Thirteen candidate genes (*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, LY6L, GML, CYP11B1, ADGRB1,* and *TSNARE1*), located on BTA14, were common to all parameters (*a*, *b*, and *c*) in heifers. The *LY6D* (lymphocyte antigen 6 family member D) gene is associated with fat yield, milk yield (Suchocki et al., 2016), and mastitis (Tiezzi et al., 2015). According to Tiezzi et al. (2015) *LY6D* gene encodes proteins that are components of the lymphocyte antigen six complex, and the control of neutrophil regulatory function also plays an important role in resistance to udder infections. The *LYNX1* (Ly6/neurotoxin 1) gene, was reported as a candidate gene associated with milk production (Hosseinzadeh et al., 2024). The *LYPD2* (LY6/PLAUR domain containing 2) gene, has been associated with cheese-making traits and MY traits in Walloon Holstein cows (Atashi et al., 2023). The *SLURP1* (secreted LY6/PLAUR domain containing 1) gene, is part of the lymphocyte-antigen-6 complex (*LY6*) known for its neutrophil regulation function, and has been linked to clinical mastitis (Tiezzi et al., 2015). The *THEM6* (thioesterase superfamily member 6) gene was associated with lactose yield (Costa et al., 2019), and encodes a thioesterase enzyme. The *TSNARE1* (T-SNARE Domain Containing 1) gene plays a role in synaptic vesicle exocytosis and intracellular protein transport, and was associated with rectal temperature during heat stress in Holstein cows (Luo et al., 2021). The *CXCL5* (C-X-C Motif Chemokine Ligand 5) and *CXCL2* (C-X-C Motif Chemokine Ligand 2) genes were associated with bovine mastitis (Islam et al., 2020). Atashi et al. (2020) reported that *THEM6, SLURP1, LYNX1*, *LYPD2*, and *LY6D* genes were associated with 305-d MY and peak yield in multiparous cows. The most significant terms identified for heifers and cows are related to LY6 UPA recep-like. In the murine proteome, there are about 60 such proteins that belong to the Ly6/uPAR family. These proteins are anchored to the cell membrane, and are associated with a rich repertoire of functions, including binding to receptors and adhesion. The Ly6/uPAR proteins modulate cell signaling in the context of brain functions and cells of the innate immune system (Tirosh et al., 2013).

Ali and Schaeffer Parameters. A total of 286 candidate genes were identified, 149 for heifers and 137 for cows (Supplemental File S2: Tables S1 and S2). Ten candidate genes (*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1,* and *TSNARE1*) were common to all parameters (*a*, *b*, *c*, *d*, and *f*) in heifers. In the most significant regions for *d* and *f* parameters, BTA11 harbors the genes *C1D* (nuclear receptor corepressor), *DNAAF10* (dynein axonemal assembly factor 10), *PNO1* (partner of NOB1 homolog), *PPP3R1* (protein phosphatase 3 regulatory subunit B, alpha), and *CNRIP1* (cannabinoid receptor interacting protein 1). The *C1D* gene is associated with biological processes of apoptotic and regulation of gene expression (Rothbarth et al., 2001). Apoptotic process can be defined as a programmed cell death process which begins when a cell receives an internal or external signal, and proceeds through a series of biochemical events which trigger an execution phase. The *DNAAF10* gene have been shown to encode proteins that are involved in axonemal dynein assembly (Braschi et al., 2022). *DNAAF10* may also be related to apoptotic process. The *PNO1* gene regulates autophagy and apoptosis of hepatocellular carcinoma cells (Han et al., 2021). *PNO1* is a key component in ribosome biogenesis and was also identified as a candidate gene related with fat deposition (Moscarelli et al., 2020). The *PPP3R1* gene encodes a phosphatase protein, and was associated with muscle growth, development, and degradation in Nelore cattle (Santos Silva et al., 2020). The *CNRIP1* gene seems to be involved in modulating part of the *CNR1* (cannabinoid receptor 1), thereby signaling receptor activity. *CNRIP1* was also identify in the synepitheliochorial placenta (dos Santos Silva et al., 2024).

Ten candidate genes (*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1,* and *TSNARE1*) were common to all the parametric functions. Atashi et al. (2020) using the Milkbot model, reported 8 of those genes (excepted *ADGRB1* and *TSNARE1*). In addidtion, Strucken et al. (2012) reported 4 significant markers for all the 3 WL curve parameters. As already mentioned, we used the stricter Bonferroni threshold and the modified Bonferroni genome-wide threshold to avoid excessive false-negative results. This approach account for the linkage disequilibrium among SNP.

To date, few studies have reported GWAS for LCP in dairy cattle. Strucken et al. (2012) performed GWAS for milk production traits in German Holstein cattle using the WL model; Atashi et al. (2020) performed GWAS for LCP in Holstein dairy cows (distributed in 4 countries: Belgium, Netherlands, Great Britain, and Denmark) using the MilkBot model; and Lázaro et al. (2024) performed GWAS for mozzarella and milk production traits in Murrah buffaloes using the WD model. Among the random regression

coefficients used in this current study, AS is the less explored, and to our best knowledge, this is the first GWAS for LCP using the AS random regression coefficients, which presents a good opportunity to explore the genomic background of LCP.

CONCLUSIONS

The genomic architecture of lactation curve parameters revealed numerous SNP significantly associated with Wood, Wilmink, and Ali-Schaeffer parameters, distributed across several chromosomes, including BTA1, BTA2, BTA3, BTA5, BTA6, BTA7, BTA8, BTA9, BTA11, BTA13, BTA14, BTA15, BTA17, BTA18, BTA19, BTA20, BTA23, and BTA29. Ten candidate genes (*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1,* and *TSNARE1*) were common to all the parametric functions (Wood, Wilmink, and Ali-Schaeffer). The identified candidate genes are involved in diverse biological pathways. The genomic regions identified overlap with various previously reported QTL. These findings enhance the understanding of the genomic basis of lactation curve parameters and provide valuable insights into the underlying mechanisms.

ACKNOWLEDGMENTS

This research was funded by the Agriculture and Food Research Initiative Competitive Grant number 2022-67021-37022 from the USDA National Institute of Food and Agriculture. The authors would like to thank Homestead Dairy, LLC (Plymouth, IN), and especially Brian and Jill Houin, for allowing our research group to use their farm records. We also thank Lely for giving us access to their system for extracting the datasets. In addition, the first author is thankful to the “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” (CAPES, Brasília, DF, Brazil; Finance Code 001) for granting financial assistance for doctoral studies.

TABLES AND FIGURE

Table 1. Descriptive statistics for number of animals, number of records, average records per animal, CG number, average daily milk yield, minimum and maximum daily milk yield, average age at calving, minimum and maximum age at calving, and milking systems in American Holstein cattle.

Variable	Heifers	Cows
Number of animals	11,139	6,735
Number of records	2,754,840	1,642,653
Average (SD) records per animal	247.31 (69.38)	243.90 (68.60)
CG number	30	25
Average (SD) daily milk yield, kg	32.86 (08.11)	39.88 (10.73)
Min (Max) daily milk yield, kg	0.22 (68.21)	0.21 (82.90)
Average (SD) age at calving, d	713.5 (40.04)	1,106 (69.01)
Min (Max) age at calving, d	640 (962)	916 (1,556)
Milking systems		
Number of records from the Lely Astronaut robots, n (%)	459,505 (16.68)	485,473 (29.55)
Number of records from the Afimilk milking parlors, n (%)	2,295,335 (83.32)	1,157,180 (70.45)

CG: contemporary group; Max: maximum; Min: minimum.

Table 2. Mean ± standard error (SE), minimum, and maximum values for each parameter (P) from the Wood, Wilmink, and Ali-Schaeffer models for daily milk yield (DMY) in American Holstein cattle.

Models	P	Mean ± SE	Minimum	Maximum
		Heifers		
Wood	a	25.310 ± 0.011	18.500	34.670
	b	2.708 ± 0.001	2.038	3.563
	c	$4.789 \times 10^{-4} \pm 2.37 \times 10^{-7}$	0.313×10^{-6}	0.662×10^{-3}
Wilmink	a	12.399 ± 0.005	9.735	16.390
	b	23.175 ± 0.010	17.410	30.360
	c	$0.239 \times 10^{-3} \pm 1.16 \times 10^{-7}$	0.169×10^{-3}	0.328×10^{-3}
Ali-Schaeffer	a	26.770 ± 0.011	20.010	36.170
	b	17.750 ± 0.013	11.190	27.320
	c	33.614 ± 0.026	20.210	51.370
	d	11.233 ± 0.006	7.615	15.930
	f	0.286 ± 0.0002	0.169	0.430
		Cows		
Wood	a	30.110 ± 0.022	16.860	47.280
	b	3.526 ± 0.002	2.094	5.274
	c	$7.36 \times 10^{-4} \pm 5.14 \times 10^{-7}$	0.465×10^{-3}	0.115×10^{-2}
Wilmink	a	19.526 ± 0.010	12.660	27.350
	b	30.866 ± 0.020	20.080	47.430
	c	$0.352 \times 10^{-3} \pm 2.45 \times 10^{-7}$	0.205×10^{-3}	0.554×10^{-3}
Ali-Schaeffer	a	26.505 ± 0.032	14.870	49.370
	b	9.095 ± 0.007	5.698	13.460
	c	30.050 ± 0.038	14.830	55.790
	d	11.996 ± 0.017	5.685	24.530
	f	0.391 ± 0.001	0.191	0.780

P: Curve's parameters; a (Wood, Wilmink) is an approximation of the initial milk yield after calving; b (Wood, Wilmink) is the increasing slope parameter up to lactation peak yield; c (Wood, Wilmink) is the decreasing slope parameter after the lactation peak yield; a (Ali-Schaeffer) is associated with lactation peak yield; b (Ali-Schaeffer) and c (Ali-Schaeffer) are associated with the decreasing slope; d (Ali-Schaeffer) and f (Ali-Schaeffer) are associated with the increasing slope.

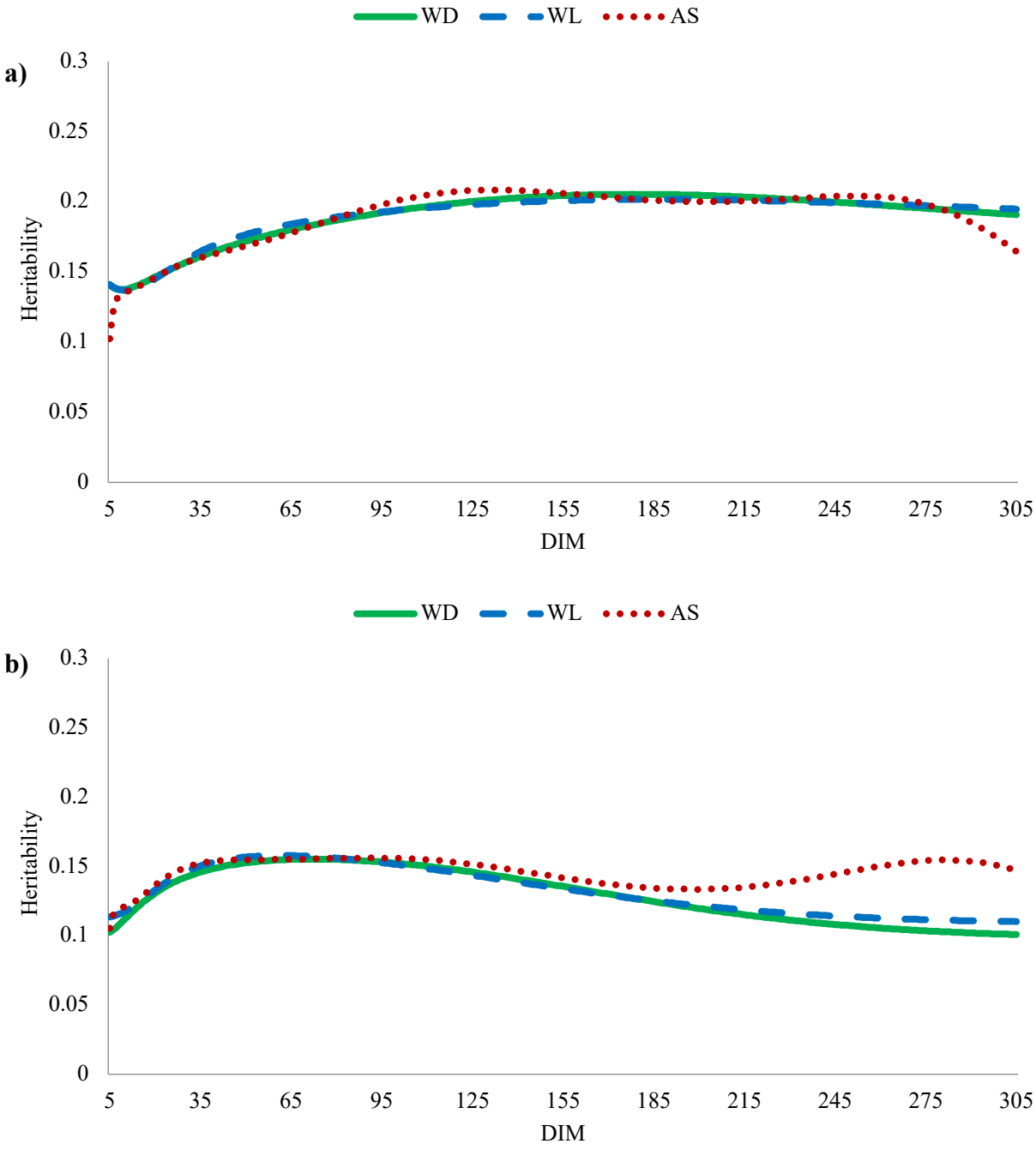

Figure 1. Heritability estimated using the Wood (WD), Wilmink (WL), and Ali-Schaeffer (AS) models. (a) represents heifers and (b) represents cows.

Table 3. Summary of the most significant SNP associated with each lactation curve parameters in American Holstein. The complete list is presented in the Supplemental Tables S1 and S2.

Models	P	Chr	Location (bp)	*P*-value	Genes	QTL
Heifers						
Wood	a	BTA14	1,801,116	1.84×10^{-8}	*PSCA, JRK, ARC, ADGRB1*	123
	b	BTA14	1,801,116	3.98×10^{-24}	*PSCA, JRK, ARC, ADGRB1*	123
	c	BTA14	35,732,773	1.37×10^{-6}	*TRPA1*	28
Wilmink	a	BTA14	1,801,116	5.28×10^{-24}	*PSCA, JRK, ARC, ADGRB1*	123
	b	BTA14	1,801,116	1.31×10^{-23}	*PSCA, JRK, ARC, ADGRB1*	123
	c	BTA14	1,736,599	7.41×10^{-11}	*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1*	115
Ali-Schaeffer	a	BTA14	1,801,116	3.95×10^{-34}	*PSCA, JRK, ARC, ADGRB1*	123
	b	BTA14	1,801,116	2.46×10^{-28}	*PSCA, JRK, ARC, ADGRB1*	123
	c	BTA14	1,801,116	2.82×10^{-25}	*PSCA, JRK, ARC, ADGRB1*	123
	d	BTA14	1,801,116	3.50×10^{-13}	*PSCA, JRK, ARC, ADGRB1*	123
	f	BTA14	1,801,116	4.23×10^{-7}	*PSCA, JRK, ARC, ADGRB1*	123
Cows						
Wood	a	BTA14	1,801,116	3.24×10^{-10}	*PSCA, JRK, ARC, ADGRB1*	123
	b	BTA14	1,736,599	1.65×10^{-30}	*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1*	115
	c	BTA14	1,736,599	3.55×10^{-14}	*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1*	115
Wilmink	a	BTA14	1,736,599	4.95×10^{-25}	*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1*	115
	b	BTA14	1,736,599	5.68×10^{-28}	*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1*	115
	c	BTA6	88,919,352	1.27×10^{-7}	*CXCL5, CXCL2*	29
Ali-Schaeffer	a	BTA14	1,736,599	1.94×10^{-31}	*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1*	115
	b	BTA14	1,801,116	1.55×10^{-9}	*PSCA, JRK, ARC, ADGRB1*	123
	c	BTA14	1,736,599	1.93×10^{-21}	*LY6D, LYNX1, LYPD2, SLURP1, THEM6, PSCA, JRK, ARC, ADGRB1*	115
	d	BTA11	66,602,594	1.16×10^{-7}	*C1D, DNAAF10, PNO1, PPP3R1, CNRIP1*	14
	f	BTA11	66,602,594	7.92×10^{-6}	*C1D, DNAAF10, PNO1, PPP3R1, CNRIP1*	14

P: Curve's parameters; Chr: Chromosome; BTA: *Bos taurus* autosome; a (Wood, Wilmink) is an approximation of the initial milk yield after calving; b (Wood, Wilmink) is the increasing slope parameter up to lactation peak yield; c (Wood, Wilmink) is the decreasing slope parameter after the lactation peak yield; a (Ali-Schaeffer) is associated with lactation peak yield; b (Ali-Schaeffer) and c (Ali-Schaeffer) are associated with the decreasing slope; d (Ali-Schaeffer) and f (Ali-Schaeffer) are associated with the increasing slope. QTL: number of QTL previously reported in Animal QTLdb, Release 56.

Table 4A. Summary of the most significant terms in the functional annotation analysis for group of genes identified in American Holstein heifers. The full list is presented in the Supplemental Tables S3, S4, and S5.

P	Category	Term	FDR	Genes
		Wood Model		
a	UP_SEQ_FEATURE	DOMAIN: UPAR/Ly6	5.9×10^{-7}	*LYPD2, LYNX1, LY6D, PSCA, SLURP1*
b	UP_SEQ_FEATURE	CROSSLNK: Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in UFM1); alternate	1.6×10^{-29}	*H4C1, H4C11, H4C12, H4C13, H4C14, H4C19, H4C2, H4C3, H4C6, H4C7, H4C8, H4C16*
c	-	-	-	-
		Wilmink Model		
a	INTERPRO	LY6 UPA recep-like	1.3×10^{-17}	*LYPD2, LYNX1, GPIHBP1, GML, LY6D, LY6E, LY6H, LY6L, PSCA, SLURP1*
b	UP_SEQ_FEATURE	CROSSLNK: Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in UFM1); alternate	4.8×10^{-30}	*H4C1, H4C11, H4C12, H4C13, H4C14, H4C19, H4C2, H4C3, H4C6, H4C7, H4C8, H4C16*
c	INTERPRO	LY6 UPA recep-like	2.0×10^{-16}	*LYPD2, LYNX1, GPIHBP1, GML, LY6D, LY6H, LY6L, PSCA, SLURP1*
		Ali and Schaeffer Model		
a	UP_SEQ_FEATURE	CROSSLNK: Glycyl lysine isopeptide (Lys-Gly) (interchain with G-Cter in UFM1); alternate	7.9×10^{-29}	*H4C1, H4C11, H4C12, H4C13, H4C14, H4C19, H4C2, H4C3, H4C6, H4C7, H4C8, H4C16*
b	INTERPRO	LY6 UPA recep-like	1.5×10^{-16}	*LYPD2, GPIHBP1, LY6D, LY6E, LY6H, LY6K, LY6L, PSCA, SLURP1*
c	INTERPRO	LY6 UPA recep-like	1.7×10^{-17}	*LYPD2, GPIHBP1, LY6D, LY6E, LY6H, LY6K, LY6L, PSCA, SLURP1*
d	UP_SEQ_FEATURE	DOMAIN: UPAR/Ly6	1.2×10^{-6}	*LYPD2, LY6D, LY6K, PSCA, SLURP1*
f	UP_SEQ_FEATURE	DOMAIN: UPAR/Ly6	6.9×10^{-6}	*LYPD2, LY6D, PSCA, SLURP1*

P: Curve's parameters; a (Wood, Wilmink) is an approximation of the initial milk yield after calving; b (Wood, Wilmink) is the increasing slope parameter up to lactation peak yield; c (Wood, Wilmink) is the decreasing slope parameter after the lactation peak yield; a (Ali-Schaeffer) is associated with lactation peak yield; b (Ali-Schaeffer) and c (Ali-Schaeffer) are associated with the decreasing slope; d (Ali-Schaeffer) and f (Ali-Schaeffer) are associated with the increasing slope.

Table 4B. Summary of the most significant terms in the functional annotation analysis for group of genes identified in American Holstein cows. The full list is presented in the Supplemental Tables S6, S7, and S8.

P	Category	Term	FDR	Genes
		Wood Model		
a	UP_SEQ_FEATURE	DOMAIN: UPAR/Ly6	1.4×10^{-10}	*LYPD2, LYNX1, GML, LY6D, PSCA, SLURP1*
b	INTERPRO	LY6 UPA recep-like	2.0×10^{-19}	*LYPD2, LYNX1, GPIHBP1, GML, LY6D, LY6E, LY6H, LY6L, PSCA, SLURP1*
c	INTERPRO	LY6 UPA recep-like	6.9×10^{-20}	*LYPD2, LYNX1, GPIHBP1, GML, LY6D, LY6E, LY6H, LY6L, PSCA, SLURP1*
		Wilmink Model		
a	INTERPRO	LY6 UPA recep-like	8.6×10^{-19}	*LYPD2, LYNX1, GPIHBP1, GML, LY6D, LY6E, LY6H, LY6L, PSCA, SLURP1*
b	INTERPRO	LY6 UPA recep-like	1.3×10^{-19}	*LYPD2, LYNX1, GPIHBP1, GML, LY6D, LY6E, LY6H, LY6L, PSCA, SLURP1*
c	PIR_SUPERFAMILY	Serum albumin subgroup	1.6×10^{-6}	*AFM, ALB, AFP*
		Ali and Schaeffer Model		
a	INTERPRO	LY6 UPA recep-like	2.9×10^{-19}	*LYPD2, LYNX1, GPIHBP1, GML, LY6D, LY6E, LY6H, LY6L, PSCA, SLURP1*
b	UP_SEQ_FEATURE	DOMAIN: UPAR/Ly6	1.6×10^{-10}	*LYPD2, LYNX1, GML, LY6D, PSCA, SLURP1*
c	INTERPRO	LY6 UPA recep-like	3.2×10^{-19}	*LYPD2, LYNX1, GPIHBP1, GML, LY6D, LY6E, LY6H, LY6L, PSCA, SLURP1*
d	-	-	-	-
f	-	-	-	-

P: Curve's parameters; a (Wood, Wilmink) is an approximation of the initial milk yield after calving; b (Wood, Wilmink) is the increasing slope parameter up to lactation peak yield; c (Wood, Wilmink) is the decreasing slope parameter after the lactation peak yield; a (Ali-Schaeffer) is associated with lactation peak yield; b (Ali-Schaeffer) and c (Ali-Schaeffer) are associated with the decreasing slope; d (Ali-Schaeffer) and f (Ali-Schaeffer) are associated with the increasing slope.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. https://doi.org/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. Pages 1–3 in Proceedings, 10th World Congr. Genet. Appl. Livest. Prod. Vancouver, BC. https://doi.org/10.13140/2.1.4801.5045.

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. https://doi.org/10.4141/cjas87-067.

Aponte, P. F. C., P. L. S. Carneiro, A. C. Araujo, V. B. Pedrosa, P. R. Fotso-Kenmogne, D. A. Silva, F. Miglior, F. S. Schenkel, and L. F. Brito. 2024. Investigating the genomic background of calving-related traits in Canadian Jersey cattle. J. Dairy Sci. 107:11195-11213. https://doi.org/10.3168/jds.2024-24768.

Atashi, H., M. Salavati, J. De Koster, J. Ehrlich, M. Crowe, G. Opsomer, GplusE Consortium, and M. Hostens. 2020. Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. J. Anim. Breed. Genet. 137:292–304. https://doi.org/10.1111/jbg.12442.

Atashi, H., Y. Chen, H. Wilmot, C. Bastin, S., Vanderick, X., Hubin, and N. Gengler. 2023. Single-step genome-wide association analyses for selected infrared-predicted cheese-making traits in Walloon Holstein cows. J. Dairy Sci. 106:7816-7831. https://doi.org/10.3168/jds.2022-23206.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Braschi, B., H. Omran, G. B. Witman, G. J. Pazour, K. K. Pfister, E. A. Bruford, and S. M. King. 2022. Consensus nomenclature for dyneins and associated assembly factors. J. Cell Biol. 221:e202109014. https://doi.org/10.1083/jcb.202109014.

Brito, L. F., Bedere, N., Douhard, F., Oliveira, H. R., Arnal, M., Peñagaricano, F., Schinckel, A. P., Baes, C. F., Miglior, F. 2021. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly

changing world. Animal 15:100292. https://doi.org/10.1016/j.animal.2021.100292.

Chen, S.-Y., J. P. Boerman, L. S. Gloria, V. B. Pedrosa, J. Doucette, and L. F. Brito. 2023. Genomic-based genetic parameters for resilience across lactations in North American Holstein cattle based on variability in daily milk yield records. J. Dairy Sci. 106:4133–4146. https://doi.org/10.3168/jds.2022-22754.

Chen, S.-Y., L. S. Gloria, V. B. Pedrosa, J. Doucette, J. P. Boerman, and L. F. Brito. 2024. Unraveling the genomic background of resilience based on variability in milk yield and milk production levels in North American Holstein cattle through genome-wide association study and Mendelian randomization analyses. J. Dairy Sci. 107:1035-1053. https://doi.org/10.3168/jds.2023-23650.

Costa, A., H. Schwarzenbacher, G. Mészáros, B. Fuerst-Waltl, C. Fuerst, J. Sölkner, and M. Penasa. 2019. On the genomic regions associated with milk lactose in Fleckvieh cattle. J. Dairy Sci. 102:10088-10099. https://doi.org/10.3168/jds.2019-16663.

dos Santos Silva, P., G. Kra, Y. Butenko, J. R. Daddam, Y. Levin and M. Zachut. 2024. Maternal supplementation with n-3 fatty acids affects placental lipid metabolism, inflammation, oxidative stress, the endocannabinoid system, and the neonate cytokine concentrations in dairy cows. J. Anim. Sci. Biotechnol. 15:74. https://doi.org/10.1186/s40104-024-01033-4.

Durinck, S., P. T. Spellman, E. Birney, and W. Huber. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4:1184–1191. https://doi .org/10.1038/nprot.2009.97.

Fotso-Kenmogne, P. R., P. L. S. Carneiro, D. A. Silva, J. A. Cobuci, P. F. C. Aponte, R. O. Hinayah, and L. F. Brito. 2025. Genomic-based genetic parameters for daily milk yield and various lactation persistency traits in American Holstein cattle. J. Dairy Sci. 108:7329-7344. https://doi.org/10.3168/jds.2024-25836.

Geweke, J. 1992. Comment: Inference and prediction in the presence of uncertainty and determinism. Stat. Sci. 7:94-101. https://doi.org/10.1214/ss/1177011446.

Goddard, M. E., Hayes, B. J., and Meuwissen, T. H. E. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed. Genet. 128:409–421. https://doi:10.1111/j.1439-0388.2011.00964.x.

Gualdrón Duarte, J. L., R. J. Cantet, R. O. Bates, C. W. Ernst, N. E. Raney, and J. P. Steibel. 2014. Rapid screening for phenotype-genotype associations by linear

transformations of genomic evaluations. BMC Bioinformatics 15:246. https://doi.org/10.1186/1471-2105-15-246.

Han, Z., D. Liu, L. Chen, Y. He, X. Tian, L. Qi, L. Chen, Y. Luo, Z. Chen, X. Hu, G. Li, L. Zhan, Y. Wang, Q. Li, P. Chen, Z. Liu, and H. Guo. 2021. PNO1 regulates autophagy and apoptosis of hepatocellular carcinoma via the MAPK signaling pathway. Cell Death Dis. 12:552. https://doi.org/10.1038/s41419-021-03837-y.

Hosseinzadeh, S., S. A. Rafat, A. Javanmard, and L. Fang. 2024. Identification of candidate genes associated with milk production and mastitis based on transcriptome-wide association study. Anim. Genet. 55:430-439. https://doi.org/10.1111/age.13422.

Hu, Z-L., C. A. Park, and J. M. Reecy. 2022. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res. D956–D961. https://doi:10.1093/nar/gkab1116.

Huang, D. W., B. T, Sherman, R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nat. Protoc. 4:44-57. https://doi.org/10.1038/nprot.2008.211.

Islam, M. A., M. Takagi, K. Fukuyama, R. Komatsu, L. Albarracin, T. Nochi, Y. Suda, W. Ikeda-Ohtsubo, V. Rutten, W. van Eden, J. Villena, H. Aso, and H. Kitazawa. 2020. Transcriptome analysis of the inflammatory responses of bovine mammary epithelial cells: exploring immunomodulatory target genes for bovine mastitis. Pathogens, 9:200. https://doi.org/10.3390/pathogens9030200.

Lázaro, S. F., H. Tonhati, H. R. Oliveira, A. A. Silva, D. C. B. Scalez, A. V. Nascimento, D. J. A. Santos, G. Stefani, I. S. Carvalho, A. F. Sandoval, and L. F. Brito. 2024. Genetic parameters and genome-wide association studies for mozzarella and milk production traits, lactation length, and lactation persistency in Murrah buffaloes. J. Dairy Sci. 107:992–1021. https://doi.org/10.3168/jds.2023-23284.

Li, X., A. J. Buitenhuis, M. S. Lund, C. Li, D. Sun, Q. Zhang, N.A. Poulsen, and G. Su. 2015. Joint genome-wide association study for milk fatty acid traits in Chinese and Danish Holstein populations. J. Dairy Sci. 98:8152–8163. https://doi:10.3168/jds.2015-9383.

Luo, H., X. Li, L. Hu, W. Xu, Q. Chu, A. Liu, G. Guo, L. Liu, L. F. Brito, and Y. Wang. 2021. Genomic analyses and biological validation of candidate genes for rectal temperature as an indicator of heat stress in Holstein cattle. J. Dairy Sci. 104:4441-4451. https://doi.org/10.3168/jds.2020-18725.

Makanjuola, B. O., F. Miglior, E. A. Abdalla, C. Maltecca, F. S. Schenkel, and C. F. Baes. 2020. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. J. Dairy Sci. 103:5183–5199. https://doi:10.3168/jds.2019-18013.

Miles, A. M., C. J. Posbergh, and H. J. Huson. 2021. Direct phenotyping and principal component analysis of type traits implicate novel QTL in bovine mastitis through genome-wide association. Animals, 11:147. https://doi.org/10.3390/ani11041147.

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for BLUPF90 family of programs. University of Georgia, Athens. Accessed Oct. 12, 2024. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all8.pdf/.

Moscarelli, A., M. T. Sardina, M. Cassandro, E. Ciani, F. Pilla, G. Senczuk, B. Portolano, and S. Mastrangelo. 2020. Genome-wide assessment of diversity and differentiation between original and modern Brown cattle populations. Anim. Genet. 52: 21-31. https://doi.org/10.1111/age.13019.

Narayana, S. G., E. de Jong, F. S. Schenkel, P. A. S. Fonseca, T. C. S. Chud, D. Powell, G. Wachoski-Dark, P. E. Ronksley, F. Miglior, K. Orsel, and H. W. Barkema. 2023. Underlying genetic architecture of resistance to mastitis in dairy cattle: A systematic review and gene prioritization analysis of genome-wide association studies. J. Dairy Sci. 106:323-351. https://doi.org/10.3168/jds.2022-21923.

Oliveira, H. R., D. A. L. Lourenco, Y. Masuda, I. Misztal, S. Tsuruta, J. Jamrozik, L. F. Brito, F. F. Silva, J. P. Cant, and F. S. Schenkel. 2019c. Single-step genome-wide association for longitudinal traits of Canadian Ayrshire, Holstein, and Jersey dairy cattle. J. Dairy Sci. 102:9995–10011. https://doi.org/10.3168/jds.2019-16821.

Oliveira, H. R., G. S. Campos, S. F. Lazaro, J. Jamrozik, A. Schinckel, and L. F. Brito. 2024. Phenotypic and genomic modeling of lactation curves: A longitudinal perspective. JDS communications 5: 241-246. https://doi.org/10.3168/jdsc.2023-0460.

Oliveira, H. R., J. P. Cant, L. F. Brito, F. L. B. Feitosa, T. C. S. Chud, P. A. S. Fonseca, J. Jamrozik, F. F. Silva, D. A. L. Lourenco, and F. S. Schenkel. 2019a. Genome-wide association for milk production traits and somatic cell score in different

lactation stages of Ayrshire, Holstein, and Jersey dairy cattle. J. Dairy Sci. 102:8159–8174. https://doi.org/10.3168/jds.2019-16451.

Oliveira, H. R., L. F. Brito, F. F. Silva, D. A. L. Lourenco, J. Jamrozik, and F. S. Schenkel. 2019b. Genomic prediction of lactation curves for milk, fat, protein, and somatic cell score in Holstein cattle. J. Dairy Sci. 102:452–463. https://doi.org/10.3168/jds.2018-15159.

Pedrosa, V. B., F. S. Schenkel, S. Y. Chen, H. R. Oliveira, T. M. Casey, M. G. Melka, and L. F. Brito 2021. Genome wide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. Genes, 12:1830. https://doi.org/10.3390/genes12111830.

Rothbarth, K., A. Hunziker, H. Stammer, and D. Werner. 2001. Promoter of the gene encoding the 16 kDa DNA-binding and apoptosis-inducing C1D protein. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 1518:271-275. https://doi.org/10.1016/S0167-4781(01)00198-1.

Santos Silva, D. B., L. F. S. Fonseca, A. F. B. Magalhães, M. M. M. Muniz, F. Baldi, J. A. Ferro, L. A. L Chardulo, D. G. Pinheiro, L. G. Albuquerque. 2020. Transcriptome profiling of muscle in Nelore cattle phenotypically divergent for the ribeye muscle area. Genomics, 112:1257-1263. https://doi.org/10.1016/j.ygeno.2019.07.012.

Sherman, B. T., M. Hao, J. Qiu, X. Jiao, M.W. Baseler, H. C. Lane, T. Imamichi and W. Chang. 2022. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. W216-W221. https://doi:10.1093/nar/gkac194.

Smith, B. J. 2007. boa: An R package for MCMC output convergence assessment and posterior inference. J. Stat. Softw. 21:1–37. https: //doi.org/10.18637/jss.v021.i11.

Sousa Junior, L. P. B., L. F. B. Pinto, V. A. R. Cruz, G. A. Oliveira Junior, H. R. Oliveira, T. S. Chud, V. B. Pedrosa, F. Miglior. F. S. Schenkel and L. F. Brito. 2024. Genome-wide association and functional genomic analyses for body conformation traits in North American holstein cattle. Front. Genet. 15:1478788. https://doi.org/10.3389/fgene.2024.1478788.

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. https://doi.org/10.1111/j.1365-2052.2011.02278.x.

Suchocki, T., K. Wojdak-Maksymiec, and J. Szyda. 2016. Using gene networks to identify genes and pathways involved in milk production traits in Polish Holstein dairy cattle. Czech J. Anim. Sci. 61: 526–538. https://doi.org/10.17221/43/2015-CJAS.

Tiezzi, F., K. L. Parker-Gaddis, J. B. Cole, J. S. Clay, and C. Maltecca. 2015. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PLoS ONE, 10:e0114919. https://doi.org/10.1371/journal.pone.0114919.

Tirosh, Y. D. Ofer, T. Eliyahu, and M. Linial. 2013. Short toxin-like proteins attack the defense line of innate immunity. Toxins, 5:1314-1331. https://doi.org/10.3390/toxins5071314.

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

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

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. https://doi.org/10.1016/0301-6226(87)90003-0.

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

Wu, J., J. Wang, B. Yue, F. Xing-Tang, C. Zhang, Y. Ma, L. Hu, Y. Bai, X. Qi, and H. Chen. 2019. Research on association between variants and haplotypes of *TRPV1* and *TRPA1* genes with growth traits in three cattle breeds. Anim. Biotechnol. 30:202-211. https://doi.org/10.1080/10495398.2018.1470530.

SUPPLEMENTAL MATERIAL

Supplemental File S1:

Figure S1. Q-Q plots and λ values of the genome-wide association analysis results for Wood's parameters in American Holstein heifers. WDx: Wood model, in which "x" indicates the Wood parameters (a, b, or c).

Figure S2. Q-Q plots and λ values of the genome-wide association analysis results for Wood's parameters in American Holstein cows. WDx: Wood model, in which "x" indicates the Wood parameters (a, b, or c).

Figure S3. Q-Q plots and λ values of the genome-wide association analysis results for Wilmink's parameters in American Holstein heifers. WLx: Wilmink model, in which "x" indicates the Wilmink parameters (a, b, or c).

Figure S4. Q-Q plots and λ values of the genome-wide association analysis results for Wilmink's parameters in American Holstein cows. WLx: Wilmink model, in which "x" indicates the Wilmink parameters (a, b, or c).

Figure S5. Q-Q plots and λ values of the genome-wide association analysis results for Ali-Schaeffer's parameters in American Holstein heifers. ASx: Ali-Schaeffer model, in which "x" indicates the Ali-Schaeffer parameters (a, b, c, d, or f).

Figure S6. Q-Q plots and λ values of the genome-wide association analysis results for Ali-Schaeffer's parameters in American Holstein cows. ASx: Ali-Schaeffer model, in which "x" indicates the Ali-Schaeffer parameters (a, b, c, d, or f).

Figure S7. Manhattan plots of the genome-wide association analysis for Wood's parameters in American Holstein heifers. The statistically significant SNP after a genome-wide modified Bonferroni correction are above the green dotted line. The statistically significant SNP after Bonferroni correction are above the red line. WDx: Wood model, in which "x" indicates the Wood parameters (a, b, or c).

Figure S8. Manhattan plots of the genome-wide association analysis for Wood's parameters in American Holstein cows. The statistically significant SNP after a genome-wide modified Bonferroni correction are above the green dotted line. The statistically

significant SNP after Bonferroni correction are above the red line. WDx: Wood model, in which "x" indicates the Wood parameters (a, b, or c).

Figure S9. Manhattan plots of the genome-wide association analysis for Wilmink's parameters in American Holstein heifers. The statistically significant SNP after a genome-wide modified Bonferroni correction are above the green dotted line. The statistically significant SNP after Bonferroni correction are above the red line. WLx: Wilmink model, in which "x" indicates the Wilmink parameters (a, b, or c).

Figure S10. Manhattan plots of the genome-wide association analysis for Wilmink's parameters in American Holstein cows. The statistically significant SNP after a genome-wide modified Bonferroni correction are above the green dotted line. The statistically significant SNP after Bonferroni correction are above the red line. WLx: Wilmink model, in which "x" indicates the Wilmink parameters (a, b, or c).

Figure S11. Manhattan plots of the genome-wide association analysis for Ali-Schaeffer's parameters in American Holstein heifers. The statistically significant SNP after a genome-wide modified Bonferroni correction are above the green dotted line. The statistically significant SNP after Bonferroni correction are above the red line. ASx: Ali-Schaeffer model, in which "x" indicates the Ali-Schaeffer parameters (a, b, c, d, or f).

Figure S12. Manhattan plots of the genome-wide association analysis for Ali-Schaeffer's parameters in American Holstein cows. The statistically significant SNP after a genome-wide modified Bonferroni correction are above the green dotted line. The statistically significant SNP after Bonferroni correction are above the red line. ASx: Ali-Schaeffer model, in which "x" indicates the Ali-Schaeffer parameters (a, b, c, d, or f).

Supplemental File S2:

Table S1. Identified candidate genes within ±100 kb of significant windows for lactation curve parameters in American Holstein heifers.

Table S2. Identified candidate genes within ±100 kb of significant windows for lactation curve parameters in American Holstein cows.

Table S3. Significant terms in the functional annotation analysis for group of genes identified for Wood's parameters (a, b, and c) in American Holstein heifers.

Table S4. Significant terms in the functional annotation analysis for group of genes identified for Wilmink's parameters (a, b, and c) in American Holstein heifers.

Table S5. Significant terms in the functional annotation analysis for group of genes identified for Ali-Schaeffer's parameters (a, b, c, d, and f) in American Holstein heifers.

Table S6. Significant terms in the functional annotation analysis for group of genes identified for Wood's parameters (a, b, and c) in American Holstein cows.

Table S7. Significant terms in the functional annotation analysis for group of genes identified for Wilmink's parameters (a, b, and c) in American Holstein cows.

Table S8. Significant terms in the functional annotation analysis for group of genes identified for Ali-Schaeffer's parameters (a, b, c, d, and f) in American Holstein cows.

Table S9. Heritability ± SD, and highest probability density (HPD) values for each parameter (P) from the Wood and Wilmink models for daily milk yield (DMY) in American Holstein cattle.

Supplemental File S3. List of the QTL for each parameter from the Wood, Wilmink, and Ali-Schaeffer models in American Holstein heifers.

Supplemental File S4. List of the QTL for each parameter from the Wood, Wilmink, and Ali-Schaeffer models in American Holstein cows.

Supplemental File S5. List of the significant SNP for each parameter from the Wood, Wilmink, and Ali-Schaeffer models in American Holstein heifers.

Supplemental File S6. List of the significant SNP for each parameter from the Wood, Wilmink, and Ali-Schaeffer models in American Holstein cows.

V – CAPÍTULO III

This paper will be submitted to the 13th World Congress on Genetics Applied to Livestock Production (WCGALP), July 12-17, 2026

Genotype-by-milking type interaction (milking robots x milking parlors) for daily milk yield in American Holstein cattle

Patrick R. Fotso-Kenmogne[1,2], Paulo L. S. Carneiro[1], Delvan A. Silva[3], Johnny I. M. Araujo[1], Carlos H. M. Malhado[1], and Luiz F. Brito[4*]

[1]Postgraduate Program in Animal Science, State University of Southwest Bahia, Itapetinga, 45700-000, BA, Brazil; [2]Ministry of Employment and Vocational Training (MINEFOP), Yaoundé, P.O. Box 16273, CE, Cameroon; [3]Department of Animal Sciences, Federal University of Viçosa, Viçosa, 36570-000, MG, Brazil; [4]Department of Animal Sciences, Purdue University, West Lafayette, 47907, IN, USA; [*]Presenting and corresponding author: britol@purdue.edu

Abstract

Investigating genotype-by-milking type interaction for daily milk yield is paramount for optimizing dairy cattle breeding programs. Automatic milking systems have been wildly adopted because of their various advantages. Therefore, we estimated genomic-based variance components and genetic parameters for daily milk yield (milking robots x milking parlors) based on a bi-trait genomic model. The systematic effects in the model included calving year, calving season, days in milk as linear covariate, and calving age as linear and quadratic covariates. Random effects included additive genetic, permanent environment, and residual effects. The data contained records of 11,788 genotyped animals, with 60,277 SNPs. The traits have reasonable additive genetic variance and low to moderate heritability (from 0.10 ± 0.01 to 0.14 ± 0.01 for milking parlors, and from 0.20 ± 0.02 to 0.21 ± 0.02 for milking robots), indicating that genetic progress for these traits is feasible. The genetic correlation estimates for milking type (milking robots x milking parlors) were higher than 0.80 in parity 1 (0.88 ± 0.03) and parity 2 (0.92 ± 0.05),

indicating no potential differential responses across milking type. The results observed in this study indicate substantial variation in daily milk yield in milking robots, potentially due to milking frequency (i.e., number of daily visits which is highly related to the animal welfare leading to superior performance) and management practices.

Introduction

Since 1992, the dairy sector has incorporated technology to optimize yield and profit. The pressure to minimize inputs and maximize efficiency has led to novel approaches to managing and milking dairy herds, including the implementation of automatic milking systems (**AMS**; Jacobs and Siegford, 2012). Automatic milking systems, also known as milking robots (**MR**), reduce the labor associated with milking, increase milk production, improve dairy cow welfare (Samoré *et al*., 2011), and generate a large volume of daily phenotypic records, such as daily milk yield (**DMY**). Multiple differences exist between milking parlors (**MP**) and MR (Jacobs and Siegford, 2012), highlighting the need for targeted research on genotype-by-milking type interaction. Bérat *et al.* (2025) stated that behavioral responses are influenced by the animal's environmental conditions and their genetics. Therefore, their performance is expected to differ across environments. From a genomic selection perspective, there is a need to assess if animals that are genomically selected to perform well in MP environments are also the best performers under MR conditions, and vice versa.

The multi-trait model is commonly used when environmental conditions can be categorized into a few distinct groups, as it allows for the evaluation of genotype-by-environment (**GxE**) interactions, and the estimation of genetic parameters and breeding values across different environments (i.e., MR and MP). To date, it is unknown if there are any GxE effects that should be considered when genetically evaluating performance, longevity, and welfare traits across these environmental conditions (MR and MP). Therefore, the objective of this study was to investigate the level of GxE for DMY in different milking system (MR and MP) in American Holstein cattle.

Materials & methods

Dataset. The data were obtained from a large commercial dairy cattle farm in North-Central Indiana. The phenotype and genomic datasets used were described in our previous study (Fotso-Kenmogne *et al*., 2025). In this study, only DMY records within the range of 5 and 305 days in milk (**DIM**) were considered, and DMY records deviating ±3.5 phenotypic standard deviation (**SD**) from the mean within the contemporary groups (**CG**,

defined by concatenating calving year and season of calving) were removed from further analyses. Finally, 4,397,493 records from 17,874 lactations of 11,788 animals remained for further analyses, including 2,754,840 DMY of 11,139 cows in parity 1; and 1,642,653 DMY of 6,735 cows in parity 2.

The genomic dataset contained 72,820 SNPs for 22,090 animals that were imputed from multiple SNPs panels as described by Chen *et al*. (2023). A quality control (**QC**) was performed to remove SNPs and animals with call rate lower than 0.90, minor allele frequency lower than 0.01, duplicated samples, and extreme deviation from the Hardy-Weinberg equilibrium ($P < 10^{-8}$). After the QC, the genomic dataset contained 60,277 SNPs and 14,464 animals. Genotyped animals with no phenotypic records were removed from subsequent analyses.

Genetic analyses. A bi-trait genomic model was used to perform the analyses. The model included the systematic effects (CG), DIM as linear covariate, calving age as linear and quadratic covariates, additive genetic, permanent environmental, and residual effects. The model can be described as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wpe} + \mathbf{e} \quad (1)$$

where $\mathbf{y}$ represents the vector containing the repeated phenotypic records for each trait (DMY robots or DMY parlors); $\mathbf{b}$ is the vector of systematic effects containing CG, calving age, and DIM; $\boldsymbol{a}$ is the vector of random additive genetic effects with $\mathbf{a} \sim \text{MVN}(0, \mathbf{G} \otimes \mathbf{\Sigma_a})$, where $\mathbf{G}$ represents the genomic relationship matrix based on the first method proposed by VanRaden (2008) and $\mathbf{\Sigma_a}$ is the covariance matrix for additive genetic effects; $\mathbf{pe}$ is the vector of random permanent environmental effects with $\mathbf{pe} \sim \text{MVN}(0, \mathbf{I_P} \otimes \mathbf{P_o})$, where $\mathbf{I_P}$ represents an identity matrix and $\mathbf{P_o}$ the covariance matrix associated with the permanent environmental effects; $\mathbf{e}$ is the vector containing the random residual matrix effects with $\mathbf{e} \sim \text{MVN}(0, \mathbf{I} \otimes \mathbf{R_o})$, where $\mathbf{I}$ represents an identity matrix, $\mathbf{R_o}$ is the residual covariance matrix, and $\otimes$ denotes the Kronecker product. $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{W}$ are the incidence matrices relating $\mathbf{b}$, $\mathbf{a}$, $\mathbf{pe}$ to $\mathbf{y}$. The analyses were based on GBLUP rather than single-step GBLUP because no pedigree information was available. The (co)variance components were estimated based on Bayesian inference using the Gibbs sampler algorithm implemented in the GIBBSF90+ software from the BLUPF90 family of programs (Misztal *et al*., 2018). Genetic correlations lower than 0.80 indicated significant GxE effect (Robertson, 1959).

Results

Phenotypic averages estimated for DMY in parity 1 and 2 according to DIM are illustrated in Figure 1. The phenotypic curve patterns were similar in parity 1 (a) and parity 2 (b). However, the average DMY was consistently higher in MR compared to MP throughout the entire lactation period.

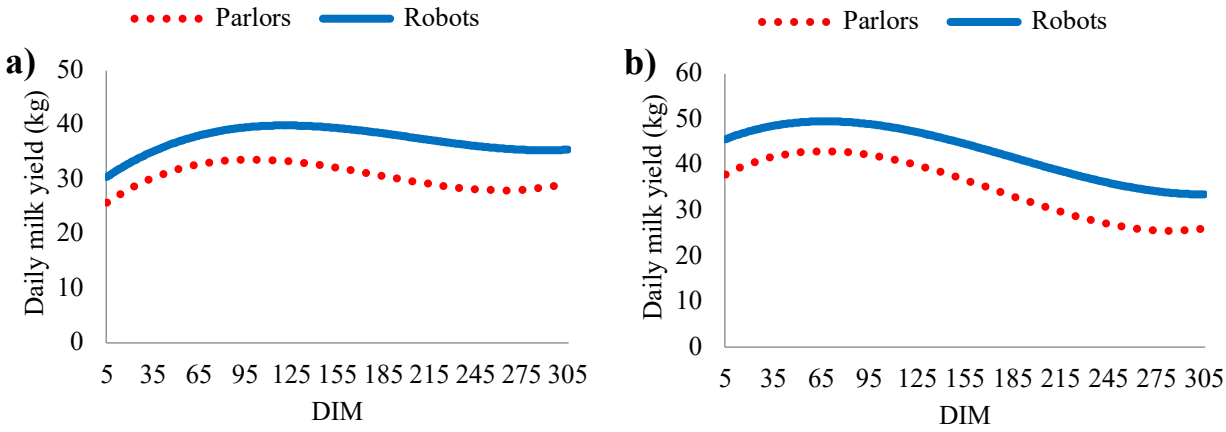

Figure 1. Phenotypic average estimated for daily milk yield (DMY). (a) represents parity 1 and (b) represents parity 2.

Genomic-based variance components, heritability, repeatability, and genetic correlation for DMY in parity 1 and parity 2 are presented in Table 1. The permanent environmental variance ranged from 14.38 ± 0.28 for MP to 23.49 ± 0.92 for MR in parity 1, and from 25.34 ± 0.71 for MP to 30.73 ± 1.63 MR in parity 2. In both parities (1 and 2), the genomic heritability estimates for DMY were low to moderate in MP and MR, respectively. The genomic-based genetic correlations between milking types (MR x MP) were greater than 0.80, with estimates of 0.88 ± 0.03 in parity 1 and 0.92 ± 0.05 in parity 2.

Table 1. Posterior means (SD in parentheses) of variance components, heritability, repeatability, and genetic correlation (milking robots x milking parlors) for daily milk yield (DMY) in American Holstein cattle.

Variable	Parity 1		Parity 2	
	Milking parlors	Milking robots	Milking parlors	Milking robots
Additive genetic variance	7.933 (0.470)	12.088 (1.125)	9.052 (0.870)	16.933 (2.254)
Permanent environmental variance	14.378 (0.281)	23.485 (0.920)	25.335 (0.707)	30.732 (1.631)
Residual variance	33.948 (0.032)	21.433 (0.045)	57.805 (0.077)	35.953 (0.074)
Heritability	0.141 (0.007)	0.208 (0.017)	0.098 (0.008)	0.202 (0.024)
Repeatability	0.397	0.624	0.373	0.570
Genetic correlation (milking robots x milking parlors)	0.875 (0.031)		0.919 (0.054)	

Discussion

Automatic milking systems record a huge amount of data related to both production and individual cows. Different trends in average DMY over DIM were observed based on the milking type (DMY higher in MR than MP). These differences can be attributed to the frequency of daily visits (i.e., milking frequency) and management practices. Furthermore, the increased in milking frequency is highly related to the animal welfare which leads to superior performance. In general, additive genetic and permanent environmental variances were higher in MR than in MP, suggesting that AMS introduces greater variability, probably due to differences in milking frequency. It is important to note that the population analyzed in this study consisted of animals from a single, large commercial farm in the United States (Chen *et al*., 2023; Fotso-Kenmogne *et al*., 2025). In contrast, Santana *et al.* (2025), who investigated genetic responses for yearling weight in Nellore cattle across different pasture and feedlot environments, reported no substantial differences in additive genetic variances. The performance of high-producing dairy cows is strongly dependent on the interaction between their genetic background and the environmental conditions under which they are raised. The genomic-based genetic correlations between milking types (MR x MP) observed in this study were high (> 0.80), indicating no potential differential genetic responses across milking systems. However, the differences in heritabilities between MR and MP were considerable. To date, genetic responses of DMY across milking types do not appear to have been reported in the literature. Therefore, a key next type will be to investigate genotype-by-milking type interaction for milk-related traits and lactation persistency using datasets from multiple countries.

Conclusions

The use of a bi-trait genomic model to evaluate genotype-by-milking type interaction (milking robots x milking parlors) for daily milk yield indicated no potential differential genetic responses across milking systems.

Acknowledgments

The authors would like to thank Homestead Dairy, LLC (Plymouth, IN, USA) and especially Brian and Jill Houin for allowing our research group to utilize their farm records. We also thank Lely for giving us access to their system for extracting the datasets.

References

Bérat H., Gengler N., Maskal J. M., Boerman J. P., and Brito, L. F. (2025). Journal of Dairy Science, 108(7): 7262-7277. https://doi.org/10.3168/jds.2024-25597

Chen S.-Y., Boerman J. P., Gloria L. S., Pedrosa V. B., Doucette J. et al. (2023). Journal of dairy science, 106(6):4133–4146. https://doi.org/10.3168/jds.2022-22754

Fotso-Kenmogne P. R., Carneiro P. L. S., Silva D. A., Cobuci J. A., Aponte P. F. C. et al. (2025). Journal of dairy science, 108(7):7329-7344. https://doi.org/10.3168/jds.2024-25836

Jacobs J. A., and Siegford J. M. (2012). Journal of dairy science, 95(5):2227-2247. http://dx.doi.org/10.3168/jds.2011-4943

Misztal I., Tsuruta S., Lourenco D. A. L., Masuda Y., Aguilar I. et al. (2018). University of Georgia, Athens. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all8.pdf/, visited on: 18-07-2025.

Robertson A. (1959). Biometrics, 15(3):469–485. https://doi.org/10.2307/2527750

Samoré A. B., Román-Ponce S. I., Vacirca F., Frigo E., Canavesi F. et al. (2011). Journal of dairy science, 94(8): 4081-4089. https://doi.org/10.3168/jds.2010-3611

Santana M. L., Bignardi A. B., Rodrigues G. R. D., Cyrillo J. N. S. G., Brito L. F. et al. (2025). Journal of Animal Science, 103:1-13. https://doi.org/10.1093/jas/skaf169

VanRaden P. M., (2008). Journal of dairy science, 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

VI – CONCLUSÕES FINAIS

Lactation persistency is a key component of the lactation curve and one of the most economically important production traits in the dairy industry. We aimed to optimize lactation persistency in high-producing dairy cows, in order to shift from traditional lactation to extended lactation and highly impact animal welfare, productivity, and sustainability of the dairy sector. Genomic-based estimates of genetic parameters for daily milk yield and various indicators of lactation persistency were reported, as well as the identification of genomic regions, candidate genes and quantitative trait loci, metabolic pathways associated with lactation curve parameters, and the genotype-by-milking type interaction (milking robots x milking parlors) for daily milk yield in American Holstein cattle. Random regression model, increasingly adopted in genetic evaluations, were employed in these analyses. The third lactation persistency measurement showed moderate heritability and low genetic correlation with 305-day milk yield. Ten candidate genes (*LY6D*, *LYNX1*, *LYPD2*, *SLURP1*, *THEM6*, *PSCA*, *JRK*, *ARC*, *ADGRB1*, and *TSNARE1*) were found to be common across all parametric functions (Wood, Wilmink, and Ali-Schaeffer). Furthermore, we used longitudinal daily milk yield records collected with milking robotic systems (Lely Astronaut robots) and milking parlor (Afimilk milking parlors). The genomic-based genetic correlations between milking types (milking robots x milking parlors) were high (> 0.80), indicating no potential differential genetic responses across milking systems. The implementation of automatic milking systems, which reduce labor associated with milking, increase milk production, and improve dairy cow welfare, opens room for the investigation of milking system interactions for milk-related traits and lactation persistency in future studies. In brief, the productive performance, longevity, and welfare of high-producing dairy cows are highly dependent on the interaction between the environments in which they are raised and their genetic background. It is unknown if there are any genotype-by-milking type interaction effects that should be considered when genetically evaluating longevity and welfare traits across milking systems. Additionally, given that the fourth-order Legendre orthogonal polynomials model is well-suited for the analysis of longitudinal dataset of daily milk records, a key next step will be the genome-wide association studies for days in milk.