SÃO PAULO STATE UNIVERSITY
SCHOOL OF AGRICULTURAL AND VETERINARIAN SCIENCES

SELECTION SIGNATURES IN DAIRY CATTLE

Larissa Graciano Braga

Veterinary medicine
MsC. in Genetics and Animal Breeding

2025

SÃO PAULO STATE UNIVERSITY
SCHOOL OF AGRICULTURAL AND VETERINARIAN SCIENCES

SELECTION SIGNATURES IN DAIRY CATTLE

Larissa Graciano Braga

Advisor: Prof. Dr. Danísio Prado Munari
Co-advisors: Dra. Tatiane Cristina Seleguim Chud
Dr. Marcos Vinicius Gualberto Barbosa da Silva

Thesis presented to the São Paulo State University - School of Agricultural and Veterinary Sciences, Campus of Jaboticabal, in partial fulfillment of requirements for the degree of Ph.D. in Animal Science

2025

B813s

Braga, Larissa Graciano
Selection signatures in dairy cattle / Larissa Graciano Braga. -- Jaboticabal, 2025
116 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Danísio Prado Munari
Coorientador: Marcos Vinicius Gualberto Barbosa Silva

1. Animais melhoramento genético. 2. Genética animal. 3. Genômica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: SELECTION SIGNATURES IN DAIRY CATTLE

AUTORA: LARISSA GRACIANO BRAGA
ORIENTADOR: DANISIO PRADO MUNARI
COORIENTADOR: MARCOS VINÍCIUS GUALBERTO BARBOSA DA SILVA
COORIENTADORA: TATIANE CRISTINA SELEGUIM CHUD

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

gov.br Documento assinado digitalmente
DANISIO PRADO MUNARI
Data: 12/09/2025 15:07:18-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. DANISIO PRADO MUNARI (Participaçao Virtual)
Departamento de Ciencias Exatas / FCAV UNESP Jaboticabal

GOV.BR Documento assinado digitalmente
PRISCILA ARRIGUCCI BERNARDES
Data: 12/09/2025 21:21:48-0300
CPF: ***.913.148-**
Verifique as assinaturas em https://v.ufsc.br

Profa. Dra. PRISCILA ARRIGUCCI BERNARDES (Participaçao Virtual)
Departamento de Zootecnia e Desenvolvimento Rural / Universidade Federal de Santa Catarina (UFSC) - Florianópolis/SC

gov.br Documento assinado digitalmente
LENIRA EL FARO ZADRA
Data: 16/09/2025 11:26:09-0300
Verifique em https://validar.iti.gov.br

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

Profa. Dra. AMANDA MARCHI MAIORANO (Participaçao Virtual)
Departamento de Zootecnia / Universidade Federal de Uberlândia (UFU) - Uberlândia/MG

gov.br Documento assinado digitalmente
AMANDA MARCHI MAIORANO
Data: 15/09/2025 09:19:56-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. SALVADOR BOCCALETTI RAMOS (Participaçao Virtual)
Departamento de Ciencias Exatas / FCAV UNESP Jaboticabal

gov.br Documento assinado digitalmente
SALVADOR BOCCALETTI RAMOS
Data: 15/09/2025 17:14:53-0300
Verifique em https://validar.iti.gov.br

Jaboticabal, 12 de setembro de 2025

AUTHOR CURRICULUM INFORMATION

Larissa Graciano Braga was born in Porto Velho, Rondônia, on May 29, 1996. She is the daughter of Bárbara Braga Graciano and Sebastião Graciano de Souza. She earned her degree in Veterinary Medicine from the Federal University of Goiás - School of Veterinary and Animal Science (EVZ/ UFG), in 2019. From 2016 to 2017, she served as a volunteer research assistant (PIVIC) in the Undergraduate Research Program (PIP/UFG), under the supervision of Prof. Dr. Paulo Henrique Jorge da Cunha. In 2016, she was Director of Projects at CONPAVet Jr., a junior enterprise based at EVZ/UFG. In 2017, she was awarded a scholarship through the MARCA/MERCOSUR academic mobility program and completed an exchange semester at the Facultad de Ciencias Veterinarias of the Universidad Nacional del Litoral in Esperanza, Santa Fé Province, Argentina. From August 2019 to July 2021, she pursued her Master's degree in Animal Breeding and Genetics at the School of Agricultural and Veterinary Sciences of São Paulo State University "Júlio de Mesquita Filho" (FCAV/UNESP), Jaboticabal Campus, under the supervision of Prof. Dr. Danísio Prado Munari and co-supervision of Dr. Tatiane Cristina Seleguim Chud and Dr. Marcos Vinicius Gualberto Barbosa da Silva. In August 2021, she began her Ph.D. in Animal Science at the same institution, under the same supervisor and co-supervisors. In June 2023, she started a 12-month doctoral research internship at the University of Guelph, in Guelph, Ontario, Canada, under the supervision of Prof. Dr. Flávio Schramm Schenkel. In February 2025, she began working as a Sessional Professor affiliated with the Department of Exact Sciences at FCAV, teaching courses related to Statistics and Informatics.

"*The credit belongs to the man who is actually in the arena, whose face is marred by dust and sweat and blood; who strives valiantly; who errs, who comes short again and again, because there is no effort without error and shortcoming; but who does actually strive to do the deeds; who knows the great enthusiasms, the great devotions; who spends himself in a worthy cause; who at the best knows in the end the triumph of high achievement, and who at the worst, if he fails, at least fails while daring greatly, so that his place shall never be with those cold and timid souls who neither know victory nor defeat.*"

Man in the Arena, 1910, Theodore Roosevelt

ACKNOWLEDGEMENTS

Gostaria de agradecer a Deus por ter me permitido realizar grandes sonhos, agraciando-me com saúde e por ter me guiado durante toda a caminhada. À mãe Maria, que sempre passou na frente dos meus medos, dificuldades e desafios.

Agradeço aos meus pais, Sebastião e Bárbara, por serem minha fortaleza e por me amarem incondicionalmente. Sempre me deram força e acolhimento nos momentos mais difíceis e sempre acreditaram no meu potencial. Vocês fizeram o que podiam para que eu tivesse uma excelente educação e me ensinaram o significado de responsabilidade e honestidade. A distância tem sido muito pesada, mas, mesmo assim, vocês sempre me apoiaram.

À Lucas Gazolla Rodrigues Soares, por vibrar a cada conquista minha, acreditar em mim (por muitas vezes mais do que eu mesma) e me incentivar. O seu apoio e suporte tem sido um acalento, especialmente nos dias mais tempestuosos. À Alzira e Sidney Rodrigues Soares, por todo cuidado, apoio, torcida e carinho que vocês têm comigo.

À Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista (FCAV/UNESP) e à University of Guelph, por ser minha segunda casa e pela excelência em pesquisa e ensino. Aos professores e funcionários por darem o melhor de si para minha formação acadêmico-profissional. Agradeço ao Programa de Pós-Graduação em Ciência Animal e à CAPES (001 e 88887.802720/2023-00) pelo financiamento da minha bolsa de pesquisa e de sanduíche. Especialmente agradeço:

Ao meu orientador, Prof. Dr. Danísio Prado Munari, aos meus coorientadores Dr. Marcos Vinicius Gualberto da Silva e Dr.ª Tatiane Cristina Seleguim Chud, e ao meu supervisor durante estágio sanduíche, Prof. Dr. Flávio Schenkel. Eu sempre achei a frase “se pude ver mais longe é porque estava em ombros de gigantes” um clichê, mas não há melhor maneira de descrever o que foram esses anos de doutorado. Vocês sempre foram pacientes e solícitos comigo, agradeço por todas as oportunidades, críticas, discussões científicas, ensinamentos acadêmicos e de vida, e pela confiança depositada em mim. Vocês foram fundamentais na minha formação e continuam a me inspirar a ser uma pesquisadora melhor.

Aos membros da banca do exame geral de qualificação, Prof. Dr. Marcos Buzankas e Prof.ª Dr.ª Amanda Maiorano, e aos membros da banca de defesa da dissertação, Dra. Lenira El Faro, Prof.ª Dr.ª Priscila Bernardes, Prof.ª Dr.ª Amanda Maiorano e Prof. Dr. Salvador Ramos, e pelas críticas, reflexões e sugestões a este trabalho.

Agradeço à Embrapa (SEG 02.13.05.011.00.00), CNPq (310199/2015-8 e 431629/2016-1) e FAPEMIG (CVZ PPM 00606/16 e APQ-02750-23) pelo financiamento das amostras utilizadas neste trabalho. Agradeço também a Associação de Criadores de Gir Leiteiro (ABCGIL), Embrapa Gado de Leite e 1000 Bulls Project Consortium pelos dados cedidos e ao Laboratório Multiusuário de Bioinformática da Embrapa Informática Agropecuária e Centre for Genetic Improvement of Livestock (CGIL) pela infraestrutura computacional e recurso de TI.

Aos companheiros de pós-graduação e do Estatística Aplicada a Genética e Melhoramento Animal (EAGMA) pelo apoio durante toda a trajetória. Com vocês compartilhei muitas dúvidas, momentos de estudo, alegria e confraternização que fizeram meus dias mais leves e felizes. Em especial agradeço a Amanda Baldi, Ana Carolina de Jesus, Ana Carolina Paz, Bruna Salatta, Camilla Arantes, Edimar Gabiati, Gabriel Gubiani, Ivan Filho, Júlia Rodrigues, Júlia Valente, Larissa Temp, Rafael Watanabe, Tádia Stivanin, Roney Teixeira, Thomaz Sena e Viviane Ligori.

À Adriana e Shirley, funcionárias do departamento de Engenharia e Ciências Exatas, por toda atenção e disposição em me auxiliar.

I thank the CGIL group, especially Alice Vanzin, Bruno Galindo, Chiara Gini, Colin Lynch, Emanueli Da Silva, Gabriella Dodd, Isis Hermisdorff, Ivan Campos, Krishani Sinhalage, Kristin Lee, Lucas Lopes, Ricarda Jahnel, Samla Cunha, Sirlene Lazzaro, and Stephanie Lam, for the amazing coffees, interactions in lab meetings, and for making the period abroad more fun. I also thank Dr. Filippo Miglior and Dr. Christine Baes for the scientific discussions and support.

To my friends in Canada. Melissa Nagy (*in memoriam*), who became one of my greatest friends. Thanks for making my life cozier and reminding me that some things “look like a tomorrow problem” and that we don’t need to take life so seriously. To Barbara Mann and Digant Purandare, with whom I shared great adventures and learned more about life and kindness. To Mélanie Gagné, Magalie, Mélianne,

Médérick, and Martin Gregoire for opening their home, farm, and hearts to me. At Triple G Farm, I learned so much about dairy farming and the beauty of Canadian endangered breeds.

Agradeço aos demais que, de alguma forma, contribuíram com a minha formação e com a realização deste trabalho, mas que não foram citados nominalmente.

Também agradeço pelas oportunidades, direitos não universalizados e privilégios que me foram concedidos ao longo de toda minha vida acadêmica. Aos meus objetivos, erros e acertos que me ensinaram muito.

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

TABLE OF CONTENTS

CHAPTER 01 – GENERAL CONSIDERATIONS 1
1 INTRODUCTION 1
2 LITERATURE REVIEW 2
2.1 Breeds 2
2.2 Selection signatures 4
2.3 Approaches to detect selection signatures 7
2.4 Copy number variations (CNV) under selection in dairy cattle 13
2.5 Biodiversity variables 14
3 REFERENCES 16
CHAPTER 02 – SELECTION SIGNATURES IN GIR AND HOLSTEIN CATTLE ... 23
1 INTRODUCTION 24
2 MATERIALS AND METHODS 25
2.1 Samples and Sequencing 25
2.2 Genomic Diversity 27
2.3 Selection Signatures 28
2.4 Functional Analysis 30
3 RESULTS 31
3.1 Alignment and Variant Calling 31
3.2 Genomic Diversity 32
3.3 Selection Signatures and Genes Under Selection 37
3.4 Overlapping Selection Signatures 45
4 DISCUSSION 47
4.1 Population Differentiation and Genetic Diversity 47
4.2 Minor Allele Frequency (MAF) 49
4.3 Heterozygosity and Inbreeding Coefficient 49
4.4 Selection Signatures and Genes Under Selection 50
5 CONCLUSIONS 56
6 REFERENCES 57
CHAPTER 03 – SELECTION SIGNATURES IN DAIRY CATTLE: A SYSTEMATIC REVIEW 67
1 INTRODUCTION 68

2 MATERIAL AND METHODS 69
2.1 Literature search and inclusion criteria 69
2.2 Exclusion criteria 70
2.3 Functional analysis 70
3 RESULTS AND DISCUSSION 70
3.1 Studies and genes under selection 70
3.2 Fertility 76
3.3 Health 77
3.4 Milk production and components 79
3.5 Environmental adaptation 82
3.6 Color patterns 85
3.7 Possible pleiotropic genes under selection 87
4 PROSPECTS 88
4.1 Molecular technology 88
4.2 Integrative approaches 88
4.3 Future hot topics 89
6 CONCLUSIONS 89
7 REFERENCES 90

ASSINATURAS DE SELEÇÃO EM BOVINOS DE LEITE

RESUMO – Tanto a seleção natural como a seleção artificial deixam marcas no genoma, conhecidas como assinaturas de seleção, que podem revelar regiões relacionadas a características adaptativas e economicamente importantes para a produção de leite. Além disso, a seleção pode alterar a diversidade genética de uma população. Elevada intensidade de seleção e uso recorrente de um número reduzido de reprodutores podem diminuir a diversidade genética, com possíveis impactos negativos em programas de melhoramento a longo prazo. Como exemplo de população leiteira, a raça Holandesa tem sido fortemente selecionada para alta produtividade no Canadá e nos Estados Unidos, enquanto a raça Gir desempenha papel relevante na produção de leite em regiões tropicais, como Índia e América Latina. No Brasil, o gado Gir foi introduzido entre os séculos XIX e XX e, desde 1985, é selecionado para produção de leite. O objetivo desta tese foi caracterizar assinaturas de seleção em populações de bovinos leiteiros. O capítulo 2 teve como objetivo avaliar a diversidade genética e identificar assinaturas de seleção em bovinos Holandeses da América do Norte (HOL), Gir da Índia (GIR_IN) e Gir Leiteiro (GIR_BR). A diversidade genética foi medida por meio da nucleotide diversity, densidade de variantes de nucleotídeo único, frequência do alelo menor, heterozigosidade observada e esperada, e coeficiente de endogamia. As assinaturas de seleção foram detectadas utilizando Tajima's D, iHS (integrated haplotype score), F_{ST} (fixation index) e XP-EHH (cross-population extended haplotype homozygosity) nos cromossomos autossomos, além de Tajima's D e iHS no cromossomo X. A população HOL apresentou menor diversidade, enquanto GIR_IN e GIR_BR apresentaram maior variabilidade genética. Genes previamente associados a características de importância econômica, incluindo crescimento, reprodução, mastite, produção de leite, tolerância ao calor, saúde e adaptação, foram identificados em regiões sob seleção. No capítulo 3, uma revisão sistemática sobre assinaturas de seleção em bovinos leiteiros é apresentada. O total de 21 estudos em 19 raças (taurinas, zebuínas e compostas) foram selecionados, com assinaturas de seleção que sobrepuseram 5.099 genes. Genes sob seleção relacionados a características como fertilidade, saúde, produção e composição do leite, adaptação ambiental, padrões de pelagem e genes com possível pleiotropia, foram discutidos. Genes foram relatados recorrentemente nos estudos de assinaturas de seleção, como *PLAG1 e HMGA2*. Embora *DGAT1* seja um gene importante para características do leite, não apresentou evidências de seleção na literatura revisada. A revisão também discutiu perspectivas futuras envolvendo ferramentas moleculares e abordagens integrativas. Como uma conclusão geral, a seleção em bovinos leiteiros é moldada não somente por características econômicas, mas também por pressões ecológicas e adaptativas. Em ambos os capítulos 2 e 3, os genes sob seleção foram relacionados a diversas características de importância econômica. Por fim, o estudo de assinaturas de seleção fornece informações valiosas para o melhoramento genético e a sustentabilidade da pecuária leiteira.

Palavras-chave: assinatura de seleção, bovinos de leite, diversidade genética, genes

SELECTION SIGNATURES IN DAIRY CATTLE

ABSTRACT – Natural and artificial selection leave footprints on the genome, known as selection signatures, which can reveal regions related to adaptive and economically important traits for the dairy sector. Additionally, selection may also alter the genetic diversity of a population. Strong selection pressure and the recurrent use of a limited number of sires may decrease genetic diversity, with possible negative effects on long-term breeding programs. As an example of dairy cattle, the Holstein breed has been strongly selected for high productivity in Canada and the United States. In contrast, the Gir breed has an important role in milk production in tropical regions such as India and Latin America. In Brazil, Gir cattle were introduced during the 19th and 20th centuries and, since 1985, have been selected for milk production. This doctoral dissertation aimed to characterize selection signatures in dairy cattle populations. The second chapter aimed to assess genetic diversity and identify selection signatures in Holstein cattle from North America (HOL), Gir cattle from India (GIR_IN), and Dairy Gir cattle from Brazil (GIR_BR). Genetic diversity was measured using nucleotide diversity, single-nucleotide variant density, minor allele frequency, observed and expected heterozygosity, and the inbreeding coefficient. Selection signatures were detected using Tajima's D, integrated haplotype score (iHS), fixation index (FST), and cross-population extended haplotype homozygosity (XP-EHH) on autosomes, as well as Tajima's D and iHS on the X chromosome. The HOL population showed lower diversity, while GIR_IN and GIR_BR presented greater variability. Genes previously associated with traits of economic importance were identified, including traits related to growth, reproduction, mastitis, milk production, heat tolerance, health, and adaptation. In addition, in the third chapter, we systematically reviewed studies published from 2016 to 2023 on selection signatures in dairy cattle. A total of 21 studies in 19 breeds (taurine, indicine, and composite) identified selection signatures overlapping with 5,099 genes. We discussed genes under selection related to economically important traits, such as fertility, health, milk production and composition, environmental adaptation, color patterns, and potential pleiotropic genes. Several genes were reported in numerous selection signature studies, such as *PLAG1* and *HMGA2*. Although DGAT1 is an important gene for milk production and composition, it did not show evidence of selection in recent literature. The systematic review also discussed future perspectives involving molecular tools and integrative approaches. As a general conclusion, dairy cattle selection is shaped not only by economic traits but also by ecological and adaptive pressures. In both chapters, the genes under selection were linked to several traits, including growth, reproduction, milk production and components, heat tolerance, health, and adaptation. The study of selection signatures provides valuable information for genetic improvement and sustainable breeding.

Keywords: dairy cattle, genes, genetic diversity, selection signatures

CHAPTER 01 – GENERAL CONSIDERATIONS

1 INTRODUCTION

The identification of regions under selection and their consequences on the modification of species is one of the interests of genetic and animal breeding area. The combination of geographic separation and natural and artificial selection has shaped the genomic architecture of cattle, leading to genetic differentiation among bovine breeds. Therefore, this differentiation can shape patterns of genetic variation of selected regions, leading to selection signatures that provide information on genome-changing mechanisms and can be used to identify loci subject to selection (Campos et al. 2017). The study of selection signatures and genetic diversity contributes to understanding processes related to genome evolution (Nielsen et al. 2005). Furthermore, it may help the identification of causal variants that control the economically important phenotypes in cattle, such as milk production, reproduction, body conformation, diseases and parasites resistance, and behavior.

In Brazil, the Dairy Gir cattle in the main dairy zebu breed, and Holstein is the primary dairy taurine breed. However, the genetic mechanism underlying selection signatures in these populations remains incompletely elucidated. Taurine cattle have higher performance but also larger environmental demands. Zebu cattle are better adapted to breeding conditions in the tropics. Taurine breeds have been selected for economic traits for a longer period and with greater intensity than zebu breeds, especially when referring to zebu populations from developed and sub-developed countries.

The strong selection pressure may be the reason for the higher linkage disequilibrium (LD) level observed in *Bos taurus* than in *Bos indicus* animals (Thévenon et al. 2007; O'Brien et al. 2014). The detection of selection signatures is based on the frequency of alleles that are under selection or in adjacent haplotypes. When the frequency is higher, it may generate segments of consecutive homozygous genotypes (Ghoreishifar et al. 2020b, a).

With the widespread use of reproductive technologies and increasing selection pressure, monitoring genetic diversity and selection signatures in dairy genetic resources is fundamental. This importance is recognized in Sustainable Development

Goal (SDG) 2.5, which aims to maintain genetic diversity. The aim of this thesis is to characterize selection signatures in dairy cattle populations. In Chapter 2 of this thesis, a study on selection signature and genetic diversity in Holstein and Dairy Gir cattle is presented. In Chapter 3, a systematic review of recent studies on selection signatures in dairy cattle is reported.

2 LITERATURE REVIEW

2.1 Breeds

2.1.1 Gir

The Gir zebu breed originates from India, specifically from the Kathiawar peninsula, in the southwest of Gujarat state, bordering the Arabian Sea. The importation of animals from India occurred between 1870 and 1962, during which approximately 6,000 zebu cattle were brought to Brazil, and fewer than 700 were Gir. Due to their adaptation to Brazilian conditions, both Gir cattle and their crosses with local criollo animals spread rapidly (O'Brien et al. 2015).

In 1938, the establishment of the herdbook for zebu breeds by the Brazilian Association of Zebu Breeders (ABCZ), in Uberaba, contributed to the regulation and dissemination of the Gir breed in the country. During the 1960s, some breeders selected animals for dual purpose, while a smaller group focused exclusively on milk production. After decades of phenotypic selection, the National Breeding Program of Dairy Gir (PNMGL) was created in 1985. In the same year, the progeny test program was implemented, contributing to population growth and to increased milk production of the breed. In the same year, the progeny test program was implemented, contributing to both the increase in population size and the enhancement of milk production within the breed (Santana et al. 2014).

Molecular studies in the breed began in 2001, with the collection of blood and semen to establish the Dairy Gir DNA bank. Following this, molecular studies using microsatellite markers were conducted on milk protein genes. Currently, the PNMGL reports the genotypes of kappa-casein, beta-lactoglobulin, and beta-casein. In 2013, analyses for hereditary diseases were introduced, including DUMPS (Deficiency of Uridine Monophosphate Synthase), CVM (Complex Vertebral Malformation), and BLAD (Bovine Leukocyte Adhesion Deficiency). Since 2016, genomic information

has been utilized for parentage verification. After 2018, genotypes have also been applied to predict genomic estimated breeding values (GEBV) for females and bulls, including progeny-tested animals (Panetto et al. 2025). These advances highlight the importance of genomic studies in the breed.

There has been a decline in the effective population size (N_e) of Dairy Gir cattle, with the last five generations showing a N_e of just 11 animals (Garcia et al. 2023). These findings raise concerns about the genetic diversity within this population. The reduction may be associated with the extensive usage of the bull Sansão and his offspring. Sansão has ranked as the leader for milk production in the PNMGL summary for ten years, and many of his sons have also been featured among the top bulls for milk production in the PNMGL reports (Compre Rural and Peruchi 2019).

According to the PNMGL genetic evaluation report from 2025 (Panetto et al. 2025), the average and standard deviation (±SD) were 3,743 ± 2,089 kg for 305 days milk yield, 135 ± 82 kg for fat yield, 4.08 ± 0.85% for fat rate, 121 ± 71 kg for protein yield, 3.36 ± 0.32% for protein rate, 459 ± 284 kg for total solids, 12.54 ± 1.27% for total solids rate, 273 ± 83 days for lactation length, and 1,166 ± 195 days for age at first calving. As a strategy to improve milk yield in Brazil, Dairy Gir are crossed with taurine breeds, particularly Holstein cattle, resulting in the Girolando cattle, which contribute to 80% of the national milk production.

2.1.2 Holstein

The Holstein breed originated approximately 2,000 years ago in the Netherlands. Migrant European tribes, specifically the Batavians with their black cattle and the Friesians with their white cows, raised animals using limited feed resources. Dutch exports to European countries and North America began to flourish after 1880, which led to the emergence of the highly productive Holstein-Friesian cattle (https://www.holsteinusa.com/holstein_breed/breedhistory.html). Currently, Holstein is the dominant dairy breed worldwide, known for producing the highest milk yield.

The establishment of the Holstein breed in the United States of America (USA) began when dairy breeders sought cattle from Holland due to the growing market for

milk. Between 1852 and 1905, a total of 7,757 black and white cattle, primarily from the Netherlands, were imported, forming the genetic foundation of the current Holstein population. In 1885, the Holstein-Friesian Association of America was established, and it was later renamed the Holstein Association USA, Inc. in 1994. The genomic evaluation in the USA began in 2008, and the number of genotyped animals rapidly expanded (https://www.holsteinusa.com/holstein_breed/breedhistory.html). In 2024, the Council on Dairy Cattle Breeding advertised that they had achieved 10 million dairy cattle genotypes, with the majority being Holstein breed (https://uscdcb.com/database-stats/).

Holstein cattle were brought to Canada in 1881, starting with Michael Cook and Son's herd in Ontario. Herman Bollert acquired foundation stock from Cook in 1883, and most of Canada's Holsteins trace back to Bollert's two animals. A depression in 1896 led to the dispersal of Bollert herd, which helped spread Holstein genetics across Canada. Further genetic development occurred after 1924, when the Montvic Farms in Quebec acquired stock. Animals from this superior Montvic herd had a lasting impact on the Holstein breed in Canada, with many of their descendants still found in Canada's Holstein Herdbook today (https://www.holstein.ca/Public/en/About_Us/The_Canadian_Dairy_Industry/The_Holstein_Breed).

According to the USA National Performance Metrics in 2023 (https://webconnect.uscdcb.com/#/national-performance-metrics), the standardized lactation averages at the age of 36 months, average and SD metrics were 11,868 ± 1,831 kg for 305 days milk yield, 498 ± 83 kg for fat yield, 4.21% for fat rate, 390 ± 57 kg for protein yield, 3.29% for protein rate, and 23.9 ± 2.2 months for age at first calving. In Canada, the average in 2023 was 11,253 kg for 305 days milk yield, 4.08% for fat rate, and 3.32% for protein rate (https://agriculture.canada.ca/en/sector/animal-industry/canadian-dairy-information-centre/statistics-market-information/dairy-animal-genetics/production-breed/per-cow)

2.2 Selection signatures

2.2.1 Molecular technologies and assembly

One of the first molecular techniques used in studies to detect selection signatures in cattle was microsatellites (Wiener et al. 2003). However, with the advancement of molecular and bioinformatic technologies, single nucleotide polymorphism (SNP) arrays and whole-genome sequencing (WGS) data have become the primary sources of genomic information for studies on the detection of selection signatures. The identification of selection signals from SNP arrays depends on the density and distribution of markers across the genome. The capacity to detect selection signatures is influenced by the density of the array, particularly for low to medium coverage. This limitation is due to an insufficient number of markers to properly identify a selection signature, which may also exist in areas without marker coverage. Gaps in marker coverage can inflate the statistical signals, causing false positives (Gautason et al. 2021). The use of WGS data enables more accurate detection of selection signatures owing to its resolution, but the cost of sequencing numerous individuals is still high, and the computational demand is a limiting factor. Other molecular techniques allow the identification of selection signatures, including genotype by sequencing (GBS) (Liu et al. 2021) and double digest restriction-site associated DNA sequencing (ddRAD) (Sukhija et al. 2024).

Genomic studies in cattle benefit from updates in the reference genome, with better assemblies and functional annotations than those of other domestic animals. The choice of reference genome can affect selection signature results (Lloret-Villas et al. 2021). In the same manner, studies using the latest cattle reference genome (ARS_UCD) (Rosen et al. 2020) are more accurate than studies that used the previous versions. As ARS_UCD versions are completer and contiguous, with fewer gaps and larger contigs, it provides more confidence in the position of variants (Rosen et al. 2020) and, consequently, selection signatures. The continued improvement of the bovine reference genome assembly, such as through the telomere-to-telomere project or the use of pangenomes, will enhance the investigation of selection signatures in dairy cattle. Additionally, better functional annotation through the study of genes, quantitative trait *loci* (QTL), and non-coding RNA, as well as the provision and curation of information in genomic databases, is essential for identifying causal regions under selection.

2.2.2 Theoretical framework

The synonyms for selection signature are signatures of selection and selective sweep. The distribution of selection signatures across the genome can be concentrated in single regions or spread across multiple *loci*. The spread of genetic variation in a population is influenced by the origin of the variation and the rate of adaptive allele sweep (Qanbari and Simianer 2014; Utsunomiya et al. 2015; Saravanan et al. 2020; Panigrahi et al. 2023). Two selection signature classifications are observed: hard sweeps and soft sweeps, and interesting illustrations are presented in Saravanan et al. (2020) (Figure 1).

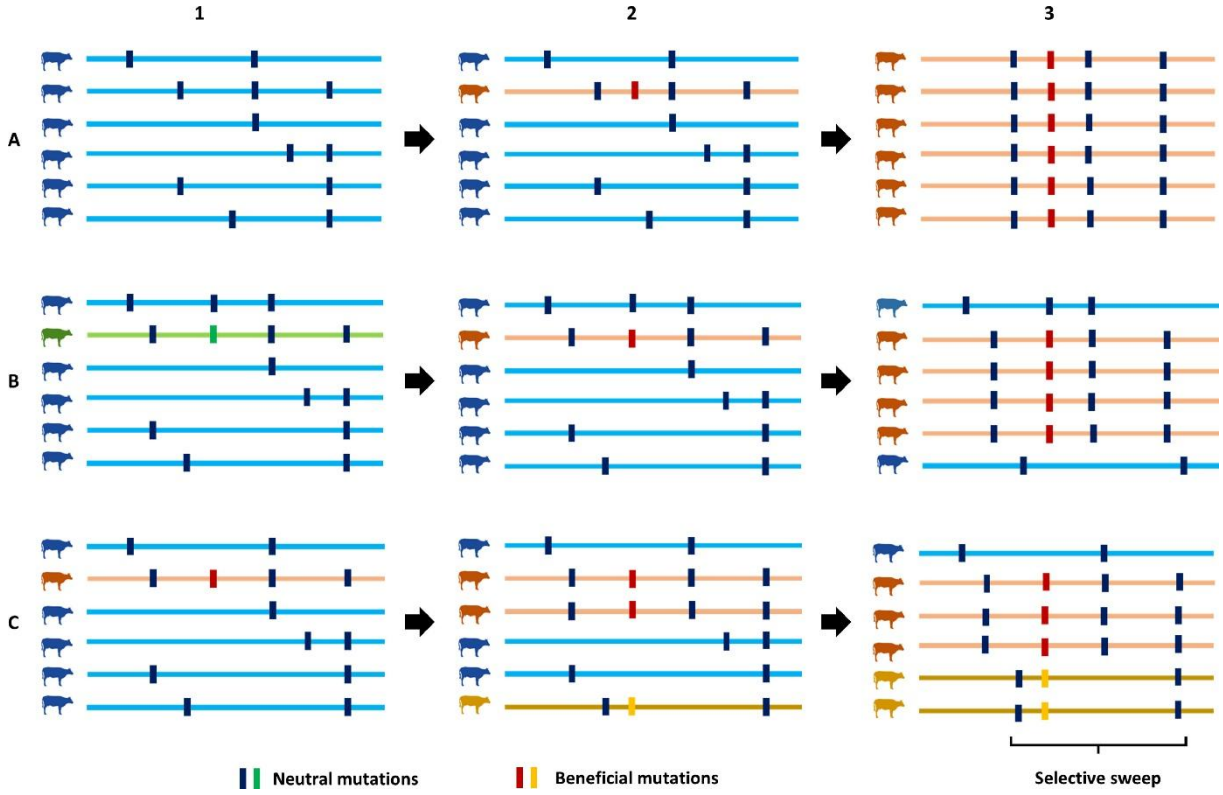

Figure 1. Types of selective sweeps (Saravanan et al. 2020). A). Hard/ classic selective sweep: 1) DNA sequences from a population with neutrally segregating alleles (blue) prior to selection; 2) a beneficial mutation in the second sequence (red) causes genetic variability around the favorable allele; 3) the beneficial allele increases in frequency in subsequent generations during the selection process and eventually reaches fixation in the population. B) Soft selective sweep due to standing

genetic variation: 1) DNA sequences from a population with neutral mutations (blue and green); 2) some genetic or environmental changes cause neutral alleles that segregate in a population (green) to become beneficial (red); 3) selection fixes the beneficial alleles in the population while leaving some genetic variation surrounding them. C) Multiple origin soft sweep (due to current independent mutations): 1) The first advantageous mutation (red) appears in the second DNA sequence; 2) both the associated neutral sites and the advantageous allele (red) become more frequent during selection. At the same locus, a second advantageous mutation (yellow) takes place; 3) the frequency of both advantageous mutations rises, but none of them can become fixed.

A hard sweep occurs when a beneficial allele is rapidly selected, causing the favored allele to reach fixation and inadvertently increasing the frequency of linked alleles. Soft sweep is either a single-origin or multiple-origin type. In the single-origin soft sweep, selection acts on standing genetic variation that was previously neutral or deleterious but has become beneficial or adaptive due to environmental or genetic changes (Hermisson and Pennings 2005). In the multiple-origin soft sweep, multiple beneficial mutations occur at a single locus on different genomic backgrounds, and they all increase in frequency simultaneously without reaching fixation (Pennings and Hermisson 2006). The softness of the sweep makes it more difficult to detect, as patterns of diversity around the selected site begin to resemble those expected under neutrality (Harris et al. 2018).

2.3 Approaches to detect selection signatures

Selecting an appropriate statistical approach for detecting selection signatures in dairy cattle should consider the sample type, study objectives, and a good understanding of how different tests detect selection signatures. Their statistical power depends on allele or haplotype frequency, age and objective of selection, number of individuals, and genotyping density (Abbasi-Moshaii et al. 2023).

Various tests have been employed to detect selection signatures and can be primarily classified into between- and within-population. For between-populations analysis with different breeds, the distinct domestication and selection histories are

explicit. However, for subpopulations within the same breed, it becomes important to consider phenotypic divergence, supported by statistical tests or evidence of substructure in genotype data, often shown through PCA or discriminant analysis (Jombart et al. 2010). Additionally, it is important to control relatedness among sampled individuals in order to capture population variation and minimize confounding signals resulting from family patterns. Such analyses confirm genetic divergence between the two groups, forming a robust basis for detecting selection signatures.

A new beneficial genetic variant that has been under positive selection pressure creates distinct signatures in the corresponding genomic region, including (i) an increase in the allele frequency spectrum towards extreme (high or low) frequencies; (ii) an excess of homozygous genotypes; (iii) the occurrence of haplotypes with high frequency, and (iv) an unusually high level of local population differentiation (Ma et al. 2015).

2.3.1 Within-population tests

In this section, we will discuss common approaches used to identify selection signatures on dairy cattle populations, such as those based on site frequency spectrum, LD, reduced local variability, and haplotype homozygosity.

2.3.1.1 Approaches based on linkage disequilibrium (LD)

Integrated haplotype score

Integrated haplotype score (iHS) is a highly sensitive test based on identifying high LD regions. It is a standard measure of decay in Extended Haplotype Homozygosity (EHH), and its score reflects how unusual the haplotypes carrying the ancestral (positive values) and derived (negative values) allele are, compared to the whole genome (Voight et al. 2006; Utsunomiya et al. 2013b). The iHS test is particularly effective at detecting selection signature when the selected allele has risen to intermediate frequencies, but it requires that the derived allele must have existed solely on a distinct background prior to selection and must not have reached fixation (Voight et al. 2006; Sabeti et al. 2006).

2.3.1.2 Approaches based on site frequency spectrum

All approaches based on site frequency spectrum assume the neutral evolutionary model, which assumes that genetic variation within a population arises only through random genetic drift and mutation and is not influenced by non-random forces.

Tajima's D

The Tajima's D statistic compares two measures of genetic diversity, the number of segregating sites (S) and the nucleotide diversity (π), which is the average number of pairwise nucleotide differences between sequences in a sample. Tajima's D value is expected to be zero under the standard neutral evolutionary model. Tajima's D may deviate from zero in populations that have experienced selection or other non-neutral forces. Negative values suggest the fixation of rare alleles, and positive values denote balancing selection, which reflects an abundance of intermediate allele frequencies (Tajima 1989; Korneliussen et al. 2013).

Fu and Li's D* statistic

As an extension of Tajima's D, Fu and Li's D* statistic detects selection signature by comparing the number of segregating sites and the number of singleton mutations (alleles appearing only once in the sample) in a genomic region. It assumes that these two estimates are correlated under the neutral evolutionary model (Fu and Li 1993).

CLR

The composite likelihood ratio (CLR) measures the skewness of the frequency spectrum across multiple loci. It compares data under a model assuming neutral evolutionary versus selection, while also incorporating information on the recombination rate to differentiate selection from other demographic events (Kim and Stephan 2002; Nielsen et al. 2005). This approach is not influenced by different demographic events, as Tajima's D, and has high power to detect selective alleles that have achieved fixation recently (Rajawat et al., 2022; Stella et al., 2010).

2.3.1.3 Approaches based on reduced local variability

ROH

Runs of homozygosity (ROH) are contiguous stretches of homozygous genotypes within an individual that arise when two haplotypes share a recent common ancestor (Gibson et al. 2006; Rebelato and Caetano 2018). Because recombination is random, the distribution of ROH across samples is expected to be highly heterogeneous under neutrality, and a high frequency of ROH detected around a specific locus in most individuals, such as a ROH island, can indicate positive selection (Curik et al. 2014). Populations with non-random mating are expected to have more and longer ROH segments. In admixed populations, the number of ROH is reduced and remains short due to the introgression of different haplotypes (Rajawat et al. 2022). Recent studies in dairy cattle have shown that breeds subject to higher selection pressure and inbreeding have more medium and longer ROH segments than breeds with lower selection pressure (Saravanan et al. 2021; Jin et al. 2022; Rajawat et al. 2022; Zhang et al. 2022).

2.3.1.4 Approaches based on haplotype homozygosity

H1, H2 and H12

Garud et al. (2015) developed three haplotype-based homozygosity approaches that focus on the detection of soft sweeps: H1, H2, and H12. H1 measures expected haplotype homozygosity, with hard sweeps expected to have a higher H1 value than soft sweeps. H2 is a measure of expected haplotype homozygosity omitting the most frequent haplotype, and it is expected to be larger for soft sweeps than hard sweeps. H12 combines the first and second most frequent haplotype in a sample into a single combined haplotype frequency, and then homozygosity is recalculated using the revised haplotype frequency distribution. H12 is expected to have more power to detect hard and soft sweeps than H1.

2.3.2 Between-population tests

In this section, two approaches used to identify selection signatures between dairy cattle populations, such as those based on single-site differentiation and haplotype differentiation, will be discussed.

2.3.2.1 Approaches based on single-site differentiation

F_{ST}

The fixation index F_{ST} (Wright 1949) is an SNP-specific approach that measures population genetic differentiation and can detect actual genetic variants under selection. F_{ST} values vary between zero, which implies no differences in allelic frequencies between two populations, to one, meaning each population is fixed for a different allele. The power of F_{ST} approach is limited when studying distantly related populations, and F_{ST} values are subject to bias and false positives when the population does not have the same effective size and is hierarchically structured (Price et al. 2010; Fariello et al. 2013).

FLK

The FLK approach is an extension of Lewontin and Krakauer's (LK) statistic (Lewontin and Krakauer 1973) that considers historical branching (population subdivision) and heterogeneity of genetic drift. Unlike F_{ST}, the FLK approach integrates the hierarchical population characteristics. It accounts for the population history through a kinship or coancestry matrix, based on the Reynolds distance, which captures (i) differences in effective population sizes (N_e) between populations and (ii) possible shared ancestry between populations (Bonhomme et al. 2010).

2.3.2.2 Approaches based on haplotype differentiation

hapFLK

hapFLK is a haplotype-based extension of the FLK approach that also accounts for the hierarchical structure and heterogeneity of different populations. hapFLK uses haplotype clustering to calculate how different the haplotype frequencies are between two populations, which allows it to detect selection signatures. This approach has high power to detect selection signatures with high and medium frequencies (Fariello et al. 2013).

XP-EHH

The cross-population extended haplotype homozygosity (XP-EHH) test is based on LD, a non-random association of alleles between two or more *loci*, and identifies alleles that have rapidly increased to high frequency and are still associated with nearby polymorphism. XP-EHH compares the integrated extended haplotype homozygosity (EHH) between two populations at the same SNP and detects selected alleles that have approached or achieved fixation in one population but not in the other (Sabeti et al. 2007). XP-EHH is indicated for detecting ongoing or nearly fixed selection signatures. In addition, this test was found to have more power than F_{ST} but less power than hapFLK for detecting hard sweeps (Fariello et al. 2013).

2.3.3 Using complementary approaches and combining approaches

In addition to the signatures of selection tests mentioned above, several other approaches have been employed in dairy cattle. Other approaches were reviewed by Saravanan et al. (2020), Utsonomiya et al. (2015), Qanbari and Simianer (2014), and Panigrahi et al. (2023). It is a frequent practice to utilize multiple approaches in a study to take advantage of complementary tests. Statistical power can be increased with the use of

distinct approach-based tests, which rely on different information to detect selection signatures.

Some methods have previously been proposed to combine various tests. Grossman et al. (Grossman et al. 2010) proposed a method that involves combining various p-values to produce a composite of signals (CMS). However, CMS relies on demographic history, which can be a limitation for livestock species with complex and unknown demographic histories (Illa et al. 2021). To address this, Utsunomiya et al. (2013b) suggested a combined statistic called meta-SS, which merges different genome-wide scan statistics without relying on demographic simulations. Randhawa et al. (2016) proposed a method called Composite Selection Signals (CSS), which integrates multiple pieces of selection evidence from constituent tests' rank distribution. However, meta-SS and CSS do not account for covariance among different single approaches. Ma et al. (2015) proposed a de-correlated composite of multiple selection signals (DCMS) to combine the outcomes of complementary tests while accounting for covariance structure among a single approach. DCMS is more

effective than earlier methods and facilitates the prioritization of candidate genes affecting major economic traits in animal breeding (Illa et al. 2021). DCMS has been used in several studies (Peripolli et al. 2020; Illa et al. 2021; Rajawat et al. 2022). However, most studies use complementary tests, but do not use composite approaches.

2.4 Copy number variations (CNV) under selection in dairy cattle

Structural variants, such as copy number variations (CNV), have contributed to evolution, adaptation, and production. CNV are gains or losses of DNA segments that can occur at multiple sites in the genome (Redon et al. 2006). CNV are heritable and most CNV follows Mendelian inheritance (Locke et al. 2006). CNV can reveal selective sweeps beyond SNP; however, research on the relationship between selection signatures and CNV is still in its initial stage.

CNV are related to selection signatures in several ways. A CNV may be swept and arise as a selection signature. A CNV might provide beneficial advantages and increase frequency due to directional selection or be hitch-hiked due to positive selection. For example, a strongly selected 12 kb multiallelic CNV with pleiotropic effects was identified in Holstein cattle on chromosome 6. This CNV, with a segment number of 4-6, had a positive effect on milk yield but a negative effect on fertility and resistance to mastitis (Lee et al. 2021).

The identification of selection signatures from CNV is possible in two distinct groups, using population frequencies of allelic CNV segment numbers (0, 1, 2, 3, 4, >4) (McCarroll et al. 2008; Poptsova et al. 2013) or Log R Ratio (LRR) from SNP arrays instead of allele frequencies (Redon et al. 2006). In dairy cattle, few studies have identified selection signatures from CNV data. In a study between Jersey and Holstein populations, the authors found segments related to starch digestion and body size (Lee et al. 2020) using F_{ST} based on CNV regions (CNVR).

Bickhart et al. (2016) identified a dairy-specific (Holstein, Jersey, and Gir) CNV under selection in the *RICTOR* gene using a heuristic approach with beef-purpose breeds. This gene encodes the RICTOR protein, a key component of the mTORC2 complex, which regulates lipogenic gene expression, lipid synthesis, and secretion in bovine mammary epithelial cells (Guo et al. 2019). Strillacci et al. (2018) used V_{ST},

which is correlated to F_{ST}, and identified the lysosome gene (*LYZ*) under selection with Valdostana Red Pied (dual purpose) and Italian Brown-Swiss cattle pair analyses. The authors hypothesized that this gene is involved in bacterial infections, including mastitis, and highlighted that Valdostana is less susceptible to mastitis than Holstein cattle.

Although CNV roles are associated with traits such as production, fertility, and disease resistance (Lee et al. 2020, 2021; Butt et al. 2021), their mechanisms remain poorly understood and require further investigation in genomic studies. Integrating CNV into selection signature analyses aids in identifying regions under selective pressure. Advances in high-resolution genotyping and sequencing techniques provide new opportunities to detect CNV more effectively. Therefore, these discoveries are expected to enhance genomic selection strategies, improving the efficiency of dairy cattle production through the exploitation of CNV.

2.5 Biodiversity variables

2.5.1 Genetic Diversity

Genetic diversity offers resilience against abrupt changes and enables species and ecosystems to adapt to changing environments, climates, and other challenges. Ultimately, genetic variation allows species to develop into distinct and new lineages.

Genetic diversity refers to the level of genetic variability present within a population. To properly capture genetic diversity, both richness and evenness should be evaluated. Richness denotes the count of genetic variants within a population, adjusted for sample size (e.g., an expectation for a given sample size). Nucleotide diversity (π) and allelic richness are two examples of richness metrics. The only way to increase genetic richness is through mutations or, locally, through gene flow. Evenness of genetic diversity is the expected proportion of heterozygotes in a population at equilibrium, a common example of which is the expected heterozygosity (H_E). This metric indicates the probability that two randomly selected alleles from a sample are different. Like richness, evenness will help a population respond rapidly to changes in the environment through adaptation (Hoban et al. 2022).

2.5.2 Genetic Differentiation

Genetic differentiation refers to the divergence in genetic composition, specifically the allele frequencies, between multiple populations of the same species. This divergence can be caused by natural selection, genetic drift, and restricted gene flow. The degree of genetic differentiation among populations can be quantified by fixation index (F_{ST}) (Hoban et al. 2022). The F_{ST} ranges from zero to one. A value of zero indicates no differences in allele frequencies between the two subpopulations, while one suggests that the allele frequencies are completely different in the two subpopulations. In the latter case, each of the two populations is fixed for a different allele; thus, within-population diversity must be zero (Ajmone-Marsan et al. 2023). Genetic drift occurring within a single population may result in an increased F_{ST} value in between-population analysis, reflecting greater genetic differentiation between the two populations under consideration.

2.5.3 Inbreeding

Inbreeding occurs when related individuals' mate. Inbreeding metrics are used to determine the level of diversity within individuals. As such, these metrics can reveal potential risks related to genetic diversity, which indicates the need for management interventions. In selecting for the improvement of a particular trait, the accumulation of homozygosity at favorable variants is the primary objective. However, the unintended result of how selection is conducted in breeding programs is the accumulation of inbreeding depression (Maltecca et al. 2020). Inbreeding increases the frequency of homozygotes for deleterious alleles because of selection and drift, which results in inbreeding depression (Falconer and Mackay 1996). Highly inbred individuals typically exhibit lower fitness, which can compromise population growth and persistence by reducing demographic rates (e.g., decreased fecundity). Inbreeding rates tend to be reduced when populations are large and there is sufficient gene flow both within and between populations. Inbreeding is often a result of small population size, increasing rapidly when N_e is less than approximately 50 and in fragmented populations where gene flow dynamics are limited (Hoban et al. 2022).

3 REFERENCES

Abbasi-Moshaii, B, Moradi, MH, Yin, T, Rahimi-Mianji, G, Nejati-Javaremi, A & König, S. 2023. Genome-wide scan for selective sweeps identifies novel loci associated with resistance to mastitis in German Holstein cattle. Journal of Animal Breeding and Genetics. 140(1):92–105. doi.org/10.1111/JBG.12737.

Ajmone-Marsan, P, Colli, L, Ginja, C, Kantanen J & Lenstra, JA. 2023. Genomic characterization of animal genetic resources – Practical guide. V. 32. Rome: FAO Animal Production and Health Guidelines. doi.org/10.4060/cc3079en.

Bickhart, DM, Xu, L, Hutchison, JL, Cole, JB, Null, DJ, Schroeder, SG, Song, J, Garcia, JF, et al. 2016. Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. DNA Research. 23(3):253–262. doi.org/10.1093/dnares/dsw013.

Bonhomme, M, Chevalet, C, Servin, B, Boitard, S, Abdallah, J, Blott, S & SanCristobal, M. 2010. Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. Genetics. 186(1):241–262. doi.org/10.1534/GENETICS.110.117275.

Campos, BM, do Carmo, AS, da Silva, TBR, Verardo, LL, de Simoni Gouveia, JJ, Mendes Malhado, CH, Barbosa da Silva, MVG & Souza Carneiro, PL. 2017. Identification of artificial selection signatures in Caracu breed lines selected for milk production and meat production. Livestock Science. 206:82–87. doi.org/10.1016/J.LIVSCI.2017.10.014.

Compre Rural & Peruchi, M. 2019. Sansão reina absoluto no topo da produção de leite - CompreRural. Available from: https://www.comprerural.com/sansao-reina-absoluto-no-topo-da-producao-de-leite/ [Accessed 30 September 2025].

Council on Dairy Cattle Breeding (CDCB). 2024. US Dairy Database Stats. Available from: https://uscdcb.com/database-stats/ [Accessed 30 September 2025].

Curik, I, Ferenčaković, M & Sölkner, J. 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livestock Science. 166(1):26–34. doi.org/10.1016/J.LIVSCI.2014.05.034.

Falconer, DS & Mackay, TFC. 1996. Introduction to Quantitative Genetics. Fourth ed. Upper Saddle River: Prentice Hall.

Fariello, MI, Boitard, S, Naya, H, SanCristobal, M & Servin, B. 2013. Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. Genetics. 193(3):929–941. doi.org/10.1534/genetics.112.147231.

Fu, YX & Li, WH. 1993. Statistical tests of neutrality of mutations. Genetics. 133(3):693–709. doi.org/10.1093/GENETICS/133.3.693.

Garcia, AO, Otto, PI, Glatzl Junior, LA, Rocha, R de FB, dos Santos, MG, de Oliveira, DA, da Silva, MVGB, Panetto, JC do C, et al. 2023. Pedigree reconstruction and population structure using SNP markers in Gir cattle. Journal of Applied Genetics. 64(2):329–340. doi.org/10.1007/S13353-023-00747-X.

Garud, NR, Messer, PW, Buzbas, EO & Petrov, DA. 2015. Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. PLOS Genetics. 11(2):e1005004. doi.org/10.1371/JOURNAL.PGEN.1005004.

Gautason, E, Schönherz, AA, Sahana, G & Guldbrandtsen, B. 2021. Genomic inbreeding and selection signatures in the local dairy breed Icelandic Cattle. Animal Genetics. 52(3):251–262. doi.org/10.1111/AGE.13058.

Ghoreishifar, SM, Moradi-Shahrbabak, H, Fallahi, MH, Jalil Sarghale, A, Moradi-Shahrbabak, M, Abdollahi-Arpanahi, R & Khansefid, M. 2020. Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, Bubalus bubalis. BMC Genetics. 21(1):16. doi.org/10.1186/s12863-020-0824-y.

Ghoreishifar, SM, Eriksson, S, Johansson, AM, Khansefid, M, Moghaddaszadeh-Ahrabi, S, Parna, N, Davoudi, P & Javanmard, A. 2020. Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. Genetics Selection Evolution. 52(1):1–15. doi.org/10.1186/s12711-020-00571-5.

Gibson, J, Morton, NE & Collins, A. 2006. Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics. 15(5):789–795. doi.org/10.1093/HMG/DDI493.

Grossman, SR, Shylakhter, I, Karlsson, EK, Byrne, EH, Morales, S, Frieden, G, Hostetter, E, Angelino, E, et al. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science. 327(5967):883–886. doi.org/10.1126/SCIENCE.1183863/

Guo, Z, Zhao, K, Feng, X, Yan, D, Yao, R, Chen, Y, Bao, L & Wang, Z. 2019. mTORC2 Regulates Lipogenic Gene Expression through PPARγ to Control Lipid Synthesis in Bovine Mammary Epithelial Cells. BioMed Research International. 2019(1):5196028. doi.org/10.1155/2019/5196028.

Harris, AM, Garud, NR & Degiorgio, M. 2018. Detection and Classification of Hard and Soft Sweeps from Unphased Genotypes by Multilocus Genotype Identity. Genetics. 210(4):1429–1452. doi.org/10.1534/GENETICS.118.301502.

Hermisson, J & Pennings, PS. 2005. Soft SweepsMolecular Population Genetics of Adaptation From Standing Genetic Variation. Genetics. 169(4):2335–2352. doi.org/10.1534/GENETICS.104.036947.

Hoban, S, Archer, FI, Bertola, LD, Bragg, JG, Breed, MF, Bruford, MW, Coleman, MA, Ekblom, R, et al. 2022. Global genetic diversity status and trends: towards a

suite of Essential Biodiversity Variables (EBVs) for genetic composition. Biological Reviews. 97(4):1511–1538. doi.org/10.1111/BRV.12852

Holstein Canada. n.d. The Holstein Breed. Available from: https://www.holstein.ca/Public/en/About_Us/The_Canadian_Dairy_Industry/The_Holstein_Breed [Accessed 30 September 2025].

Illa, SK, Mukherjee, S, Nath, S & Mukherjee, A. 2021. Genome-Wide Scanning for Signatures of Selection Revealed the Putative Genomic Regions and Candidate Genes Controlling Milk Composition and Coat Color Traits in Sahiwal Cattle. Frontiers in Genetics. 12:1193. doi.org/10.3389/fgene.2021.699422.

Jang, J, Terefe, E, Kim, K, Lee, YH, Belay, G, Tijjani, A, Han, J, Hanotte, O, et al. 2021. Population differentiated copy number variation of Bos taurus, Bos indicus and their African hybrids. BMC Genomics. 22(1):531. doi.org/10.1186/s12864-021-07808-7.

Jin, L, Qu, K, Hanif, Q, Zhang, J, Liu, J, Chen, N, Suolang, Q, Lei, C, et al. 2022. Whole-Genome Sequencing of Endangered Dengchuan Cattle Reveals Its Genomic Diversity and Selection Signatures. Frontiers in Genetics. 13:611. doi.org/10.3389/fgene.2022.833475.

Kim, Y & Stephan, W. 2002. Detecting a Local Signature of Genetic Hitchhiking Along a Recombining Chromosome. Genetics. 160(2):765–777. doi.org/10.1093/GENETICS/160.2.765.

Korneliussen, TS, Moltke, I, Albrechtsen, A & Nielsen, R. 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics. 14(1):289. doi.org/10.1186/1471-2105-14-289.

Lee, YL, Bosse, M, Mullaart, E, Groenen, MAM, Veerkamp, RF & Bouwman, AC. 2020. Functional and population genetic features of copy number variations in two dairy cattle populations. BMC Genomics. 21(1):1–15. doi.org/10.1186/s12864-020-6496-1.

Lee, YL, Takeda, H, Moreira, GCM, Karim, L, Mullaart, E, Coppieters, W, Appeltant, R, Veerkamp, RF, et al. 2021. A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle. PLOS Genetics. 17(7):e1009331. doi.org/10.1371/JOURNAL.PGEN.1009331.

Lewontin, RC & Krakauer, J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics. 74(1):175–195. doi.org/10.1093/GENETICS/74.1.175.

Liu, D, Chen, Z, Zhao, W, Guo, L, Sun, H, Zhu, K, Liu, G, Shen, X, et al. 2021. Genome-wide selection signatures detection in Shanghai Holstein cattle population identified genes related to adaption, health and reproduction traits. BMC Genomics. 22(1):747. doi.org/10.1186/s12864-021-08042-x.

Lloret-Villas, A, Bhati, M, Kadri, NK, Fries, R & Pausch, H. 2021. Investigating the impact of reference assembly choice on genomic analyses in a cattle breed. BMC Genomics. 22(1):1–17. doi.org/10.1186/S12864-021-07554-W.

Ma, Y, Ding, X, Qanbari, S, Weigend, S, Zhang, Q & Simianer, H. 2015. Properties of different selection signature statistics and a new strategy for combining them. Heredity. 115(5):426–436. doi.org/10.1038/hdy.2015.42.

Maltecca, C, Tiezzi, F, Cole, JB & Baes, C. 2020. Symposium review: Exploiting homozygosity in the era of genomics—Selection, inbreeding, and mating programs. Journal of Dairy Science. 103(6):5302–5313. doi.org/10.3168/JDS.2019-17846.

McCarroll, SA, Kuruvilla, FG, Korn, JM, Cawley, S, Nemesh, J, Wysoker, A, Shapero, MH, De Bakker, PIW, et al. 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nature Genetics 2008 40:10. 40(10):1166–1174. doi.org/10.1038/ng.238.

Nielsen, R, Williamson, S, Kim, Y, Hubisz, MJ, Clark, AG & Bustamante, C. 2005. Genomic scans for selective sweeps using SNP data. Genome Research. 15(11):1566–1575. doi.org/10.1101/GR.4252305.

O'Brien, AMP, Utsunomiya, YT, Meszaros, G, Bickhart, DM, Liu, GE, Van Tassell, CP, Sonstegard, TS, Da Silva, MVB, et al. 2014. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. Genetics Selection Evolution. 46:19. doi.org/10.1186/1297-9686-46-19.

O'Brien, AMP, Höller, D, Boison, SA, Milanesi, M, Bomba, L, Utsunomiya, YT, Carvalheiro, R, Neves, HH, et al. 2015. Low levels of taurine introgression in the current Brazilian Nelore and Gir indicine cattle populations. Genetics Selection Evolution. 47(1):1–7. doi.org/10.1186/s12711-015-0109-5.

Panetto, JC do C, Silva, MVGB, Machado, MA & Martins, MF. 2025. Programa Nacional de Melhoramento do Gir Leiteiro: sumário brasileiro de touros: 8a avaliação genômica de touros: resultado do teste de progênie. Juiz de Fora. Available from: https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1175223/1/Programa-Nacional-de-Melhoramento-do-Gir-Leiteiro-sumario-brasileiro-de-touros-8a-avaliacao-genomica.pdf [Accessed 17 July 2025].

Panigrahi, M, Rajawat, D, Nayak, SS, Ghildiyal, K, Sharma, A, Jain, K, Lei, C, Bhushan, B, et al. 2023. Landmarks in the history of selective sweeps. Animal Genetics. 54(6):667–688. doi.org/10.1111/AGE.13355.

Pennings, PS & Hermisson, J. 2006. Soft Sweeps II—Molecular Population Genetics of Adaptation from Recurrent Mutation or Migration. Molecular Biology and Evolution. 23(5):1076–1084. doi.org/10.1093/MOLBEV/MSJ117.

Peripolli, E, Reimer, C, Ha, N-T, Geibel, J, Machado, MA, Panetto, JC do C, do Egito, AA, Baldi, F, et al. 2020. Genome-wide detection of signatures of selection in indicine

and Brazilian locally adapted taurine cattle breeds using whole-genome re-sequencing data. BMC Genomics. 21(1):624. doi.org/10.1186/s12864-020-07035-6.

Poptsova, M, Banerjee, S, Gokcumen, O, Rubin, MA & Demichelis, F. 2013. Impact of constitutional copy number variants on biological pathway evolution. BMC Evolutionary Biology. 13(1):19. doi.org/10.1186/1471-2148-13-19.

Price, AL, Zaitlen, NA, Reich, D & Patterson, N. 2010. New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics 2010 11:7. 11(7):459–463. doi.org/10.1038/nrg2813.

Qanbari, S & Simianer, H. 2014. Mapping signatures of positive selection in the genome of livestock. Livestock Science. 166:133–143. doi.org/https://doi.org/10.1016/j.livsci.2014.05.003.

Rajawat, D, Panigrahi, M, Kumar, H, Nayak, SS, Parida, S, Bhushan, B, Gaur, GK, Dutt, T, et al. 2022. Identification of important genomic footprints using eight different selection signature statistics in domestic cattle breeds. Gene. 816:146165. doi.org/10.1016/J.GENE.2021.146165.

Randhawa, IAS, Khatkar, MS, Thomson, PC & Raadsma, HW. 2016. A Meta-Assembly of Selection Signatures in Cattle. PLOS ONE. 11(4):e0153013. doi.org/10.1371/JOURNAL.PONE.0153013.

Rebelato, AB & Caetano, AR. 2018. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. Pesquisa Agropecuária Brasileira. 53(9):975–984. doi.org/10.1590/S0100-204X2018000900001.

Redon, R, Ishikawa, S, Fitch, KR, Feuk, L, Perry, GH, Andrews, TD, Fiegler, H, Shapero, MH, et al. 2006. Global variation in copy number in the human genome. Nature. 444(7118):444–454. doi.org/10.1038/nature05329.

Rosen, BD, Bickhart, DM, Schnabel, RD, Koren, S, Elsik, CG, Tseng, E, Rowan, TN, Low, WY, et al. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience. 9(3):1–9. doi.org/10.1093/gigascience/giaa021.

Sabeti, PC, Schaffner, SF, Fry, B, Lohmueller, J, Varilly, P, Shamovsky, O, Palma, A, Mikkelsen, TS, et al. 2006. Positive natural selection in the human lineage. Science. 312(5780):1614–1620. doi.org/10.1126/SCIENCE.1124309.

Sabeti, PC, Varilly, P, Fry, B, Lohmueller, J, Hostetter, E, Cotsapas, C, Xie, X, Byrne, EH, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature. 449(7164). doi.org/10.1038/nature06250.

Santana, ML, Pereira, RJ, Bignardi, AB, El Faro, L, Tonhati, H & Albuquerque, LG. 2014. History, structure, and genetic diversity of Brazilian Gir cattle. Livestock Science. 163(1):26–33. doi.org/10.1016/j.livsci.2014.02.007.

Saravanan, KA, Panigrahi, M, Kumar, H, Bhushan, B, Dutt, T & Mishra, BP. 2020. Selection signatures in livestock genome: A review of concepts, approaches and

applications. Livestock Science. 241:104257. doi.org/https://doi.org/10.1016/j.livsci.2020.104257.

Saravanan, KA, Panigrahi, M, Kumar, H, Parida, S, Bhushan, B, Gaur, GK, Dutt, T, Mishra, BP, et al. 2021. Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. Genomics. 113(3):955–963. doi.org/10.1016/J.YGENO.2021.02.009.

Stella, A, Ajmone-Marsan, P, Lazzari, B & Boettcher, P. 2010. Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. Genetics. 185(4):1451–1461. doi.org/10.1534/genetics.110.116111.

Strillacci, MG, Gorla, E, Cozzi, MC, Vevey, M, Genova, F, Scienski, K, Longeri, M & Bagnato, A. 2018. A copy number variant scan in the autochthonous Valdostana Red Pied cattle breed and comparison with specialized dairy populations. PLOS ONE. 13(9):e0204669. doi.org/10.1371/journal.pone.0204669.

Sukhija, N, Malik, AA, Devadasan, JM, Dash, A, Bidyalaxmi, K, Ravi Kumar, D, Kousalaya Devi, M, Choudhary, A, et al. 2024. Genome-wide selection signatures address trait specific candidate genes in cattle indigenous to arid regions of India. Animal Biotechnology. 35(1). doi.org/10.1080/10495398.2023.2290521.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123(3):585–595.

Thévenon, S, Dayo, GK, Sylla, S, Sidibe, I, Berthier, D, Legros, H, Boichard, D, Eggen, A, et al. 2007. The extent of linkage disequilibrium in a large cattle population of western Africa and its consequences for association studies. Animal Genetics. 38(3):277–286. doi.org/https://doi.org/10.1111/j.1365-2052.2007.01601.x.

Utsunomiya, YT, Pérez O'Brien, AM, Sonstegard, TS, Van Tassell, CP, do Carmo, AS, Mészáros, G, Sölkner, J & Garcia, JF. 2013. Detecting Loci under Recent Positive Selection in Dairy and Beef Cattle by Combining Different Genome-Wide Scan Methods. PLOS ONE. 8(5):e64280-. Available from: https://doi.org/10.1371/journal.pone.0064280.

Utsunomiya, YT, Pérez O'Brien, AM, Sonstegard, TS, Sölkner, J & Garcia, JF. 2015. Genomic data as the "hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome. Frontiers in Genetics. 6. doi.org/10.3389/fgene.2015.00036.

Voight, BF, Kudaravalli, S, Wen, X & Pritchard, JK. 2006. A map of recent positive selection in the human genome. PLoS Biology. 4(3):0446–0458. doi.org/10.1371/journal.pbio.0040072.

Wiener, P, Burton, D, Ajmone-Marsan, P, Dunner, S, Mommens, G, Nijman, IJ, Rodellar, C, Valentini, A, et al. 2003. Signatures of selection? Patterns of microsatellite diversity on a chromosome containing a selected locus. Heredity 2003 90:5. 90(5):350–358. doi.org/10.1038/sj.hdy.6800257.

Wright, S. 1949. The genetical structure of populations. Annals of Eugenics. 15(1):323–354. doi.org/https://doi.org/10.1111/j.1469-1809.1949.tb02451.x.

Zhang, Q, Schönherz, AA, Lund, MS & Guldbrandtsen, B. 2022. Positive Selection and Adaptive Introgression of Haplotypes from Bos indicus Improve the Modern Bos taurus Cattle. Agriculture. 12(6):844. doi.org/10.3390/AGRICULTURE12060844.

CHAPTER 02 – SELECTION SIGNATURES IN GIR AND HOLSTEIN CATTLE

ABSTRACT - Natural and artificial selection leave footprints on the genome, known as selection signatures, that can indicate regions related to adaptive and economically important traits. The recurrent use of a limited number of sires and increased selection pressure may affect genetic diversity, potentially affecting long-term breeding programs. Among dairy cattle, the Holstein breed has been intensively selected to maximize productivity, particularly in Canada and the United States. In the dairy industry, the Gir breed plays an important role in milk production in tropical regions such as India and Latin America. Gir cattle were introduced into Brazil in the 19th and 20th centuries, and since 1985, this breed has been intensively selected for milk production. This study aimed to assess the genetic diversity and characterize the selection signatures in Holstein cattle from the United States and Canada (HOL), Gir cattle from India (GIR_IN), and Dairy Gir cattle from Brazil (GIR_BR). Genetic diversity was assessed by nucleotide diversity, single nucleotide variant density analysis, minor allele frequency, observed and expected heterozygosity, and the inbreeding coefficient. Selection signatures were identified via Tajima's D, the integrated haplotype score (iHS), the fixation index, and the cross-population extended haplotype homozygosity test for autosomes. Additionally, the analysis of selection signatures using Tajima's D and iHS was conducted for the X chromosome. Lower genetic diversity was observed in the HOL population, whereas the GIR_IN and GIR_BR populations presented greater diversity. Several genes previously related to economically important traits were identified as being under selection, including *DNAJC18*, *FSHR*, *HELB*, *HMGA2, PLAG1*, *GAB3*, and *PTEN*. In conclusion, the genes identified within the selection signatures were linked to several traits, including growth, reproduction, mastitis, milk production, heat tolerance, health, and adaptation.

Keywords: milk, selection signature, selective sweep, genetic diversity

1 INTRODUCTION

The selection process, either natural or artificial, leaves footprints on the genome that can be detected. Unique genetic patterns or marks in the genomes of selected individuals are called selection signatures or selective sweeps (Jensen et al., 2016; Fay and Wu, 2000). Selection signature studies can identify genes and beneficial mutations linked to adaptive and economically important traits in dairy cattle, including reproduction, health, and milk production. Additionally, studying selection signatures can contribute to understanding genetic architecture diversity within and between populations.

Taurine cattle breeds have been selected for economic traits for a longer period and have been subjected to greater selection pressure than indicine breeds, especially when referring to indicine populations from developing and subdeveloping countries. As a major example, the Holstein–Friesian breed (Bos taurus) is the most widely distributed breed of cattle in more than 150 countries. This breed originated in the northern provinces of North Holland and West Friesland of the Netherlands (McGuffey and Shirley, 2011). It was exported to the United States in the 1870s and 1880s and was strongly selected to increase milk yield (Theunissen, 2012). After the implementation of genomic selection, there has been a notable increase in the rate of inbreeding and co-ancestry in Holstein populations (Makanjuola et al., 2020). Owing to the intense selection that has been applied, monitoring both the genetic diversity and the impact of selection on the genome of Holstein cattle is important.

The Gir breed (Bos indicus) is an indicine breed adapted to tropical conditions that originated in the Indian state of Gujarat. It is one of the most important local dairy breeds in India. Between 1870 and 1962, approximately 6,000 zebu cattle were imported from India to Brazil, including fewer than 700 Gir individuals (O'Brien et al., 2015). After their arrival in Brazil, the Gir cattle were initially bred for dual purposes (meat and milk). In 1985, the Brazilian Dairy Gir Breeding Program (PNMGL) was established, which resulted in an increase in milk production from 2,200 kg to 5,500 kg per 305 d of lactation on average. The Dairy Gir is well adapted to the harsh conditions of tropical dairy production, and it is considered the main dairy indicine breed in Brazil (Prata et al., 2015). It can also be found in most other Latin American

countries (Corredor et al., 2023). Dairy Gir cattle are also used in crossbreeding with taurine breeds, such as Holstein cattle or locally adapted breeds.

A motivation for investigating selection signatures is that these regions can pinpoint genes and mutations with large phenotypic effects in a population even if they are no longer segregating. Additionally, selection signatures can complement GWAS. Genetic diversity is an important element to consider in breeding programs because selection often affects allele and genotype frequency and may affect the capacity to respond to environmental changes. Careful planning and monitoring are required to maintain diversity and offer alternatives for long-term breed improvement or the selection of animals for specific immediate purposes (Makanjuola et al., 2020; Carrier et al., 2023). Studying selection signatures and genetic diversity can provide valuable insights into the progress of selection and aid in genetic improvement in livestock populations. This study aimed to assess genetic diversity and characterize selection signatures in Holstein and Gir cattle breeds.

2 MATERIALS AND METHODS

The institutional research ethics board of the São Paulo State University (Jaboticabal, Brazil), University of Guelph (Guelph, Canada), and the Brazilian Agricultural Research Corporation (Embrapa; Juiz de Fora, Brazil) did not require ethics approval for this study because the data used in this investigation were assessed from previous projects

2.1 Samples and Sequencing

For the Gir breed, 14 samples from Gir bulls from India (GIR_IN) and 42 samples from Dairy Gir sires from Brazil (GIR_BR) born between 1960 and 2007 were used. In brief, the sires that had the most progeny in PNMGL, representing most lineages in the population, were selected for whole-genome sequencing (WGS).

For GIR_BR, DNA was extracted via either the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA), following the recommendations of the manufacturer, or a saline buffer and phenol/chloroform purification protocol (Machado et al., 2010). The

concentration and quality of the isolated DNA were evaluated via a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE) or a Qubit fluorometer 2.0 (Life Technologies, Grand Island, NY). For library preparation, the Illumina TruSeq Nano Kit (Illumina Inc., San Diego, CA) was used according to the manufacturer's recommendations. Sequencing was performed via the Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA), producing reads measuring 2 × 100 bp and 2 × 200 bp, with an average sequencing coverage of 13.9X, or via the NovaSeq 6000 platform (Illumina Inc., San Diego, CA), with an average sequencing coverage of 16X and reads measuring 2 × 100 bp and 2 × 200 bp.

The GIR_IN population consisted of 14 samples of bulls and cows from Gujarat state in India raised for milk production. DNA extraction was performed by following the Sambrook and Russel adapted protocol (Sambrook and Russel, 2001). The sequencing was performed on the Illumina HiSeq2500 platform (Illumina Inc., San Diego, CA), producing reads measuring 2 × 125 bp, with an average sequencing coverage of 13.5X.

The WGS processing for the GIR_BR and GIR_IN samples followed the same pipeline, according to the recommendations of the 1000 Bull Genomes Project protocol (http://www.1000bullgenomes.com, last accessed on 11/20/2020; Hayes and Daetwyler, 2019). In summary, read quality was evaluated using the FastQC tool, and trimming was performed using SeqyClean software (Zhbannikov et al., 2017). The reads were aligned to the bovine reference genome ARS-UCD 1.2 using the mem option of the BWA algorithm (Li and Durbin, 2009). Conversion to binary format, sorting, and indexing were completed via SAMtools (Li et al., 2009). Duplicates were removed using the MarkDuplicates option of the Picard tool software (2019; https://broadinstitute.github.io/picard/). Base quality score recalibration (BQSR) was performed via BaseRecalibrator and PrintReads of the Genome Analysis Toolkit (GATK; v. 3.8-1-0-gf15c1c3ef; https://gatk.broadinstitute.org/hc/en-us). The set of known variants provided by the 1000 Bull Genomes project (Hayes and Daetwyler, 2019) was applied for BQSR. Variant calling was performed using the HaplotypeCaller option from GATK according to the same protocol (-ERC GVCF -variant_index_type LINEAR -variant_index_parameter 128000).

Following BQSR, samples were combined using the option genotypeGVCFs from the GATK software. The variant filtering followed the variant quality score recalibration (VQSR) from GATK software (v. 4.4.0) according to the parameters in the Run 9 pipeline from 1000 Bull Genomes (Hayes and Daetwyler, 2019). Variant quality score recalibration uses machine learning algorithms to learn from training datasets the annotation profile of good variants and bad variants, integrating information from multiple dimensions. The training and validation sets of the variants were provided by the 1000 Bull Genomes Project (Hayes and Daetwyler, 2019). Only variants occurring in the tranche 99% were used (Neumann et al., 2023). Duplicate samples were checked through the KING-robust estimator (Manichaikul et al., 2010) with a cutoff of 0.354 using Plink2 software (Chang et al., 2015).

Single nucleotide variants (SNVs) from WGS data from 307 Holstein bulls and cows from Canada and the United States (HOL) were accessed from the 1000 Bull Genomes project (Run 9; Hayes and Daetwyler, 2019). Only samples with a minimum coverage of 6X and autosomal SNVs occurring in the tranche 99% (from the VQSR performed by the 1000 Bull Genomes Project) were used (Neumann et al., 2023). Duplicate samples were excluded through the KING-robust estimator (Manichaikul et al., 2010) with a cutoff of 0.354 via Plink2 software. For GIR_BR, GIR_IN, and HOL, autosomes were retrieved for diversity and selection signature analysis, and the Bos taurus X chromosome (BTX) was scanned for within-population selection signatures.

2.2 Genomic Diversity

Principal component analysis (PCA) was performed using the genotype matrix as follows. For each population, one marker of pairs with a high level of linkage disequilibrium (LD) was filtered out using an r^2 threshold of 0.2, a window size of 50 SNVs, and a step size of 5 SNVs (–indep-pairwise 50 5 0.2). Additionally, variants with missing call rates exceeding 0.05 (–geno 0.05) were filtered out. These quality control steps were applied only for PCA. For the population differentiation metric, the ratio of averages of the fixation index (F_{ST}) based on Hudson's method (Hudson et al., 1992) was computed via the option –fst from Plink2 between each pair of populations.

The genomic variation within populations was assessed separately using nucleotide diversity (π), SNV density analysis, minor allele frequency (MAF), observed (H_O) and expected (H_E) heterozygosity, and inbreeding coefficient based on the methods-of-moments estimator (F_{HOM}). Nucleotide diversity is defined as the average number of nucleotide differences per site between 2 DNA sequences sampled from a population, in which higher values indicate greater diversity (Nei and Li, 1979). The software VCFtools v. 0.1.15 (Danecek et al., 2011) was used to estimate the nucleotide diversity of each population in nonoverlapping window sizes of 50 kb with a step of 20 kb containing at least 5 SNVs. Descriptive statistics of total (π_{Total}) and chromosome (π_{Chr}) nucleotide diversity were subsequently calculated. Biallelic marker distribution was evaluated with an SNV density plot through the CMplot package (Yin et al., 2021) in R software (https://www.r-project.org/), which uses a bin of 1 Mb.

For heterozygosity, MAF, and inbreeding coefficient, only variants with a minimum mean depth of 10X were retrieved to avoid biased estimates (Kardos and Waples, 2024). The MAF was calculated via Plink 1.9 (--freq) for each population. SNVs with MAF greater than 0.1 (10%) were classified as polymorphic SNVs (Rajawat et al., 2024). VCFtools v. 0.1.15 was used to estimate H_O and H_E (--het). Heterozygosity is a measure often used to track the amount of genetic variation retained by a population. The F_{HOM} was computed according to the method-of-moments estimator (Li and Horvitz, 1953) using the following equation: $F_{HOM} = \frac{O_{HOM} - E_{HOM}}{N_{VAR} - E_{HOM}}$, where O_{HOM} represents the number of observed homozygous genotypes, E_{HOM} the number of expected homozygous genotypes, and N_{VAR} the total number of variants.

2.3 Selection Signatures

For autosome chromosomes, within- and between-population analyses were performed using the total number of variants after VQSR procedures. For chromosome X, only within-population analyses were conducted. All variants in the 99% tranche in the VQSR were used in the analysis.

2.3.1 Within-population

Integrated Haplotype Score Test Statistic: The R package rehh v. 3.2.2 (Gautier et al., 2017) was used to perform the integrated haplotype score (iHS) test within each population with unpolarized markers. Phasing within each population was performed via Beagle v. 5.1 (Browning and Browning, 2007) with default parameters. To identify potential selection signatures, |iHS| was calculated for windows of 50 kb containing a minimum of 5 SNVs. For autosomes, the top 1%, and for chromosome X, the top 0.5% were considered potential selection signatures, and adjacent windows were merged. Integrated haplotype score is a highly sensitive test that identifies high-LD regions, and it is a standard measure of decay in extended haplotype homozygosity (EHH). The iHS detects selection when the selected allele has achieved intermediate frequencies, but before selection; the derived allele must have existed on a distinct background and not yet reached fixation (Voight et al., 2006; Utsunomiya et al., 2013). In the case of unpolarized markers, this test uses major and minor alleles instead (Klassmann and Gautier, 2022), which was applied in this study.

Tajima's D Test Statistic: Tajima's D statistic was computed within the population in window sizes of 50 kb via VCFtools software. Only windows with at least 5 markers were considered. For autosomes, the bottommost 1%, and for chromosome X, the bottommost 0.5% were retrieved, and adjacent selection signatures were merged. Tajima's D statistic compares 2 measures of genetic diversity, the number of segregating sites (S) and the nucleotide diversity, which is the average number of pairwise nucleotide differences between sequences in a sample (Tajima, 1989). The purpose of Tajima's D test is to distinguish between a DNA sequence evolving randomly and one evolving under a nonrandom process, which may include, for instance, selection, demographic expansion, contraction, or introgression. Tajima's D statistic value is expected to be zero under the standard neutral evolutionary model. However, deviations from zero may indicate nonneutral forces or selection. Negative Tajima's D values suggest the fixation of alleles or rare alleles, whereas positive values denote balancing selection, which reflects an abundance of intermediate allele frequencies (Tajima, 1989; Korneliussen et al., 2013).

2.3.2 Between-populations

Fixation Index: Pairwise autosomal Wright's F_{ST} values were estimated based on Hudson's method (Hudson et al., 1992) using the scikit-allel v1.3.7 library (Miles et al., 2023) in Python with windows of 50 kb and a minimum of 5 markers. Only the top 1% of windows were selected as signature candidates. The fixation index (Wright, 1949) is a test that measures population genetic differentiation and can detect actual genetic variants under selection. The F_{ST} values vary from zero, which implies no differences in allelic frequencies between 2 populations, to one, meaning that each population is fixed for a different allele. Hudson's method performs better than traditional F_{ST} calculation when there is a large variation in sample size across populations and accounts for changes in the demographic history of the populations being compared (Bhatia et al., 2013).

Cross-Population Extended Haplotype Homozygosity (XP-EHH): Phasing within each population was performed using Beagle v. 5.1 (Browning and Browning, 2007), and selection signatures between populations were subsequently investigated via the XP-EHH test with windows of 50 kb containing a minimum of 5 SNPs via the rehh (Gautier et al., 2017) R package. The top and bottom 0.5% extreme windows were considered potential selection signatures, and adjacent windows were merged. Cross-population EHH compares the integrated EHH between 2 populations at the same SNP and detects selected alleles that have approached or achieved fixation in one population but not in the other. This test is indicated for detecting ongoing or nearly fixed selection signatures because this test is based on LD and identifies alleles that have rapidly increased to high frequency and are still associated with nearby polymorphisms (Sabeti et al., 2007).

2.4 Functional Analysis

The GALLO package (Fonseca et al., 2020) from R software was used to identify genes and QTL within selection signatures. Genes and QTL were retrieved from the Ensembl Genes database (Ensembl Release 110, ARS_UCD1.2; https://www.ensembl.org) and the Animal Genome database (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index, ARS_UCD1.2), respectively. Additionally, colocalized genes were also evaluated for the top 10

windows with the highest nucleotide diversity and the regions with the highest and lowest SNV density. For genes within selection signatures, terms from the Gene Ontology (GO) database and metabolic pathways predicted by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database were enriched (false discovery rate [FDR] <0.05) using the WebGestaltR package (Wang and Liao, 2020) in R software. The enrichment analysis was performed via hypergeometric overrepresentation analysis. Gene Ontology terms are divided into 3 groups: Cellular Components, Biological Processes, and Molecular Functions. Metabolic pathways predicted by the Reactome database (https://reactome.org/) were also enriched (FDR <0.05) using the website of the repository. Gene under selection were used to perform protein-protein interaction network using STRING database (Szklarczyk et al., 2023). The networks were clustered using the k-mean algorithm to improve visualization.

3 RESULTS

3.1 Alignment and Variant Calling

After duplicate removal, 14 GIR_IN, 42 GIR_BR, and 307 HOL were used in the analysis. Overall, the coverage ranged from 6.38X to 51.78X. The alignment metrics and coverage per sample are presented in Supplemental Table S1 (see Notes). The number of consensus markers for the HOL and GIR_IN pair was 31,229,604, for the HOL and GIR_BR pair, 37,296,579, and for the GIR_IN and GIR_BR pair, 30,395,111 (Table 1).

Table 1 Summary of number of samples, minimum coverage, mean and standard deviation, maximum coverage, and numbers of autosome biallelic SNVs and multiallelic SNVs after quality control, as well as transition/transversion rate (Ts/Tv) per population[1].

Population	N	Min. (X)	Mean (X)	Max. (X)	Biallelic SNVs	Multiallelic SNVs	Ts/Tv
GIR_IN	14	6.60	17.77 (5.47)	29.20	31,253,205	352,760	2.26
GIR_BR	42	10.20	16.09 (3.02)	25	31,264,796	353,009	2.26
HOL	307	6.38	17.55 (8.08)	51.78	57,106,559	2,538,795	2.18

[1]N = number of samples, Min. (X) = minimum coverage, Mean (X) = mean and standard deviation (parentheses), Max. (X) = maximum coverage.

3.2 Genomic Diversity

The PCA results are shown in Figure 1A, in which the 3 populations exhibit distinct clusters. The ratio-of-average F_{ST} values obtained via the Hudson method can be found in Table 2.

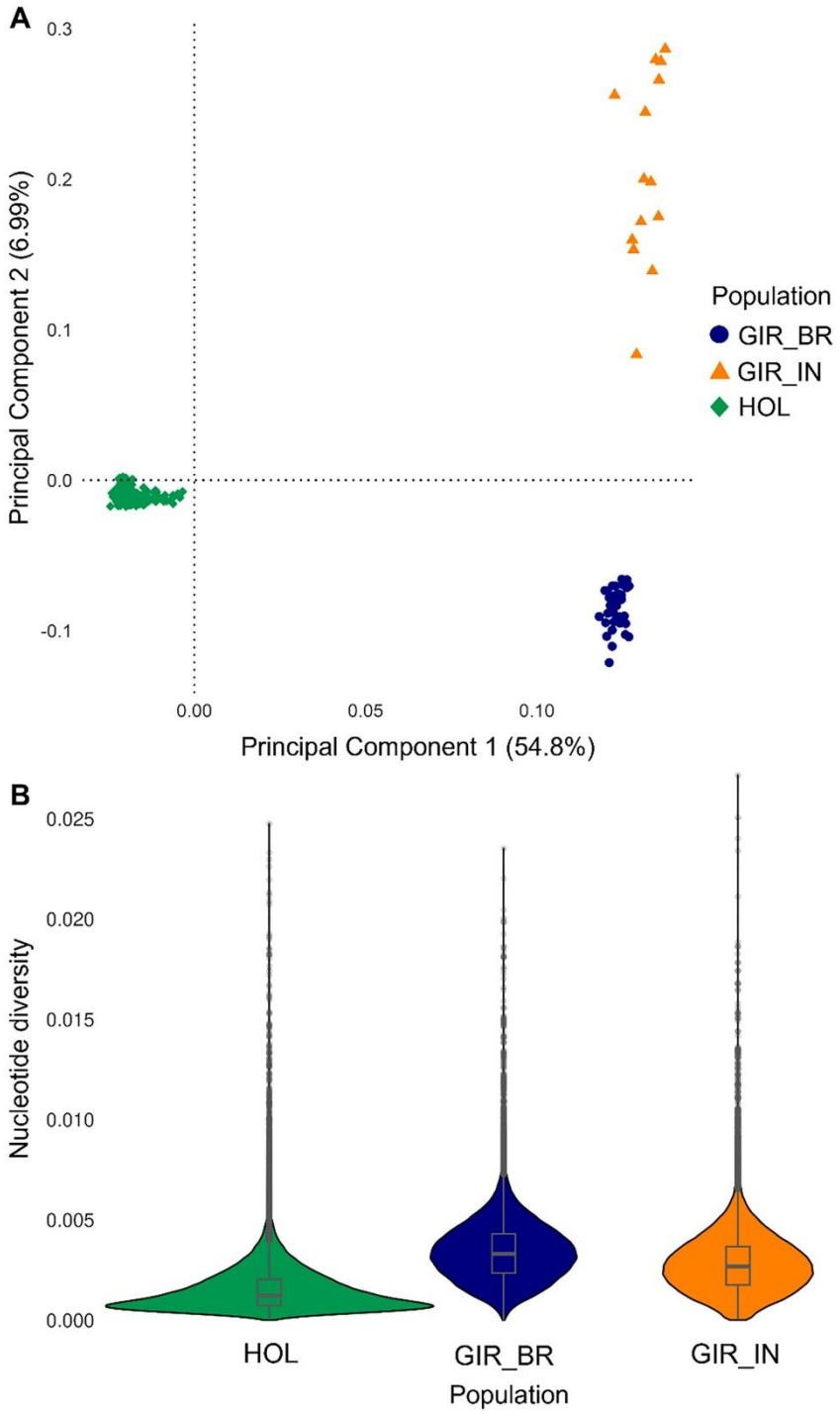

Figure 1. (A) Principal component analysis of individuals. After filtering for missing call rate, 279,397 variants were used for PCA. (B) Violin and boxplot graphs of nucleotide diversity for populations of Holstein bulls and cows from Canada and the United States (HOL), Dairy Gir sires from Brazil (GIR_BR), and Gir bulls from India (GIR_IN). The midline indicates the median; the upper and lower edges of the box represent the first and third quartiles; the whiskers extend to 1.5 times the

interquartile range; and the dots represent outliers. The shape of each violin illustrates the kernel density estimation of the distribution of nucleotide diversity for each population.

Table 2. Pairwise ratio-of-averages of F_{ST} based on Hudson's method among Holstein bulls and cows from Canada and the United States (HOL), Dairy Gir sires from Brazil (GIR_BR), and Gir bulls from India (GIR_IN) populations.

Population 1	Population 2	Ratio-of-averages F_{ST} value
HOL	GIR_IN	0.5411
HOL	GIR_BR	0.4671
GIR_IN	GIR_BR	0.1062

The GIR_BR population exhibited the highest π_{Total}, whereas the HOL population presented the lowest mean value despite having a larger sample size (Table 3; Figure 1B; Supplemental Figure S1; see Notes). For the GIR_IN population, a higher mean π_{Chr} was observed in BTA 28, whereas the lowest was noted in BTA2 and 11 (Supplemental Table S2; see Notes). In the GIR_BR population, the highest mean π_{Chr} was observed in BTA27 and 28, whereas the lowest was found in BTA3, 5, 13, and 19 (Supplemental Table S3; see Notes). For the HOL population, the highest mean π_{Chr} was observed in BTA23, with the lowest in BTA13 and 22 (Supplemental Table S4; see Notes). The top 10 windows with the highest nucleotide diversity and colocalized genes across the 3 populations are reported in Supplemental Table S5 (see Notes).

Table 3. Minimum, first quartile, median, mean, third quartile, and standard deviation (SD) of π_{Total} for HOL, GIR_BR, and GIR_IN populations. All π_{Total} measures are shown as percentages (%).

Population	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum	SD
GIR_IN	2.17e-04	0.175	0.267	0.278	0.366	2.716	0.143
GIR_BR	1.07e-04	0.233	0.330	0.338	0.431	2.351	0.146
HOL	2.34e-05	0.073	0.122	0.152	0.203	2.475	0.112

The SNV density plots for each population are displayed in Supplemental Figures S2–S4 (see Notes). There was an absence of SNVs in BTA10:23.78–24.78 Mb across all populations. No gene is described in the ARS_UCD1.2 bovine annotation, and no variants are reported for this region (https://www.ensembl.org/). We hypothesize that this region is highly conserved in bovines. The region BTA23:24.00–30.00 Mb presented a greater density of SNVs for all populations. The annotated genes for this region are presented in Supplemental Table S6 (see Notes). The H_O, H_E, F_{HOM}, MAF, and proportion of polymorphic SNVs results are reported in Table 4 and Figure 2A.

Table 4. Mean and standard deviation in brackets of observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient based on the method-of-moments (F_{HOM}), minor allele frequency (MAF), and proportion of polymorphic SNVs for three cattle populations.

Population	H_O	H_E	F_{HOM}	MAF	Proportion of polymorphic SNVs
GIR_IN	0.3727 (0.0327)	0.3412 (0.0002)	-0.0924 (0.0955)	0.1576 (0.1643)	0.5179
GIR_BR	0.2761 (0.0232)	0.2791 (0.0002)	0.0109 (0.0828)	0.1912 (0.1496)	0.6272
HOL	0.2186 (0.0189)	0.2256 (0.0003)	0.0309 (0.08)	0.0448 (0.1084)	0.1460

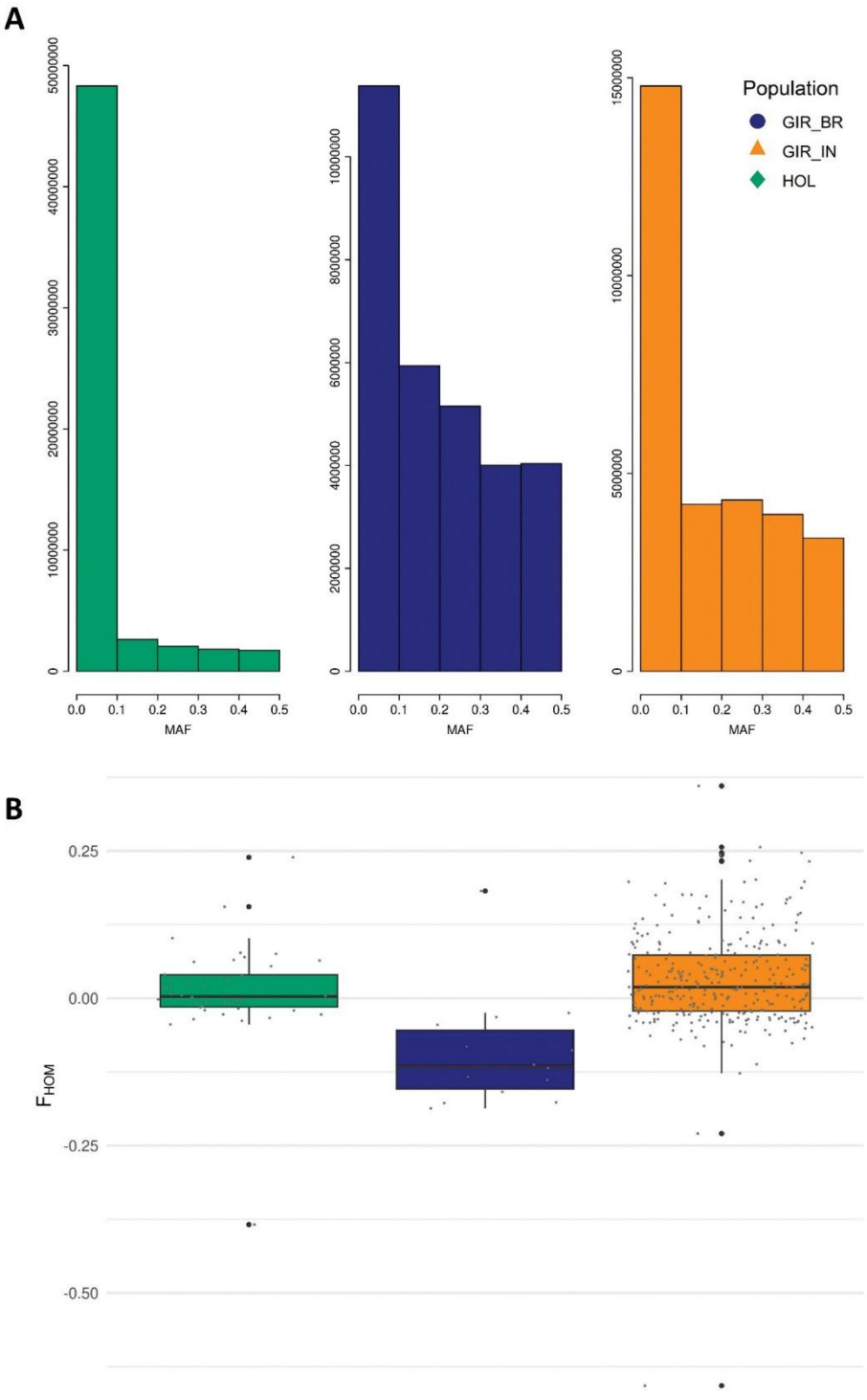

Figure 2. (A) Histogram plot of minor allele frequencies (MAF). The y-axis represents the allele count within each MAF bin. (B) Boxplots of inbreeding coefficient based on method-of-moments (F_{HOM}) for Holstein bulls and cows from Canada and the United States (HOL), Dairy Gir sires from Brazil (GIR_BR), and Gir bulls from India (GIR_IN). The midline indicates the median; the upper and lower edges of the box

represent the first and third quartiles; the whiskers extend to 1.5 times the interquartile range; and the dots represent individual animals.

3.3 Selection Signatures and Genes Under Selection

Supplemental Table S7 lists the top 10 extreme windows from each test and population. No metabolic pathway from the Reactome database was enriched for the list of genes under selection in all 3 populations. Protein-protein interactions networks are listed in Supplemental Figures S5–S19.

3.3.1 Within-population

- iHS Test Statistic for Autosomes

The Manhattan plots of autosomal windows are shown in Figure 3, and for BTX, in Supplemental Figure S20 (see Notes). In GIR_IN, after the top windows were merged, 326 nonoverlapping selection signatures were revealed. The largest length of a selection signature was BTA22:23.55–23.95 Mb. Across 252 selection signatures, a total of 354 genes and 1,315 QTL were found, with 39 enriched QTL (FDR <0.01; Supplemental Tables S8 and S9; see Notes).

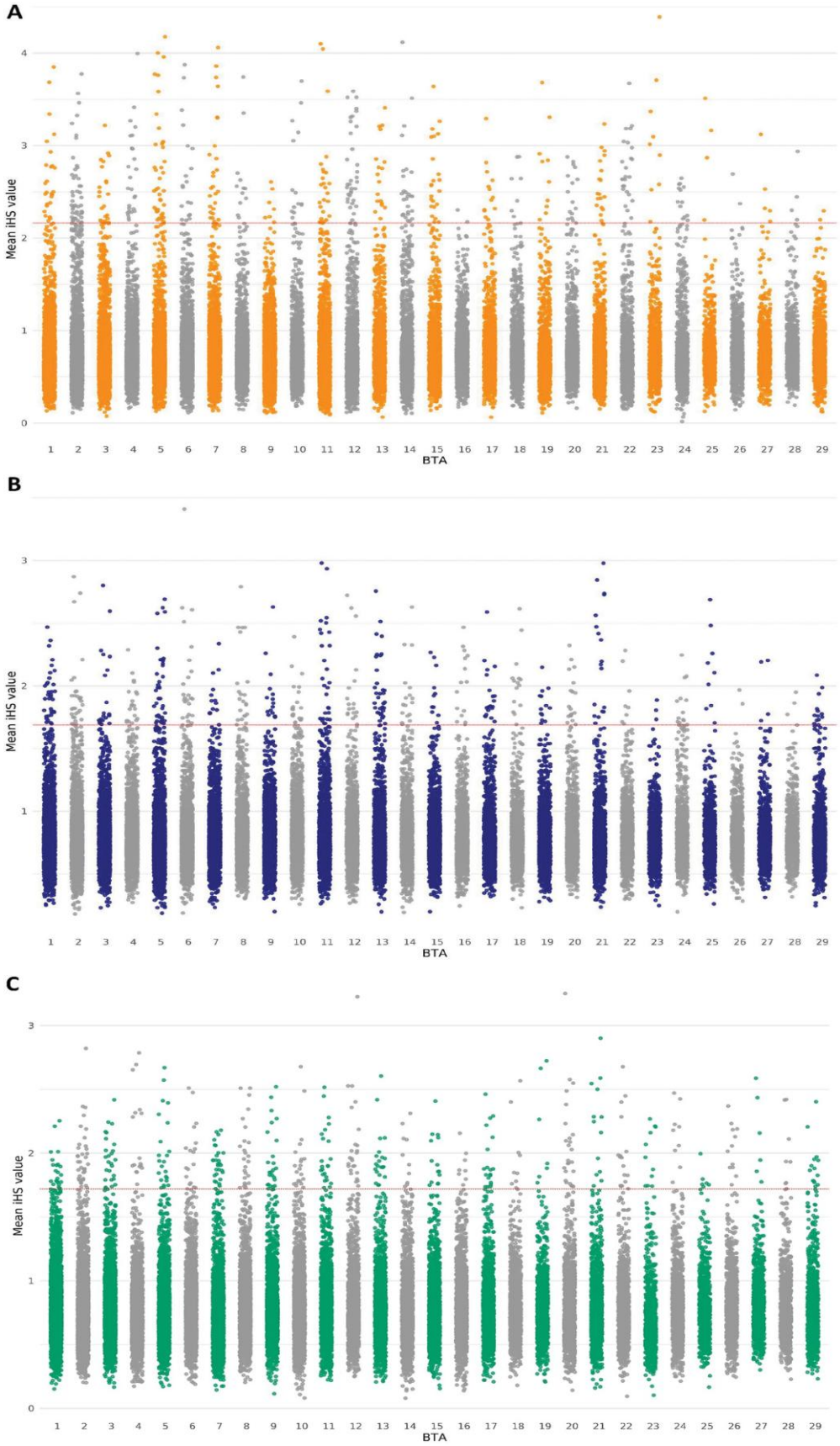

Figure 3. Manhattan plots of the mean |iHS| values for windows in 3 dairy cattle

populations: (A) Gir bulls from India (GIR_IN), (B) Dairy Gir sires from Brazil (GIR_BR), and (C) Holstein bulls and cows from Canada and the United States (HOL). The red dashed line indicates the threshold values of 2.16, 1.69, and 1.72 for the identification of potential selection signatures in each population, respectively.

For the GIR_BR population, the largest selection signature in length was BTA11:11.65–12.4 Mb. After merging, 320 non-overlapping selection signatures were identified. A total of 884 QTL and 314 genes were under selection in 231 selection signature regions in the functional analysis (Supplemental Table S10; see Notes). Furthermore, QTL enrichment revealed 27 enriched QTL (FDR <0.01; Supplemental Table S9).

For the HOL population, the largest length of a selection signature was found in BTA20:26.75–27.1 Mb. After merging, 385 non-overlapping selection signatures were identified (Supplemental Table S11; see Notes). For the functional analysis, 274 regions revealed 374 candidate genes and 1,342 QTL in selection signature regions, 22 of which were enriched (FDR <0.01; Supplemental Table S9).

- Tajima's D Statistic Test for Autosomes

The Manhattan plots of autosomal windows are shown in Figure 3, and for the BTX, in Supplemental Figure S21 (see Notes). In GIR_IN, the autosomal mean Tajima's D was 1.24, and the SD was 1.12 (Figure 4). After merging overlapping windows, 292 selection signatures were identified. The largest length of a selection signature was found in BTA14:39.25–39.70 Mb. In the functional analysis, 378 candidate genes and 1,203 QTL were under selection across 232 selection signature regions (Supplemental Table S12; see Notes). Furthermore, QTL enrichment analysis revealed 32 enriched QTL (FDR <0.01; Supplemental Table S9). The KEGG term “axon guidance” was significantly enriched (FDR <0.05) for the genes *PPP3R1*, *ROCK2*, *ABLIM3*, *TRPC4*, *PARD3*, *CFL2*, *PARD6G*, *NTN3*, *PRKCZ*, and

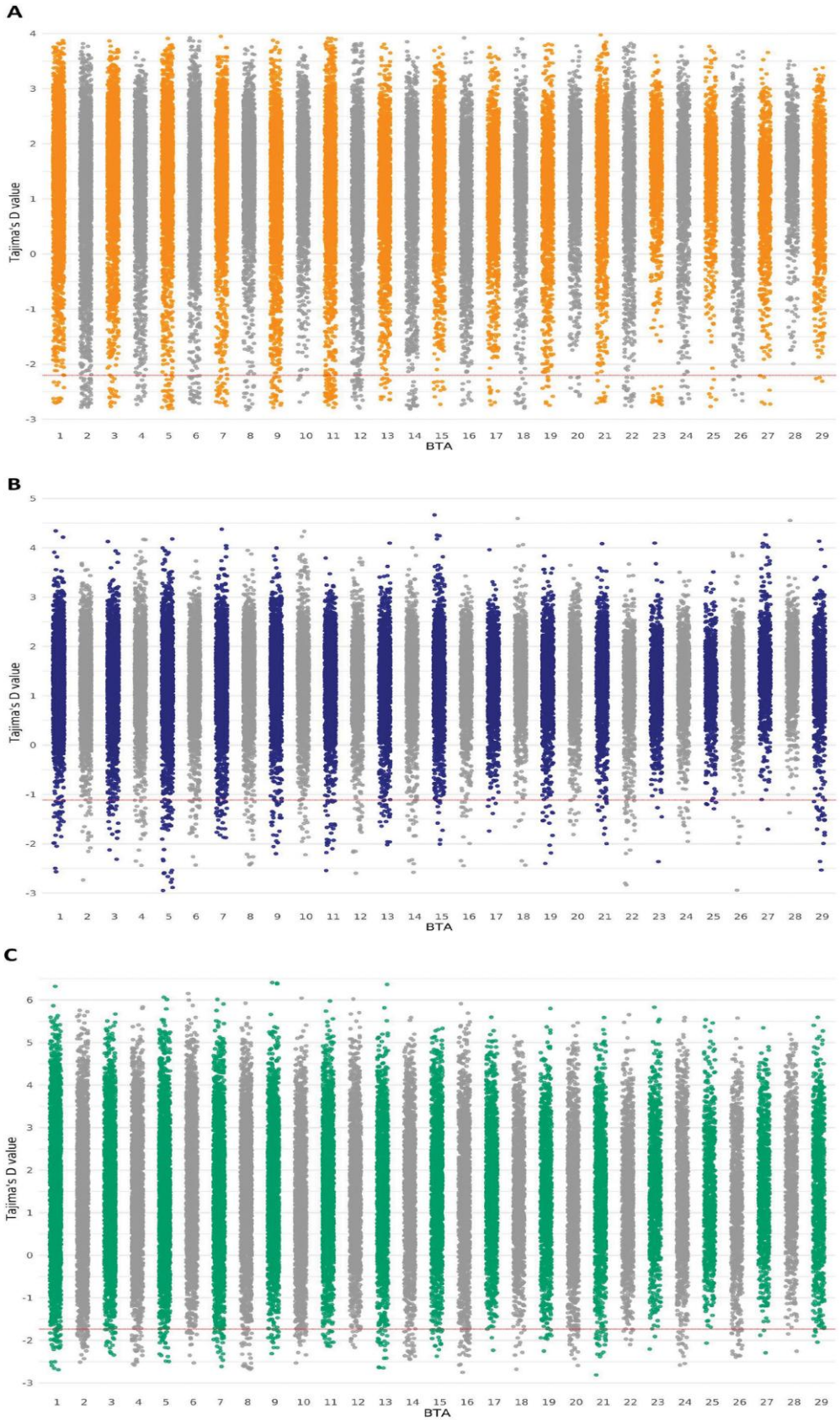

Figure 4. Manhattan plot of Tajima's D values for 3 dairy cattle populations: (A) Gir bulls from India (GIR_IN), (B) Dairy Gir sires from Brazil (GIR_BR), and (C) Holstein

bulls and cows from Canada and the United States (HOL). The red dashed line indicates the thresholds of −2.20, −1.11, and −1.74 used to identify potential selection signatures in each population, respectively.

For GIR_BR, the autosomal Tajima's D mean was 1.31 (SD: 0.82). A total of 304 non-overlapping selection signatures were observed. The selection signature with the greatest length was BTA11:11.85–12.4 Mb. We identified 1,392 QTL and 531 genes under selection in 264 selection signatures for the functional analysis (Supplemental Table S14; see Notes). Among the QTL, 33 were enriched (FDR <0.01; Supplemental Table S9).

In the HOL population, the autosomal Tajima's D mean was 1.48 (SD: 1.46). A total of 357 non-overlapping selection signatures were detected after merging. The largest selection signature was found in BTA8:105.9–106.4 Mb. In the functional analysis, 283 regions revealed 457 candidate genes and 1,713 QTL under selection (Supplemental Table S15; see Notes), 40 of which were enriched (FDR <0.01; Supplemental Table S9).

3.3.2 Between-populations

- Fixation Index

For the HOL and GIR_IN population pair, the F_{ST} values ranged from 0 to 0.991 (Figure 5A). By merging adjacent windows, 324 non-overlapping selection signatures were identified. In the functional analysis, 268 putative selection signatures were colocalized with 558 genes and 1,417 QTL (Supplemental Table S19; see Notes), of which 32 QTL were enriched (FDR <0.01; Supplemental Table S9).

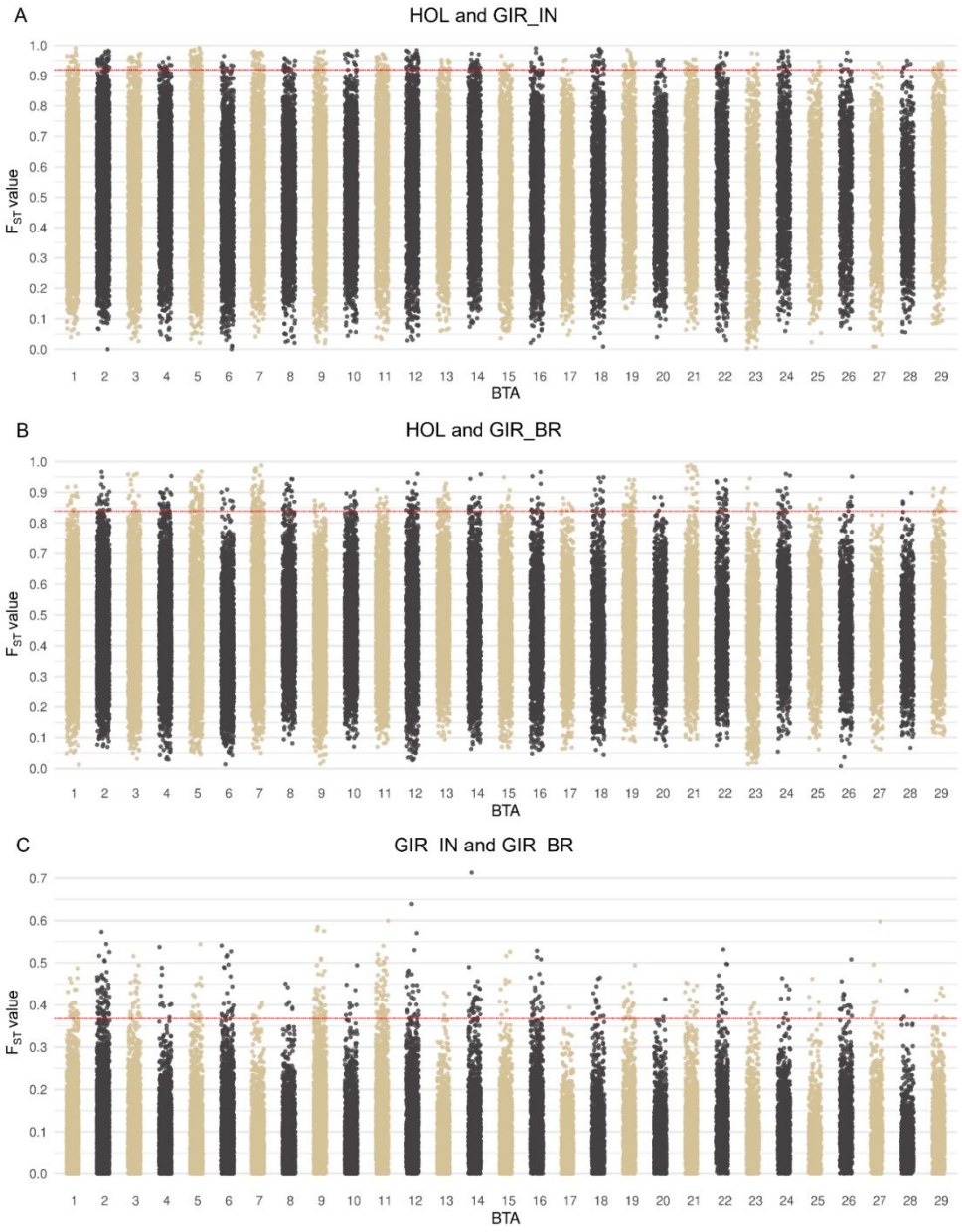

Figure 5. (A) Manhattan plot of F_{ST} values for HOL (Holstein bulls and cows from Canada and the United States) and GIR_IN (Gir bulls from India) population pair; the red dashed line represents the threshold of 0.92 to consider potential selection signatures. (B) Manhattan plot of F_{ST} values for HOL and GIR_BR (Dairy Gir sires from Brazil) population pair; the red dashed line represents the threshold of 0.84 to consider potential selection signatures. (C) Manhattan plot of F_{ST} values for GIR_IN and GIR_BR population pair; the red dashed line represents the threshold of 0.37 to

consider potential selection signatures.

For the HOL and GIR_BR population pair, the F_{ST} values ranged from 0.008 to 0.988 (Figure 5B). After adjacent windows were merged, 326 non-overlapping selection signatures were identified. The functional analysis revealed 1,962 QTL and 633 genes under selection in 288 selection signatures (Supplemental Table S20; see Notes). Furthermore, QTL enrichment revealed 37 enriched QTL (FDR <0.01; Supplemental Table S9). The KEGG terms “EGFR tyrosine kinase inhibitor resistance,” “Bacterial invasion of epithelial cells,” “Chronic myeloid leukemia,” “Cellular senescence,” and “Thyroid hormone signaling pathway” were significantly enriched (FDR <0.05) for several genes (Supplemental Table S13).

For the GIR_IN and GIR_BR pair, the F_{ST} values ranged from 0 to 0.713 (Figure 5C). Following the merging of adjacent windows, a total of 341 non-overlapping selection signatures were identified. The functional analysis revealed that 1,268 QTL and 376 genes are potentially under selection within 269 selection signatures (Supplemental Table S21; see Notes). Quantitative trait loci enrichment revealed 29 enriched QTL (FDR <0.01; Supplemental Table S9).

- Cross-Population Extended Haplotype Homozygosity

For the HOL and GIR_IN population pair, positive values denote selection in HOL, and negative values indicate selection in GIR_IN (Figure 6A). Following the merging of adjacent windows, 144 top and 185 bottom nonoverlapping selection signatures were identified for the HOL GIR_IN pair. A total of 119 genes and 464 QTL under selection in 98 selection signatures were found in the functional analysis for HOL (Supplemental Table S16; see Notes), and 8 QTL were enriched (FDR <0.01; Supplemental Table S9). The molecular function GO term “calcium ion binding” was significantly enriched (FDR <0.05) for the genes *CHP1*, *EHD4*, *PLA2G4D*, *PLA2G4F*, *S100A2*, *S100A3*, *S100A4*, *S100A5*, and *HPCA* (Supplemental Table S13). For selection signatures related to the GIR_IN population, 166 genes and 390 QTL were under selection in 131 windows (Supplemental Table S16). The results of the QTL enrichment analysis revealed that 21 QTL were significant (FDR <0.01; Supplemental Table S9). The biological process GO term

“regulation of lipase activity” was significantly enriched (FDR <0.05) for the genes *PLA2G5*, *SORT1*, *LPAR1*, and *ANXA8L1* (Supplemental Table S13).

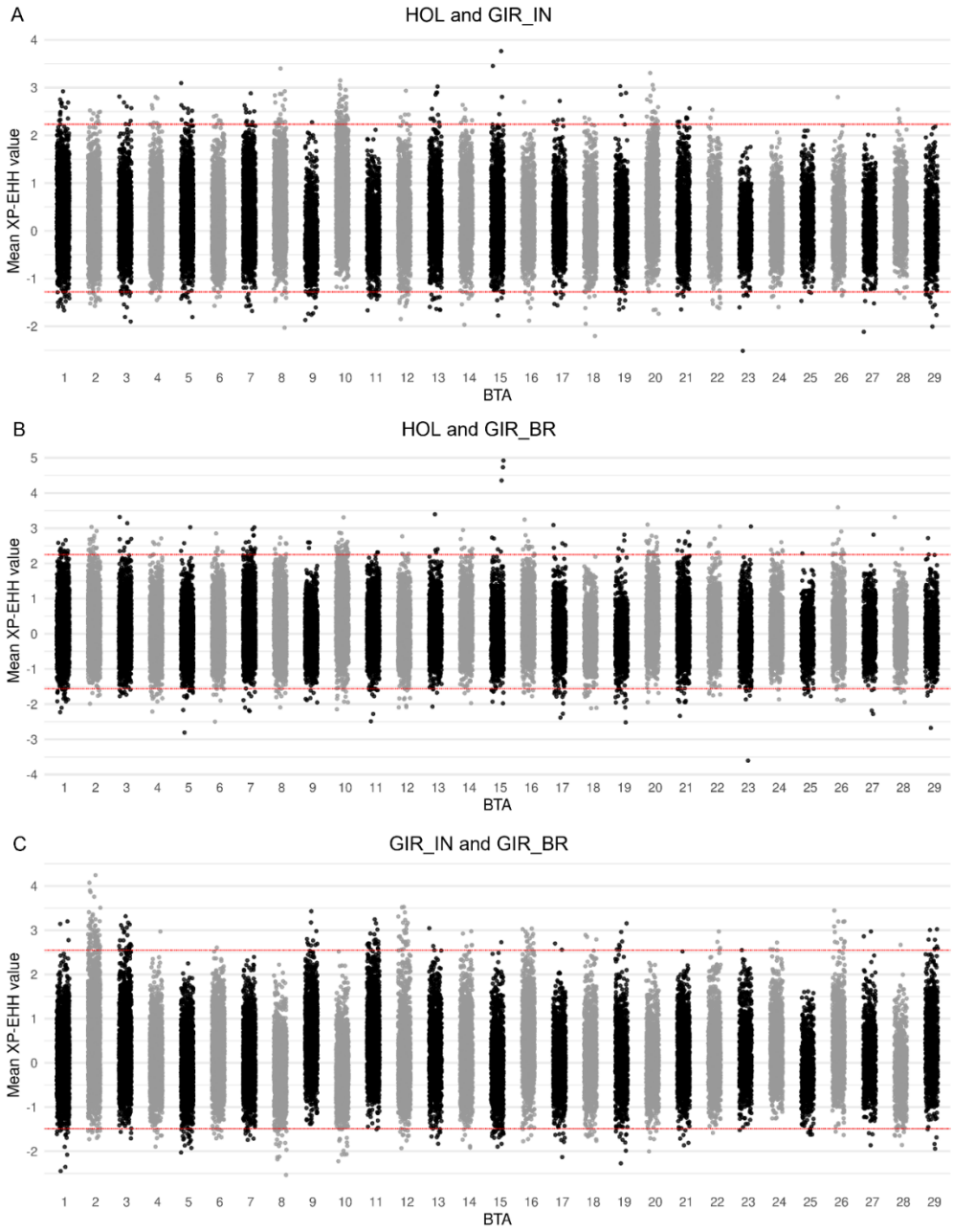

Figure 6. (A) Manhattan plot of XP-EHH value for HOL (Holstein bulls and cows from Canada and the United States) and GIR_IN (Gir bulls from India) pair. The red dashed lines represent thresholds of 2.232 and −1.280 to consider potential selection

signatures for HOL and GIR_IN populations, respectively. (B) Manhattan plot of XP-EHH value for HOL and GIR_BR (Dairy Gir sires from Brazil) pair. The red dashed lines represent thresholds of 2.252 and −1.561 to consider potential selection signatures for HOL and GIR_BR populations, respectively. (C) Manhattan plot of XP-EHH value for GIR_IN and GIR_BR pair. The red dashed lines represent thresholds of 2.550 and −1.486 to consider potential selection signatures for GIR_IN and GIR_BR populations, respectively.

For the HOL and GIR_BR population pair, positive values indicate selection in HOL, and negative values indicate selection in GIR_BR (Figure 6B). Following the cutoff and combination, 153 top and 198 bottom nonoverlapping selection signatures were observed. A total of 110 putative selection signatures overlapped with 160 genes and 631 QTL (Supplemental Table S17, see Notes), 24 of which were enriched (FDR <0.01) in the HOL population (Supplemental Table S9). The molecular function GO term “calcium ion binding” was significantly enriched (FDR <0.05) for the genes *S100A1*, *S100A13*, *S100A14*, *S100A16*, *S100A2*, *S100A3*, *S100A4*, *S100A5*, *EHD4*, *PLA2G4D*, and *PLA2G* (Supplemental Table S13). For the GIR_BR population, 160 genes and 587 QTL were located on 160 putative selection signatures (Supplemental Table S17), and 15 QTL were enriched (FDR <0.01; Supplemental Table S9).

For the GIR_IN and GIR_BR population pair, positive values indicate selection in GIR_IN, and negative values indicate selection in GIR_BR (Figure 6C). After combination, 101 top and 201 bottom nonoverlapping selection signatures were found. The functional analysis revealed 83 selection signature regions harbored 125 genes and 492 QTL (Supplemental Table S18; see Notes), with 6 enriched QTL (FDR <0.01) for GIR_IN (Supplemental Table S9). For the GIR_BR population, 146 selection signatures with 151 genes and 402 QTL (Supplemental Table S18), of which 10 were enriched (FDR <0.01) for GIR_BR (Supplemental Table S9).

3.4 Overlapping Selection Signatures

Figure 7 illustrates the overlapping autosomal selection signatures among populations and tests. In the within-pulation analyses (Figure 7A), GIR_IN and

GIR_BR shared more selection signatures with each other than with the HOL population, as expected, given that they are different populations of the same breed. However, in the between-population analyses (Figure 7B), the highest number of shared selection signatures was observed between GIR_BR and HOL and between GIR_IN and HOL.

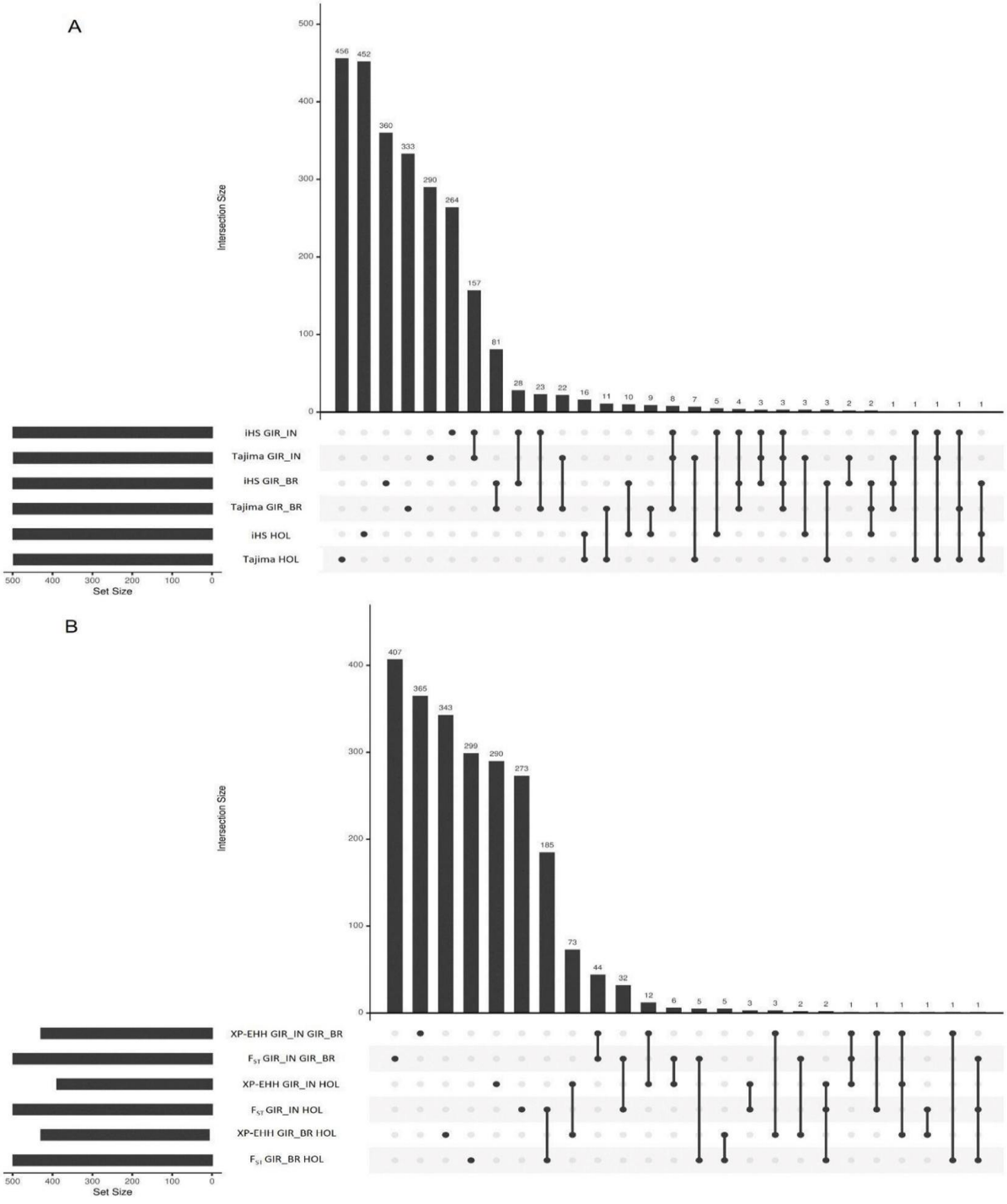

Figure 7. (A) UpSet plot of overlapping autosomal selection signatures among Gir bulls from India (GIR_IN), Dairy Gir sires from Brazil (GIR_BR), and Holstein bulls and cows from Canada and the United States (HOL) populations and Tajima's D and iHS tests. (B) UpSet plot of overlapping selection signatures among pairwise population analyses and XP-EHH and F_{ST} tests. Single dots indicate the total number of selection signatures identified by each test and population; connected dots represent regions shared across tests or populations.

4 DISCUSSION

This study assessed genetic diversity and selection signatures in 3 cattle populations: a Holstein population from the United States and Canada, Gir from India, and Dairy Gir from Brazil, which were originally imported from India and have been selected for milk production in Brazil over the past 40 years. We assessed several diversity metrics and detected regions under selection via within-population and between-population tests.

4.1 Population Differentiation and Genetic Diversity

As expected, a high ratio of averages of F_{ST} based on Hudson's method was observed when the HOL population was included in the pair. Specifically, we observed a similar genetic differentiation between the HOL and GIR_BR populations (0.4671) compared with the findings (0.436) reported by Tijjani et al. (2022) when Holstein cows from Europe and Dairy Gir cows from the same GIR_BR population that were studied in our research. Moderate (0.1062) genetic differentiation was observed between the GIR_BR and GIR_IN pair. Despite belonging to the same breed, these populations have diverse backgrounds and have experienced distinct selection processes, with greater selection pressure on the GIR_BR population.

The Gir populations, particularly GIR_BR, presented higher values of mean π_{Total} than did the HOL population, indicating considerably greater genetic diversity for Gir populations. In contrast to HOL, GIR_BR, and GIR_IN populations experienced weaker selection intensity, likely resulting in more alleles segregating and sequences more divergent, creating more different sites per sequence (Nei and

Li, 1979). Masharing et al. (2023) reported a greater nucleotide diversity value (0.356) for Gir from India using ddRAD sequencing than our study did (mean π_{Total} = 0.278). Differences in sampling methods and the density of the molecular sequencing technique used may have contributed to this discrepancy. The HOL population showed the lowest mean π_{Total} (0.152%), likely due to the stronger selection pressure for maximizing productivity. Neumann et al. (2023) reported a similar value for nucleotide diversity in the Holstein population (Mean π_{Total} = 0.147%). The authors used a set of Holstein samples from Run 9 of the 1000 Bull Genomes project, including 60 individuals from Canada and the United States.

The regions with the highest levels of nucleotide diversity across all populations were BTA4 (113.1–113.17 Mb, 105.56–105.63 Mb), BTA12 (71.16–71.27 Mb, 72.8 - 72.85 Mb), BTA15 (45.96–46.01 Mb, 46.2–46.25 Mb), and BTA23 (26.04–26.17 Mb, 27.1–27.15 Mb). The top 10 regions with relatively high nucleotide diversity were associated with uncharacterized genes (LOC genes; Supplemental Table S5). Among the characterized genes with higher nucleotide diversity, *GIMAP7*, on BTA4, is a member of the GTPase family that is conserved in both plants and vertebrates (Krücken et al., 2004). Regions with high genetic diversity are known to be responsible for regulating immunity (Li et al., 2023). Future research could further unravel the genomic architecture and impact of regions with high nucleotide diversity.

The region with the highest SNV density (Supplemental Figures S2–S4) aligns with our analysis of nucleotide diversity on BTA23 (regions 26.04–26.17 Mb and 27.1–27.15 Mb) across all populations. When a region has a high density of SNVs and exhibits much higher-than-average nucleotide diversity, this suggests that the region has accumulated many mutations over time. These mutations contribute to increased genetic variation within each population, because nucleotide diversity measures the degree of variation between the sequences of different alleles (Nei and Li, 1979). Several gene families related to immunity (i.e., major histocompatibility complex, *BOLA*), olfaction (olfactory receptor family), and reproduction (i.e., glutathione S-transferase α; Rabahi et al., 1999) were annotated in this region (Supplemental Table S6).

4.2 Minor Allele Frequency (MAF)

The GIR_BR cows showed the highest mean MAF (0.191) among the studied populations. This may be because the animals sampled for this population were strategically chosen to represent the different lineages of the Brazilian Dairy Gir. A previous study including a sample of 5 Peruvian Gir descendants from Brazilian Dairy Gir reported a lower MAF of 0.13 (Corredor et al., 2023). An MAF score higher than 0.1 indicates sufficient polymorphism in a population for effective genomic selection (Rajawat et al., 2024). The proportions of polymorphic markers (markers with an MAF >0.1) in GIR_IN and GIR_BR were 0.52 and 0.63, respectively. The GIR_IN individuals were not part of an animal breeding program. The lower MAF (0.044) observed in the HOL population corroborates the intense selection pressure applied on this breed in Canada and the United States.

4.3 Heterozygosity and Inbreeding Coefficient

The H_O and H_E in the HOL population (0.219 and 0.226, respectively) were lower than those reported for other Holstein populations. Previous studies reported H_O ranging from 0.36 to 0.38 and H_E between 0.31 and 0.37 (Makina et al., 2014; Vostry et al., 2023). The inbreeding coefficient in HOL was 0.031, similar to values previously reported for Holsteins from Canada and the United States (Vostry et al., 2023). Our results suggest that, despite the lower levels of heterozygosity, the overall level of inbreeding remains relatively low. The reduced genetic diversity is likely associated with the intensive selection pressure applied to Holsteins, which increased after the implementation of genomic selection (Makanjuola et al., 2020). The average inbreeding coefficient increased by 0.26% between 2010 and 2020 in Canadian Holsteins (Van Doormaal, 2024) while maintaining high productivity, which agrees with our findings.

The GIR_IN population displayed the lowest mean of the inbreeding coefficient (−0.092) and the highest difference between H_O (0.373) and H_E (0.341). The lower levels of selection pressure, compared with the other populations, may have contributed to these results. However, this comparison should be taken with caution due to the GIR_IN population sample size. A small sample size (<20 animals) can

lead to underestimation of H_E and, thus, alter the genetic distances to other populations (Ajmone-Marsan et al., 2023).

The GIR_BR population displayed an F_{HOM} of 0.011. Using a part of the population used in this study, Peripolli et al. (2020) reported an inbreeding coefficient based on runs of homozygosity (F_{ROH}) of 0.04, and Garcia et al. (2023) reported an F_{ROH} of 0.19; however, those coefficients rely on distinct methods, and low to high correlations were observed between F_{ROH} and F_{HOM} in Dairy Gir cattle (Peripolli et al., 2018).

4.4 Selection Signatures and Genes Under Selection

This study used several tests to identify selection signatures within and between Holstein and 2 populations of Gir cattle. The genes colocalized with the selection signatures highlight positive selection pressure related to adaptation and the selection criteria applied, especially for Holstein and Dairy Gir cattle. More specifically, the genes under selection were related to immunity, milk production, mammary gland function, embryogenesis, reproduction, and environmental adaptation.

We assumed the extremes of the 1% windows as candidate selection signatures, and we merged adjacent windows to identify signs of potential recent selection. This is because there has not been enough time for crossover events during meiosis to disrupt and recombine the chromosomal segments exhibiting these signatures. The positive mean Tajima's D value observed in the genome across all genome regions is expected because only approximately 5% of the mammalian genome is under selection (Mouse Genome Sequencing Consortium, 2002), and selection is applied for numerous traits in breeding programs. Although the 3 population sets had a large difference in number of individuals, the chosen tests for between-population analyses (XP-EHH and F_{ST}) were suitable for unbalanced sample sizes.

Employing multiple methods enhances the reliability of selection signature detection (Qanbari and Simianer 2014) and strengthens confidence in the results, highlighting candidates for future studies. However, because these strategies are

grounded in varying statistical assumptions, they may generate different findings that can complement one another (Rodrigues et al. 2025).

4.4.1 Environmental Adaptation

Selection signatures between Holstein (*Bos taurus taurus*) and Gir (*Bos taurus indicus*) populations may reflect the long-term evolutionary divergence that led to differentiation between the 2 subspecies. We detected selection signatures through multiple tests (iHS GIR_IN, iHS GIR_BR, Tajima's D GIR_BR, F_{ST} GIR_IN and HOL, and F_{ST} GIR_BR and HOL) that coincided with the regions reported by Naval-Sánchez et al. (2020) and Balbi et al. (2023). These regions are also colocalized with the *HELB* gene. Naval-Sánchez et al. (2020) identified exclusive indicine variants and a selection signature in BTA5:47.67–48.1 Mb, whereas Balbi et al. (2023) identified a main peak on BTA5:46.94–48.03 Mb in a GWAS for coat score and hair length in Brangus animals. The *HELB* gene is an ATP-dependent DNA helicase involved in DNA replication, damage repair, and replication stress responses (Hazeslip et al., 2020). Thus, the *HELB* gene may have contributed to the adaptation of indicine cattle to the environmental conditions of tropical regions (Naval-Sánchez et al., 2020).

Members of the heat shock protein 70 (HSP70) family were found in the selection signatures for the Gir populations. This family is the largest and most abundant family of heat shock proteins (HSP). These proteins are chaperones that provide protection against cellular stress by maintaining protein folding and decreasing apoptosis (Velichko et al., 2013). The *HSPA4* gene was found in the GIR_IN population through the iHS test. *HSPA4* also regulates the progression of spermatogenesis, directly affecting male fertility (Held et al., 2011). Additionally, the *HSPA12A* gene was detected in the analysis between GIR_IN and GIR_BR via the F_{ST} test, which could be explained by the particular differences between Brazil and India, because the expression of this gene may be linked to environmental stressors and adaptation between Brazil and India.

Members of the heat shock protein 40 (HSP40) family were also found in our analyses, especially for Gir populations. This protein collaborates with other HSP to maintain protein homeostasis, forming an important chaperone system (Liu et al.,

2020). The *DNAJC18* gene was found in the GIR_IN and HOL pair and in the GIR_BR and HOL pair in the F_{ST} analysis and in the GIR_BR population through the iHS test. This gene is involved in thermogenesis and cold and heat adaptation (Huang et al., 2023). The *DNAJC3* gene was observed under selection pressure in the HOL population via the iHS test and in HOL (GIR_IN and HOL pair) and GIR_BR (GIR_IN and GIR_BR pair) through the XP-EHH test. This gene was previously implicated in endoplasmic reticulum stress in hepatocytes exposed to nonesterified fatty acids (NEFA), indicating that this gene might be directly related to negative energy balance (Fang et al., 2022).

Many members of the olfactory family were identified via intra- and between-population analyses. Olfaction is a primary sense in cattle and is essential for the recognition of toxic plants; behavioral interactions, such as mother–offspring and social relationships; and the communication and recognition of fear, danger, estrus, disease, and death (Salesse, 2017). Several olfactory genes are subjected to copy number variation in Holstein and Dairy Gir cattle (Butty et al., 2020; Braga et al., 2023). For example, the gene *OR2AJ9* was found in the GIR_IN and GIR_BR pairwise analyses and was previously found to be subjected to deletion in Dairy Gir animals from the same population as in this study (Braga et al., 2023).

Several phospholipase A2 (PLA2) family genes were found under selection across the 3 populations. *PLA2G4E*, *PLA2G4D*, and *PLA2G4F* were identified via the XP-EHH statistic for HOL with both Gir populations. Phospholipase A2 is involved in several biological events, such as the production of lipid mediators, the immune response, membrane remodeling, and skin barrier maintenance (Murakami et al., 2020). The *PLA2G* gene family was previously reported to be subjected to copy number variation in cattle and related to the immune response (Golik et al., 2006; Stothard et al., 2011; Seroussi et al., 2013). The *PLA2R1* gene was identified via within-population tests for the GIR_IN population and between-population tests for the GIR_IN and GIR_BR pair, highlighting that this gene is under selection in the GIR_IN population. This gene has been linked to stress, the immune response, and oxidative stress in heat-stressed Holstein cows (Zheng et al., 2014), and may be related to heat tolerance in this population

4.4.2 Milk Production and Mammary Gland Function in Relation to Embryogenesis

The greater selection pressure for milk traits in GIR_BR led by PNMGL, compared with GIR_IN, may have left a distinct genetic footprint related to milk production. In the F_{ST} analysis, the window with the highest value in the GIR_BR and GIR_IN pair was 0.713 (BTA 14:0.05–0.1 Mb). This region harbors QTL related to milk yield (243502) and *ACER3*, *OR5D70P*, and *U6* genes. Further studies are encouraged to take a closer look at this region in Gir cattle because this region may be related to increased milk production. The QTL colocalized with ACER3, insulin level (193405), was significantly enriched (FDR <0.01) in our analysis. Insulin levels are directly related to milk yield (Horst et al., 2021). However, no study comparing insulin levels in Brazilian Gir and Gir from India has been published to our knowledge. Another gene that might affect lactation is *PTEN*. This gene was found in the HOL population via the iHS test. *PTEN* regulates fatty acid metabolism (Fu et al., 2012) and can inhibit mammary gland development and lactation in dairy cows (Wang et al., 2014).

Genes previously found to affect embryogenesis and mammary gland function were found under selection. The *CAV1* gene is downregulated in Holstein cows with hemorrhagic mastitis (Zhang et al., 2022). The authors also observed strong detection of *CAV1* protein in vascular cells of healthy mammary glands and minor evidence of the same protein in hemorrhagic mastitis. Moreover, this gene was found to be upregulated in bovine oviduct epithelial cells exposed to elevated NEFA compared with basal levels (Jordaens et al., 2017), as well as in granulosa cells during ovulation and peri-ovulation (Lussier et al., 2017). These findings suggest that the *CAV1* gene influences fertility and mammary gland health.

The *AQP1* gene was found in a Tajima's D selection signature in HOL, and *AQP8* was found in an iHS selection signature in GIR_BR. Aquaporins (AQP) are integral transmembrane proteins that regulate water transport across cell membranes (Laloux et al., 2018). *AQP1* and *AQP8* play essential roles in early bovine embryo development. It was previously found to be expressed in the cumulus–oocyte complex and blastocyst embryo stage (Petano-Duque et al., 2022). The AQP are expressed in the bovine mammary gland and might be responsible for increasing permeability and, to a lesser extent, milk secretion (Mobasheri et al., 2011).

4.4.3 Fertility

Several genes and QTL related to reproduction were under selection in the BTX. The *GAB3* gene was found to be under selection via Tajima's D in the HOL population. This gene has been previously associated with spontaneous abortion in Holstein cattle (Oliver et al., 2019). *GAB3* mutant mice exhibited defects in their ability to successfully complete pregnancies (Sliz et al., 2019). *DACH2* was found in the GIR_BR population via iHS, and it is associated with premature ovarian failure in humans (Bione et al., 2004) and puberty in cattle (Cánovas et al., 2014).

The *FSHR* gene was found to be under selection in the GIR_IN population in the XP-EHH, GIR_IN and HOL pair analysis. This gene was previously related to early reproductive traits in the Brahman breed (Tahir et al., 2021). It encodes the follicle-stimulating hormone receptor, which controls follicle development, spermatogenesis, steroidogenesis, ovulation, and hormonal balance (Ulloa-Aguirre et al., 2018). Several genes of the family spermatogenesis-associated proteins (*SPATA2L*, *SPATA13*, *SPATA24*, and *SPATA33*) were identified via iHS for the HOL population and between-population pairwise analyses with the HOL population.

Genes previously related to reproduction traits and embryogenesis were also identified via GIR_IN and GIR_BR pair analyses. The *SAMD5* and *SNORA70* genes were found to be under selection pressure in the F_{ST} analysis. The *SAMD5* gene is related to the sire-conception rate in Holstein cattle (Fang et al., 2019), and *SNORA70* is linked to scrotal circumference at 18 months of age in Nelore, Brahman, and Tropical Composite breeds (Melo et al., 2019, Miles et al., 2023), underscoring the importance of reproductive traits under selection pressure. In pairwise XP-EHH analysis, the *WDR36* gene was found to be under selection pressure. The *WDR36* gene is related to blastocyst development in cattle through miR-146b regulation (Pavani et al., 2024). Knockdown of *WDR36* can lead to preimplantation embryonic lethality in mice, and its knockout causes the apoptotic death of blastomeres (Gallenberger et al., 2011).

4.4.4 Growth Traits and Development

Selection signatures overlapping genes affecting growth traits and milk production were found in this study. The *PLAG1* gene was found to be under selection via the F_{ST} test in the GIR_IN and HOL pair and in the HOL population via Tajima's D test. The region harboring the *PLAG1* gene is fixed or near fixation in Holstein cattle, as supported by our findings and its identification under selection in multiple studies (Signer-Hasler et al., 2017; Taye et al., 2017; Zinovieva et al., 2020; Bertolini et al., 2022; Jin et al., 2022). *PLAG1* regulates embryonic and fetal development in cattle (D'Occhio et al., 2024) and is ubiquitously and highly expressed in fetuses compared with adult cattle (Li et al., 2020). In Holstein cattle, this gene was previously associated with milk composition traits (Fink et al., 2017), stature (Karim et al., 2011; Littlejohn et al., 2012; Bouwman et al., 2018), growth traits (Littlejohn et al., 2012; Fink et al., 2017), and heifer livability (Gao et al., 2023).

Genes related to stature in cattle, such as *IGF2*, *HMGA2*, *CCND2*, and *NCAPG*, were found in selection signatures for the GIR_BR population (Bouwman et al., 2018). *HMGA2* regulates *IGF2*, a paralog of *IGF1* known to regulate growth (Abi Habib et al., 2018). The *HMGA2* gene is associated with ovarian length and height in various cattle breeds (Shen et al., 2024). An indicine-specific copy number variant in the *HGMA2* gene was previously associated with navel length in Nelore cattle (Aguiar et al., 2018).

Several QTL associated with muscularity and body fat were enriched for GIR_IN, GIR_BR, and HOL, including traits such as Longissimus muscle area, Subcutaneous fat thickness, Carcass weight, Marbling score, and Body weight. These findings may be related to the fact that, especially during the early lactation period, absorbed nutrients and body reserves need to be partitioned toward milk synthesis to allow milk production and yield to increase (Piantoni and VandeHaar 2023).

4.4.5 Health and Disease Resistance

Our research also revealed numerous genes from the solute carrier family across different populations and tests. Interestingly, a selection signature region harboring *SLC11A1* (also known as *NRAMP1*) was close to fixation in the HOL population. This gene plays an important role in host innate immunity by modulating

iron metabolism in macrophages and influencing early-phase macrophage activation (Wyllie et al., 2002). Polymorphisms in the *SLC11A1* gene were previously related to susceptibility to tuberculosis in Holstein cattle (Liu et al., 2017) and resistance to several intracellular pathogen infections, including brucellosis (Cortés and González, 2015), paratuberculosis (Pinedo et al., 2009), and tuberculosis (Kadarmideen et al., 2011; Liu et al., 2017), in various breeds of cattle. Compared with indicine and crossbred cattle, Holstein cattle are more susceptible to tuberculosis (Ameni et al., 2007), whereas indicine and crossbred cattle exhibit a greater degree of diversity in the *SLC11A1* gene (Cortés and González, 2015; Soares Fioravanti et al., 2020). Therefore, increasing variability within the *SLC11A1* gene might benefit Holstein cattle health.

Another iron-related gene was found under selection in the BTX for the GIR_IN population via iHS analysis. *TMEM164* has a key role in ferroptosis, an iron-dependent regulated cell death implicated in health and disease. This gene was previously linked to *Mycobacterium bovis* infection (Correia et al., 2022) and SCS (Sanchez et al., 2023).

5 CONCLUSIONS

We investigated genetic diversity and selection signatures in a Holstein cattle population from the United States and Canada, Gir cattle from India, and Gir cattle imported from India to Brazil and selected for milk production over the past 40 years (Dairy Gir cattle). We observed that genetic diversity was lower in the Holstein population than in the Dairy Gir population. Genes linked to major economic traits in dairy cattle, including growth traits, reproduction, mastitis, milk production, heat tolerance, health, temperament, and adaptation, were found to be under selection pressure. Additionally, we also recommended target areas for increasing diversity and areas that should be further assessed as advantageous for marker-enhanced selection.

6 REFERENCES

Abi Habib W et al. (2018) Genetic disruption of the oncogenic HMGA2-PLAG1-IGF2 pathway causes fetal growth restriction. Genetics in Medicine 20:250-258.

Aguiar TS et al. (2018) Association of Copy Number Variation at Intron 3 of HMGA2 With Navel Length in Bos indicus. Frontiers in Genetics 9:627.

Ajmone-Marsan P, Boettcher PJ, Colli L, Ginja C, Kantanen J, Lenstra JA (eds) (2023) Genomic characterization of animal genetic resources – Practical guide. FAO Animal Production and Health Guidelines No. 32. Rome.

Ameni G et al. (2007) High prevalence and increased severity of pathology of bovine tuberculosis in Holsteins compared to zebu breeds under field cattle husbandry in central Ethiopia. Clinical and Vaccine Immunology 14:1356-1361.

Balbi M et al. (2023) Coat score. A possible explanation for the zebuine selective sweep located on bovine chromosome 5: 47,670,001-48,100,000 bp. Animal Biotechnology 34:1422-1428.

Bertolini F et al. (2022) Signatures of selection are present in the genome of two close autochthonous cattle breeds raised in the North of Italy and mainly distinguished for their coat colours. Journal of Animal Breeding and Genetics 139:307-319.

Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting FST: the impact of rare variants. Genome Research 23:1514-1521.

Bione S et al. (2004) Mutation analysis of two candidate genes for premature ovarian failure, DACH2 and POF1B. Human Reproduction 19:2759-2766.

Bouwman AC et al. (2018) Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nature Genetics 50:362-367.

Braga LG et al. (2023) Identification of copy number variations in the genome of Dairy Gir cattle. PLoS One 18:e0284085.

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics 81:1084-1097.

Butty AM et al. (2020) High confidence copy number variants identified in Holstein dairy cattle from whole genome sequence and genotype array data. Scientific Reports 10:8044.

Cánovas A et al. (2014) Multi-Tissue Omics Analyses Reveal Molecular Regulatory Networks for Puberty in Composite Beef Cattle. PLoS One 9:e102551.

Carrier A, Gilbert I, Leclerc P, Duchesne M, Robert C (2023) Characterization of the genetic pool of the Canadienne dairy cattle breed. Genetics Selection Evolution 55:32.

Corredor FA et al. (2023) Genetic diversity and population structure of a Peruvian cattle herd using SNP data. Frontiers in Genetics 14:1073843.

Correia CN et al. (2022) High-resolution transcriptomics of bovine purified protein derivative-stimulated peripheral blood from cattle infected with Mycobacterium bovis across an experimental time course. Tuberculosis 136:102235.

Cortés ÁV, González HR (2015) A Preliminary Study of Solute Carrier Family Gene in Adapted Bovine Breeds of Panama. Italian Journal of Animal Science 14:4057.

D'Occhio MJ et al. (2024) Pleomorphic adenoma gene 1 in reproduction and implication for embryonic survival in cattle: a review. Journal of Animal Science 102:skae103.

Fang L et al. (2019) Integrating signals from sperm methylome analysis and genome-wide association study for a better understanding of male fertility in cattle. Epigenomes 3:10.

Fang Z et al. (2022) Targeting IRE1α and PERK in the endoplasmic reticulum stress pathway attenuates fatty acid-induced insulin resistance in bovine hepatocytes. Journal of Dairy Science 105:6895-6908.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Fink T et al. (2017) Functional confirmation of PLAG1 as the candidate causative gene underlying major pleiotropic effects on body weight and milk characteristics. Scientific Reports 7:44793.

Fonseca PAS, Suárez-Vega A, Marras G, Cánovas Á (2020) GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience 9:giaa149.

Fu S et al. (2012) Increase of fatty acid oxidation and VLDL assembly and secretion overexpression of PTEN in cultured hepatocytes of newborn calf. Cellular Physiology and Biochemistry 30:1005-1013.

Gallenberger M et al. (2011) Lack of WDR36 leads to preimplantation embryonic lethality in mice and delays the formation of small subunit ribosomal RNA in human cells in vitro. Human Molecular Genetics 20:422-435.

Gao Y et al. (2023) Genome-wide association analysis of heifer livability and early first calving in Holstein cattle. BMC Genomics 24:628.

Garcia AO et al. (2023) Pedigree reconstruction and population structure using SNP markers in Gir cattle. Journal of Applied Genetics 64:329-340.

Gautier M, Klassmann A, Vitalis R (2017) rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Molecular Ecology Resources 17:78-90.

Golik M et al. (2006) Accelerated expansion of group IID-like phospholipase A2 genes in Bos taurus. Genomics 87:527-533.

Hayes BJ, Daetwyler HD (2019) 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: applications and outcomes. Annual Review of Animal Biosciences 7:89-102.

Hazeslip L, Zafar MK, Chauhan MZ, Byrd AK (2020) Genome Maintenance by DNA Helicase B. Genes 11:578.

Held T, Barakat AZ, Mohamed BA, Paprotta I, Meinhardt A, Engel W, Adham IM (2011) Heat-shock protein HSPA4 is required for progression of spermatogenesis. Reproduction 142:133–144.

Horst EA, Kvidera SK, Baumgard LH (2021) Invited review: The influence of immune activation on transition cow health and performance—A critical evaluation of traditional dogmas. Journal of Dairy Science 104:8380–8410.

Huang N, Zhao L, Wang J, Jiang Q, Ju Z, Wang X, Yang C, Gao Y, Wei X, Zhang Y, Xiao Y, Liu W, Lu S, Huang J (2023) Signatures of selection in indigenous Chinese cattle genomes reveal adaptive genes and genetic variations to cold climate. Journal of Animal Science 101:skad006.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589.

Jensen JD, Foll M, Bernatchez L (2016) The past, present and future of genomic scans for selection. Molecular Ecology 25:1–4.

Jin L, Qu K, Hanif Q, Zhang J, Liu J, Chen N, Suolang Q, Lei C, Huang B (2022) Whole-Genome Sequencing of Endangered Dengchuan Cattle Reveals Its Genomic Diversity and Selection Signatures. Frontiers in Genetics 13:833475.

Jordaens L, Van Hoeck V, Maillo V, Gutierrez-Adan A, Marei WFA, Vlaeminck B, Thys S, Sturmey RG, Bols PEJ, Leroy JLMR (2017) Maternal metabolic stress may affect oviduct gatekeeper function. Reproduction 153:759–773.

Kadarmideen HN, Ali AA, Thomson PC, Müller B, Zinsstag J (2011) Polymorphisms of the SLC11A1 gene and resistance to bovine tuberculosis in African Zebu cattle. Animal Genetics 42:656–658.

Kardos M, Waples RS (2024) Low-coverage sequencing and Wahlund effect severely bias estimates of inbreeding, heterozygosity and effective population size in North American wolves. Molecular Ecology e17415.

Karim L, Takeda H, Lin L, Druet T, Arias JA, Baurain D, Cambisano N, Davis SR, Farnir F, Grisart B et al. (2011) Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. Nature Genetics 43:405–413.

Klassmann A, Gautier M (2022) Detecting selection using extended haplotype homozygosity (EHH)-based statistics in unphased or unpolarized data. PLoS One 17:e0262024.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinformatics 14:289.

Krücken J, Schroetel RM, Müller IU, Saïdani N, Marinovski P, Benten WP, Stamm O, Wunderlich F (2004) Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. Gene 341:291–304.

Laloux T, Junqueira B, Maistriaux LC, Ahmed J, Jurkiewicz A, Chaumont F (2018) Plant and Mammal Aquaporins: Same but Different. International Journal of Molecular Sciences 19:521.

Li CC, Horvitz DG (1953) Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics 5:107–117.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Li TT, Xia T, Wu JQ, Hong H, Sun ZL, Wang M, Ding FR, Wang J, Jiang S, Li J et al. (2023) De novo genome assembly depicts the immune genomic characteristics of cattle. Nature Communications 14:6601.

Li Z, Wu M, Zhao H, Fan L, Zhang Y, Yuan T, He S, Wang P, Zhang Y, Sun X, Wang S (2020) The PLAG1 mRNA expression analysis among genetic variants and relevance to growth traits in Chinese cattle. Animal Biotechnology 31:504–511.

Littlejohn M, Grala T, Sanders K, Walker C, Waghorn G, Macdonald K, Coppieters W, Georges M, Spelman R, Hillerton E et al. (2012) Genetic variation in PLAG1 associates with early life body weight and peripubertal weight and growth in Bos taurus. Animal Genetics 43:591–594.

Liu K, Zhang B, Teng Z, Wang Y, Dong G, Xu C, Qin B, Song C, Chai J, Li Y et al. (2017) Association between SLC11A1 (NRAMP1) polymorphisms and susceptibility to tuberculosis in Chinese Holstein cattle. Tuberculosis 103:10–15.

Liu Q, Liang C, Zhou L (2020) Structural and functional analysis of the Hsp70/Hsp40 chaperone system. Protein Science 29:378–390.

Lussier JG, Diouf MN, Lévesque V, Sirois J, Ndiaye K (2017) Gene expression profiling of upregulated mRNAs in granulosa cells of bovine ovulatory follicles following stimulation with hCG. Reproductive Biology and Endocrinology 15:88.

Machado MA, Azevedo AL, Teodoro RL, Pires MA, Peixoto MG, de Freitas C, Prata MC, Furlong J, da Silva MV, Guimarães SE et al. (2010) Genome wide scan for quantitative trait loci affecting tick resistance in cattle (Bos taurus × Bos indicus). BMC Genomics 11:280.

Makanjuola BO, Miglior F, Abdalla EA, Maltecca C, Schenkel FS, Baes CF (2020) Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. Journal of Dairy Science 103:5183–5199.

Makina SO, Muchadeyi FC, van Marle-Köster E, MacNeil MD, Maiwashe A (2014) Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Frontiers in Genetics 5:333.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867–2873.

Mao X, Sahana G, De Koning DJ, Guldbrandtsen B (2016) Genome-wide association studies of growth traits in three dairy cattle breeds using whole-genome sequence data. Journal of Animal Science 94:1426–1437.

Masharing N, Sodhi M, Chanda D, Singh I, Vivek P, Tiwari M, Kumari P, Mukesh M (2023) ddRAD sequencing based genotyping of six indigenous dairy cattle breeds of India to infer existing genetic diversity and population structure. Scientific Reports 13:9379.

McGuffey RK, Shirley JE (2011) History of Dairy Farming. In: Fuquay JW (ed) Encyclopedia of Dairy Sciences. Academic Press, San Diego, pp 2–11.

Miles A, bot pyup.io, Rodrigues M (2023) cggh/scikit-allel: v1.3.7.

Mobasheri A, Kendall BH, Maxwell JE, Sawran AV, German AJ, Marples D, Luck MR, Royal MD (2011) Cellular localization of aquaporins along the secretory pathway of the lactating bovine mammary gland: an immunohistochemical study. Acta Histochemica 113:137–149.

Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562.

Murakami M, Sato H, Taketomi Y (2020) Updating Phospholipase A2 Biology. Biomolecules 10:1457.

Naval-Sánchez M, Porto-Neto LR, Cardoso DF, Hayes BJ, Daetwyler HD, Kijas J, Reverter A (2020) Selection signatures in tropical cattle are enriched for promoter and coding regions and reveal missense mutations in the damage response gene HELB. Genetics Selection Evolution 52:27.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* 76:5269-5273.

Neumann GB, Korkuć P, Arends D, Wolf MJ, May K, König S, Brockmann GA (2023) Genomic diversity and relationship analyses of endangered German Black Pied cattle (DSN) to 68 other taurine breeds based on whole-genome sequencing. *Frontiers in Genetics* 13:993959.

O'Brien AM, Höller D, Boison SA, Milanesi M, Bomba L, Utsunomiya YT, Carvalheiro R, Neves HH, Silva MV, VanTassell CP, Sonstegard TS, Mészáros G, Ajmone-Marsan P, Garcia F, Sölkner J (2015) Low levels of taurine introgression in the current Brazilian Nelore and Gir indicine cattle populations. *Genetics Selection Evolution* 47:31.

Oliver KF, Wahl AM, Dick M, Toenges JA, Kiser JN, Galliou JM, Moraes JGN, Burns GW, Dalton J, Spencer TE, Neibergs HL (2019) Genomic analysis of spontaneous abortion in Holstein heifers and primiparous cows. *Genes* 10:954.

Pavani KC, XueFeng G, Chunduru J, Meese T, Peelman LJ, Van Nieuwerburgh F, Deforce D, Hendrix A, Tilleman K, Van Soom A, Smits K (2023) MicroRNA-146b negatively affects bovine embryo development and quality. *Reproduction* (Advance online publication).

Peripolli E, Stafuzza NB, Munari DP, Lima ALF, Irgang R, Machado MA, Panetto JCDC, Ventura RV, Baldi F, Silva MVGB (2018) Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. *BMC Genomics* 19:34.

Peripolli E, Reimer C, Ha NT, Geibel J, Machado MA, Panetto JCDC, Egito AA, Baldi F, Simianer H, Silva MVGB (2020) Genome-wide detection of signatures of selection in indicine and Brazilian locally adapted taurine cattle breeds using whole-genome re-sequencing data. *BMC Genomics* 21:624.

Petano-Duque JM, Castro-Vargas RE, Cruz-Mendez JS, Lozano-Villegas KJ, Herrera-Sánchez MP, Uribe-García HF, Naranjo-Gómez JS, Otero-Arroyo RJ, Rondón-Barragán IS (2022) Gene expression of aquaporins (AQPs) in cumulus oocytes complex and embryo of cattle. *Animals* 13:98.

Pinedo PJ, Buergelt CD, Donovan GA, Melendez P, Morel L, Wu R, Langaee TY, Rae DO (2009) Candidate gene polymorphisms (BoIFNG, TLR4, SLC11A1) as risk factors for paratuberculosis infection in cattle. *Preventive Veterinary Medicine* 91:189-196.

Prata MA, Faro LE, Moreira HL, Verneque RS, Vercesi Filho AE, Peixoto MG, Cardoso VL (2015) Genetic parameters for milk production traits and breeding goals for Gir dairy cattle in Brazil. *Genetics and Molecular Research* 14:12585-12594.

Qanbari S, Simianer H (2014) Mapping signatures of positive selection in the genome of livestock. *Livestock Science* 166:133-143.

Rabahi F, Brûlé S, Sirois J, Beckers JF, Silversides DW, Lussier JG (1999) High expression of bovine alpha glutathione S-transferase (GSTA1, GSTA2) subunits is mainly associated with steroidogenically active cells and regulated by gonadotropins in bovine ovarian follicles. *Endocrinology* 140:3507-3517.

Rajawat D, Ghildiyal K, Nayak SS, Sharma A, Parida S, Kumar S, Ghosh AK, Singh U, Sivalingam J, Bhushan B, Dutt T, Panigrahi M (2024) Genome-wide mining of diversity and evolutionary signatures revealed selective hotspots in Indian Sahiwal cattle. *Gene* 901:148178.

Rodrigues JL, Braga LG, Watanabe RN, Schenkel FS, Berry DP, Buzanskas ME, Munari DP (2025) Genetic diversity and selection signatures in sheep breeds. Journal of Applied Genetics. 30:1-13.

Sabeti PC, Varilly P et al. (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913-918.

Salesse R (2017) Opinion paper: Smell: an affordable way to improve livestock welfare. *Animal* 11:1425-1426.

Sambrook J, Russel DW (2001) Molecular cloning: a laboratory manual. 3rd ed. Cold Spring Harbor Laboratory Press, New York.

Sanchez MP, Escouflaire C, Baur A, Bottin F, Hozé C, Boussaha M, Fritz S, Capitan A, Boichard D (2023) X-linked genes influence various complex traits in dairy cattle. *BMC Genomics* 24:1-18.

Seroussi E, Klompus S, Silanikove M, Krifucks O, Shapiro F, Gertler A, Leitner G (2013) Nonbactericidal secreted phospholipase A2s are potential anti-inflammatory factors in the mammary gland. *Immunogenetics* 65:861-871.

Shen S, Zhu L, Yang Y, Bi Y, Li J, Wang Y, Pan C, Wang S, Lan X (2024) Exploration of the polymorphism distribution of bovine HMGA2 gene in worldwide breeds and its associations with ovarian traits. *Animals* 14:796.

Signer-Hasler H, Burren A, Neuditschko M, Frischknecht M, Garrick D, Stricker C, Gredler B, Bapst B, Flury C (2017) Population structure and genomic inbreeding in nine Swiss dairy cattle populations. *Genetics Selection Evolution* 49:83.

Sliz A, Locker KCS, Lampe K, Godarova A, Plas DR, Janssen EM, Jones H, Herr AB, Hoebe K (2019) Gab3 is required for IL-2- and IL-15-induced NK cell expansion and limits trophoblast invasion during pregnancy. *Science Immunology* 4:eav3866.

Soares Fioravanti MC, Silva Freitas TM, Moura MI, Lage Costa G, Moraes Dias J, Kim Pires LG, Gómez MM, Landi V (2020) Resistance and resilience to diseases in local ruminant breeds: a focus on South America. *Archivos de Zootecnia* 69:338-352.

Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y, Liao X, Moore SS (2011) Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics* 12:559.

Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, et al. (2023) The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* 51:D638-D646.

Tahir MS, Porto-Neto LR, Gondro C, Shittu OB, Wockner K, Tan AW, Smith HR, Gouveia GC, Kour J, Fortes MR (2021) Meta-analysis of heifer traits identified reproductive pathways in Bos indicus cattle. *Genes* 12:768.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.

Taye M, Lee W, Jeon S, Yoon J, Dessie T, Hanotte O, Mwai OA, Kemp S, Cho S, Oh SJ, Lee HK, Kim H (2017) Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mammalian Genome* 28:528-541.

Theunissen B (2012) Breeding for nobility or for production? Cultures of dairy cattle breeding in the Netherlands, 1945-1995. *Isis* 103:278-309.

Tijjani A, Salim B, Silva MVB, Eltahir HA, Musa TH, Marshall K, Hanotte O, Musa HH (2022) Genomic signatures for drylands adaptation at gene-rich regions in African zebu cattle. *Genomics* 114:110423.

Ulloa-Aguirre A, Zariñán T, Jardón-Valadez E, Gutiérrez-Sagal R, Dias JA (2018) Structure-Function Relationships of the Follicle-Stimulating Hormone Receptor. Frontiers in Endocrinology 9:707.

Utsunomiya YT, Pérez O'Brien AM, Sonstegard TS, Van Tassell CP, do Carmo AS, Mészáros G, Sölkner J, Garcia JF (2013) Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. PLoS One 8:64280.

Van Doormaal, B. 2024. Inbreeding Update - Lactanet. Accessed Feb. 27, 2025. https://lactanet.ca/en/inbreeding-update/.

Velichko AK, Markova EN, Petrova NV, Razin SV, Kantidze OL (2013) Mechanisms of heat shock response in mammals. Cellular and Molecular Life Sciences 70:4229-4241.

Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biology 4:72.

Vostry L, Vostra-Vydrova H, Moravcikova N, Kasarda R, Cubric-Curik V, Brzakova M, Solkner J, et al. (2023) Genomic diversity and population structure of the Czech Holstein cattle. Livestock Science 273:105261.

Wang J, Liao Y (2020) WebGestaltR: Gene Set Analysis Toolkit. WebGestaltR.

Wang Z, Hou X, Qu B, Wang J, Gao X, Li Q (2014) Pten regulates development and lactation in the mammary glands of dairy cows. PLoS One 9:102118.

Wright S (1951) The genetical structure of populations. Annals of Eugenics 15:323-354.

Wyllie S, Seu P, Goss JA (2002) The natural resistance-associated macrophage protein 1 Slc11a1 (formerly Nramp1) and iron metabolism in macrophages. Microbes and Infection 4:351-359.

Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X (2021) rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. Genomics, Proteomics & Bioinformatics 19:619-628.

Zhang Q, Bai X, Shi J, Wang X, Zhang B, Dai L, Lin T, Gao Y, Zhang Y, Zhao X (2022) DIA proteomics identified the potential targets associated with angiogenesis in the mammary glands of dairy cows with hemorrhagic mastitis. Frontiers in Veterinary Science 9:980963.

Zhbannikov IY, Hunter SS, Foster JA, Settles ML (2017) SeqyClean. ACM 3107411:3107446.

Zheng Y, Chen KL, Zheng XM, Li HX, Wang GL (2014) Identification and bioinformatics analysis of microRNAs associated with stress and immune response in serum of heat-stressed and normal Holstein cows. Cell Stress and Chaperones 19:973-981.

Zinovieva NA, Dotsev AV, Sermyagin AA, Deniskova TE, Abdelmanova AS, Kharzinova VR, Sölkner J, et al. (2020) Selection signatures in two oldest Russian native cattle breeds revealed using high-density single nucleotide polymorphism analysis. PLoS One 15:242200.

CHAPTER 03 – SELECTION SIGNATURES IN DAIRY CATTLE: A SYSTEMATIC REVIEW

ABSTRACT - Dairy cattle have undergone both natural and artificial selection, which has shaped the genomes of current populations. Throughout this process, individuals under selection acquire genomic footprints called selection signatures. Identifying footprints of selection is helpful in the study of artificial selection, as it can provide insights into the genetic improvement of dairy breeds, particularly in identifying genes and beneficial mutations associated with adaptive and economically important traits in dairy production. This systematic review aimed to characterize research findings from 2016-2023 on selection signatures in dairy cattle. We analyzed 21 studies conducted in19 breeds (taurine, indicine, and composite), in which 5,099 genes were colocalized with selection signatures. This review examined the potential of such selection signatures to identify genes and beneficial mutations associated with traits under selection in dairy production, including fertility, health, milk production and composition, environmental adaptation, and color traits. We highlighted genes commonly found under selection, such as PLAG1. While DGAT1 is an important gene for dairy production that was found to be under selection in previous studies, it did not present evidence of selection within the reviewed literature. Additionally, we emphasized the relationship between copy number variations (CNV) and selection signatures and explored prospects related to molecular technology, integrative approaches, and new frontiers of research. Our review has demonstrated that dairy cattle selection is driven not only by economic traits but also by ecological pressures and desirable color patterns.
Keywords: artificial selection, genes, natural selection, selection pressure, selective sweep

1 INTRODUCTION

Dairy cattle production is one of the pillars of the agribusiness industry worldwide. The dairy production sector will continue to expand, supported by the growing demand for milk and dairy products. Efforts are being focused on optimizing milk production systems to improve productivity and quality, animal health, and feed efficiency in herds (OECD-FAO 2022) by selecting genetically superior animals for economic traits under selection. The natural and artificial selection process leaves footprints on the genome that can be detected through genetic analysis. Unique genetic patterns or marks in the genome of selected individuals are called selection signatures or selective sweeps (Fay and Wu 2000; Jensen et al. 2016).

Selection signature analyses can provide insights into the history and evolution of dairy cattle populations. The ancestors of modern cattle are the extinct aurochs (*Bos primigenius*), from which both taurine and indicine (or zebu) cattle have evolved (Pitt et al. 2019), resulting in the formation of more than 1,000 contemporary breeds (FAO 2015). Archaeological and genomic evidence suggests that taurine cattle (*Bos taurus*) were domesticated from *Bos primigenius primigenius* in the Fertile Crescent during the Neolithic period, more than 10,000 years ago, and are currently inhabiting the Persian Gulf region to the Nile River Valley (Bruford et al. 2003; Ajmone-Marsan et al. 2010; MacHugh et al. 2017). Approximately 1,500 years later, a second domestication event occurred in the Indus Valley, current regions of India and Pakistan, from *Bos primigenius nomadicus*, resulting in indicine cattle (*Bos indicus*) (Loftus et al. 1994). During domestication, animals were exposed to heterogeneous environmental pressures related to temperature, parasites, altitude, and water and food shortage.

In addition to revealing ancient selection signals, the motivation for investigating selection signatures is that these regions can pinpoint genes and mutations with large phenotypic effects regardless of whether they are currently segregating within a population. Variants that are no longer segregating can be detected using classical quantitative genetics methods, but this requires designing linkage experiments through specific crosses (Ramey et al. 2013; Utsunomiya et al. 2015). Given that genes and beneficial mutations are linked to adaptive and

economically important traits, selection signature studies provide valuable insights into the changes in and progress of artificial selection, aiding in genetic improvement in livestock populations (Cesarani et al. 2018). Single-nucleotide polymorphisms (SNP) harboring or located within a selection signature related to economic traits can serve as population-informative markers and, therefore, can be included in arrays focused on specific breeds (Schiavo et al. 2020). Additionally, selection signatures can be used to validate the results of genome-wide association studies (GWAS) for phenotypic variation (Gonzalez-Prendes et al. 2022).

Foundational studies on selection signatures have focused on taurine animals and milk yield traits (Gutiérrez-Gil et al. 2015; Randhawa et al. 2016). However, with the increasing demand for resilient and disease-resistant animals, studies have focused on health and environmental adaptation traits. Numerous studies, including those involving indicine and indigenous cattle breeds, have also been published due to the widespread use of SNP arrays and whole-genome sequencing (WGS). This review aimed to characterize research findings from 2016-2023 on selection signatures in dairy cattle, highlighting genes under selection that are related to economically important traits in the dairy industry, including fertility, health, milk production and composition, environmental adaptation, and color traits. Additionally, we addressed fundamental concepts of selection signatures, the molecular technologies employed to detect them and discussed prospects in this field.

2 MATERIAL AND METHODS

2.1 Literature search and inclusion criteria

To guide this systematic review, we selected studies between 2016 and 2023 that focused on selection signatures of dairy-purpose cattle, which were available in the Web of Science and Scopus databases. For the literature search, we used a combination of the terms "selection signatures", "selective sweep", "signatures of selection", “dairy”, “milk*”, “cattle”, “bovine”, and “cow”.

To select relevant scientific articles, we initially checked if the study contained any of the search terms in the abstract, in title or in keywords. After removing duplicate scientific articles, we applied the additional inclusion criteria: (1) peer-reviewed articles published in English; (2) the study used methods to identify

selection signatures within- or between-population; (3) the study used dairy purpose cattle as the experimental model; (4) availability of full-text articles. Articles meeting these criteria proceeded to the full-text review stage. In cases where articles conducted multiple analyses, including both dairy and non-dairy cattle, the results were screened to extract only those related to dairy-purpose cattle.

2.2 Exclusion criteria

Duplicates, review articles, and studies not published in English were not included. Additionally, analysis that mixed dairy-purpose cattle with dual or beef purposes and runs of homozygosity (ROH) results were not included in the review due to their multifactorial nature. Studies published before 2016 were not reviewed, as they were previously addressed in earlier meta-analyses and reviews on selection signatures in cattle (Gutiérrez-Gil et al. 2015; Randhawa et al. 2016).

2.3 Functional analysis

Although within and between-population approaches were considered, only genes identified in within-population analyses were compiled in this systematic review. Previously uncharacterized genes (ENSAGBTA) were checked for gene name updates via the gprofiler2 R package (Peterson et al. 2020). For the studies that did not display gene search results, the corresponding genomic windows were annotated via genes retrieved from the Ensembl Genes database (Ensembl Release 113, https://www.ensembl.org/, last accessed 28/01/2024) and the GALLO R package (Fonseca et al. 2020b). For studies that reported genomic windows in previous genome reference versions (i.e., UMD_3.1), the genomic coordinates were converted to ARS_UCD1.2 via the UCSC Lift Over tool (Raney et al. 2024) using default parameters.

3 RESULTS AND DISCUSSION

3.1 Studies and genes under selection

After the exclusion of duplicated and non-dairy selection signatures articles, 35 studies of selection signatures in dairy cattle were screened for this systematic review, including within- and between-population studies (Table 1).

Table 1. Details on selection signature studies in dairy cattle published from 2016 to 2023.

Author	Number of retrieved genes	Breed (number of animals)	DNA molecular technology*	Tests
Persichilli et al., 2023	Not applied	Italian Holstein (74), Parmigiano Reggiano Holstein (105), North American Holstein (101)	SNP array (Imputed 86K)	F_{ST}, wF_{ST}, XP-EHH, Rsb
Rajawat et al., 2023	1045	Gir (24), Hariana (10), Nelore (24), Ongole (20), Red Sindhi (10), Sahiwal (17), Tharparkar (72)	SNP array (50K)	iHS, ROH, Tajima's D, F_{ST}
Velayudhan et al., 2023	Not applied	Crossbred Indian+Exotic (37), Holstein-Friesian (50), Jersey (7)	SNP array (62K)	XP-EHH
Nayak et al., 2023a	292	Tharparkar (72), Vrindavani (72), Frieswal (14), Gir (24), Sahiwal (17), Red Sindhi (10), Holstein-Friesian (63), Jersey (28), Ongole (20)	SNP array (50K)	iHS, CLR, ROH, Tajima's D, F_{ST}, XP-EHH, Rsb
Velayudhan et al., 2023	Not applied	Not reported	SNP array (62K)	XP-EHH
Rocha et al., 2023	Not applied	Gir (2093)	SNP array (50K)	ROH, F_{ST}
Nayak et al., 2023b	Not applied	Tharparkar (72), Frieswal (14), Vrindavani (72), Gir (24), Sahiwal (17), Red Sindhi (10), Kankrej (10), Ongole (20), Hariana (10), Nelore (24), Holstein Friesian (63), Jersey (28), Ayrshire (18), Guernsey (21), Brown Swiss (24)	SNP array (50K)	ROH, F_{ST}, XP-EHH
Abbasi-Moshaii et al., 2022	Not applied	Holstein (266)	SNP array (50K)	XP-EHH
Ruvinsky et al., 2022	Not applied	Yaroslavl (22), Kholmogory (20), Yakut (20) and Holstein (20), Run 9 of 1000Bulls genome project (4779)	WGS (>6X)	wF_{ST}, HapFLK
Jin et al., 2022	1553	Dengchuan (10), Angus (9), Holstein (8)	WGS (8.2-23X)	ROH, nucleotide diversity (pi-

				ratio), CLR, XP-CLR, F_{ST}
Rajawat et al., 2022	Not applied	Tharparkar (72), Ayrshire (18), Brown Swiss (24), Gir (24), Guernsey (21), Hariana (10), Holstein Friesian (30), Jersey (28), Kankrej (10), Ongole (20), Red Sindhi (10), and Sahiwal (17)	SNP array (50K)	Tajima's D, Fu & Li's D*, CLR, ROH, iHS, F_{ST}, FLK, hapFLK
Tijjani et al., 2022	1239	Gir (10), Kenana (10), Butana (10), Aryashai (10), Baggara (10), Gash (10), Fulani (10), Ogaden (9), Kenyan Boran (10), Savannah Muturu (10), N'Dama (10), Ankole (10)	WGS (NA)	Hp, F_{ST}, XP-EHH
Gonzalez-Prendes et al., 2022	110	Holstein Friesian (5), Groningen White Headed (21), Meuse-Rhine-Yssel (23), Dutch Belted (23), Dutch Friesian (24), Deep Red (24)	WGS (7X to 13X)	CLR, ω-statistic
Illa et al., 2021	Not applied	Sahiwal (193)	SNP chip (53K)	F_{ST}, H1, H12, Tajima's D, nucleotide diversity
Gautason et al., 2021	6	Icelandic (8064)	SNP array (777K, 50K)	ROH, iHS
Schmidtmann et al., 2021	118	Red and White Dual-Purpose (744), Red Holstein (88), Meuse-Rhine-Yssel (292)	SNP array (777K, 50K)	iHS, XP-EHH
Saravanan et al., 2021	357	Tharparkar (24), Gir (24), Sahiwal (17), Red Sindhi (10), Kankrej (10), Ongole (20), Hariana (10), Holstein Friesian (30), Jersey (28), Ayrshire (18), Guernsey (21), Brown Swiss (24)	SNP array (50K)	ROH, iHS, F_{ST}
Liu et al., 2021	254	Holstein (1092)	Genotyping by genome reducing and sequencing (GGRS)	iHS, ROH
Lloret-Villas et al., 2021	25	Brown Swiss (161)	WGS (8.8-62.4X)	CLR
Singh et al., 2020	68	Vrindavani (96), Hariana (18), Holstein Friesian	SNP array (777K, 50K)	iHS, XP-EHH, F_{ST}

		(30), Jersey (21), Brown Swiss (24)		
Ghoreishifar et al., 2020	Not applied	Fjällnära (16), Fjäll (23), Swedish Holstein–Friesian (24), Swedish Red (25), Swedish Red Polled (18)	SNP array (150K)	DCMS with F_{ST}, H1, H12, Tajima's D, nucleotide diversity
Zinovieva et al., 2020	Not applied	Holstein (25), Yaroslavl (31), Kholmogor (25)	SNP array (777K)	ROH, F_{ST}, hapFLK
Huson et al., 2020	Not applied	Island Jersey (49), nonisland Jersey (46)	SNP array (777K)	F_{ST}, ROH
Moradian et al., 2020	9	Sarabi (20)	SNP array (777K)	iHS
Tijjani et al., 2019	882	Jersey (10), Holstein (10), N'Dama (10), Muturu (10)	WGS (2-45X)	iHS, Rsb
Refoyo-Martínez et al. 2019	3	Brown Swiss (4), Holstein (10), Jersey (9), Kerry Cattle (4)	WGS (NA) and SNP array (777K)	Sb
Mekonnen et al., 2019	Not applied	Sheko (12), East African Shorthorn Zebu (92), Ankole (25), Karamojong (16), Nganda (23), Serere (12), Fogera (11)	SNP array (777K)	iHS, CLR, Rsb
Bahbahani et al., 2018	53	Kenana (25), Butana (25), Small East African shorthorn zebu (92), Holstein Friesian (59), Jersey (32), N'Dama (24), Nelore (35), Sheko (18)	SNP array (777K)	iHS, Rsb
Zhang et al., 2018	Not applied	Holstein (32), Brown Swiss (20), Danish Red (45)	WGS (>~10X)	iHS, F_{ST}, ROH
Cheruiyot et al., 2018	52	Lushoto (485), Rungwe (346), Ndama (24), East African Shorthorn Zebu (50)	SNP array (150K, 777K)	iHS, XP-EHH, pcadapt
Signer-Hasler et al., 2017	Not applied	Brown Swiss (281), Braunvieh (3386), Original Braunvieh (167), Holstein (2568), Red Holstein (1960), Swiss Fleckvieh (547), Simmental (248), Eringer (36), Evolèner (21)	SNP array (50K, 777K)	F_{ST}
Taye et al., 2017	209	Holstein (10), Hanwoo (11), N'Dama (10)	WGS (average of ~11X)	Tajima's D, XP-CLR, XP-EHH
Iso-Touru et al., 2016	Not applied	Finnish Ayrshire (28), Western Finncattle (40),	SNP array (50K)	XP-EHH

		Nothern Finncattle (25), Eastern Finncattle (40), Yarovslavskaya (20), Busha (35), Kalmykian cattle (22), Podolian cattle (16), Ukrainian Grey (24), Yakutian (40)		
Khayatzadeh et al., 2016	81	Red Holstein-Friesian (101), Simmental (91), Swiss Fleckvieh (308)	SNP array (50K)	iHS, XP-EHH
Chen et al., 2016	37	Holstein (96), Simmental (447)	SNP array (777K)	LRH, XP-EHH, F_{ST}

*For SNP array, density is in brackets. For whole-genome sequencing (WGS), coverage is in brackets. Whether the information was not found in the article or supplementary material, not available (NA) was written.

In total, 5,001 genes and 98 miRNA were found to be under selection in 20 studies. Little overlap was observed among studies and breeds (not shown data), likely due to differences in methodology, marker density and position in SNP arrays, varying natural selection, selection objective depending on population purpose and history, and potential divergence among subpopulations of the same breed.

The number of genes found in studies that used WGS differed from those that relied on SNP arrays, which can be explained by the greater power of sequencing technology detection. Holstein cattle are the most studied breed (Table 1, Figure 1). Notably, 41.8% of the studies explored Holstein selection, leading to a greater number of retrieved genes. However, despite the relatively lower number of individuals with available genomic information, indigenous and locally adapted breeds are in the spotlight owing to their dairy productivity and unique ability to cope with environmental pressures, such as diseases, heat, and cold stress. The most popular tests for selection signatures between-populations were F_{ST} (19 studies) and XP-EHH (14 studies). For the within-population tests, iHS was the most popular (16 studies) (Table 1).

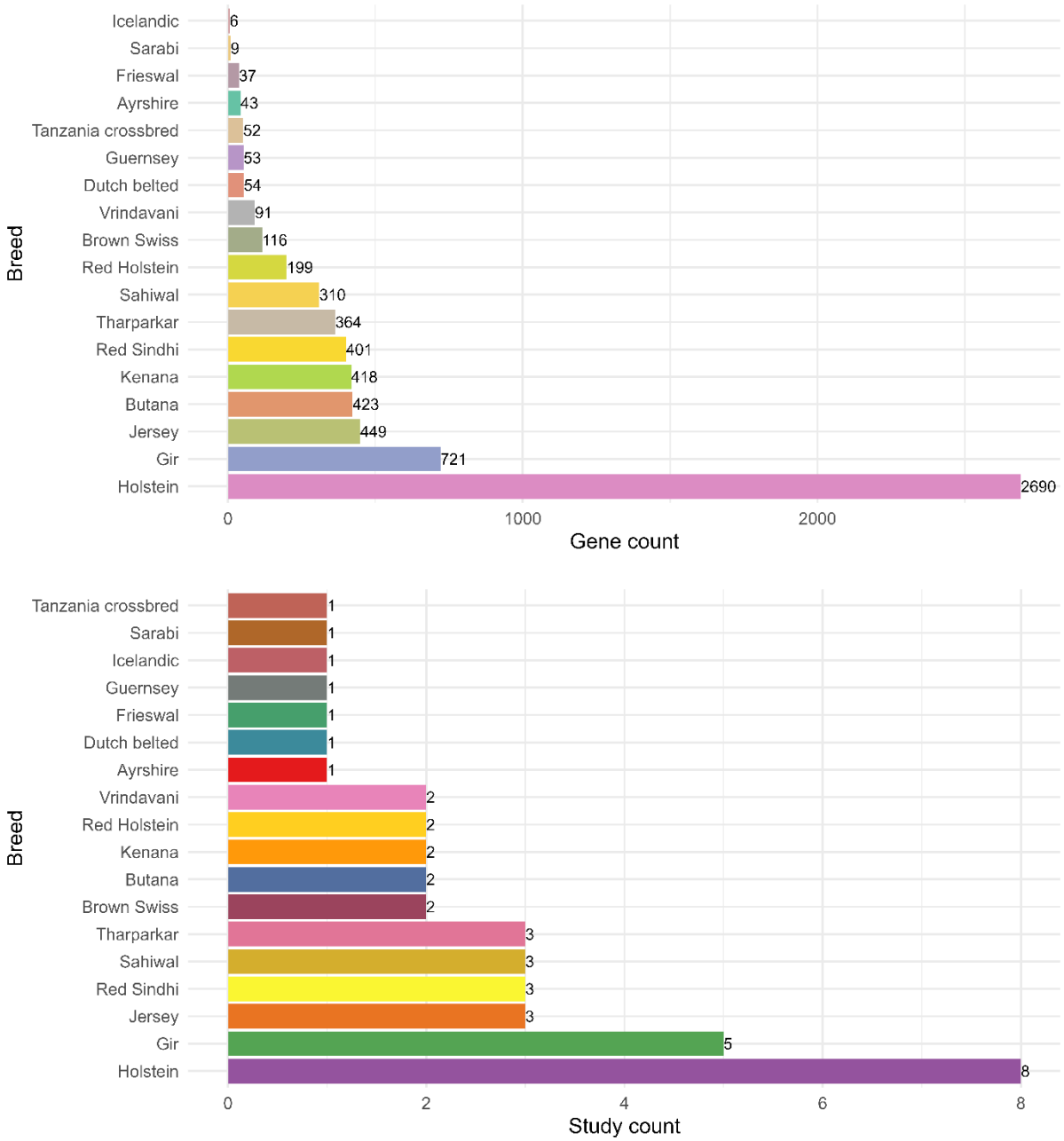

Figure 1. (A) Number of genes under selection per breed identified by within-population tests in studies published from 2016 to 2023 for dairy-purpose cattle. (B) Number of studies for each breed, highlighting that the most studied breeds are Holstein and Gir. Molecular technology and the number of studies, animals, and tests impact the number of genes identified under selection.

Selection signature analysis provides valuable insights into the genetic basis of complex traits selected in dairy cattle breeding worldwide. The following sections were split into specific traits, such as fertility, health, milk production and composition,

environmental adaptation, color patterns, and potential pleiotropic genes for better clarity and understanding.

3.2 Fertility

The identification of selection signatures provides insights into the genetic architecture of fertility traits in dairy cattle. Rocha et al. (2023) identified selection signatures in two groups of Dairy Gir cattle, which differed in their breeding values for the total number of oocytes, viable oocytes, and number of embryos. Using the F_{ST} approach, these authors identified selection signals covering transcription factors and genes that could contribute to the performance of these traits. For instance, the *KNTC1* gene was related to ovarian dysfunction in mice (Kaune et al. 2022), and the *ABL2* gene was differentially co-expressed between highly fertile and sub fertile cows (Fonseca et al. 2020a).

Reproductive hormones, such as choriogonadotropin, luteinizing hormone, and estrogens, play a core role in reproductive processes, and artificial selection pressure may have left patterns in genes related to steroid hormones and steroidogenesis. Velayudhan et al. (2023b) identified several genes involved in the synthesis of steroids and encoding the receptor of choriogonadotropin and luteinizing hormones via the XP-EHH test in crossbreds (Indian native dairy cows with exotic milk breeds) and Holstein-Friesian cattle, such as *CYP17A1* (Singh et al. 2022) and *LHCGR* (Yuan et al. 2017).

In the dairy cattle industry, it is widely recognized that strong directional selection for milk production has negatively impacted fertility (Misztal and Lourenco 2024), resulting in effects not only on the phenotypic average of herds but also at the DNA level. Huson et al. (2020) applied the F_{ST} test to compare genetic differentiation between Jersey from the island and non-island Jersey (from the United States, Canada, New Zealand, and Denmark). The strongest signal (0.74) harboured *DSC1*, *DSC2*, and *DSC3,* which are members of the desmocolin gene family. The *DSC2* gene has been associated with several fertility traits in the literature, such as age at first calving, productive life, and cow, daughter, and heifer conception rate (Cochran et al. 2013; Ortega et al. 2016; Zuleica Trujano-Chavez et al. 2022; Gonzalez Berrios

et al. 2024). These findings illustrate the effect of distinct selective objects and pressure on subpopulations of the same breed.

However, as they share a common genetic background, different populations of the same breed are also expected to exhibit similar genomic patterns. For example, Persichilli et al. (2023) found selection signatures in common with three subpopulations of Holstein cattle breed (i.e., individuals from the United States and Canada, ii. Italian Holstein cattle not involved in Parmigiano Reggiano cheese production, and iii) Italian Holstein cattle used for Parmigiano Reggiano cheese production) using F_{ST}. The overlapping genes under selection between Italian involved or not Parmigiano Reggiano cheese production and non-Italian Holsteins are related to fertility, such as *LEO1*, *FER*, and *ZFPM1*. The *LEO1* gene was related to oocyte and sperm production (Chu et al. 2012; Lillepea et al. 2024). The *FER* gene encodes the FER protein, a tyrosine-protein kinase involved in the process of sperm capacitation (sperm ability to fertilize) (Alvau et al. 2016; Bathala et al. 2018). The *ZFPM1* gene has been associated with dizygotic twin births (Mbarek et al. 2024).

Understanding the genetic basis of fertility can help address the challenges faced in dairy cattle production, such as reduced fertility and decreased reproductive efficiency, which lead to economic losses for dairy farmers. However, studies on gene expression are needed to validate these findings. Furthermore, weighted single step GWAS (WssGWAS) and genomic selection methodologies that account for varying weights of selected variants in specific regions should enhance our understanding of how these regions impact the genetic variation in fertility traits. This will enable more informed decisions for more effective breeding in dairy cattle.

3.3 Health

The ability of a cattle population to cope with specific diseases can be identified in the genome through selection signature approaches. Health and immunity traits are essential for improving disease resistance and overall animal health in dairy cattle, leading to increased milk production and profitability.

Mastitis is an important disease affecting dairy cattle worldwide, resulting in substantial economic losses to the dairy industry, and selection signatures may contribute to understanding the genomic architecture underlying this disease.

Persichilli et al. (2023) studied selection signatures between Italian and North American Holsteins using F_{ST} test and identified the *MAPK6* gene, which is related to mastitis (Ghahramani et al. 2021).

Specifically, experimental designs focused on identifying selection signatures associated with mastitis resistance and susceptibility can aid in the development of breeding strategies to improve mastitis resistance in dairy cattle by pinpointing regions involved in this disease. For instance, Abbasi-Moshaii et al. (2023) performed a genomic scan for selection signatures related to mastitis via the XP-EHH approach on German Holstein cows with one or more records for clinical or subclinical mastitis (sick group) and without any entry (healthy group) in their first lactation. The miRNA *MIR455* gene, identified in the healthy group, was downregulated in infected mammary gland tissue in another study (Li et al. 2015). These findings suggest that *MIR455* may be a potential candidate for mastitis resistance.

Natural selection may have shaped immune-related and disease-resistance genes in indicine and indigenous cattle breeds. Comparing these populations with high-productivity dairy breeds through between-populations analysis may reveal their specific adaptations and evolutionary history. In this section, we provide three cases as examples of this narrative.

Nayak et al. (2023a) studied selection signatures in the major histocompatibility complex (MHC) region for indicine (Gir, Hariana, Kankrej, Nelore, Ongole, Red Sindhi, Sahiwal, and Tharparkar), taurine (Holstein Friesian, Jersey, Ayrshire, Guernsey, and Brown Swiss), and composite cattle (Frieswal and Vrindavani). A selection signature involving the *BOLA* and *MIC1* genes in Red Sindhi and Jersey was reported using the F_{ST} test. *BOLA* belongs to the bovine MHC class II gene family, which mediates disease resistance (Behl et al. 2012). The MHC is important for initiating the adaptive immune response; class I and class II molecules encoded by MHC genes communicate infection by presenting peptides derived from pathogens to receptors CD8+ and CD4+ on T-cells (Behl et al. 2012). *MIC1* is the macrophage inhibitory cytokine 1 gene, and macrophages are part of the inflammatory response.

Interleukin family genes are known for their involvement in the immune system and inflammatory processes. Iso-Touru et al. (2016) studied selection signatures

between Finnish Ayshire and rare and endangered Eurasian taurine breeds. These authors found a selection signature between Finnish Ayshire and native dairy breeds (Yarovslavskaya, Eastern, and Northern Finncattle) encompassing several genes from the interleukin family on chromosome 16 via the XP-EHH test. Using de-correlated composite of multiple signals (DCMS), Ghoreishifar et al. (2020a) identified several genes under selection associated with host resistance to gastrointestinal nematodes, including *CYP4F2* and *TMEM37* (Li et al. 2011), in Fjäll and Swedish Red Polled breeds, respectively. The northern cattle were initially raised in woodlands in the summer and may have adapted to their local environment and its challenges.

Bahbahani et al. (2018) reported genes under selection related to trypanotolerance in African cattle from Sudan. The authors analyzed the selection signatures between the trypanotolerant breeds Butana and Kenana and Holstein and Jersey breeds (Butana_Holstein+Jersey, Kenana_Holstein+Jersey) via the Rsb test. The authors identified the *IRAK3* (also known as *IRAK-M*) gene under selection, which is related to the activation of innate immunity following Trypanosoma infection (Campos and Gazzinelli 2004). The IRAK3 protein plays an essential role in initiating the innate immune response as a part of the signaling cascade of Toll-like receptors and interleukin-1 receptors (Pereira and Gazzinelli 2023).

Within-population selection signature may also reveal evolutionary history of indigenous breeds. Gautason et al. (2021) identified the *TRAM2* gene as being under selection in Icelandic dairy cattle via iHS test. This gene was previously associated with leukemia virus susceptibility (Carignano et al. 2018). This disease has never been reported in Iceland. Thus, the authors hypothesize that selection for the *TRAM2* gene may have led to the extinction of the virus in the country or that selection for the gene occurred prior to the migration of the breed to Iceland. However, further research is needed to elucidate the exact mechanism behind this identified selection pattern.

3.4 Milk production and components

Milk and its components are major economic factors in dairy cattle production. Identifying selection signatures related to these traits can help pinpoint causal

regions under selection in dairy purpose breeds. In many countries, milk fat is a major selection criterion due to the profitability of fat (Mariani et al. 2022).

Holstein cattle are the most popular breed worldwide because of their high productivity and strong selection pressure, which have left remarkable footprints in the breed genome. Using the long-range haplotype (LRH) (Sabeti et al. 2002) test, Chen et al. (2016) reported a selection signature encompassing the *MGST1* gene in Holstein cattle. The *MGST1* gene has been previously associated with milk fat production in QTL mapping studies (Littlejohn et al. 2016) and GWAS (Raven et al. 2014; Cruz et al. 2019). On one hand, different tests can be complementary, as they allow the detection of different patterns in the genome, even in different subpopulations of the same breed. On the other hand, the same test can spot similar selection signatures in different subpopulations. For instance, genes involved in the lactation pathway, such as *GHR* (*growth hormone receptor*) and *GHRHR* (*growth hormone-releasing hormone receptor*), were found to be under selection in Holstein cattle by Tijjani et al. (2019) and Liu et al. (2021) via the iHS test. *GHR* encodes a transmembrane receptor for growth hormone and affects milk yield, fat, and protein components (Blottet al. 2003; Waters et al. 2011).

In Red Holsteins, Schmidtmann et al. (2021) reported a set of selection signatures on chromosome 18 between 16.5 and 33.5 Mb related to milk production traits. One of the strongest signals observed by the authors was located near *FTO* gene, which was previously linked to milk protein and fat yield in Holstein cattle (Zielke et al. 2013).

The major protein in milk is casein; it accounts for approximately 80% of the total protein, and its quality and abundance are essential for dairy products. The casein group is mainly composed of four unique proteins, including α_{s1}, α_{s2}, β, and κ-casein (Rajput and Sharma 2023). Taye et al. (2017) studied selection signatures in Holstein, Hanwoo, and N'Dama cattle. The authors identified selection signatures via Tajima's D in Holstein cattle spanning the *CSN1*, *CSN2*, and *CSN3* genes from the casein family. *CSN3* regulates the milk kappa casein profile in cattle (Martin et al. 2002; Caroli et al. 2009; Amalfitano et al. 2022). Additionally, the authors reported that the SNP rs43703015 and rs43703016 in the *CSN3* gene were exclusive to the

Holstein population. Therefore, the authors hypothesized that these signals might be related to dairy specialization.

Over the genomic era, commercial breeds have benefited from relevant genetic improvements in milk production. In contrast, indicine and indigenous populations worldwide still face challenges in this regard due to their relatively recent artificial selection compared with the most popular dairy breeds. However, it is possible to detect selection signatures related to several aspects of dairy production in indicine and indigenous populations, such as milk yield, composition, and lactation persistency.

Illa et al. (2021) identified selection signatures encompassing several genes related to milk production and composition in Sahiwal cattle. These authors reported genes possibly related to fat milk content (*SLC25A1*, *HADH*, *CYP2U1*, and *SGMS2*), milk yield (*GHR* and *ADGRL4*), the mammary gland (*CCL28*), and lactation persistency (*HMGCS1*). Compared with Holstein cattle, Sahiwal cattle have a slightly more desirable profile of fatty acid composition (Sharma et al. 2018). The findings highlight the potential of native breeds and their crosses for the dairy industry.

Nayak et al. (2023b) identified several genes related to milk production and fat and protein content in indicine (Tharparkar, Gir, Sahiwal, Red Sindhi), crossbred (Vrindavani, Frieswal) and taurine (Holstein–Friesian, Jersey) dairy breeds via multiple within- and between-population tests. The authors highlighted the *EPHA6* gene identified by iHS, XP-EHH, Rsb, and F_{ST} in Holstein–Friesian cattle, which is related to milk fat production (Liu et al. 2020; Laodim et al. 2023). Rajawat et al. (2023) studied selection signatures in indicine (Gir, Hariana, Nelore, Ongole, Red Sindhi, Tharparkar, and Sahiwal) and dairy taurine breeds (Holstein Friesian, Jersey, Brown Swiss, and Ayrshire). The author found that the gene CACNG2 was under selection via the F_{ST} test in Gir and dairy taurine animals. This gene was previously associated with fat percentage in Holsteins (Kolbehdari et al. 2009).

Cheruiyot et al. (2018) applied the iHS test and identified several genes under selection related to lactation persistence and milk yield, such as *ABCG2* (Cohen-Zinder et al. 2005; Do et al. 2017), *HERC5* (Do et al. 2017; Pedrosa et al. 2021), and *HERC6* (Cohen-Zinder et al. 2005; Do et al. 2017), in Tanzanian crossbred cattle. The Tanzanian crossbreeds are the product of crossing local East African Shorthorn

Zebu with imported exotic dairy breeds (mainly Holstein, Friesian, and Ayrshire). This practice started during the colonial era and was further encouraged by government initiatives in the 1980s (Mwenya 1993). Rajawat et al. (2022) also reported that *HERC5* and *HERC6* were under selection in Gir via Tajima's D test.

Although *DGAT1* is a gene of major interest in milk production and has been mentioned under selection in a past study in dairy cows (Qanbari et al. 2010), this gene did not appear in any selection signal in the analyzed literature. *DGAT1* is part of the Diacylglycerol acyltransferases (DGAT) family, which are key enzymes in lipid storage. The *K232A* allele has a widely studied influence on milk production traits, in which K and A allele frequency varies among breeds and localities (Da Silva et al. 2010; Parra-Bracamonte et al. 2025). However, Esfahani et al. (2025) reported 146 missense mutations in the bovine *DGAT1* gene that could also affect milk production but need to be further studied. While it has been a focus in dairy systems, we can hypothesize several reasons for *DGAT1* was not shown to be under selection, including: i) the selection pressure on *DGAT1* alleles is relatively recent and it has not yet left a detectable pattern in a region of the genome, as the first studies on DGAT1 were published in 2002 (Grisart et al. 2002) or *DGAT1* is not under directional selection; ii) there are differences among subpopulations, such as genomic differences within DGAT1 and the genotyped population is not currently under selection, especially in studies that used the WIDDE database (http://widde.toulouse.inra.fr/widde/); iv) the current tests may not effectively capture selection signals within this gene.

3.5 Environmental adaptation

Thermotolerance is an important physiological mechanism that can influence milk production in cattle. Heat stress impacts the expression of gluconeogenic genes, which reduces the synthesis of hepatic glucose and the levels of nonesterified fatty acids in the blood. These changes disrupt lactose synthesis in the mammary gland, resulting in a decrease in milk yield (Wheelock et al. 2010; Rhoads et al. 2011).

Due to climate change, indicine and indigenous breeds, as well as their crossbreeds, are valuable genetic resources and can be used to establish breeding strategies for developing resilient and productive dairy cattle populations in tropical

and subtropical regions. Knowledge of selection signatures helps in understanding the underlying mechanism of heat tolerance and disease resistance, which are the main adaptable traits of indigenous cattle breeds (Saravanan et al. 2021). The genetic architecture of heat tolerance indicators via selection signature studies in indicine and indigenous cattle can pinpoint candidate genes related to environmental adaptability. This highlights the potential for conservation and genetic improvement in these breeds. In addition, genes under selection related to heat stress can be used in marker-enhanced selection for poorly thermotolerant cattle.

The most recognized biological markers of thermal stress responses are heat shock proteins (HSP) and heat shock factors (HSF), the genes responsible for their activation and expression. Saravanan et al. (2021) detected selection signatures harboring genes related to thermal stress in indicine breeds using iHS test. The authors identified the *HSPA1L* gene in Gir cattle and the *HSPA1B* gene in Tharparkar and Gir breeds.

Members of the DNAJ family serve as cofactors for heat shock protein 70 (Hsp70). Illa et al. (2021) identified selection signatures in Sawihal cattle harboring the *DNAJB4* gene. Liu et al. (2021) identified another member of this family, *DNAJB5*, via iHS in Shangai Holsteins. The investigated Holstein population was exposed to a subtropical climate with long, hot, and humid summers and conditions of thermal discomfort. Tijjani et al. (2022) identified heat shock protein genes (*HSPA9* and *DNAJC18*) via the pooled heterozygosity (Hp) test (Lewontin and Krakauer 1973) in Kenana and Dairy Gir cattle.

Heat stress activates HSF genes, which are responsible for the transcription of heat shock protein genes (Wu 1995). Bahbahani et al. (2018) identified the *HSF5* gene under selection via a pairwise Rsb test between the Butana breed and European taurine dairy cattle (Holstein–Friesian and Jersey, combined).

Extremely arid conditions may leave specific patterns in the genome related to survival and drought tolerance. Tijjani et al. (2022) used the pooled heterozygosity (Hp), F_{ST}, and XP-EHH approaches to identify selection signatures related to adaptation to Sudan desert and semi-desertic regions in African Sudanese indicine and other indicine breeds. The authors identified selection signatures in the Kenana and Butana breeds, including genes involved in glucose homeostasis and insulin

signaling (*FAAP20*, *PRKCZ*, *SKI*, and *PLCH2*). These biological processes contribute to animal survival during nutritional shortages in deserts, where periods of dietary restriction increase insulin sensitivity and glucose absorption (Tijjani et al. 2022).

Given their distinct subspecies and selection history, differences in selection signatures between taurine and indicine animals are expected. Rajawat et al. (2022) conducted a study to detect selection signatures in seven indicine breeds (dairy and dual purpose) and five taurine dairy breeds. They combined different approaches in the DCMS framework (Ma et al. 2015). Interestingly, they found selection signatures encompassing the *ITPR2* gene via the F_{ST} approach in the Jersey-Tharparkar pair. This finding may indicate differences in the selection for heat tolerance between taurine and indicine cattle, as *ITPR2* is related to heat stress (Sigdel et al. 2019; Cheruiyot et al. 2021).

In the matter of environmental adaptation, selection signatures can reveal the direction of selection in composite breeds. Singh et al. (2020) employed iHS, XP-EHH, and F_{ST} to study selection signatures in Vrindavani, an Indian composite dairy cattle breed, and its parental breeds (Hariana, Holstein Friesian, Jersey, and Brown Swiss). Despite milk yield being the main goal of artificial selection in the Vrindavani breed, their findings imply that selection in this breed was more pronounced in genomic regions related to environmental adaptation. This finding indicates that the selection pressure in this composite breed was more influenced by environmental factors, such as heat stress, rather than only by selective breeding for milk production.

Extreme cold climates and limited forage availability are the main challenges for dairy cattle breeds in arctic regions. These abilities can be studied at the genomic level through selection signatures that can identify causal genomic regions for adaptation to harsh environments. Ruvinsky et al. (2022) applied the F_{ST} and HapFLK tests to detect selection signatures in Yaroslavl cattle, a native Russian dairy breed known for its small stature, low productivity, and ability to cope with cold climates and lack of forage. The authors compared this breed with two related breeds (Russian Kholmogory and Holstein) and the taurine set of the 1000 Bull Genomes Project. Their analysis revealed a haplotype at the BTA28:30056005-30201884 position that encompassed the *MSS51* and *KAT6B* genes, which were previously

linked with skeletal muscle (Moyer and Wagner 2015) and leg morphology (Van Den Berg et al. 2014), respectively. This haplotype was associated with lower weight and height in adult life in the Yaroslavl breed in the study by Ruvinsky et al. (2022), which may have been historically beneficial under harsh environmental conditions and a low feeding base in Russia.

Furthermore, the haplotype was present at a low frequency in the two other breeds, Modern Angler and Lithuanian Red, which are also adapted to cold climates. High-altitude environments require that their inhabitants adapt to low levels of oxygen. In a Swedish mountain breed, Fjäll, Ghoreishifar et al. (2020a) reported genes related to hypoxia and high-altitude adaptation, such as *PPP1R12A* (or *MYPT1*) (Lopez et al. 2016) and *SLC16A3* (Goyal et al. 2013).

Artificial selection is not the only anthropogenic action that impacts genome patterns; pollution can also impact dairy cattle. Velayudhan et al. (2023a) studied selection signatures among dairy cattle populations in India, considering their exposure to heavy metals (cadmium and lead) and their geographical locations (rural, urban, and a combination of urban and rural areas). For heavy metal exposure analysis, the authors identified genes related to stress response and inflammatory response, such as *OPDR1*, *KDR* and *ADORA2A*.

3.6 Color patterns

Regions controlling eye pigmentation are under selection in cattle breeds. A lack of ambilateral circumocular pigmentation (pigmentation surrounding the eyes) is a possible etiology for bovine ocular squamous cell carcinoma (eye cancer) (Pazhanivel and Umarani 2022). Illa et al. (2021) reported a selection signature in the Sawihal breed harboring the *LEF1* gene, and Rajawat et al. (2023) reported a selection signature encompassing the *CRIM1* gene in Gir cattle, both of which display ambilateral circumocular pigmentation. *LEF1* and *CRIM1* were related to ambilateral circumocular pigmentation in Fleckvieh cattle (Pausch et al. 2012). Gonzalez-Prendes et al. (2022) performed a GWAS for ambilateral circumocular pigmentation and complemented this approach with signature analysis using various white- and nonwhite-headed breeds. The regions with the most significant associations obtained by both methods were found on BTA3 (115–118 Mb) in Dutch

Belted cattle and BTA3 (12–13 Mb), BTA11 (93–94 Mb), and BTA22 (45–48 Mb) in Holstein Friesian cattle, which are both nonwhite-headed breeds. These genes and regions might be especially important for the selective breeding of animals that have increased susceptibility to eye tumors, i.e., European taurine cattle breeds (Pazhanivel and Umarani 2022).

Well-studied genes (e.g., *MC1R*, *KIT*, *ASIP* (Girardot et al. 2006), and *TYRP1)* (Berryere et al. 2003)) are targeted in cattle for the investigation of coat color. The *KIT* gene was found to be under selection in Yaroslavl, Kholmogory, and Holstein cattle via the F_{ST} and HapFLK tests (Zinovieva et al. 2020; Ruvinskiy et al. 2022). All three breeds exhibit black-and-white patterns. Rajawat et al. (2022) reported several genes related to red and brown coat color, such as *MC1R* via the VarLD test in Sahiwal cattle and *ASIP* and *TYRP1* in Red Sindhi, both of which are reddish breeds. Jin et al. (2022) identified the *MC1R* gene under selection in Holstein cattle via the CLR test, and Rajawat et al. (2023) identified the same gene in Sahiwal cattle via Tajima's D test. A selection signature encompassing the *PMEL* gene was found in Butana cattle in a between-population analysis with Kenana (Bahbahani et al. 2018). Kenana cattle are born with red coats that turn grey as they mature. Butana cattle are distinguished by their deep red coat color, and *PMEL* may be connected to this color (Bahbahani et al. 2018).

Coat color is a prominent phenotypic trait in cattle and other species and is frequently influenced by both natural and artificial selection pressures. Its importance spans practical factors, such as adaptation to environmental conditions and sociocultural considerations, including breeders' preferences for specific aesthetic types. Compared with other quantitative traits, such as milk yield or milk fat percentage, the genetic architecture of coat color is generally less complex. However, dominance and epistatic interactions may still contribute to phenotypic variation, especially for crossed breeds. It is expected that advances in selection signatures with sequencing data and omics approaches will enable a more detailed understanding of the genetic basis of coat color in cattle, facilitating the identification, functional validation, and characterization of novel genetic variants and their interactions.

3.7 Possible pleiotropic genes under selection

Several studies using taurine and composite breeds have identified selection signatures encompassing the *PLAG1* gene. For example, this gene was identified under selection within Chinese Holstein cattle via the CLR test (Jin et al. 2022), Swedish cattle (Ghoreishifar et al. 2020a), Tanzanian crossbred cattle via the iHS test (Cheruiyot et al. 2018), and between Holstein and Dengchuan cattle via the F_{ST} (Jin et al. 2022), and several Swiss dairy breeds (Brown Swiss, Red Holstein, Holstein, and Braunvieh) via the F_{ST} (Signer-Hasler et al. 2017).

The *PLAG1* gene is a widely recognized gene with pleiotropic effects in cattle. In Holstein cattle, the *PLAG1* gene was previously associated with milk composition traits (Fink et al. 2017), stature (Karim et al. 2011; Littlejohn et al. 2012; Bouwman et al. 2018), growth traits (Littlejohn et al. 2012; Fink et al. 2017) and heifer livability (Gao et al. 2023). A mutation in the *PLAG1* gene was found to be introgressed in indicine cattle (Utsunomiya et al. 2017), and in addition to regulating growth traits and stature, it may also regulate puberty. The *PLAG1* gene was previously related to birth weight (Utsunomiya et al. 2013a, 2017) and heifer fertility in Nelore cattle (Mota et al. 2020, 2022) and puberty in Brahman cattle (Hawken et al. 2012; Fortes et al. 2013; Engle and Hayes 2022).

Another possible pleiotropic gene, *HMGA2*, was identified in a selection signature in Kenana, Butana, and Dairy Gir cattle by Tijjani et al. (2022). This gene is known to be involved in growth in mammals (Habib et al. 2018) and was previously linked to ovarian length and height in a various cattle breed (Shen et al. 2024) and navel length in Nelore cattle (Aguiar et al. 2018). Aguiar et al. (2018) reported a copy number gain in this gene that is indicine-specific and has ancient origins. Together, these studies provide evidence for the important role of *HMGA2* in the evolution and adaptation of indicine breeds.

Genes affecting mammary gland tissue are important not only for milk production but also for mastitis. Zhang et al. (2018) studied the modern Danish Red Dairy Cattle employing the iHS test and identity-by-descent approach to identify selection signatures resulting from the strategic crossbreeding of traditional Danish Red Dairy Cattle with high-yielding breeds such as Holstein and Brown Swiss. The authors identified regions under selection that were introgressed from the Brown

Swiss flaking *LAP3* gene. Polymorphisms in this gene have been related to milk production traits and mastitis resistance (Zheng et al. 2011; Worku et al. 2022, 2023). Jin et al. (2022) conducted a selection signature analysis in Holstein, Angus, and Dengchuan, an endangered yellow cattle breed in China. The authors identified the *RERE* gene under selection via the use of a π-ratio and CLR. The *RERE* gene was reported to be associated with milk fat percentage (Buaban et al. 2022) and was differentially expressed between healthy cows and cows with mastitis (Asselstine et al. 2022). The RERE protein regulates apoptosis and cell proliferation and differentiation (Waerner et al. 2001; Wang et al. 2017), which may explain the role of the *RERE* gene in multiple traits related to the mammary gland.

4 PROSPECTS

4.1 Molecular technology

With the ongoing reduction in the cost of long-read sequencing, this technology has been used more widely in selection signature studies. However, underrepresentation in SNP arrays of markers from local and indicine breeds persists as most of these breeds are raised in developing and sub-developed countries. Sequencing usage allows for more accurate identification of genetic variants, their extensions, and associated functional effects, providing insights into the genetic basis of traits in dairy cattle. Additionally, the use of pangenomes can provide a better scenario for between-population studies.

4.2 Integrative approaches

The integration of selection signatures to complement GWAS should be further explored. The validation of results through functional genomics is necessary but may face challenges in livestock due to practical complexities, high costs, and the management of animal populations. Combining GWAS and selection signature approaches will provide a comprehensive understanding of the genetic basis of complex traits in dairy cattle, enabling the identification of genes and genomic regions under selection linked to important traits for the dairy industry.

The design of experimental studies influences the creation of a robust database for exploring polygenic traits. Joint studies on CNV and selection

signatures require further exploration. CNV are an important source of genetic variation and have been associated with several simple and complex genetic traits in cattle. However, their relationship with selection signatures remains poorly understood and requires further investigation.

4.3 Future hot topics

Due to climate change, studies about the selection process in indigenous and locally adapted dairy cattle will continue to be a topic of interest. Understanding how these cattle have adapted to their local environments and identifying the favored alleles that have enabled them to produce in challenging conditions may help develop strategies to mitigate the impacts of climate change on dairy production.

6 CONCLUSIONS

This review provided a comprehensive overview of the current scenario of selection signatures in dairy cattle studies, including the selection of the most relevant traits for milk production worldwide and specific adaptive and disease resilience traits of dairy breeds. We highlighted important genes found to be under selection in taurine, indicine, and composite dairy cattle breeds. However, there is a lack of functional validation of candidate genes; these studies will enhance understanding in multiple fields of animal breeding and genetics. An alternative that can provide more information is the integration of GWAS and selection signature studies, as this may lead to gains in marker-enhanced selection and functional understanding of complex traits.

Dairy cattle populations have not been selected only for traits that are present in the selection indices; rather, their genomes are also shaped by environmental pressures and the selection of desirable exterior traits, such as human preference regarding coat color. As selection indices evolve to incorporate sustainable traits such as feed efficiency and methane emission, the identification of selection signatures linked with these traits is encouraged in future generations.

7 REFERENCES

Abbasi-Moshaii, B, Moradi, MH, Yin, T, Rahimi-Mianji, G, Nejati-Javaremi, A & König, S. 2023. Genome-wide scan for selective sweeps identifies novel loci associated with resistance to mastitis in German Holstein cattle. Journal of Animal Breeding and Genetics. 140(1):92–105. doi.org/10.1111/JBG.12737.

Aguiar, TS, Torrecilha, RBP, Milanesi, M, Utsunomiya, ATH, Trigo, BB, Tijjani, A, Musa, HH, Lopes, FL, et al. 2018. Association of Copy Number Variation at Intron 3 of HMGA2 With Navel Length in Bos indicus. Frontiers in Genetics. 9:627. doi.org/10.3389/FGENE.2018.00627.

Ajmone-Marsan, P, Garcia, JF & Lenstra, JA. 2010. On the origin of cattle: How aurochs became cattle and colonized the world. Evolutionary Anthropology. 19(4):148–157. doi.org/10.1002/evan.20267.

Alvau, A, Battistone, MA, Gervasi, MG, Navarrete, FA, Xu, X, Sánchez-Cárdenas, C, de la Vega-Beltran, JL, da Ros, VG, et al. 2016. The tyrosine kinase FER is responsible for the capacitationassociated increase in tyrosine phosphorylation in murine sperm. Development (Cambridge). 143(13):2325–2333. doi.org/10.1242/DEV.136499/263999/AM/THE-TYROSINE-KINASE-FER-IS-RESPONSIBLE-FOR-THE.

Amalfitano, N, Macedo Mota, LF, Rosa, GJM, Cecchinato, A & Bittante, G. 2022. Role of CSN2, CSN3, and BLG genes and the polygenic background in the cattle milk protein profile. Journal of Dairy Science. 105(7):6001–6020. doi.org/10.3168/JDS.2021-21421.

Asselstine, V, Medrano, JF & Cánovas, A. 2022. Identification of novel alternative splicing associated with mastitis disease in Holstein dairy cows using large gap read mapping. BMC Genomics. 23(1):1–15. doi.org/10.1186/S12864-022-08430-X/FIGURES/5.

Bahbahani, H, Salim, B, Almathen, F, Enezi, F Al, Mwacharo, JM & Hanotte, O. 2018. Signatures of positive selection in African Butana and Kenana dairy zebu cattle. PLOS ONE. 13(1):e0190446. doi.org/10.1371/JOURNAL.PONE.0190446.

Bathala, P, Fereshteh, Z, Li, K, Al-Dossary, AA, Galileo, DS & Martin-DeLeon, PA. 2018. Oviductal extracellular vesicles (oviductosomes, OVS) are conserved in humans: murine OVS play a pivotal role in sperm capacitation and fertility. Molecular Human Reproduction. 24(3):143–157. doi.org/10.1093/MOLEHR/GAY003.

Behl, JD, Verma, NK, Tyagi, N, Mishra, P, Behl, R & Joshi, BK. 2012. The Major Histocompatibility Complex in Bovines: A Review. International Scholarly Research Notices. 2012:1–12. doi.org/10.5402/2012/872710.

Van Den Berg, I, Fritz, S, Rodriguez, S, Rocha, D, Boussaha, M, Lund, MS & Boichard, D. 2014. Concordance analysis for QTL detection in dairy cattle: A case

study of leg morphology. Genetics Selection Evolution. 46(1):1–14. doi.org/10.1186/1297-9686-46-31.

Berryere, TG, Schmutz, SM, Schimpf, RJ, Cowan, CM & Potter, J. 2003. TYRP1 is associated with dun coat colour in Dexter cattle or how now brown cow? Animal Genetics. 34(3):169–175. doi.org/10.1046/J.1365-2052.2003.00985.X.

Blott, S, Kim, JJ, Moisio, S, Schmidt-Küntzel, A, Cornet, A, Berzi, P, Cambisano, N, Ford, C, et al. 2003. Molecular Dissection of a Quantitative Trait Locus: A Phenylalanine-to-Tyrosine Substitution in the Transmembrane Domain of the Bovine Growth Hormone Receptor Is Associated With a Major Effect on Milk Yield and Composition. Genetics. 163(1):253–266. doi.org/10.1093/GENETICS/163.1.253.

Bouwman, AC, Daetwyler, HD, Chamberlain, AJ, Ponce, CH, Sargolzaei, M, Schenkel, FS, Sahana, G, Govignon-Gion, A, et al. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nature Genetics 2018 50:3. 50(3):362–367. doi.org/10.1038/s41588-018-0056-5.

Bruford, MW, Bradley, DG & Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics. 4(11):900–910. doi.org/10.1038/nrg1203.

Buaban, S, Lengnudum, K, Boonkum, W & Phakdeedindan, P. 2022. Genome-wide association study on milk production and somatic cell score for Thai dairy cattle using weighted single-step approach with random regression test-day model. Journal of Dairy Science. 105(1):468–494. doi.org/10.3168/JDS.2020-19826.

Campos, MA & Gazzinelli, RT. 2004. Trypanosoma cruzi and its components as exogenous mediators of inflammation recognized through Toll-like receptors. Mediators of Inflammation. 13(3):139–143. doi.org/10.1080/09511920410001713565.

Carignano, HA, Roldan, DL, Beribe, MJ, Raschia, MA, Amadio, A, Nani, JP, Gutierrez, G, Alvarez, I, et al. 2018. Genome-wide scan for commons SNPs affecting bovine leukemia virus infection level in dairy cattle. BMC Genomics 2018 19:1. 19(1):1–15. doi.org/10.1186/S12864-018-4523-2.

Caroli, AM, Chessa, S & Erhardt, GJ. 2009. Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. Journal of Dairy Science. 92(11):5335–5352. doi.org/10.3168/JDS.2009-2461.

Cesarani, A, Sorbolini, S, Criscione, A, Bordonaro, S, Pulina, G, Battacone, G, Marletta, D, Gaspa, G, et al. 2018. Genome-wide variability and selection signatures in Italian island cattle breeds. Animal Genetics. 49(5):371–383. doi.org/10.1111/AGE.12697.

Chen, M, Pan, D, Ren, H, Fu, J, Li, J, Su, G, Wang, A, Jiang, L, et al. 2016. Identification of selective sweeps reveals divergent selection between Chinese

Holstein and Simmental cattle populations. Genetics Selection Evolution. 48(1):76. doi.org/10.1186/s12711-016-0254-5.

Cheruiyot, EK, Bett, RC, Amimo, JO, Zhang, Y, Mrode, R & Mujibi, FDN. 2018. Signatures of Selection in Admixed Dairy Cattle in Tanzania. Frontiers in Genetics. 9:405292. doi.org/10.3389/fgene.2018.00607.

Cheruiyot, EK, Haile-Mariam, M, Cocks, BG, MacLeod, IM, Xiang, R & Pryce, JE. 2021. New loci and neuronal pathways for resilience to heat stress in cattle. Scientific Reports 2021 11:1. 11(1):1–16. doi.org/10.1038/s41598-021-95816-8.

Chu, T, Dufort, I & Sirard, MA. 2012. Effect of ovarian stimulation on oocyte gene expression in cattle. Theriogenology. 77(9):1928–1938. doi.org/10.1016/J.THERIOGENOLOGY.2012.01.015.

Cochran, SD, Cole, JB, Null, DJ & Hansen, PJ. 2013. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genetics. 14(1):49. doi.org/10.1186/1471-2156-14-49.

Cohen-Zinder, M, Seroussi, E, Larkin, DM, Loor, JJ, Everts-Van Der Wind, A, Lee, JH, Drackley, JK, Band, MR, et al. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Research. 15(7):936–944. doi.org/10.1101/GR.3806705.

Cruz, VAR, Oliveira, HR, Brito, LF, Fleming, A, Larmer, S, Miglior, F & Schenkel, FS. 2019. Genome-Wide Association Study for Milk Fatty Acids in Holstein Cattle Accounting for the DGAT1 Gene Effect. Animals 2019, Vol. 9, Page 997. 9(11):997. doi.org/10.3390/ANI9110997.

Do, DN, Bissonnette, N, Lacasse, P, Miglior, F, Sargolzaei, M, Zhao, X & Ibeagha-Awemu, EM. 2017. Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. Journal of Dairy Science. 100(3):1955–1970. doi.org/10.3168/JDS.2016-11910.

Engle, BN & Hayes, BJ. 2022. Genetic variation in PLAG1 is associated with early fertility in Australian Brahman cattle. Journal of Animal Science. 100(4). doi.org/10.1093/JAS/SKAC084.

Esfahani, EN, Ansari Mahyari, S, Davoodi, P, Ghaderi-Zefrehei, M & Lesch, BJ. 2025. Decoding cattle (Bos taurus) diacylglycerol acyltransferase (DGAT) gene families: A pathway to functional understanding. Journal of Heredity. (February, 11). doi.org/10.1093/JHERED/ESAE079.

FAO. 2015. The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture. B.D. Scherf & D. Pilling, Eds. Rome: FAO Commission on Genetic Resources for Food and Agriculture Assessments. doi.org/10.4060/I4787E.

Fay, JC & Wu, C-I. 2000. Hitchhiking Under Positive Darwinian Selection. Genetics. 155(3):1405–1413. doi.org/10.1093/genetics/155.3.1405.

Fink, T, Tiplady, K, Lopdell, T, Johnson, T, Snell, RG, Spelman, RJ, Davis, SR & Littlejohn, MD. 2017. Functional confirmation of PLAG1 as the candidate causative gene underlying major pleiotropic effects on body weight and milk characteristics. Scientific Reports 2017 7:1. 7(1):1–8. doi.org/10.1038/srep44793.

Fonseca, PAS, Suárez-Vega, A, Marras, G & Cánovas, Á. 2020. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience. 9(12):1–9. doi.org/10.1093/gigascience/giaa149.

Fonseca, PAS, Suárez-Vega, A & Cánovas, A. 2020. Weighted Gene Correlation Network Meta-Analysis Reveals Functional Candidate Genes Associated with High- and Sub-Fertile Reproductive Performance in Beef Cattle. Genes 2020, Vol. 11, Page 543. 11(5):543. doi.org/10.3390/GENES11050543.

Fortes, MRS, Kemper, K, Sasazaki, S, Reverter, A, Pryce, JE, Barendse, W, Bunch, R, McCulloch, R, et al. 2013. Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. Animal Genetics. 44(6):636–647. doi.org/10.1111/AGE.12075.

Gao, Y, Marceau, A, Iqbal, V, Torres-Vázquez, JA, Neupane, M, Jiang, J, Liu, GE & Ma, L. 2023. Genome-wide association analysis of heifer livability and early first calving in Holstein cattle. BMC Genomics. 24(1):628. doi.org/10.1186/s12864-023-09736-0.

Gautason, E, Schönherz, AA, Sahana, G & Guldbrandtsen, B. 2021. Genomic inbreeding and selection signatures in the local dairy breed Icelandic Cattle. Animal Genetics. 52(3):251–262. doi.org/10.1111/AGE.13058.

Ghahramani, N, Shodja, J, Rafat, SA, Panahi, B & Hasanpur, K. 2021. Integrative Systems Biology Analysis Elucidates Mastitis Disease Underlying Functional Modules in Dairy Cattle. Frontiers in Genetics. 12:712306. doi.org/10.3389/FGENE.2021.712306.

Ghoreishifar, SM, Eriksson, S, Johansson, AM, Khansefid, M, Moghaddaszadeh-Ahrabi, S, Parna, N, Davoudi, P & Javanmard, A. 2020. Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. Genetics Selection Evolution. 52(1):1–15. doi.org/10.1186/s12711-020-00571-5.

Girardot, M, Guibert, S, Laforet, MP, Gallard, Y, Larroque, H & Oulmouden, A. 2006. The insertion of a full-length Bos taurus LINE element is responsible for a transcriptional deregulation of the Normande Agouti gene. Pigment Cell Research. 19(4):346–355. doi.org/10.1111/J.1600-0749.2006.00312.X.

Gonzalez Berrios, CL, Bowden, CF, Saad, HM, Bishop, J V., Van Campen, H, Pinedo, P, Hansen, TR & Thomas, MG. 2024. Identification of candidate SNPs associated with embryo mortality and fertility traits in lactating Holstein cows. Frontiers in Genetics. 15:1409335. doi.org/10.3389/FGENE.2024.1409335.

Gonzalez-Prendes, R, Ginja, C, Kantanen, J, Ghanem, N, Kugonza, DR, Makgahlela, ML, Groenen, MAM & Crooijmans, RPMA. 2022. Integrative QTL mapping and selection signatures in Groningen White Headed cattle inferred from whole-genome sequences. PLOS ONE. 17(10):e0276309. doi.org/10.1371/JOURNAL.PONE.0276309.

Goyal, R, Van Wickle, J, Goyal, D, Matei, N & Longo, LD. 2013. Antenatal Maternal Long-Term Hypoxia: Acclimatization Responses with Altered Gene Expression in Ovine Fetal Carotid Arteries. PLOS ONE. 8(12):e82200. doi.org/10.1371/JOURNAL.PONE.0082200.

Grisart, B, Coppieters, W, Farnir, F, Karim, L, Ford, C, Berzi, P, Cambisano, N, Mni, M, et al. 2002. Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition. Genome Research. 12(2):222–231. doi.org/10.1101/GR.224202.

Gutiérrez-Gil, B, Arranz, JJ & Wiener, P. 2015. An interpretive review of selective sweep studies in Bos taurus cattle populations: Identification of unique and shared selection signals across breeds. Frontiers in Genetics. 6(MAY):167. doi.org/10.3389/FGENE.2015.00167.

Habib, WA, Brioude, F, Edouard, T, Bennett, JT, Lienhardt-Roussie, A, Tixier, F, Salem, J, Yuen, T, et al. 2018. Genetic disruption of the oncogenic HMGA2–PLAG1–IGF2 pathway causes fetal growth restriction. Genetics in Medicine. 20(2):250–258. doi.org/10.1038/GIM.2017.105.

Hawken, RJ, Zhang, YD, Fortes, MRS, Collis, E, Barris, WC, Corbet, NJ, Williams, PJ, Fordyce, G, et al. 2012. Genome-wide association studies of female reproduction in tropically adapted beef cattle. Journal of Animal Science. 90(5):1398–1410. doi.org/10.2527/JAS.2011-4410.

Huson, HJ, Sonstegard, TS, Godfrey, J, Hambrook, D, Wolfe, C, Wiggans, G, Blackburn, H & VanTassell, CP. 2020. A Genetic Investigation of Island Jersey Cattle, the Foundation of the Jersey Breed: Comparing Population Structure and Selection to Guernsey, Holstein, and United States Jersey Cattle. Frontiers in Genetics. 11:519719. doi.org/10.3389/FGENE.2020.00366.

Illa, SK, Mukherjee, S, Nath, S & Mukherjee, A. 2021. Genome-Wide Scanning for Signatures of Selection Revealed the Putative Genomic Regions and Candidate Genes Controlling Milk Composition and Coat Color Traits in Sahiwal Cattle. Frontiers in Genetics. 12:1193. doi.org/10.3389/fgene.2021.699422.

Iso-Touru, T, Tapio, M, Vilkki, J, Kiseleva, T, Ammosov, I, Ivanova, Z, Popov, R, Ozerov, M, et al. 2016. Genetic diversity and genomic signatures of selection among

cattle breeds from Siberia, eastern and northern Europe. Animal Genetics. 47(6):647–657. doi.org/10.1111/AGE.12473.

Jensen, JD, Foll, M & Bernatchez, L. 2016. The past, present and future of genomic scans for selection. Molecular Ecology. 25(1):1–4. doi.org/https://doi.org/10.1111/mec.13493.

Jin, L, Qu, K, Hanif, Q, Zhang, J, Liu, J, Chen, N, Suolang, Q, Lei, C, et al. 2022. Whole-Genome Sequencing of Endangered Dengchuan Cattle Reveals Its Genomic Diversity and Selection Signatures. Frontiers in Genetics. 13:611. doi.org/10.3389/fgene.2022.833475.

Karim, L, Takeda, H, Lin, L, Druet, T, Arias, JAC, Baurain, D, Cambisano, N, Davis, SR, et al. 2011. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. Nature Genetics 2011 43:5. 43(5):405–413. doi.org/10.1038/ng.814.

Kaune, H, Montiel, JF, Fenwick, M & Williams, SA. 2022. Rapid ovarian transcript changes during the onset of premature ovarian insufficiency in a mouse model. Reproduction and Fertility. 3(3):173–186. doi.org/10.1530/RAF-22-0036.

Kolbehdari, D, Wang, Z, Grant, JR, Murdoch, B, Prasad, A, Xiu, Z, Marques, E, Stothard, P, et al. 2009. A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. Journal of Animal Breeding and Genetics. 126(3):216–227. doi.org/10.1111/J.1439-0388.2008.00793.X.

Laodim, T, Koonawootrittriron, S, Elzo, MA, Suwanasopee, T, Jattawa, D & Sarakul, M. 2023. Genetic factors influencing milk and fat yields in tropically adapted dairy cattle: insights from quantitative trait loci analysis and gene associations. Animal Bioscience. 37(4):576–590. doi.org/10.5713/AB.23.0246.

Lewontin, RC & Krakauer, J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics. 74(1):175–195. doi.org/10.1093/GENETICS/74.1.175.

Li, R, Zhang, CL, Liao, XX, Chen, D, Wang, WQ, Zhu, YH, Geng, XH, Ji, DJ, et al. 2015. Transcriptome MicroRNA Profiling of Bovine Mammary Glands Infected with Staphylococcus aureus. International Journal of Molecular Sciences 2015, Vol. 16, Pages 4997-5013. 16(3):4997–5013. doi.org/10.3390/IJMS16034997.

Li, RW, Rinaldi, M & Capuco, A V. 2011. Characterization of the abomasal transcriptome for mechanisms of resistance to gastrointestinal nematodes in cattle. Veterinary Research. 42(1):1–11. doi.org/10.1186/1297-9716-42-114.

Lillepea, K, Juchnewitsch, AG, Kasak, L, Valkna, A, Dutta, A, Pomm, K, Poolamets, O, Nagirnaja, L, et al. 2024. Toward clinical exomes in diagnostics and management of male infertility. The American Journal of Human Genetics. 111(5):877–895. doi.org/10.1016/J.AJHG.2024.03.013.

Littlejohn, M, Grala, T, Sanders, K, Walker, C, Waghorn, G, MacDonald, K, Coppieters, W, Georges, M, et al. 2012. Genetic variation in PLAG1 associates with early life body weight and peripubertal weight and growth in Bos taurus. Animal Genetics. 43(5):591–594. doi.org/10.1111/J.1365-2052.2011.02293.X.

Littlejohn, MD, Tiplady, K, Fink, TA, Lehnert, K, Lopdell, T, Johnson, T, Couldrey, C, Keehan, M, et al. 2016. Sequence-based Association Analysis Reveals an MGST1 eQTL with Pleiotropic Effects on Bovine Milk Composition. Scientific Reports 2016 6:1. 6(1):1–14. doi.org/10.1038/srep25376.

Liu, D, Chen, Z, Zhao, W, Guo, L, Sun, H, Zhu, K, Liu, G, Shen, X, et al. 2021. Genome-wide selection signatures detection in Shanghai Holstein cattle population identified genes related to adaption, health and reproduction traits. BMC Genomics. 22(1):747. doi.org/10.1186/s12864-021-08042-x.

Liu, L, Zhou, J, Chen, CJ, Zhang, J, Wen, W, Tian, J, Zhang, Z & Gu, Y. 2020. GWAS-Based Identification of New Loci for Milk Yield, Fat, and Protein in Holstein Cattle. Animals 2020, Vol. 10, Page 2048. 10(11):2048. doi.org/10.3390/ANI10112048.

Loftus, RT, MacHugh, DE, Bradley, DG, Sharp, PM & Cunningham, P. 1994. Evidence for two independent domestications of cattle. Proceedings of the National Academy of Sciences. 91(7):2757–2761. doi.org/10.1073/pnas.91.7.2757.

Lopez, NC, Ebensperger, G, Herrera, EA, Reyes, R V., Calaf, G, Cabello, G, Moraga, FA, Beñaldo, FA, et al. 2016. Role of the RhoA/ROCK pathway in high-altitude associated neonatal pulmonary hypertension in lambs. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 310(11):R1053–R1063. doi.org/10.1152/ajpregu.00177.2015.

Ma, Y, Ding, X, Qanbari, S, Weigend, S, Zhang, Q & Simianer, H. 2015. Properties of different selection signature statistics and a new strategy for combining them. Heredity. 115(5):426–436. doi.org/10.1038/hdy.2015.42.

MacHugh, DE, Larson, G & Orlando, L. 2017. Taming the Past: Ancient DNA and the Study of Animal Domestication. Annual Review of Animal Biosciences. 5(1):329–351. doi.org/10.1146/annurev-animal-022516-022747.

Mariani, E, Malacarne, M, Cipolat-Gotet, C, Cecchinato, A, Bittante, G & Summer, A. 2022. Prediction of fresh and ripened cheese yield using detailed milk composition and udder health indicators from individual Brown Swiss cows. Frontiers in Veterinary Science. 9:1012251. doi.org/10.3389/FVETS.2022.1012251.

Martin, P, Szymanowska, M, Zwierzchowski, L & Leroux, C. 2002. The impact of genetic polymorphisms on the protein composition of ruminant milks. Reproduction Nutrition Development. 42(5):433–459. doi.org/10.1051/RND:2002036.

Mbarek, H, Gordon, SD, Duffy, DL, Hubers, N, Mortlock, S, Beck, JJ, Hottenga, JJ, Pool, R, et al. 2024. Genome-wide association study meta-analysis of dizygotic

twinning illuminates genetic regulation of female fecundity. Human Reproduction. 39(1):240–257. doi.org/10.1093/HUMREP/DEAD247.

Misztal, I & Lourenco, D. 2024. Potential negative effects of genomic selection. Journal of Animal Science. 102. doi.org/10.1093/JAS/SKAE155.

Mota, LFM, Lopes, FB, Fernandes Júnior, GA, Rosa, GJM, Magalhães, AFB, Carvalheiro, R & Albuquerque, LG. 2020. Genome-wide scan highlights the role of candidate genes on phenotypic plasticity for age at first calving in Nellore heifers. Scientific Reports. 10(1):6481. doi.org/10.1038/s41598-020-63516-4.

Mota, LFM, Carvajal, AB, Bernardes, PA, Buzanskas, ME, Baldi, F, Lôbo, RB & Munari, DP. 2022. Integrating genome-wide association study and pathway analysis reveals physiological aspects affecting heifer early calving defined at different ages in Nelore cattle. Genomics. 114(4):110395. doi.org/10.1016/J.YGENO.2022.110395.

Moyer, AL & Wagner, KR. 2015. Mammalian Mss51 is a Skeletal Muscle-Specific Gene Modulating Cellular Metabolism. Journal of Neuromuscular Diseases. 2(4):371–385. doi.org/10.3233/JND-150119.

Mwenya, WNM. 1993. The impact of the introduction of exotic cattle in east and southern Africa. In: Proceedings of the Future of Livestock Industries in East and Southern Africa Workshop Held at Kadoma Ranch Hotel, Zimbabwe. Addis Ababa: International Livestock Centre for Africa.

Nayak, SS, Panigrahi, M, Kumar, H, Rajawat, D, Sharma, A, Bhushan, B & Dutt, T. 2023. Evidence for selective sweeps in the MHC gene repertoire of various cattle breeds. Animal Biotechnology. 34(8):4167–4173. doi.org/10.1080/10495398.2023.2196317.

Nayak, SS, Panigrahi, M, Rajawat, D, Ghildiyal, K, Sharma, A, Parida, S, Bhushan, B, Mishra, BP, et al. 2023. Comprehensive selection signature analyses in dairy cattle exploiting purebred and crossbred genomic data. Mammalian Genome. 34(4):615–631. doi.org/10.1007/S00335-023-10021-4.

OECD-FAO. 2022. OECD-FAO Agricultural Outlook 2022-2031. (OECD-FAO Agricultural Outlook). Paris: OECD Publishing. doi.org/10.1787/f1b0b29c-en.

Ortega, MS, Denicol, AC, Cole, JB, Null, DJ & Hansen, PJ. 2016. Use of single nucleotide polymorphisms in candidate genes associated with daughter pregnancy rate for prediction of genetic merit for reproduction in Holstein cows. Animal Genetics. 47(3):288–297. doi.org/10.1111/AGE.12420.

Parra-Bracamonte, GM, Vázquez-Armijo, JF, Rosa-Reyna, XFD la, Magaña-Monforte, JG, Martínez-González, JC, Lopez-Villalobos, N, Herrera-Ojeda, JB & Pacheco-Contreras, VI. 2025. Frequency of genetic variations associated with milk yield and composition in seven Zebu breeds. The Journal of Agricultural Science. (June, 27):1–22. doi.org/10.1017/S0021859625100154.

Pausch, H, Wang, X, Jung, S, Krogmeier, D, Edel, C, Emmerling, R, Götz, KU & Fries, R. 2012. Identification of QTL for UV-Protective Eye Area Pigmentation in Cattle by Progeny Phenotyping and Genome-Wide Association Analysis. PLOS ONE. 7(5):e36346. doi.org/10.1371/JOURNAL.PONE.0036346.

Pazhanivel, N & Umarani, R. 2022. Current Status of Bovine Ocular Squamous Cell Carcinoma. Indian Journal of Veterinary Pathology. 46(3):183–192. doi.org/10.5958/0973-970X.2022.00031.1.

Pedrosa, VB, Schenkel, FS, Chen, SY, Oliveira, HR, Casey, TM, Melka, MG & Brito, LF. 2021. Genomewide association analyses of lactation persistency and milk production traits in holstein cattle based on imputed whole-genome sequence data. Genes. 12(11):1830. doi.org/10.3390/GENES12111830/S1.

Pereira, M & Gazzinelli, RT. 2023. Regulation of innate immune signaling by IRAK proteins. Frontiers in Immunology. 14:1133354. doi.org/10.3389/FIMMU.2023.1133354.

Persichilli, C, Senczuk, G, Mastrangelo, S, Marusi, M, van Kaam, JT, Finocchiaro, R, Di Civita, M, Cassandro, M, et al. 2023. Exploring genome-wide differentiation and signatures of selection in Italian and North American Holstein populations. Journal of Dairy Science. 106(8):5537–5553. doi.org/10.3168/JDS.2022-22159.

Peterson, H, Kolberg, L, Raudvere, U, Kuzmin, I & Vilo, J. 2020. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Research 2020 9:709. 9:709. doi.org/10.12688/f1000research.24956.2.

Pitt, D, Sevane, N, Nicolazzi, EL, MacHugh, DE, Park, SDE, Colli, L, Martinez, R, Bruford, MW, et al. 2019. Domestication of cattle: Two or three events? Evolutionary Applications. 12(1):123–136. doi.org/https://doi.org/10.1111/eva.12674.

Qanbari, S, Pimentel, ECG, Tetens, J, Thaller, G, Lichtner, P, Sharifi, AR & Simianer, H. 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. Animal Genetics. 41(4):377–389. doi.org/10.1111/J.1365-2052.2009.02016.X.

Rajawat, D, Panigrahi, M, Kumar, H, Nayak, SS, Parida, S, Bhushan, B, Gaur, GK, Dutt, T, et al. 2022. Identification of important genomic footprints using eight different selection signature statistics in domestic cattle breeds. Gene. 816:146165. doi.org/10.1016/J.GENE.2021.146165.

Rajawat, D, Panigrahi, M, Nayak, SS, Ghildiyal, K, Sharma, A, Kumar, H, Parida, S, Bhushan, B, et al. 2023. Uncovering genes underlying coat color variation in indigenous cattle breeds through genome-wide positive selection. Animal Biotechnology. 34(8):3920–3933. doi.org/10.1080/10495398.2023.2240387.

Rajput, YS & Sharma, R. 2023. Enzymes Beyond Traditional Applications in Dairy Science and Technology. Elsevier. doi.org/10.1016/C2021-0-02498-6.

Ramey, HR, Decker, JE, McKay, SD, Rolf, MM, Schnabel, RD & Taylor, JF. 2013. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genomics. 14(1):1–18. doi.org/10.1186/1471-2164-14-382.

Randhawa, IAS, Khatkar, MS, Thomson, PC & Raadsma, HW. 2016. A Meta-Assembly of Selection Signatures in Cattle. PLOS ONE. 11(4):e0153013. doi.org/10.1371/JOURNAL.PONE.0153013.

Raney, BJ, Barber, GP, Benet-Pagès, A, Casper, J, Clawson, H, Cline, MS, Diekhans, M, Fischer, C, et al. 2024. The UCSC Genome Browser database: 2024 update. Nucleic Acids Research. 52(D1):D1082–D1088. doi.org/10.1093/NAR/GKAD987.

Raven, LA, Cocks, BG & Hayes, BJ. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics. 15(1):1–14. doi.org/10.1186/1471-2164-15-62.

Refoyo-Martínez, A, Da Fonseca, RR, Halldórsdóttir, K, Árnason, E, Mailund, T & Racimo, F. 2019. Identifying loci under positive selection in complex population histories. Genome Research. 29(9):1506–1520. doi.org/10.1101/GR.246777.118.

Rhoads, RP, La Noce, AJ, Wheelock, JB & Baumgard, LH. 2011. Short communication: Alterations in expression of gluconeogenic genes during heat stress and exogenous bovine somatotropin administration. Journal of Dairy Science. 94(4):1917–1921. doi.org/10.3168/JDS.2010-3722.

Rocha, R de FB, Garcia, AO, Otto, PI, da Silva, MVB, Martins, MF, Machado, MA, Panetto, JC do C & Guimarães, SEF. 2023. Runs of homozygosity and signatures of selection for number of oocytes and embryos in the Gir Indicine cattle. Mammalian Genome 2023. 1:1–15. doi.org/10.1007/S00335-023-09989-W.

Ruvinskiy, D, Igoshin, A, Yurchenko, A, Ilina, A V. & Larkin, DM. 2022. Resequencing the Yaroslavl cattle genomes reveals signatures of selection and a rare haplotype on BTA28 likely to be related to breed phenotypes. Animal Genetics. 53(5):680–684. doi.org/10.1111/AGE.13230.

Sabeti, PC, Reich, DE, Higgins, JM, Levine, HZP, Richter, DJ, Schaffner, SF, Gabriel, SB, Platko, J V., et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature. 419(6909):832–837. doi.org/10.1038/nature01140.

Saravanan, KA, Panigrahi, M, Kumar, H, Parida, S, Bhushan, B, Gaur, GK, Dutt, T, Mishra, BP, et al. 2021. Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. Genomics. 113(3):955–963. doi.org/10.1016/J.YGENO.2021.02.009.

Schiavo, G, Bertolini, F, Galimberti, G, Bovo, S, Dall'Olio, S, Nanni Costa, L, Gallo, M & Fontanesi, L. 2020. A machine learning approach for the identification of

population-informative markers from high-throughput genotyping data: application to several pig breeds. Animal. 14(2):223–232. doi.org/10.1017/S1751731119002167.

Schmidtmann, C, Schönherz, A, Guldbrandtsen, B, Marjanovic, J, Calus, M, Hinrichs, D & Thaller, G. 2021. Assessing the genetic background and genomic relatedness of red cattle populations originating from Northern Europe. Genetics Selection Evolution 2021 53:1. 53(1):1–18. doi.org/10.1186/S12711-021-00613-6.

Sharma, R, Ahlawat, S, Aggarwal, RAK, Dua, A, Sharma, V & Tantia, MS. 2018. Comparative milk metabolite profiling for exploring superiority of indigenous Indian cow milk over exotic and crossbred counterparts. Journal of Food Science and Technology. 55(10):4232–4243. doi.org/10.1007/s13197-018-3360-2.

Shen, S, Zhu, L, Yang, Y, Bi, Y, Li, J, Wang, Y, Pan, C, Wang, S, et al. 2024. Exploration of the Polymorphism Distribution of Bovine HMGA2 Gene in Worldwide Breeds and Its Associations with Ovarian Traits. Animals. 14(5):796. doi.org/10.3390/ANI14050796/S1.

Sigdel, A, Abdollahi-Arpanahi, R, Aguilar, I & Peñagaricano, F. 2019. Whole Genome Mapping Reveals Novel Genes and Pathways Involved in Milk Production Under Heat Stress in US Holstein Cows. Frontiers in Genetics. 10:473150. doi.org/10.3389/FGENE.2019.00928.

Signer-Hasler, H, Burren, A, Neuditschko, M, Frischknecht, M, Garrick, D, Stricker, C, Gredler, B, Bapst, B, et al. 2017. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genetics Selection Evolution. 49(1):83. doi.org/10.1186/s12711-017-0358-6.

Da Silva, MVGB, Sonstegard, TS, Thallman, RM, Connor, EE, Schnabel, RD & Van Tassell, CP. 2010. Characterization of DGAT1 Allelic Effects in a Sample of North American Holstein Cattle. Animal Biotechnology. 21(2):88–99. doi.org/10.1080/10495390903504625.

Singh, A, Mehrotra, A, Gondro, C, Romero, AR da S, Pandey, AK, Karthikeyan, A, Bashir, A, Mishra, BP, et al. 2020. Signatures of Selection in Composite Vrindavani Cattle of India. Frontiers in Genetics. 11:1687. doi.org/10.3389/FGENE.2020.589496.

Singh, H, Kumar, R, Mazumder, A, Salahuddin, Mazumder, R & Abdullah, MohdM. 2022. Insights into Interactions of Human Cytochrome P450 17A1: A Review. Current Drug Metabolism. 23(3):172–187. doi.org/10.2174/1389200223666220401093833.

Taye, M, Lee, W, Jeon, S, Yoon, J, Dessie, T, Hanotte, O, Mwai, OA, Kemp, S, et al. 2017. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. Mammalian Genome. 28(11):528–541. doi.org/10.1007/s00335-017-9715-6.

Tijjani, A, Utsunomiya, YT, Ezekwe, AG, Nashiru, O & Hanotte, O. 2019. Genome sequence analysis reveals selection signatures in endangered trypanotolerant West

African Muturu cattle. Frontiers in Genetics. 10(JUN):434317. doi.org/10.3389/FGENE.2019.00442.

Tijjani, A, Salim, B, da Silva, MVB, Eltahir, HA, Musa, TH, Marshall, K, Hanotte, O & Musa, HH. 2022. Genomic signatures for drylands adaptation at gene-rich regions in African zebu cattle. Genomics. 114(4):110423. doi.org/10.1016/J.YGENO.2022.110423.

Utsunomiya, YT, do Carmo, AS, Carvalheiro, R, Neves, HH, Matos, MC, Zavarez, LB, Pérez O'Brien, AM, Sölkner, J, et al. 2013. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. BMC Genetics. 14(1):52. doi.org/10.1186/1471-2156-14-52.

Utsunomiya, YT, Pérez O'Brien, AM, Sonstegard, TS, Sölkner, J & Garcia, JF. 2015. Genomic data as the "hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome. Frontiers in Genetics. 6. doi.org/10.3389/fgene.2015.00036.

Utsunomiya, YT, Milanesi, M, Utsunomiya, ATH, Torrecilha, RBP, Kim, E-S, Costa, MS, Aguiar, TS, Schroeder, S, et al. 2017. A PLAG1 mutation contributed to stature recovery in modern cattle. Scientific Reports. 7(1):17140. doi.org/10.1038/s41598-017-17127-1.

Velayudhan, SM, Yin, T, Alam, S, Brügemann, K, Sejian, V, Bhatta, R, Schlecht, E & König, S. 2023. Unraveling the Genomic Association for Milk Production Traits and Signatures of Selection of Cattle in a Harsh Tropical Environment. Biology. 12(12):1483. doi.org/10.3390/BIOLOGY12121483/S1.

Velayudhan, SM, Alam, S, Yin, T, Brügemann, K, Buerkert, A, Sejian, V, Bhatta, R, Schlecht, E, et al. 2023. Selective Sweeps in Cattle Genomes in Response to the Influence of Urbanization and Environmental Contamination. Genes. 14(11):2083. doi.org/10.3390/GENES14112083/S1.

Waerner, T, Gardellin, P, Pfizenmaier, K, Weith, A & Kraut, N. 2001. Human RERE Is Localized to Nuclear Promyelocytic Leukemia Oncogenic Domains and Enhances Apoptosis. Cell Growth & Differentiation. 12(4):201–210.

Wang, H, Gui, H, Rallo, MS, Xu, Z & Matise, MP. 2017. Atrophin protein RERE positively regulates Notch targets in the developing vertebrate spinal cord. Journal of Neurochemistry. 141(3):347–357. doi.org/10.1111/JNC.13969.

Waters, SM, McCabe, MS, Howard, DJ, Giblin, L, Magee, DA, MacHugh, DE & Berry, DP. 2011. Associations between newly discovered polymorphisms in the Bos taurusgrowth hormone receptor gene and performance traits in Holstein–Friesian dairy cattle. Animal Genetics. 42(1):39–49. doi.org/10.1111/J.1365-2052.2010.02087.X.

Wheelock, JB, Rhoads, RP, VanBaale, MJ, Sanders, SR & Baumgard, LH. 2010. Effects of heat stress on energetic metabolism in lactating Holstein cows. Journal of Dairy Science. 93(2):644–655. doi.org/10.3168/JDS.2009-2295.

Worku, D, Gowane, GR, Mukherjee, A, Alex, R, Joshi, P & Verma, A. 2022. Associations between polymorphisms of LAP3 and SIRT1 genes with clinical mastitis and milk production traits in Sahiwal and Karan Fries dairy cattle. Veterinary Medicine and Science. 8(6):2593–2604. doi.org/10.1002/VMS3.924.

Worku, D, Gowane, G & Verma, A. 2023. Genetic variation in promoter region of the bovine LAP3 gene associated with estimated breeding values of milk production traits and clinical mastitis in dairy cattle. PLOS ONE. 18(5):e0277156. doi.org/10.1371/JOURNAL.PONE.0277156.

Wu, C. 1995. Heat Shock Transcription Factors: Structure and Regulation. Annual Review of Cell and Developmental Biology. 11(1):441–469. doi.org/10.1146/annurev.cb.11.110195.002301.

Yuan, P, He, Z, Zheng, L, Wang, W, Li, Y, Zhao, H, Zhang, VW, Zhang, Q, et al. 2017. Genetic evidence of ‘genuine’ empty follicle syndrome: a novel effective mutation in the LHCGR gene and review of the literature. Human Reproduction. 32(4):944–953. doi.org/10.1093/HUMREP/DEX015.

Zhang, Q, Calus, MPL, Bosse, M, Sahana, G, Lund, MS & Guldbrandtsen, B. 2018. Human-Mediated Introgression of Haplotypes in a Modern Dairy Cattle Breed. Genetics. 209(4):1305–1317. doi.org/10.1534/GENETICS.118.301143.

Zheng, X, Ju, Z, Wang, J, Li, Q, Huang, J, Zhang, A, Zhong, J & Wang, C. 2011. Single nucleotide polymorphisms, haplotypes and combined genotypes of LAP3 gene in bovine and their association with milk production traits. Molecular Biology Reports. 38(6):4053–4061. doi.org/10.1007/s11033-010-0524-1.

Zielke, LG, Bortfeldt, RH, Reissmann, M, Tetens, J, Thaller, G & Brockmann, GA. 2013. Impact of Variation at the FTO Locus on Milk Fat Yield in Holstein Dairy Cattle. PLOS ONE. 8(5):e63406. doi.org/10.1371/JOURNAL.PONE.0063406.

Zinovieva, NA, Dotsev, AV, Sermyagin, AA, Deniskova, TE, Abdelmanova, AS, Kharzinova, VR, Sölkner, J, Reyer, H, et al. 2020. Selection signatures in two oldest Russian native cattle breeds revealed using high-density single nucleotide polymorphism analysis. PLOS ONE. 15(11):e0242200. doi.org/10.1371/JOURNAL.PONE.0242200.

Zuleica Trujano-Chavez, M, Ruíz-Flores, A, López-Ordaz, R, Pérez-Rodríguez, P, Chapingo, A, de México, E & en Producción, P. 2022. Genetic diversity in reproductive traits of Braunvieh cattle determined with SNP markers. Veterinary Medicine and Science. 8(4):1709–1720. doi.org/10.1002/VMS3.836.