REVIEW ARTICLE

GWAS analysis in Holstein cattle: current knowledge and future directions

Mihaela Oroz[1], Marko Oroz[1], Mario Shihabi[2], Ino Čurik[2], Marija Spehar[3], Nikola Raguž[1], Boris Lukić[1]

[1]*Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, Osijek, Croatia (mioroz@fazos.hr)*
[2]*University of Zagreb Faculty of Agriculture, Svetošimunska cesta 25, Zagreb, Croatia*
[3]*Centre for Livestock Breeding, Svetošimunska cesta 25, Zagreb, Croatia*

Summary
Genomic selection, combined with traditional selection methods, improves the efficiency of animal selection and allows for a more precise assessment of their production potential. The Genome-Wide Association Studies (GWAS) method is crucial for analysing the association of genetic markers with milk yield traits, whereby increasing the sample size and marker density improves the accuracy of the genomic association. GWAS analyses have identified key genes associated with milk production in cattle, as well as traits such as productive life, age at first calving and disease resistance in Holstein cows. In addition, genes associated with mastitis, one of the major health problems in Holstein cows, together with other health traits, represent potential targets for further research.

Key words: selection, genomics, GWAS, Holstein cattle, milk production

Introduction

Milk has been valued as a high-quality food of animal origin since ancient times, and cow's milk is the most widely used food in the human diet, accounting for 81% of global milk production (Solarczyk et al., 2024). The dominant breed in milk production is the Holstein-Friesian cattle, the most widespread dairy breed in the world, mainly due to its high milk production. In addition to high milk production, milk quality is another key factor in cattle breeding programs (Liu et al., 2020). Profitability and qualityof dairy cattle production also depends on the health and fertility of cows and their productive lifespan, and improving traits such as calving ease and conformation reduces the herd's elimination rate and improves the production. (Abo-Ismail et al., 2017, Liang et al., 2024). Selection and breeding programs are crucial for achieving and improving the production traits of dairy cattle (Ozdemir et al., 2024). Since the traits related to milk production are quantitative, they are influenced by the additive genetic effect and environmental factors, so traditional selection methods have proven to be ineffective (Liu et al., 2020). While traditional selection methods remain in use, advances in genetics and genomics have transformed dairy cattle breeding by enabling much greater selection accuracy. Genomic tools allow breeders to identify animals with superior genetic potential at a very early age, often before any phenotypic traits are expressed, thereby shortening the generation interval and accelerating genetic gain.

Brief history of genotyping

Genome analysis is used to assess the potential of domestic animal production. This analysis involves genotyping using DNA markers distributed throughout the genome. A genetic marker is a marker based on a DNA polymorphism caused by a mutation (Peka and Balatsky, 2024). The addition of DNA information allows for a more accurate assessment of breeding values of breeding animals and a more precise selection. Selection at the DNA and quantitative trait loci

(QTL) levels accelerates genetic progress by identifying key genes responsible for milk production (Ozdemir et al., 2024). Although the use of DNA markers to improve genetic progress in livestock breeding was proposed several decades ago, their application was limited for along time due to the high cost of genotyping and the complexity of calculating breeding values. Today, these obstacles have been completely overcome, so that genomic evaluation is officially accepted and widely used in dairy farming due to its advantages over traditional selection methods (Palombo et al., 2020). The first GWAS results were already published in 2005 and 2006, but the starting point is the Wellcome Trust Case Control Consortium (WTCCC) document from 2007 (Wellcome Trust Case Control Consortium, 2007). This case-control study in humans aimed to identify variants associated with common diseases and represented a major breakthrough at the time. GWAS analyses a number of genetic variants in the genome to determine the associations between genomic variants and specific traits and aims to identify which genomic locations are linearly associated with complex traits in the population. GWAS analyzes are based on the concept of genetic linkage disequilibrium or LD, i.e. the non-random association of alleles at different loci, and a particular focus is on assessing the association between common single nucleotide polymorphisms or SNPs. LD facilitates the identification of genomic regions associated with specific traits through markers that are correlated with causal variants. Therefore, the pattern and extent of LD in a population have impact on the accuracy and effectiveness of detecting significant genetic associations (Visscher et al., 2012). In dairy cattle genomics, a various SNP panels (Illumina Bovine SNP50K, GGP Bovine 100K etc.) of different densities are commonly used for analysis (Abo-Ismail et al., 2017). The sample size in GWAS analyses is crucial for the accuracy of the results. Increasing the sample size in combination with advanced multivariate analysis methods (e.g. Multivariate Linear Mixed Models or mvLLM; Multi-Trait Mixed Model or MTMM; Multi-Trait Bayesian Models etc.) has improved the precision of estimates, which has led to a wider acceptance of polygenic predictors in all research disciplines (Abdellaoui et al., 2023). In addition to increasing sample size, increasing marker density from low to mid density in cattle has also contributed to the accuracy of genomic linkage and provided better insight into the genetic architecture of dairy traits in cows (Liu et al., 2020).

GWAS and genetic markers in selection for milk yield and longevity traits

The adoption of GWAS has enabled genomic research to successfully identify several key genes responsible for milk production in cattle (Chen et al., 2018). GWAS analysis has also used to investigate the association between genetic factors and longevity as an important factor for productivity in Holstein cows in the USA (Wiggans and Carrilo, 2022). Two chromosomal regions with additive effects were associated with productive lifespan in Holstein breed. The SLC4A4-GC-NPFFR2 *(SGN)* region on chromosome 6, which has a significant pleiotropic effect on a total of four traits: milk yield, number of somatic cells, fertility and productive life, and a large region on chromosome 10 that affects several immune-related functions such as the development, homeostasis, and differentiation of T cells and B cells (Liang et al., 2024).

Age at first calving is a relatively new reproductive trait used in genomic analyses of Holstein cows in the USA (Wiggans and Carrilo, 2022). This trait affects the longevity and productivity of dairy cows, particularly in terms of total milk yield over their lifetime production. GWAS has been used to identify genes and chromosomal regions that influence this trait. Significant (p-value < 10^{-8}) additive effects were identified on three chromosomes: in the 7.86–8.12 Mb region on chromosome 15, 27.07–27.48 Mb region and 31.25–32.11 Mb region on chromosome 19, and in the 26.92–32.60 Mb region on chromosome 23. In these regions, two genes are associated with sex hormones, and their known biological functions may be relevant to age at first calving. These are a sex hormone-binding globulin gene *(SHBG)* and a progesterone receptor gene *(PGR).* Significant SNP variants in their proximity and their known biological function indicate a possible link between these reproductive hormones with age at first calving in Holstein cows (Prakapenka et al., 2023). The diacylglycerol O-acyltransferase 1*(DGAT1)* gene discovered by GWAS is one of the key genes influencing milk production in cows (Liu et al., 2020). The gene is located on the 14th autosomal chromosome of cattle in the centromere region. It encodes diacylglycerol-O-transferase, an enzyme essential for triglyceride synthesis, lipid absorption, and adipose tissue development. The non-conservative polymorphism K232A, caused by an A-A to G-C substitution in exon 8, leads to a lysine-to-alanine change and significantly affects milk fat and protein content (Mahmoudi and Rashidi, 2023). It is possible to use *DGAT1* as a genetic marker for the selection of milk composition traits, especially in Holstein breed (Ozdemir et al., 2024.). In addition to this gene, kappa casein gene *(CSN3)* can be used as a good marker for milk quality (protein and fat content) (Liu et al., 2020) and the prolactin gene *(PRL)* as a marker for milk quantity (Ozdemir et al., 2024). Within the *DGAT1* gene, a SNP was discovered that showed a significant association with bone quality in Holstein cows, as well as with rump width and conformation of the animal. Bone quality is an important trait that influences the functional longevity of cows, indicating a possible link between the *DGAT1* gene and functional longevity in Holstein cows (Abo-Ismail et al., 2017). In addition to this gene, which has been the most studied, several other genes identified by GWAS influence milk production in Holstein cows. Chen et al., (2018) confirmed the association of the: DLG associated protein 1 (*DLGAP1)*, adaptor related protein complex 2 subunit beta 1 (*AP2B1)*, inositol 1,4,5-trisphosphate receptor type 2 (*ITPR2)* and thrombospondin 4 (*THBS4)* genes with milk yield, and the rho guanine nucleotide exchange factor 4 (*ARHGEF4)*, tudor domain containing 1 (*TDRD1)* and kinesin family member 19 (*KIF19)* genes are currently assumed to be associated with milk production (Iung et al., 2019). The ATP-binding cassette subfamily G member 2 gene (ABCG2) located on 6th autosomal chromosome of cattle, has been identified as an important genetic factor influencing milk production traits. It is associated with both the yield and composition of milk, as well as with the excretion efficiency of certain pharmaceuticals into milk. The Y581S polymorphism within this gene plays a significant role in modulating milk quality by affecting its fat and protein content (Peka and Balatsky, 2024). The growth hormone receptor gene (*GHR)* is crucial for postnatal growth. It significantly affects metabolism, and it is also associated with meat quality in cattle, and participates in the regulation of milk production and fertility in cows (Dettori et al., 2018). Studies have shown that GHR affects milk yield, protein and fat content, as well as the overall physical condition of cows. Among several identified SNPs in this gene, the F279Y missense polymorphism in exon 8 is considered to have a direct impact on both milk quality and quantity. This effect has been observed not only in Holstein cows, but also in Ayrshire and Jersey breeds (Peka and Balatsky, 2024). The *PRL* gene, which is located on chromosome 23 showed a significant association with fatty acids in milk. The mechanism by which *PRL* influences the fatty acid composition of milk is not yet fully understood, but it appears to involve signaling pathways correlation with signal transducers and transcriptional activators. The leptin (*LEP)* gene, located on chromosome 4, is associated with nine fatty acids in milk, and in the presence of the *PRL* gene, fat synthesis is enhanced, suggesting that these two genes together influence milk quality (Pegolo et al., 2016). Do et al. (2017) identified the *PRL* gene as a possible candidate for influencing lactation persistence in Holstein cows. The casein beta (*CSN2)* regulates the synthesis of all casein fractions except κ-casein, and its relevance in selection programs stems from its distinct characteristics and strong association with specific DNA markers. Allelic variants of beta-casein represent quality traits as targeted mutations can influence the type of milk product, such as A1 milk (Kulibaba et al., 2024). A GWAS analysis was also performed in the Holstein breed to identify genes that influence the occurrence of mastitis in cows. Significant SNPs were identified on chromosomes 14, 15, 22 and 29, with the following genes being associated with the immune response and inflammatory processes: deoxyribonuclease

1L3 (*DNASE1L3),* solute carrier family 36 member 4 *(SLC36A4),* armadillo repeat containing 1 *(ARMC1),* phosphodiesterase 7A *(PDE7A),* matrix metallopeptidase 13 *(MMP13)* and CD44 molecule *(CD44)*. These genes represent potential targets for further research, which should focus on comparing their expression levels in cows affected by mastitis and in healthy individuals for better insight, since mastitis is one of the main health problems in milk production (Ashja et al., 2024).

Conclusions

GWAS results have made a significant contribution to Holstein cattle breeding by identifying genes linked to production and longevity traits, leading to improvements in milk yield and composition, fertility, and overall lifespan. More recent GWAS studies, aimed to identify genes that influence disease resistance in cattle, particularly those affecting udder health, with the goal of reducing of mastitis incidence. Identifying these genes and incorporating them into selection programs can improve animal health and lower treatment costs. GWAS results, combined with genomic selection, provide deeper insights and enable faster genetic progress by allowing breeders to select the best animals for breeding even before they produce offspring. In most cases, the majority of genetic variation is effectively captured by GBLUP models, which assume small, additive effects across the whole genome. However, for certain traits influenced by major genes or large-effect loci, incorporating specific SNPs as fixed effects detected by GWAS in evaluations can enhance prediction accuracy and biological interpretation. In addition, the estimated breeding values are more accurate than the breeding values obtained using traditional selection methods. The combination of GWAS and genomic selection is a promising strategy in modern cattle breeding to achieve sustainability and productivity.

References

Abdellaoui A., Yengo L., Verweij K.J.H., Visscher P.M. (2023). 15 years of GWAS discovery: Realizing the promise. American Journal of Human Genetics, 2, 110(2): 179-194.

Abo-Ismail M.K., Brito L.F., Miller S.P., Sargolzaei M., Grossi D.A., Moore S.S., Plastow G., Stothard P., Nayeri S., Schenkel F.S. (2017). Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle. Genetics Selection Evolution, 7, 49(1): 82.

Ashja A., Zorc M., Dovc P. (2024). Genome-Wide Association Study for Milk Somatic Cell Score in Holstein Friesian Cows in Slovenia. Animals, 19, 14(18): 2713.

Chen Z., Yao Y., Ma P., Wang Q., Pan Y. (2018). Haplotype-based genome-wide association study identifies loci and candidate genes for milk yield in Holsteins. PLoS One, 15, 13(2).

Dettori M. L., Pazzola M., Paschino P., Amills M., Vacca G. M. (2018). Association between the GHR, GHRHR, and IGF1 gene polymorphisms and milk yield and quality traits in Sarda sheep. Journal of dairy science, 101(11): 9978–9986.

Do D.N., Bissonnette N., Lacasse P., Miglior F., Sargolzaei M., Zhao X., Ibeagha-Awemu E.M. (2017). Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. Journal of Dairy Science, 100(3): 1955-1970.

Iung L.H.S, Petrini J., Ramírez-Díaz J., Salvian M., Rovadoscki G.A., Pilonetto F., Dauria B.D., Machado P.F., Coutinho L.L., Wiggans G.R., Mourão G.B. (2019). Genome-wide association study for milk production traits in a Brazilian Holstein population. Journal of Dairy Science, 102(6): 5305-5314.

Kulibaba R.O., Liashenko Y.V., Sakhatskyi M.I. (2024). Polymorphism of CSN2 and TNF-α Genes in the Population of Holstein Cattle Bred in Ukraine. Cytology and Genetics, 58: 29–38.

Liang Z., Prakapenka D., Zaabza H.B., VanRaden P.M., Van Tassell C.P., Da Y. (2024). A million-cow genome-wide association study of productive life in U.S. Holstein cows. Genetics Selection Evolution, 26, 56(1): 67.
Liu L., Zhou J., Chen C. J., 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, 10(11): 2048.
Mahmoudi P., Rashidi A. (2023). Strong evidence for association between K232A polymorphism of the DGAT1 gene and milk fat and protein contents: A meta-analysis. Journal of dairy science, 106(4): 2573–2587.
Ozdemir M., Motmain Z., Ekinci K., Saygıl E. (2024). Associations Between BLG, CSN3, DGAT1, GH, PIT1, and PRL Gene Polymorphisms and Milk Production Traits in Holstein Dairy Cows: A Meta-Analysis. Biochemical Genetics.
Palombo V., Pegolo S., Conte G., Cesarani A., Macciotta N. P. P., Stefanon B., D'Andrea, M. (2020). Genomic prediction for latent variables related to milk fatty acid composition in Holstein, Simmental and Brown Swiss dairy cattle breeds. Journal of Animal Breeding and Genetics.
Pegolo S., Cecchinato A., Mele M., Conte G., Schiavon S., Bittante G. (2016). Effects of candidate gene polymorphisms on the detailed fatty acids profile determined by gas chromatography in bovine milk. Journal of Dairy Science, 99(6): 4558-4573.
Peka M., Balatsky V. (2024). Bioinformatic approach to identifying causative missense polymorphisms in animal genomes. BMC Genomics, 19, 25(1): 1226.
Prakapenka D., Liang Z., Da Y. (2023). Genome-Wide Association Study of Age at First Calving in U.S. Holstein Cows. International Journal of Molecular Sciences, 12, 24(8): 7109.
Solarczyk P., Slósarz J., Gołębiewski M., Natalello A., Musati M., Luciano G., Priolo A., Puppel K. (2024). The Influence of Crossbreeding on the Composition of Protein and Fat Fractions in Milk: A Comparison Between Purebred Polish Holstein Friesian and Polish Holstein Friesian × Swedish Red Cows. Nutrients. 16(21): 3634.
Visscher P.M., Brown M.A., McCarthy M.I., Yang J. (2012). Five years of GWAS discovery. American Journal of Human Genetics 13;90(1):7-24.
Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447(7145): 661–678.
Wiggans G. R., Carrillo J. A. (2022). Genomic selection in United States dairy cattle. Frontiers in genetics, 13, 994466.