Animal Molecular Breeding 2024, Vol.14, No.5, 307-317
http://animalscipublisher.com/index.php/amb

Research Insight Open Access

Genome-Wide Association Studies for Milk Production in Dairy Cattle

Qineng Si

Biotechnology Reseach Center, Cuixi Academy of Biotechnology, Zhuji, 311800, Zhejiang, China

Corresponding email: qineng.si@cuixi.org

Animal Molecular Breeding, 2024, Vol.14, No.5 doi: 10.5376/amb.2024.14.0032

Received: 28 Jul., 2024

Accepted: 06 Sep., 2024

Published: 20 Sep., 2024

Preferred citation for this article:

Si Q.N., 2024, Genome-wide association studies for milk production in dairy cattle, Animal Molecular Breeding, 14(5): 307-317 (doi: 10.5376/amb.2024.14.0032)

Abstract This study synthesizes findings from multiple GWAS, highlighting key genomic regions and candidate genes associated with milk yield, fat percentage, protein percentage, and somatic cell score (SCS). Notable genes such as *DGAT1*, *ABCG2*, and *MGST1* are consistently implicated, along with novel candidates like *CCSER1* and *CUX2*. By integrating high-density SNP chips and whole-genome sequencing, these studies have enhanced the detection of quantitative trait loci (QTLs) and refined genomic selection strategies. The findings underscore the polygenic nature of milk production traits and the utility of GWAS in improving breeding accuracy. Future prospects include the integration of machine learning, epigenomics, and metabolomics to further enhance genetic predictions, optimize breeding programs, and ensure sustainable dairy farming practices.

Keywords Genome-wide association studies (GWAS); Milk production; Quantitative trait loci (QTL); Genomic selection; Dairy cattle

1 Introduction

Milk production is a critical economic trait in dairy cattle, significantly influencing the profitability and sustainability of dairy farming. The efficiency and volume of milk production are determined by a complex interplay of genetic, environmental, and management factors. Over the years, selective breeding has been employed to enhance milk yield and quality, focusing on traits such as milk fat, protein content, and overall milk volume (Pryce et al., 2010; Buaban et al., 2021; Taherkhani et al., 2022). The advent of genomic technologies has revolutionized the ability to identify and select for these traits, providing a more precise and efficient approach to dairy cattle breeding.

Understanding the genetic basis of milk production traits is essential for several reasons. Firstly, it allows for the identification of specific genes and genomic regions that influence milk yield and composition, which can be targeted in breeding programs to improve these traits (Mai et al., 2010; Dadousis et al., 2017a; Teng et al., 2023). Secondly, it provides insights into the biological mechanisms underlying milk production, which can lead to the development of new strategies for enhancing milk yield and quality (Chen et al., 2018; Otto et al., 2020). Genome-wide association studies (GWAS) have been particularly instrumental in this regard, enabling the discovery of quantitative trait loci (QTL) and candidate genes associated with milk production traits across different cattle breeds (Bolormaa et al., 2010; Jiang et al., 2010).

This study aims to synthesize current knowledge on the genetic factors influencing milk production in dairy cattle, particularly based on findings from genome-wide association studies (GWAS). It seeks to identify key genomic regions and candidate genes associated with milk production traits, discuss the comparison of different GWAS approaches and their effectiveness in identifying relevant genetic markers, and explore the implications of these genetic structures for dairy cattle breeding and milk production improvement. The study aspires to provide a comprehensive overview of the genetic information on milk production in dairy cattle and highlight potential research directions for future studies and applications in genomic selection programs.

2 Overview of Genome-Wide Association Studies (GWAS)

2.1 Introduction to GWAS

Genome-Wide Association Studies (GWAS) are a powerful tool used to identify genetic variants associated with specific traits by scanning the entire genome. This approach involves comparing the DNA of individuals with different phenotypes to find genetic markers, such as single nucleotide polymorphisms (SNPs), that occur more frequently in individuals with a particular trait. GWAS has been widely applied in both human and animal genetics to uncover the genetic basis of complex traits and diseases (Jiang et al., 2019; Ma, 2020; Wang et al., 2020).

2.2 History and evolution of GWAS in livestock research

The application of GWAS in livestock, particularly in dairy cattle, has evolved significantly since its inception. Initially, GWAS in livestock focused on simple linear regression models to identify quantitative trait loci (QTL) associated with economically important traits such as milk production. Over time, more sophisticated statistical models, including mixed models and Bayesian approaches, were developed to account for population structure and relatedness among individuals, which are common in livestock populations (Jiang et al., 2010; Ma, 2020; Bakhshalizadeh et al., 2021).

The first GWAS in dairy cattle were conducted in the early 2000s, and since then, the field has seen rapid advancements. The availability of high-density SNP chips, such as the Illumina BovineSNP50 BeadChip, has enabled researchers to perform more comprehensive and accurate genome scans. These technological advancements have led to the identification of numerous QTLs and candidate genes associated with milk production and other important traits (Jiang et al., 2010; Chen et al., 2018; Jiang et al., 2019).

2.3 Methodologies and statistical models used in GWAS

Several methodologies and statistical models are employed in GWAS to identify genetic associations with traits of interest. The choice of method depends on the study design, population structure, and the specific traits being investigated. Commonly used methods include:

Single-SNP Analysis: This approach tests each SNP individually for association with the trait. It is simple and computationally efficient but may lack power to detect associations in the presence of complex genetic architectures (Pryce et al., 2010; Wang et al., 2020).

Haplotype-Based Analysis: Instead of analyzing individual SNPs, this method considers haplotypes, which are combinations of alleles at adjacent loci. Haplotype-based analysis can increase the power to detect associations by capturing the combined effect of multiple SNPs (Pryce et al., 2010; Chen et al., 2018).

Mixed Linear Models (MLM): MLMs account for population structure and relatedness among individuals by incorporating random effects. This method is particularly useful in livestock populations where individuals are often related. MLMs have become the standard approach in GWAS for dairy cattle (Jiang et al., 2019; Ma, 2020; Wang et al., 2020).

Bayesian Approaches: These methods use prior information and probabilistic models to estimate the effects of SNPs. Bayesian approaches can handle complex genetic architectures and provide more accurate estimates of SNP effects, especially in studies with small sample sizes (Ma, 2020).

Meta-Analysis: This method combines data from multiple GWAS to increase the power to detect associations. Meta-analysis can identify QTLs with higher precision and reveal consistent genetic associations across different populations and studies (Bakhshalizadeh et al., 2021; Taherkhani et al., 2022).

3 Milk Production Traits and Their Genetic Basis

3.1 Key traits influencing milk production

Milk production in dairy cattle is influenced by several key traits, including milk yield (MY), milk fat yield (FY), milk protein yield (PY), milk fat percentage (FP), and milk protein percentage (PP). These traits are critical for the economic viability of dairy farming as they directly impact the quantity and quality of milk produced. Studies have shown that these traits are highly heritable and can be significantly influenced by genetic factors (Jiang et al., 2010; Wang et al., 2019; Teng et al., 2023). For instance, milk yield is a primary trait of interest due to its direct correlation with dairy farm profitability, while fat and protein percentages are essential for determining milk quality and processing characteristics (Chen et al., 2018; Bakhshalizadeh et al., 2021; Kim et al., 2021).

3.2 Heritability of milk production traits

Heritability estimates for milk production traits indicate a substantial genetic component, which makes these traits amenable to selection and genetic improvement. For example, heritability estimates for milk yield, fat yield, and protein yield are generally high, suggesting that a significant proportion of the variation in these traits can be attributed to genetic differences among individuals (Jiang et al., 2010; Taherkhani et al., 2022; Laodiim et al., 2023). This high heritability underscores the potential for using genomic selection to enhance these traits in dairy cattle populations. Longitudinal studies have further demonstrated that the genetic influence on these traits can be consistently observed across different stages of lactation, reinforcing the importance of genetic factors in milk production (Teng et al., 2023).

3.3 Major genes and QTLs (Quantitative Trait Loci) associated with milk yield and composition

Genome-wide association studies (GWAS) have identified numerous quantitative trait loci (QTLs) and candidate genes associated with milk production traits. Notably, the *DGAT1* gene on chromosome 14 has been repeatedly implicated in influencing milk yield, fat percentage, and protein percentage across various studies (Wang et al., 2019; Bakhshalizadeh et al., 2021; Kim et al., 2021). Other significant genes include GHR, which is associated with milk yield, and *ABCG2*, which affects milk composition traits such as fat and protein percentages (Iung et al., 2019; Teng et al., 2023).

Meta-analyses and gene network analyses have further refined the understanding of these genetic influences by integrating data from multiple studies, thereby increasing the power to detect significant QTLs and elucidate the underlying biological mechanisms (Bakhshalizadeh et al., 2021; Taherkhani et al., 2022). For instance, a meta-analysis identified 9 QTLs for milk yield, 36 QTLs for fat percentage, and 10 QTLs for protein percentage, highlighting the complex genetic architecture of these traits (Bakhshalizadeh et al., 2021). Additionally, novel candidate genes such as *PDE4B* and *ANO2* have been discovered, which may play roles in milk production through pathways related to fat metabolism and cellular signaling (Kim et al., 2021).

4 Applications of GWAS in Dairy Cattle

4.1 Identification of genetic variants associated with milk production

Genome-wide association studies (GWAS) have significantly advanced our understanding of the genetic architecture underlying milk production traits in dairy cattle. By analyzing large datasets of phenotypic and genotypic information, researchers have identified numerous single nucleotide polymorphisms (SNPs) and genomic regions associated with milk yield, fat yield, protein yield, and other related traits. For instance, a study on Brazilian Holstein cattle identified 46 genomic windows explaining more than 1% of the genetic variance for milk yield and other traits, highlighting genes such as *MGST1*, *ABCG2*, *DGAT1*, and PAEP (Iung et al., 2019). Similarly, a meta-analysis of GWAS in Holstein cows identified significant QTLs for milk yield, fat percentage, and protein percentage, with notable SNPs located near the *DGAT1* and PPP1R16A genes (Bakhshalizadeh et al., 2021). These findings underscore the polygenic nature of milk production traits and the importance of specific genomic regions in influencing these traits.

4.2 GWAS findings on milk composition (e.g., Fat, Protein, Lactose)

GWAS have also been instrumental in elucidating the genetic basis of milk composition, including fat, protein, and lactose content. In a study involving Chinese Holstein cows, researchers identified 28 candidate SNPs associated with various milk composition traits using a mixed linear model (Wang et al., 2020). Another study focused on milk protein composition traits in Chinese Holstein cows identified significant associations with genes such as *CSN1S1, CSN1S2, CSN2, CSN3*, and *DGAT1*, which are known to influence milk protein fractions like casein and lactoglobulin (Zhou et al., 2019). Additionally, research on Thai dairy cattle revealed genomic regions associated with fat, protein, and total solid percentages, further emphasizing the role of specific genes in determining milk composition (Buaban et al., 2021) (Figure 1). These studies provide valuable insights into the genetic factors that influence milk quality and composition, which are crucial for dairy production and processing.

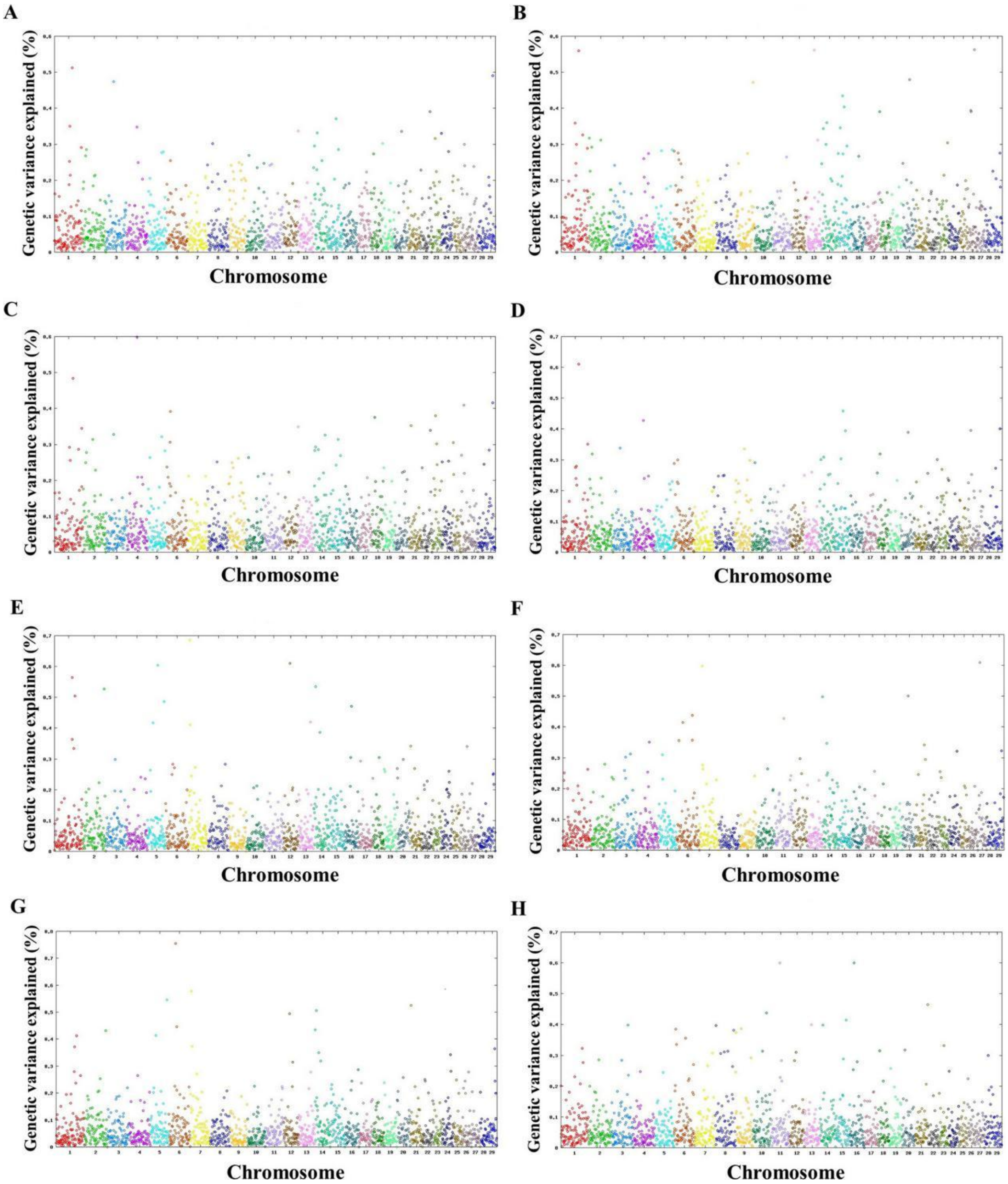

Figure 1 Manhattan plots of the additive genetic variance explained by windows of 20 adjacent SNPs for milk production traits and SCS in Thai dairy cattle: (A) milk yield, (B) fat yield, (C) protein yield, (D) TS yield, (E) fat percentage, (F) protein percentage, (G) TS percentage, and (H) SCS (Adopted from Buaban et al., 2021)

4.3 GWAS contributions to improving milk quality and yield

The application of GWAS in dairy cattle has not only enhanced our understanding of the genetic determinants of milk production and composition but also contributed to practical improvements in milk quality and yield (Yang, 2024). By identifying key genetic variants and QTLs, GWAS enables the development of genomic selection strategies that can be used to breed cattle with superior milk production traits. For example, the identification of novel genes involved in tissue repair, immune response, and glucose homeostasis in Brazilian Holstein cattle can inform the design of customized SNP arrays for genomic selection, potentially improving milk yield and quality under specific environmental conditions (Iung et al., 2019). Furthermore, the integration of GWAS findings into breeding programs has been shown to increase the accuracy of genomic predictions, as demonstrated in studies on Thai dairy cattle and Brazilian Girolando cattle (Otto et al., 2020; Buaban et al., 2021). These advancements highlight the potential of GWAS to drive genetic gains in dairy cattle, ultimately leading to more efficient and productive dairy farming practices.

5 Functional Annotation and Pathway Analysis

5.1 Linking genetic markers to biological functions

Genome-wide association studies (GWAS) have identified numerous single nucleotide polymorphisms (SNPs) associated with milk production traits in dairy cattle. These genetic markers are linked to various biological functions through gene ontology (GO) and pathway enrichment analyses. For instance, significant SNPs identified in Holstein cows were linked to genes such as *DGAT1* and PPP1R16A, which are associated with milk yield, fat percentage, and protein percentage (Bakhshalizadeh et al., 2021). Similarly, a meta-analysis of GWAS data revealed significant loci on chromosome 14, implicating genes like *DGAT1* and CPSF1 in milk production traits (Taherkhani et al., 2022). These findings highlight the importance of specific genomic regions and their associated biological functions in influencing milk production.

5.2 Pathway analysis of identified genes

Pathway analysis of genes identified through GWAS provides insights into the biological mechanisms underlying milk production traits. For example, GO term enrichment analysis has identified pathways such as regulation of cation channel activity, ion channel complex, and phosphoric diester hydrolase activity as significant for milk yield (Bakhshalizadeh et al., 2021). Additionally, pathways related to calcium signaling, metabolic processes, and carbohydrate digestion have been associated with milk coagulation properties and cheese yield (Dadousis et al., 2017b). These pathways are crucial for understanding how genetic variations influence milk production and its components.

5.3 Integrating GWAS with transcriptomic and proteomic data

Integrating GWAS findings with transcriptomic and proteomic data enhances the understanding of the genetic architecture of milk production traits. For instance, combining GWAS with gene expression data can identify candidate genes and their regulatory networks. In a study on Thai dairy cattle, genomic regions associated with milk production traits were found to contain genes related to heat tolerance, longevity, and fertility, indicating indirect selection pressures (Buaban et al., 2021). Furthermore, integrating proteomic data can reveal protein interactions and pathways involved in milk production, as demonstrated by the identification of intracellular cell transportation and protein catabolism networks. This integrative approach provides a comprehensive view of the genetic and molecular mechanisms driving milk production in dairy cattle.

6 Case Studies

6.1 Successful GWAS applications in improving dairy breeds

Genome-wide association studies (GWAS) have significantly contributed to the improvement of dairy breeds by identifying quantitative trait loci (QTL) associated with key production traits. For instance, a study comparing within-breed and multibreed GWAS in French and Danish dairy cattle demonstrated that multibreed GWAS could enhance the detection and fine mapping of QTL, thereby increasing the power and precision of genetic selection (Berg et al., 2016). Similarly, a meta-analysis of GWAS in Holstein cows across different countries identified numerous QTLs for milk yield, fat percentage, and protein percentage, highlighting the potential of integrating multiple studies to improve genetic evaluations (Bakhshalizadeh et al., 2021). These findings underscore the utility of GWAS in pinpointing genetic variants that can be targeted for breeding programs to enhance milk production traits.

6.2 GWAS in holstein cattle: a comprehensive study

Holstein cattle, being one of the most prominent dairy breeds, have been extensively studied using GWAS to understand the genetic basis of milk production traits. A longitudinal GWAS in a Chinese Holstein population identified numerous QTL regions associated with milk yield, fat percentage, and protein percentage, providing insights into candidate genes such as *DGAT1*, HSF1, and *MGST1* (Teng et al., 2023). Another study in Brazilian Holstein cattle identified genomic regions associated with traditional milk production traits under tropical conditions, confirming well-known genes like *MGST1* and *DGAT1*, and discovering novel genes involved in mammary gland repair and immune response (Iung et al., 2019). These comprehensive studies in Holstein cattle have not only validated previously known genetic markers but also uncovered new candidate genes, thereby enhancing our understanding of the genetic architecture of milk production traits.

6.3 Case study of crossbreeding programs influenced by GWAS findings

Crossbreeding programs have also benefited from GWAS findings, particularly in breeds like Girolando, which is a cross between Gir and Holstein cattle. A study on Girolando cattle identified QTL and candidate genes associated with 305-day milk yield, revealing significant genetic variance attributed to breed-specific SNP alleles (Otto et al., 2020) (Figure 2). This information is crucial for designing genomic selection panels tailored to crossbred populations, thereby optimizing milk production traits. Additionally, multibreed GWAS have shown that combining data from different breeds can refine QTL mapping, as demonstrated in a study involving Holstein and Jersey cattle, which identified small genomic intervals and potential candidate genes for milk production (Bindea et al., 2009). These case studies illustrate how GWAS findings can be leveraged to enhance crossbreeding programs, ultimately leading to improved dairy cattle performance.

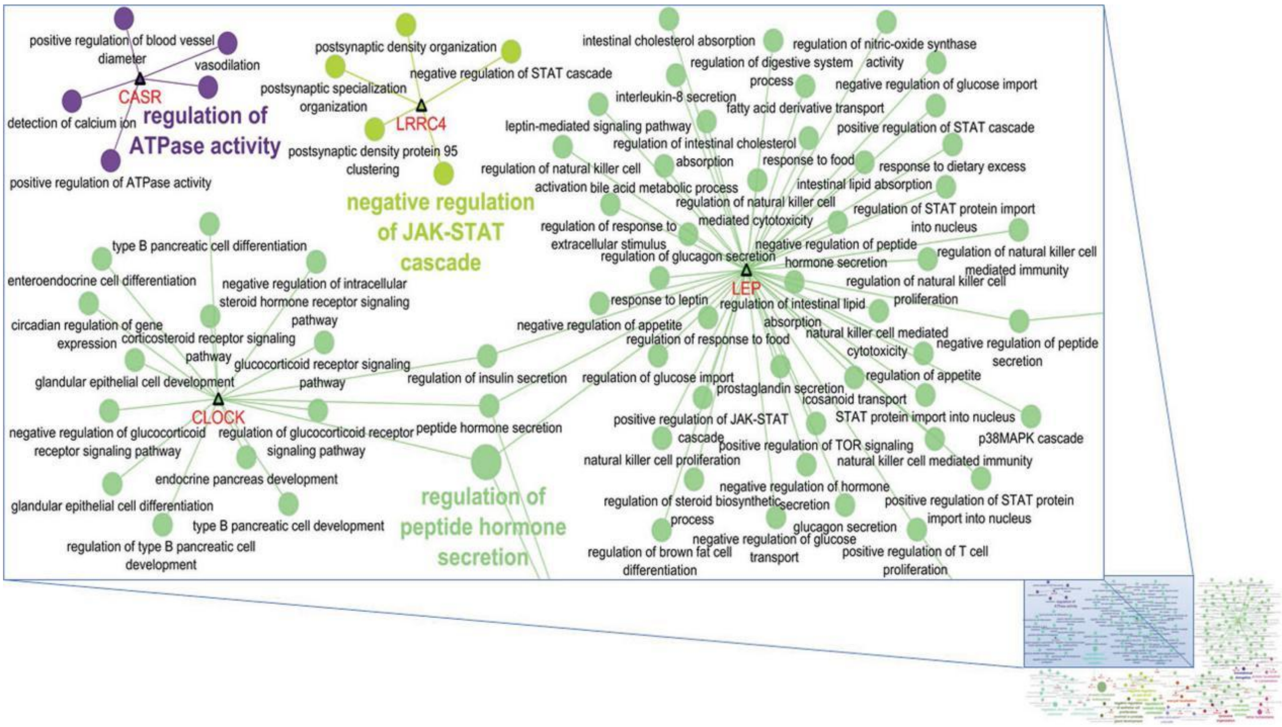

Figure 2 Functional networks showing gene interactions (triangle nodes) related to 305-d milk yield and the relationships across genes and their subnetworks related to regulation of hormone secretion, ATPase activity, and the JAK/STAT cascade (Janus kinase/signal transducers and activators of transcription) in the Girolando population. Node size denotes term enrichment significance, from the ClueGO Cytoscape plug-in (Bindea et al., 2009). The most enriched terms per group are shown in bold purple, green, and light green letters, whereas the sub-biological process terms are labeled in black (Adopted from Otto et al., 2020)

7 Challenges and Limitations of GWAS in Milk Production

7.1 Population structure and genetic diversity

One of the primary challenges in conducting genome-wide association studies (GWAS) for milk production traits in dairy cattle is accounting for population structure and genetic diversity. Different cattle breeds and even subpopulations within a breed can exhibit significant genetic variation, which can confound GWAS results. For instance, combining data from multiple populations can enhance the detection power of GWAS, but it also introduces complexity due to varying genetic backgrounds (Gebreyesus et al., 2019). Additionally, environmental conditions, such as those found in tropical climates, can influence the expression of milk production traits, making it difficult to confirm GWAS findings across different populations (Coutinho et al., 2019).

7.2 Issues of false positives and statistical power

False positives and limited statistical power are significant issues in GWAS. The detection of false positives can be exacerbated by the large number of single nucleotide polymorphisms (SNPs) tested simultaneously. Multivariate GWAS methods have been shown to improve the power to detect true associations without increasing the false discovery rate (Bolormaa et al., 2010). However, the sample size remains a critical factor affecting statistical power. Combining datasets from different populations or conducting meta-analyses can help increase sample sizes and improve the robustness of GWAS findings (Gebreyesus et al., 2019; Bakhshalizadeh et al., 2021). Despite these strategies, the risk of false positives persists, necessitating rigorous validation of identified SNPs in independent populations (Pryce et al., 2010).

7.3 Challenges in identifying causative variants

Identifying causative variants among the numerous associated SNPs remains a daunting task. Many GWAS identify regions of the genome associated with traits, but pinpointing the exact causative mutations requires further fine-mapping and functional studies. For example, while some studies have successfully narrowed down genomic intervals containing causative mutations using haplotypes, the precision of quantitative trait loci (QTL) mapping can still be limited (Pryce et al., 2010). Longitudinal GWAS and the use of whole-genome sequence data can provide more detailed insights, but they also require high-quality imputation and extensive computational resources (Teng et al., 2023). Moreover, the polygenic nature of milk production traits, where multiple genes contribute small effects, complicates the identification of individual causative variants (Buaban et al., 2021).

8 Integrating GWAS with Other Genomic Tools

8.1 Use of genomic selection and marker-assisted selection (MAS)

Genomic selection (GS) and marker-assisted selection (MAS) have revolutionized dairy cattle breeding by leveraging the power of genome-wide association studies (GWAS). GS uses genomic breeding values (GEBV) calculated from dense genetic markers across the genome, capturing the effects of quantitative trait loci (QTL) that contribute to variation in traits such as milk production. This method has shown significant improvements in the reliability of breeding values, leading to faster genetic gains compared to traditional selection methods (Hayes et al., 2009). MAS, on the other hand, focuses on specific markers associated with desirable traits identified through GWAS. For instance, the identification of SNP markers and haplotypes associated with milk production traits has been validated across different breeds, enhancing the precision of QTL mapping and the effectiveness of MAS (Pryce et al., 2010; Gutierrez-Reinoso et al., 2021). The integration of these genomic tools into breeding programs has not only increased productivity but also helped mitigate issues like inbreeding depression by providing more accurate selection criteria (Gutierrez-Reinoso et al., 2021).

8.2 Combining GWAS with epigenomics and metabolomics

The integration of GWAS with epigenomics and metabolomics offers a comprehensive approach to understanding the genetic architecture of milk production traits. Epigenomics studies the modifications on DNA and histones that affect gene expression without altering the DNA sequence, providing insights into gene regulation mechanisms. Metabolomics, which involves the large-scale study of small molecules (metabolites) within cells, tissues, or organisms, can reveal the biochemical pathways influenced by genetic variations. Combining these approaches with GWAS can identify not only the genetic variants associated with milk production but also how these variants influence metabolic pathways and gene expression. For example, studies have identified candidate genes and biological networks related to milk production and somatic cell score (SCS), implicating pathways such as intracellular cell transportation and protein catabolism (Buaban et al., 2021). This holistic approach can lead to the discovery of novel genes and pathways, offering new targets for genetic improvement and better understanding of the underlying biology of milk production traits (Marina et al., 2021).

8.3 Future prospects of integrating GWAS with machine learning techniques

The future of integrating GWAS with machine learning (ML) techniques holds great promise for enhancing the accuracy and efficiency of genomic predictions. Machine learning algorithms can handle large and complex datasets, making them ideal for analyzing the vast amount of data generated by GWAS. These techniques can identify complex patterns and interactions between genetic variants that traditional statistical methods might miss. For instance, ML can improve the prediction accuracy of GEBV by incorporating non-linear relationships and interactions among SNPs (Hayes et al., 2009). Additionally, ML can be used to refine the selection of candidate variants for genomic evaluation, as demonstrated by studies that have used meta-analysis and joint analyses to select the most relevant SNPs for milk production traits (Berg et al., 2016; Teissier et al., 2018). The integration of

ML with GWAS can also facilitate the development of more sophisticated models that account for epistatic interactions and environmental factors, ultimately leading to more precise and robust genomic selection strategies (Taherkhani et al., 2022; Sahana et al., 2023).

9 Future Perspectives and Research Directions

9.1 Emerging technologies in GWAS

The field of genome-wide association studies (GWAS) is rapidly evolving with the advent of new technologies and methodologies. One significant advancement is the use of high-throughput SNP genotyping technologies, such as the Illumina BovineSNP50 BeadChip, which has broadened the scope for identifying genes associated with milk production traits in dairy cattle (Jiang et al., 2010). Additionally, the integration of whole-genome sequencing data in GWAS has enhanced the precision and power of detecting quantitative trait loci (QTL) (Berg et al., 2016). Meta-analysis methods, which combine data from multiple independent studies, have also proven effective in increasing the accuracy of QTL detection and understanding the biological mechanisms underlying milk production traits (Bakhshalizadeh et al., 2021). Furthermore, the development of single-step genomic best linear unbiased prediction (ssGBLUP) methods has improved the accuracy of genomic predictions, even in populations with limited genotyped animals (Buaban et al., 2021).

9.2 Prospects of precision breeding in dairy cattle

Precision breeding, which leverages genomic information to select superior candidates, holds great promise for the future of dairy cattle breeding. The use of genomic breeding values (GEBVs) has already shown significant improvements in dairy cattle productivity by enabling more accurate selection of animals with desirable traits (Gutierrez-Reinoso et al., 2021). Multibreed GWAS approaches have demonstrated the potential to refine QTL mapping and identify causal variants with greater precision, especially when QTL are segregating across breeds (Raven et al., 2014). This approach can be particularly useful in identifying candidate genes and refining QTL intervals, as seen in studies involving Holstein and Jersey cattle (Jiang et al., 2010; Pryce et al., 2010). The integration of genomic selection with precision management practices on modern dairy farms can further optimize breeding programs and enhance food production efficiency (Gutierrez-Reinoso et al., 2021).

9.3 Ethical and regulatory considerations in genetic research

As the field of genetic research in dairy cattle advances, it is crucial to address the ethical and regulatory considerations associated with these technologies. One major concern is the potential for inbreeding depression due to the closed bloodlines in several milk breeds, which can negatively impact reproductive performance and overall genetic diversity (Gutierrez-Reinoso et al., 2021). To mitigate these effects, it is essential to implement strategies that promote genetic diversity and avoid excessive inbreeding. Additionally, the use of genome editing technologies, such as CRISPR, raises ethical questions regarding animal welfare and the long-term impacts on the genetic makeup of dairy cattle. Regulatory frameworks must be established to ensure that genetic research and breeding practices adhere to ethical standards and promote the well-being of the animals involved. Furthermore, transparent communication with stakeholders, including farmers, consumers, and policymakers, is necessary to build trust and ensure the responsible use of genetic technologies in dairy cattle breeding.

10 Concluding Remarks

Genome-wide association studies (GWAS) have transformed our understanding of the genetic basis of milk production traits in dairy cattle. Extensive research has uncovered numerous genomic regions and candidate genes linked to key traits such as milk yield, fat percentage, protein percentage, and somatic cell score (SCS). For instance, GWAS in Thai dairy cattle identified 210 candidate genes across 19 chromosomes associated with milk production traits and 21 genes on 3 chromosomes related to SCS. Similarly, a longitudinal GWAS in Chinese Holstein cattle pinpointed multiple quantitative trait loci (QTLs) and identified 28 candidate genes, including prominent genes such as *DGAT1* and novel ones like *CCSER1* and *CUX2*. Meta-analyses have further validated QTLs across diverse populations, enhancing the precision and reliability of genetic studies. In addition, research on Brazilian Holstein and Danish Jersey cattle has reinforced the significance of well-known genes like *MGST1*, *ABCG2*, and *DGAT1*, while also uncovering new genes involved in crucial biological pathways.

These findings offer transformative potential for the dairy industry. Identifying specific genomic regions and candidate genes associated with milk production traits enables the refinement of genomic selection strategies. Incorporating these genetic markers into breeding programs can substantially improve selection accuracy and accelerate genetic gains. For example, applying weighted single-step GWAS has demonstrated superior genomic prediction accuracy compared to traditional methods. Furthermore, novel genes linked to traits such as heat tolerance, longevity, and fertility present opportunities to breed more resilient and productive dairy cattle.

The integration of these genetic insights into breeding programs can facilitate the creation of customized SNP arrays tailored to target specific traits under varying environmental conditions. These advancements in GWAS and their practical applications hold the promise to revolutionize dairy cattle breeding. By enabling higher milk production, better animal health, and increased economic returns, GWAS findings contribute to sustainable and profitable dairy farming practices.

Acknowledgements

The author is grateful to the reviewers for their in-depth analysis and detailed feedback during the review process.

Conflict of Interest Disclosure

Author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Bakhshalizadeh S., Zerehdaran S., and Javadmanesh A., 2021, Meta-analysis of genome-wide association studies and gene networks analysis for milk production traits in Holstein cows, Livestock Science, 250: 104605.
https://doi.org/10.1016/j.livsci.2021.104605

Berg I., Boichard D., and Lund M., 2016, Comparing power and precision of within-breed and multibreed genome-wide association studies of production traits using whole-genome sequence data for 5 French and Danish dairy cattle breeds, Journal of Dairy Science, 99(11): 8932-8945.
https://doi.org/10.3168/jds.2016-11073
PMid:27568046

Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., Kirilovsky A., Fridman W.H., Pagès F., Trajanoski Z., and Galon J., 2009, ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics, 25(8): 1091-1093.
https://doi.org/10.1093/bioinformatics/btp101
PMid:19237447 PMCid:PMC2666812

Bolormaa S., Pryce J., Hayes B., and Goddard M., 2010, Multivariate analysis of a genome-wide association study in dairy cattle, Journal of Dairy Science, 93(8): 3818-3833.
https://doi.org/10.3168/jds.2009-2980
PMid:20655452

Buaban S., Lengnudum K., Boonkum W., and Phakdeedindan P., 2021, 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.
https://doi.org/10.3168/jds.2020-19826
PMid:34756438

Chen Z., Yao Y., Ma P., Wang Q., and Pan Y., 2018, Haplotype-based genome-wide association study identifies loci and candidate genes for milk yield in Holsteins. PLoS ONE, 13(2): e0192695.
https://doi.org/10.1371/journal.pone.0192695
PMid:29447209 PMCid:PMC5813974

Dadousis C., Biffani S., Cipolat-Gotet C., Nicolazzi E., Rosa G., Gianola D., Rossoni A., Santus E., Bittante G., and Cecchinato A., 2017a, Genome-wide association study for cheese yield and curd nutrient recovery in dairy cows, Journal of Dairy Science, 100(2): 1259-1271.
https://doi.org/10.3168/jds.2016-11586
PMid:27889122

Dadousis C., Pegolo S., Rosa G., Gianola D., Bittante G., and Cecchinato A., 2017b, Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield, and curd nutrient recovery in dairy cattle, Journal of Dairy Science, 100(2): 1223-1231.
https://doi.org/10.3168/jds.2016-11587
PMid:27988128

Gebreyesus G., Gebreyesus G., Buitenhuis A., Poulsen N., Visker M., Zhang Q., Valenberg H., Sun D., and Bovenhuis H., 2019, Combining multi-population datasets for joint genome-wide association and meta-analyses: the case of bovine milk fat composition traits, Journal of Dairy Science, 102(12): 11124-11141.
https://doi.org/10.3168/jds.2019-16676
PMid:31563305

Gutierrez-Reinoso M., Aponte P., and García-Herreros M., 2021, Genomic analysis, progress and future perspectives in dairy cattle selection: a review, Animals, 11(3): 599.
https://doi.org/10.3390/ani11030599
PMid:33668747 PMCid:PMC7996307

Hayes B., Bowman P., Chamberlain A., and Goddard M., 2009, Invited review: genomic selection in dairy cattle: progress and challenges, Journal of Dairy Science, 92(2): 433-443.
https://doi.org/10.3168/jds.2008-1646
PMid:19164653

Iung L., Petrini J., Ramírez-Díaz J., Salvian M., Rovadoscki G., Pilonetto F., Dauria B., Machado P., Coutinho L., Wiggans G., and Mourão G., 2019, Genome-wide association study for milk production traits in a Brazilian Holstein population, Journal of Dairy Science, 102(6): 5305-5314.
https://doi.org/10.3168/jds.2018-14811
PMid:30904307

Jiang J., Ma L., Prakapenka D., VanRaden P., Cole J., and Da Y., 2019, A large-scale genome-wide association study in U.S. holstein cattle, Frontiers in Genetics, 10: 412.
https://doi.org/10.3389/fgene.2019.00412
PMid:31139206 PMCid:PMC6527781

Jiang L., Liu J., Sun D., Ma P., Ding X., Yu Y., and Zhang Q., 2010, Genome wide association studies for milk production traits in chinese holstein population, PLoS ONE, 5(10): e13661.
https://doi.org/10.1371/journal.pone.0013661
PMid:21048968 PMCid:PMC2965099

Kim S., Lim B., Cho J., Lee S., Dang C., Jeon J., Kim J., and Lee J., 2021, Genome-wide identification of candidate genes for milk production traits in korean holstein cattle, Animals, 11(5): 1392.
https://doi.org/10.3390/ani11051392
PMid:34068321 PMCid:PMC8153329

Laodiim T., Koonawootrittriron S., Elzo M., Suwanasopee T., Jattawa D., and Sarakul M., 2023, Genetic factors influencing milk and fat yields in tropically adapted dairy cattle: insights from QTL analysis and gene associations, Animal Bioscience, 37(4): 576.
https://doi.org/10.5713/ab.23.0246
PMid:37946425 PMCid:PMC10915225

Ma L., 2020, 301 Methods of genome-wide association studies and their applications in dairy cattle, Journal of Animal Science, 98: 31-31.
https://doi.org/10.1093/jas/skaa278.055
PMCid:PMC7702888

Mai M., Sahana G., Christiansen F., and Guldbrandtsen B., 2010, A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip, Journal of Animal Science, 88(11): 3522-3528.
https://doi.org/10.2527/jas.2009-2713
PMid:20656975

Marina H., Pelayo R., Suárez-Vega A., Gutiérrez-Gil B., Esteban-Blanco C., and Arranz J., 2021, Genome-wide association studies (GWAS) and post-GWAS analyses for technological traits in Assaf and Churra dairy breeds, Journal of Dairy Science, 104(11): 11850-11866.
https://doi.org/10.3168/jds.2021-20510
PMid:34454756

Otto P., Guimarães S., Calus M., Vandenplas J., Machado M., Panetto J., and Silva M., 2020, Single-step genome-wide association studies (GWAS) and post-GWAS analyses to identify genomic regions and candidate genes for milk yield in Brazilian Girolando cattle, Journal of Dairy Science, 103(11): 10347-10360.
https://doi.org/10.3168/jds.2019-17890
PMid:32896396

Pryce J., Bolormaa S., Chamberlain A., Bowman P., Savin K., Goddard M., and Hayes B., 2010, A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes, Journal of Dairy Science, 93(7): 3331-3345.
https://doi.org/10.3168/jds.2009-2893
PMid:20630249

Raven L., Cocks B., and Hayes B., 2014, Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle, BMC Genomics, 15: 62.
https://doi.org/10.1186/1471-2164-15-62
PMid:24456127 PMCid:PMC3905911

Sahana G., Cai Z., Sanchez M., Bouwman A., and Boichard D., 2023, Invited review: Good practices in genome-wide association studies to identify candidate sequence variants in dairy cattle, Journal of Dairy Science, 106(8): 5218-5241.
https://doi.org/10.3168/jds.2022-22694
PMid:37349208

Taherkhani L., Banabazi M., Emamjomeh-Kashan N., Noshary A., and Imumorin I., 2022, The candidate chromosomal regions responsible for milk yield of cow: a GWAS meta-analysis, Animals, 12(5): 582.
https://doi.org/10.3390/ani12050582
PMid:35268150 PMCid:PMC8909671

Teissier M., Sanchez M., Boussaha M., Barbat A., Hozé C., Robert-Granié C., and Croiseau P., 2018, Use of meta-analyses and joint analyses to select variants in whole genome sequences for genomic evaluation: an application in milk production of French dairy cattle breeds, Journal of Dairy Science, 101(4): 3126-3139.
https://doi.org/10.3168/jds.2017-13587
PMid:29428760

Teng J., Wang D., Zhao C., Zhang X., Chen Z., Liu J., Sun D., Tang H., Wang W., Li J., Mei C., Yang Z., Ning C., and Zhang Q., 2023, Longitudinal genome-wide association studies of milk production traits in Holstein cattle using whole-genome sequence data imputed from medium-density chip data, Journal of Dairy Science, 106(4): 2535-2550.
https://doi.org/10.3168/jds.2022-22277
PMid:36797187

Wang D., Ning C., Liu J., Zhang Q., and Jiang L., 2019, Short communication: Replication of genome-wide association studies for milk production traits in Chinese Holstein by an efficient rotated linear mixed model, Journal of Dairy Science, 102(3): 2378-2383.
https://doi.org/10.3168/jds.2018-15298
PMid:30639022

Wang T., Li J., Gao X., Song W., Chen C., Yao D., Ma J., Xu L., and Ma Y., 2020, Genome-wide association study of milk components in Chinese Holstein cows using single nucleotide polymorphism, Livestock Science, 233: 103951.
https://doi.org/10.1016/j.livsci.2020.103951

Yang B., 2024, Enhancing dairy cow milk fat synthesis genes with CRISPR-Cas9 technology to increase dairy product yield, International Journal of Molecular Veterinary Research, 14(1): 9-16.
https://doi.org/10.5376/ijmvr.2024.14.0002

Zhou C., Li C., Cai W., Liu S., Yin H., Shi S., Zhang Q., and Zhang S., 2019, Genome-wide association study for milk protein composition traits in a chinese holstein population using a single-step approach, Frontiers in Genetics, 10: 72.
https://doi.org/10.3389/fgene.2019.00072
PMid:30838020 PMCid:PMC6389681