FROM GENOMES TO THE FUTURE: REDEFINING ANIMAL BREEDING

Prof.Dr. Sena Ardicli

Bursa Uludag University, Faculty of Veterinary Medicine, Department of Genetics, Bursa, Turkiye

Abstract

The integration of genomic information into dairy cattle breeding programs has transformed animal improvement strategies during the last two decades. The development of dense single-nucleotide polymorphism (SNP) arrays and advances in computational genetics have enabled the prediction of genomic breeding values with unprecedented accuracy, particularly in young animals lacking phenotypic records. Genomic selection has substantially accelerated genetic gain by reducing generation intervals and improving selection intensity. Furthermore, the increasing availability of whole-genome sequence data, large reference populations, and high-throughput phenotyping platforms has expanded the potential of genomic prediction beyond traditional production traits toward fertility, health, resilience, and environmental adaptation. Recent developments in genome-wide association studies, fine-mapping approaches, machine learning algorithms, and multi-omics integration have further enhanced the ability to identify biologically relevant variants and improve prediction performance. However, several challenges remain, including ascertainment bias in SNP arrays, maintenance of genetic diversity, long-term sustainability of breeding programs, and the integration of causal variants into routine evaluations. The emergence of transcriptomics, functional genomics, and artificial intelligence-driven prediction models provides new opportunities to overcome these limitations and improve the biological interpretation of genomic evaluations.

Keywords: ruminants, genomic selection, genomic prediction, GWAS, transcriptomics, breeding value

Introduction

Genetic improvement has been one of the major driving forces behind the remarkable increase in productivity observed over the last century in ruminants, especially dairy cattle. Traditional breeding programs relied primarily on phenotypic performance records and pedigree information to estimate breeding values through best linear unbiased prediction (BLUP), allowing systematic selection of genetically superior animals. Although this approach generated substantial improvements in milk production and other economically important traits, its effectiveness was constrained by long generation intervals, incomplete pedigree information, and limited accuracy in young animals lacking phenotypic records (Meuwissen et al., 2016).

Under conventional breeding schemes, selection decisions rely primarily on pedigree information and phenotypic records, resulting in moderate rates of genetic improvement due to limited selection accuracy and relatively long generation intervals. In contrast, genomic selection enables the identification of genetically superior animals at a much earlier age using dense genome-wide marker information. This approach substantially increases selection accuracy while simultaneously reducing generation intervals, thereby accelerating genetic gain. As a consequence, genetic progress accumulates more rapidly over time under genomic selection than under conventional breeding strategies (Figure 1).

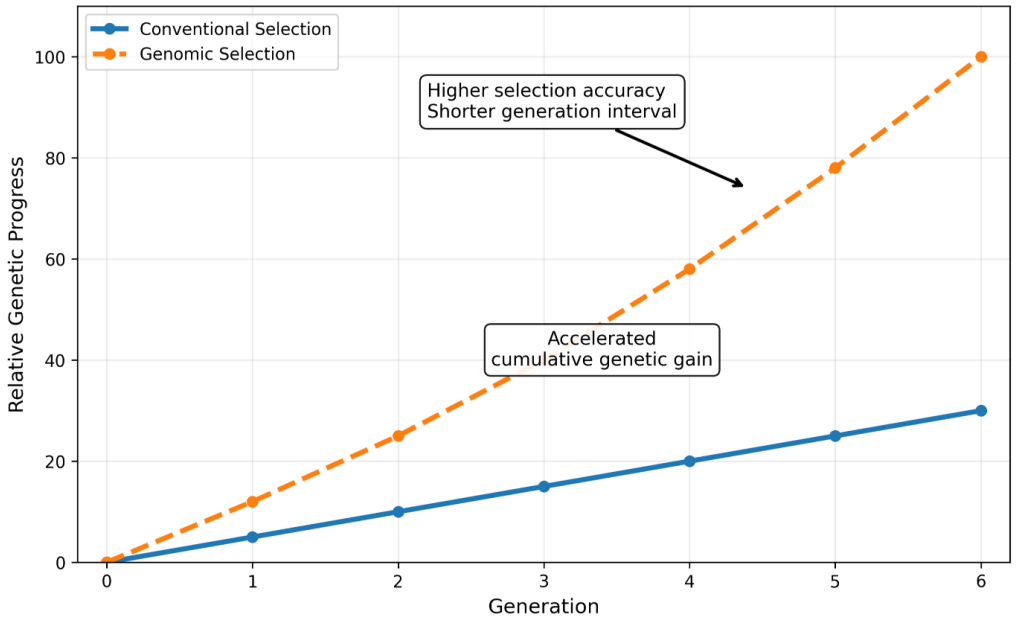


Figure 1. Conceptual comparison of conventional and genomic selection in livestock breeding.

The emergence of molecular genetics during the late twentieth century created expectations that DNA-based information would revolutionize animal breeding. Initial efforts focused on marker-assisted selection (MAS), which aimed to identify quantitative trait loci (QTL) associated with economically important traits and subsequently incorporate these loci into breeding programs. While numerous QTL mapping studies successfully identified genomic regions associated with production traits, the practical implementation of MAS was limited because most complex traits are controlled by a very large number of loci with individually small effects. Consequently, only a minor proportion of the total genetic variance could be captured by significant markers, leaving most genetic variation unexplained (Meuwissen et al., 2016).

The limitations of MAS led to the development of genomic selection, a concept that fundamentally changed the philosophy of genetic evaluation. Rather than identifying only statistically significant markers, genomic selection simultaneously estimates the effects of thousands of markers distributed across the entire genome. This approach demonstrated through simulation studies that dense marker maps could achieve breeding value prediction accuracies approaching 0.85 by capturing linkage disequilibrium between markers and QTL across the genome (Meuwissen et al., 2001). This pioneering aspect established the theoretical foundation for modern genomic prediction systems and is widely regarded as one of the most influential contributions to livestock genetics.

Parallel technological developments enabled the rapid adoption of genomic selection. The sequencing of the bovine genome facilitated the discovery of hundreds of thousands of SNP markers distributed throughout the genome. At the same time, advances in genotyping technologies dramatically reduced the cost of generating genomic data. These developments enabled the creation of high-density SNP arrays capable of providing genome-wide coverage for large populations at economically feasible costs (Hayes et al., 2009). As a result, genomic estimated breeding values (GEBVs) can be calculated using dense marker information, allowing selection decisions to be made at much earlier ages than previously possible.

The impact of genomic selection on cattle breeding has been profound. In conventional breeding programs, the genetic merit of young bulls could be accurately assessed only after progeny testing, a process that requires several years and substantial financial investment. Genomic prediction dramatically reduced this dependency by enabling accurate estimation of breeding values shortly after birth. Consequently, generation intervals decreased substantially while selection accuracy increased, leading to significantly higher rates of genetic gain (Hayes et al., 2009; Meuwissen et al., 2016). Today, genomic evaluations have become the standard approach in most major cattle breeding programs worldwide.

The success of genomic selection is largely attributable to its ability to capture the highly polygenic architecture of complex traits. Current evidence suggests that economically important dairy traits are influenced by thousands

of loci distributed across the genome rather than a small number of major genes. This realization resolved many of the limitations of earlier QTL-based approaches. It helped explain the phenomenon often described as "missing heritability," in which genome-wide association studies identified only a small proportion of the expected genetic variance despite large sample sizes (Meuwissen et al., 2016). By estimating the effects of all markers simultaneously, genomic selection captures both large and small genetic effects, thereby providing a more comprehensive representation of the genetic architecture underlying complex traits.

In parallel with advances in genomic prediction methodology, genome-wide association studies (GWAS) have become increasingly powerful tools for identifying genomic regions associated with economically important traits. The availability of large reference populations and imputed whole-genome sequence data has significantly improved the ability to detect causal variants and candidate genes. Recent large-scale analyses involving more than 50,000 Holstein bulls have identified hundreds of significant association peaks across production, fertility, health, longevity, and conformation traits, substantially expanding our understanding of the genetic architecture of dairy cattle performance (Wang et al., 2025). Furthermore, Bayesian fine-mapping approaches have improved the prioritization of candidate variants by assigning probabilities of causality to individual loci, thereby bridging the gap between association signals and biological mechanisms (Wang et al., 2025).

The increasing availability of functional genomic data has opened new opportunities to improve prediction accuracy and biological interpretation. Among these approaches, transcriptomics has emerged as a particularly promising source of information because gene expression profiles lie between genotype and phenotype. Recent studies have demonstrated that transcriptomic information can explain a substantial proportion of phenotypic variation and complement traditional genomic information when incorporated into prediction models (Haas et al., 2025). Such findings have stimulated growing interest in multi-omics approaches that integrate genomic, transcriptomic, epigenomic, proteomic, and metabolomic information within unified prediction frameworks. At the same time, advances in artificial intelligence and machine learning have created new opportunities for genomic prediction. Traditional statistical approaches, such as genomic BLUP and Bayesian regression, remain widely used for their robustness and computational efficiency. However, machine learning algorithms can model complex non-linear relationships and interactions that conventional methods may not adequately capture. Recent studies have demonstrated improved predictive performance using approaches such as gradient boosting machines, elastic net regression, and ensemble learning, particularly for complex health-related phenotypes (Mota et al., 2024). These developments suggest that artificial intelligence may play an increasingly important role in future breeding programs.

Despite the remarkable success of genomic selection, several important challenges remain. One major concern is ascertainment bias associated with SNP array design. Most commercially available bovine SNP arrays were developed primarily using *Bos taurus* reference populations, resulting in marker panels that may not adequately represent genetic diversity in other populations, particularly *Bos indicus* breeds. Such bias can influence genomic relationship matrices, reduce prediction accuracy, and complicate multi-breed evaluations (Warburton & Hayes, 2026). Addressing these limitations will be essential for expanding genomic selection across diverse cattle populations. Another important challenge concerns the long-term sustainability of genomic improvement programs. While genomic selection has substantially increased short-term genetic gain, intensive selection can accelerate inbreeding and reduce genetic diversity. Maintaining sufficient genetic variation is critical for preserving future selection potential and preventing the unfavorable consequences of inbreeding depression. Recent simulation studies have demonstrated that combining genomic prediction with diversity management strategies can improve long-term genetic gain while simultaneously reducing inbreeding rates (Pook et al., 2026). In addition to genome-wide prediction, increasing attention has been directed toward integrating specific functional markers into breeding programs. For example, polymorphisms in milk protein genes, such as beta-casein (*CSN2*) and kappa-casein (*CSN3*), have attracted considerable interest due to their associations with milk quality, processing characteristics, and consumer preferences. Recent studies in Holstein cattle have demonstrated significant relationships between casein genotypes and genomic merit indicators, including total performance index, net merit, productive life, and protein-related breeding values, highlighting the potential value of integrating functional variants into genomic selection programs (Ardicli et al., 2023; Ardicli et al., 2024).

Taken together, these developments indicate that genomic selection has evolved far beyond its original implementation based solely on dense SNP markers. Modern breeding programs increasingly combine genomic prediction with sequence-based analyses, fine-mapping strategies, functional genomics, transcriptomics, machine learning, and diversity management approaches (Figure 2).

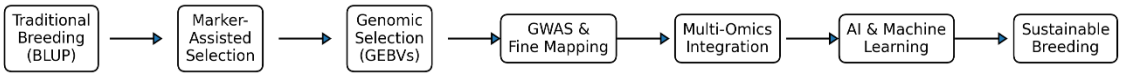


Figure 1. Evolution of animal breeding technologies from conventional pedigree-based selection to sustainable genomics-assisted breeding.

Genomic Selection Methods

The statistical foundation of genomic selection rests on the accurate estimation of marker effects across thousands of loci simultaneously. Genomic best linear unbiased prediction (GBLUP) remains the most widely implemented method in routine dairy cattle evaluations, primarily because it is computationally tractable at the population scale and robust across diverse trait architectures. In GBLUP, the pedigree-based numerator relationship matrix is replaced by a genomic relationship matrix (GRM) constructed from dense SNP markers, enabling the capture of Mendelian sampling variation that pedigree-based approaches cannot resolve (VanRaden, 2008). The method assumes that marker effects follow a normal distribution with homogeneous variance, an assumption consistent with the infinitesimal model and well suited to highly polygenic traits such as milk yield and protein content, for which no single locus explains a disproportionate fraction of the genetic variance.

Bayesian variable selection methods were developed to accommodate trait architectures that depart from the infinitesimal assumption. BayesA assigns a scaled-t before each marker, allowing locus-specific variance to differ across the genome. BayesB extends this by assuming that a proportion π of markers has no effect, effectively performing variable selection within the model. BayesCπ further generalizes this framework by estimating π from the data rather than fixing it a priori, thereby allowing the model to adapt to the observed genetic architecture (Habier et al., 2011). In practice, the advantage of Bayesian methods over GBLUP is trait-dependent, for traits influenced by loci with large individual effects, such as milk fat percentage. Bayesian approaches provide meaningfully higher prediction accuracy, whereas for traits approaching a truly polygenic architecture, the two classes of methods converge (Habier et al., 2011; Hayes et al., 2009; Meuwissen et al., 2016). A major operational advance was the development of single-step GBLUP (ssGBLUP), which integrates genotyped and non-genotyped animals into a unified evaluation through a blended H-matrix that combines the GRM with the pedigree-based relationship matrix (Aguilar et al., 2010; Christensen & Lund, 2010). This approach eliminates the need for sequential processing steps, reduces bias from preselection of genotyped animals, and has been shown to outperform two-step approaches in terms of both accuracy and unbiasedness of GEBVs. The scalability of ssGBLUP has been substantially improved by ssSNPBLUP, which reformulates the model in terms of SNP effects rather than animal effects, enabling computationally efficient evaluation of populations comprising millions of genotyped individuals. Recent large-scale implementations in beef and dairy populations involving more than 2.6 million genotyped animals have validated the operational feasibility of this approach across a diverse array of platforms (Vandenplas et al., 2023).

The composition and size of the reference population are critical determinants of GEBV reliability. Prediction accuracy increases with reference population size, the heritability of the target trait, and the extent of linkage disequilibrium between markers and causal variants. For production traits with moderate to high heritability, reliabilities of 0.50–0.85 are routinely achieved in large Holstein reference populations, whereas fertility and health traits, which carry lower heritability and more complex genetic architectures, typically yield reliabilities of 0.20–0.50 (Wiggans et al., 2017). Cross-population genomic prediction (for example, applying equations derived in Holstein to Jersey or *Bos indicus* populations) remains substantially less accurate due to differences in linkage disequilibrium phase and allele frequency between breeds, underscoring the importance of breed-specific or multi-breed reference designs (Hayes et al., 2009). Genotype imputation has been central to the cost-effective expansion of reference populations. Animals genotyped on low-density arrays (6K–50K SNPs) can be imputed to high-density (777K) or whole-genome sequence resolution using population-based algorithms, with imputation accuracy from 50K to high-density panels routinely exceeding 97% in well-structured Holstein populations. The use of imputed whole-genome sequence data as input for genomic prediction has shown modest gains in accuracy for traits with known large-effect variants, where causal mutations or their close proxies are directly represented in the sequence data. Still, it provides little benefit over high-density array genotypes for purely polygenic traits (Wang et al., 2025). Nevertheless, whole-genome sequence data are indispensable for downstream applications

such as fine-mapping and the identification of candidate causal variants, which can subsequently be incorporated into weighted GBLUP models to improve prediction for traits with identifiable major loci (Fragomeni et al., 2017).

Emerging Technologies: Integration of Multi-Omics Approaches in Genomic Prediction

The statistical foundations of genomic selection, primarily driven by dense SNP arrays and methods such as GBLUP and Bayesian models, have traditionally captured the linear relationship between genotype and phenotype. While these methods have revolutionized dairy cattle breeding, they inherently treat the intermediate biological mechanisms between DNA variation and ultimate phenotypic performance as a "black box". To overcome these limitations and improve the biological interpretation of genomic evaluations, modern livestock genetics is shifting toward systems biology by integrating multi-omics data. By incorporating transcriptomic, epigenomic, proteomic, and metabolomic layers into unified prediction frameworks, contemporary breeding programs aim to capture the dynamic physiological responses that drive complex traits in modern dairy production systems.

Transcriptomics as a Biological Bridge

Among emerging functional genomic resources, transcriptomics has proven particularly promising for expanding prediction frameworks, as gene expression profiles lie directly between genotype and phenotype. High-throughput sequencing data capture the temporal and tissue-specific activation of gene networks in response to physiological states. Recent studies have demonstrated that transcriptomic information can explain a substantial proportion of phenotypic variation and complement traditional genomic information when incorporated into prediction models. Operationally, integrating transcriptomic data allows the evaluation model to go beyond static DNA profiles. Biological priorities or weights can be assigned to specific genomic regions or neighboring SNP markers within a weighted GBLUP framework based on expression profiles, effectively prioritizing biologically relevant variants over unannotated markers. Alternatively, multi-kernel models complement the traditional GRM by incorporating a transcriptomic relationship matrix derived from expression similarities. This architecture allows the prediction equations to capture both additive genetic variances and the regulatory variations reflected in the transcriptome (Haas et al., 2025).

Epigenomic, Proteomic, and Metabolomic Layer Synergy

Beyond mRNA expression, integrating other omics layers provides a more comprehensive representation of the biological architecture underlying complex, environmentally sensitive traits.

Epigenomics: DNA methylation and histone modifications play crucial roles in regulating gene expression without altering the underlying DNA sequence. Integrating epigenomic marks into prediction models is highly promising for low-heritability traits such as fertility, health, resilience, and environmental adaptation, where conventional SNP markers often yield lower prediction reliability due to strong environmental interactions.

Proteomics and Metabolomics: The proteome and metabolome reflect the end-products of cellular processes. Critical traits such as milk protein composition, quality, and processing characteristics are directly dictated by circulating metabolites and protein expressions.

Sustainability, Genetic Diversity, and Long-Term Genetic Gain

The remarkable success of genomic selection has accelerated genetic improvement across livestock species; however, it has also raised concerns regarding the long-term sustainability of breeding programs. By increasing selection accuracy and intensity, genomic selection enables rapid accumulation of favorable alleles within populations. While this leads to substantial short-term genetic gains, it may simultaneously reduce genetic diversity and increase inbreeding rates if not carefully managed (Bertorelle et al., 2022; Doekes et al., 2021; Pook et al., 2026; Ramu et al., 2017). Consequently, modern breeding strategies must balance immediate genetic progress with the preservation of sufficient genetic variation to ensure continued improvement across future generations. Genetic diversity represents the biological foundation upon which all selection programs operate. The response to selection depends not only on the accuracy of breeding value estimation but also on the amount of additive genetic variation available within the population. Excessive reduction of genetic diversity can decrease heritability, increase the frequency of deleterious alleles, and enhance the expression of inbreeding depression, ultimately compromising both productivity and animal fitness (Hidalgo et al., 2020; Pook et al., 2026; VanRaden et al., 2014). These concerns are particularly relevant in dairy cattle breeding, where widespread use of elite sires through artificial insemination has substantially increased the genetic contribution of a relatively small number

of individuals to subsequent generations. Sustainability in animal breeding extends beyond the preservation of genetic diversity and increasingly encompasses adaptation to climate change, emerging diseases, evolving production systems, and societal demands for animal welfare. Consequently, modern breeding objectives must balance production with resilience, robustness, fertility, health, feed efficiency, and environmental adaptability. Genomic selection provides an effective framework for simultaneously improving these complex traits through multi-trait selection indices (Meuwissen et al., 2016). Furthermore, integrating whole-genome sequence data, causal variants, and multi-omics information offers new opportunities to enhance sustainable breeding by improving the identification of biological mechanisms underlying resilience and adaptation (Haas et al., 2025). Together, these advances are expected to facilitate more sustainable and future-oriented genomic evaluation systems.

Conclusions

The future success of livestock breeding programs will depend not only on maximizing genetic gain but also on preserving the evolutionary potential of animal populations. Sustainable breeding requires a holistic approach that combines genomic prediction, diversity management, functional genomics, and carefully designed mating strategies. The integration of these components will help ensure that future generations of livestock remain productive, resilient, and genetically diverse amid rapidly changing environmental and production challenges.

References

- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science, 93(2), 743-752.
- Ardicli, S., Aldevir, O., Aksu, E., & Gumen, A. (2023). The variation in the beta-casein genotypes and its effect on milk yield and genomic values in Holstein-Friesian cows. Animal Biotechnology, 34(8), 4116-4125.
- Ardicli, S., Aldevir, Ö., Aksu, E., Kucuk, K., & Gümen, A. (2024). Associations of bovine beta-casein and kappa-casein genotypes with genomic merit in Holstein Friesian cattle. Archives Animal Breeding, 67(1), 61-71.
- Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H. E., & Van Oosterhout, C. (2022). Genetic load: genomic estimates and applications in non-model animals. Nature Reviews Genetics, 23(8), 492-503.
- Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. Genetics Selection Evolution, 42(1), 2.
- Doekes, H. P., Bijma, P., & Windig, J. J. (2021). How depressing is inbreeding? A meta-analysis of 30 years of research on the effects of inbreeding in livestock. Genes, 12(6), 926.
- Fragomeni, B. O., Lourenco, D. A., Masuda, Y., Legarra, A., & Misztal, I. (2017). Incorporation of causative quantitative trait nucleotides in single-step GBLUP. Genetics Selection Evolution, 49(1), 59.
- Haas, V. P., Wellmann, R., Duenk, P., Oster, M., Ponsuksili, S., Bennewitz, J., & Calus, M. P. (2025). Incorporating transcriptomic data into genomic prediction models to improve the prediction accuracy of phenotypes of efficiency traits. Genetics Selection Evolution, 57(1), 59.
- Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. BMC bioinformatics, 12(1), 186.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science, 92(2), 433-443.
- Hidalgo, J., Tsuruta, S., Lourenco, D., Masuda, Y., Huang, Y., Gray, K. A., & Misztal, I. (2020). Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. Journal of animal science, 98(2), skaa032.
- Meuwissen, T., Hayes, B., & Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. Animal frontiers, 6(1), 6-14.
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. genetics, 157(4), 1819-1829.
- Mota, L. F., Giannuzzi, D., Pegolo, S., Sturaro, E., Gianola, D., Negrini, R., Trevisi, E., Ajmone Marsan, P., & Cecchinato, A. (2024). Genomic prediction of blood biomarkers of metabolic disorders in Holstein cattle using parametric and nonparametric models. Genetics Selection Evolution, 56(1), 31.
- Pook, T., Hassanpour, A., Niehoff, T. A., & Calus, M. P. (2026). Strategies to improve on selection based on estimated breeding values. Genetics Selection Evolution.
- Ramu, P., Esuma, W., Kawuki, R., Rabbi, I. Y., Egesi, C., Bredeson, J. V., Bart, R. S., Verma, J., Buckler, E. S., & Lu, F. (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. Nature Genetics, 49(6), 959-963.
- Vandenplas, J., Ten Napel, J., Darbaghshahi, S. N., Evans, R., Calus, M. P., Veerkamp, R., Cromie, A., Mäntysaari, E. A., & Strandén, I. (2023). Efficient large-scale single-step evaluations and indirect genomic prediction of genotyped selection candidates. Genetics Selection Evolution, 55(1), 37.

- VanRaden, P., Tooker, M., Wright, J., Sun, C., & Hutchison, J. (2014). Comparison of single-trait to multi-trait national evaluations for yield, health, and fertility. Journal of Dairy Science, 97(12), 7952-7962.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science, 91(11), 4414-4423.
- Wang, J., Gao, Y., Toghiani, S., Cole, J. B., Maltecca, C., Ma, L., & Jiang, J. (2025). Genome-wide association study and fine-mapping using imputed sequences to prioritize candidate genes for 30 complex traits in 50,309 Holstein bulls. Journal of Dairy Science.
- Warburton, C. L., & Hayes, B. J. (2026). Ascertainment Bias in Cattle SNP Arrays and Implications for Multibreed Genomic Predictions. Animal Genetics, 57(2), e70081.
- Wiggans, G. R., Cole, J. B., Hubbard, S. M., & Sonstegard, T. S. (2017). Genomic selection in dairy cattle: the USDA experience. Annual review of animal biosciences, 5, 309-327.