International Journal of Animal Biotechnology
E-ISSN 2277-4122
https://ijab.evegenis.org/

Review Article

Genomic Selection in Dairy Cattle: Principles, Implementation, and Future Perspectives

Henk Bovenhuis[1]* and Jean-Jacques Colleau[2]

[1]Animal Breeding and Genomics Centre, Wageningen University & Research, The Netherlands
[2]INRA, UMR 1313 Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France

Received 01 Nov 2014, Accepted 10 Oct 2014, Available online 01 Dec 2014, **Vol.4 (2014)**

Abstract

Genomic selection (GS) has fundamentally transformed dairy cattle breeding by enabling highly accurate prediction of breeding values in young animals long before progeny testing data become available. First proposed theoretically by Meuwissen et al. (2001) and implemented commercially in 2008–2009 in the United States and Western Europe, GS exploits the linkage disequilibrium between dense single nucleotide polymorphism (SNP) markers distributed across the genome and quantitative trait loci to estimate genomic estimated breeding values (GEBVs). This review comprehensively examines the theoretical underpinnings of GS, the statistical models employed (GBLUP, ssGBLUP, Bayesian alphabet methods), the SNP genotyping platforms currently in use, the critical role of reference population size and composition, the impact of imputation on cost reduction, across-country genomic evaluations coordinated through Interbull, and the economic consequences of adopting GS in place of progeny testing. Empirical data from Holstein, Jersey, Brown Swiss, and several Nordic breeds confirm that GS has approximately doubled the rate of genetic gain per year while simultaneously reducing progeny testing costs by over 90%. The review also addresses current limitations — including reduced accuracy for low-heritability and female traits, limited performance for minor breeds with small reference populations, and the challenge of accounting for genotype-by-environment interaction and proposes research priorities for the next decade, including whole-genome sequence-based selection, functional annotation of genomic regions through the FAANG initiative, and development of low-cost genotyping solutions for developing-country dairy systems.

Keywords: *Genomic selection, Dairy cattle, Single nucleotide polymorphism, GBLUP, ssGBLUP, Reference population, Genetic gain, Imputation, Reliability, Linkage disequilibrium*

1. Introduction

The genetic improvement of dairy cattle has constituted one of the great success stories of applied animal science. Over the five decades from 1960 to 2010, the average milk yield of Holstein cows in North America more than doubled, driven almost entirely by genetic improvement through the progeny testing system. Under conventional progeny testing, a bull born today would not receive a reliable estimated breeding value (EBV) until six to seven years later, after sufficient daughters had completed their first lactations and their performance data had been collected, edited, and incorporated into national genetic evaluations. While this system delivered substantial and sustained genetic progress, the long generation interval was its principal constraint.

*Corresponding author: Henk Bovenhuis, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
DOI: https://doi.org/10.14741/ijab/v.4.1.1

The theoretical foundation for overcoming this constraint was laid by Meuwissen, Hayes and Goddard (2001), who demonstrated through simulation that simultaneous estimation of the effects of thousands of SNP markers — one for every linkage disequilibrium (LD) block in the genome — could yield breeding value predictions of remarkable accuracy for animals with no phenotypic information of their own. The core insight was that when marker density is sufficiently high relative to the extent of LD in the population, every quantitative trait locus (QTL) will be in LD with at least one marker, and the collective effects of all markers can capture virtually all genetic variance. This approach, which became known as genomic selection, offered the tantalising prospect of selecting bulls for breeding at one to two years of age rather than six to seven, potentially doubling the annual rate of genetic gain.

The first large-scale practical implementations of GS in dairy cattle took place in the United States and several European countries in 2008–2009. Validation analyses in North American Holsteins showed that GEBVs for young bulls without daughters had reliabilities comparable to those of bulls with 20–30 daughters (VanRaden et al., 2009). Similar findings were reported in Denmark (Lund et al., 2011), the Netherlands (Calus et al., 2008), and France (Colleau et al., 2011). These results triggered a wholesale transformation of dairy cattle breeding programmes globally: AI organisations discontinued large-scale progeny testing, redirecting resources towards genotyping more young bulls and increasing the size of reference populations. Within five years of commercial implementation, GS had become the standard evaluation method in Holstein and several other major dairy breeds worldwide.

The purpose of this review is to synthesise current knowledge on GS in dairy cattle as of 2014, critically evaluating the statistical methods, genotyping platforms, reference population strategies, and economic implications that have shaped the global adoption of this technology. We also identify key research gaps and emerging methodological developments that will define the next phase of GS implementation.

2. Theoretical Framework

2.1 Linkage Disequilibrium and Marker Density

The accuracy of genomic prediction fundamentally depends on the persistence of LD between SNP markers and causative QTL across the families present in the reference population. In Holstein cattle, effective population size has historically been small (~100 individuals), resulting in extensive LD extending over distances of 100 kb or more. This means that a relatively modest density of 50,000 SNP (approximately one every 60 kb across the 2.5 Gb bovine genome) captures most QTL variation. In contrast, breeds with larger effective population sizes, such as many indigenous African breeds, have shorter LD blocks and require higher marker densities for equivalent prediction accuracy (Goddard et al., 2011).

The accuracy of GEBV is theoretically approximated by Daetwyler et al. (2008) as $r = \sqrt{(Nh^2/(Nh^2+Me))}$, where N is the effective reference population size, h^2 is trait heritability, and Me is the effective number of independently segregating chromosome segments (approximately 2NeL, where L is genome length in Morgans). For Holstein cattle with Ne ≈ 100 and L ≈ 30 Morgans, Me ≈ 6,000, implying that a reference population of ~5,000 bulls provides theoretical reliabilities of 0.7–0.8 for moderately heritable traits — consistent with observed values. This formula also demonstrates why traits with low heritability require disproportionately larger reference populations to achieve acceptable accuracy.

2.2 Statistical Models for Genomic Prediction

Multiple statistical frameworks have been proposed for estimating SNP effects from reference population data. The dominant approach in commercial dairy cattle GS is GBLUP (Genomic Best Linear Unbiased Prediction), which replaces the pedigree-based numerator relationship matrix A with a genomic relationship matrix G constructed from SNP genotypes. The G matrix captures realised relationships between individuals more precisely than pedigree relationships by accounting for Mendelian sampling deviations. The resulting GEBVs are equivalent to those from a model fitting all SNPs simultaneously with equal variance, implicitly assuming that each SNP contributes equally to total genetic variance — an assumption that is well-suited to highly polygenic traits.

Bayesian methods (BayesA, BayesB, BayesCpi, BayesR, BLASSO) allow marker-specific variances and can model the prior expectation that a proportion of SNPs have zero effect. These methods are theoretically advantageous when traits are influenced by a few QTL with large effects, but in practice provide only marginally higher accuracy than GBLUP for most dairy traits in Holstein cattle. Single-step GBLUP (ssGBLUP), developed independently by Aguilar et al. (2010) and Christensen & Lund (2010), simultaneously analyses genotyped and non-genotyped animals by combining G and A into a unified H matrix, elegantly avoiding selection bias that arises when only a non-random subset of animals is genotyped. Table 3 provides a structured comparison of the major prediction methods.

Table 3. Comparison of statistical methods for genomic prediction in dairy cattle

Method	Approach	Complexity	Characteristics
GBLUP	Whole-genome regression	Low	Assumes equal variance for all SNP; fast computation; widely used
ssGBLUP	Combined pedigree+genomic	Low-Med	Utilises all phenotypes; no selection bias; recommended for routine eval
BayesA	Marker-specific variance	High	Assumes all SNP have effect; t-distributed variances; computationally intensive
BayesB	Mixture model	High	Fraction pi of SNP assumed zero effect; powerful for oligogenic traits
BayesCpi	Mixture (pi	High	pi estimated from data; flexible; widely used in

Method	Approach	Complexity	Characteristics
	estimated)		practice
RKHS	Reproducing kernel Hilbert	High	Non-parametric; captures epistasis; lower interpretability

GBLUP = Genomic BLUP; ssGBLUP = Single-step GBLUP; RKHS = Reproducing Kernel Hilbert Space; pp = percentage points.

Table 2. Major SNP genotyping platforms used in dairy cattle genomic selection programmes

Platform	SNP Count	Cost (USD)	Use Case	Notes
BovineHD (Illumina)	777,962	~$200	High-dens. GS	Best for QTL fine-mapping
BovineSNP50 (Illumina)	54,001	~$50	Routine GS	Standard commercial array
GGP Bovine 150K (Neogen)	139,376	~$100	Mid-density GS	Good imputation anchor
BovineLD (Illumina)	6,909	~$20	Parentage + impute	Cost-effective for cows
GGP Bovine 35K (Neogen)	35,000	~$35	Commercial farms	Impute to 50K or HD

4. Reference Population: Size, Composition, and Management

The reference population — the set of genotyped animals with reliable phenotypic records (typically expressed as de-regressed proofs, DRP, or daughter trait deviations, DTD) — is the engine that drives genomic prediction. Reference population size is the single most important determinant of GEBV reliability. In North American Holstein, the reference population had grown to over 30,000 bulls with reliable proofs by 2013, enabling reliabilities of 70–80% for major production traits in young bulls. European countries including France, Germany, the Netherlands, and Denmark maintain national reference populations of 10,000–20,000 bulls and participate in international pooling initiatives that further increase effective reference size.

Multi-breed reference populations can increase prediction accuracy for minor breeds or crossbred animals, but only when LD phase between markers and QTL is consistent across breeds — a condition that generally requires direct inclusion of animals from the target breed (Erbe et al., 2012). International reference populations, coordinated through Eurogenomics and the ICBF international genomic project, now include over 60,000 bulls from multiple Holstein subpopulations across North America, Europe, and Australasia, providing a powerful shared resource that smaller national programmes could not independently assemble.

Optimal management of the reference population requires continuous updating as new genotyped bulls accumulate proven daughters, maintaining currency of SNP effect estimates as genetic drift, selection, and new mutations gradually alter allele frequencies and LD structure. Simulation studies suggest that re-estimation of SNP effects from an updated reference every one to two years is sufficient to maintain prediction accuracy, though more frequent updates may be warranted when the population is undergoing rapid genetic change due to intensive genomic selection (Mulder et al., 2012).

Conventional pathway: Bull born → Progeny test (300 daughters) → EBV at 6–7 yr, REL ~80% || GS pathway: Bull born → Genotyping → GEBV at 1–2 yr, REL 70–80% | Generation interval: 6–7 yr → 1.5–2 yr | Annual genetic gain: +50–100% improvement under GS

Figure 1. Schematic comparison of conventional progeny testing pathway versus genomic selection pathway for dairy bull evaluation, illustrating the reduction in generation interval and the flow of genetic information under each system.

5. Genetic Gain, Economic Impact, and Breed-Specific Outcomes

The primary economic justification for GS adoption is its ability to approximately double the annual rate of genetic gain by halving the generation interval. Schaeffer (2006) estimated that replacing the Canadian progeny testing programme with GS would reduce costs by ~92% while increasing genetic gain. Empirical data from the first five years of GS implementation in Holstein have largely confirmed these projections. In the United States, the genetic trend for lifetime net merit has approximately doubled since 2008 (Cole & VanRaden, 2011). Nordic Red breeds, Norwegian Red, and Montbéliarde have seen similar accelerations.

The economic value of genomic information extends beyond the breeding nucleus. Commercial dairy farmers benefit from genotyping their female calves, obtaining accurate GEBV for production, health, and longevity traits that inform culling and replacement decisions. The payback period for cow genotyping depends on the specific population, management system, and trait emphasis, but studies from Ireland, the Netherlands, and Norway suggest net positive economic returns when genomic information is integrated into culling and replacement strategies (Berry & Kearney, 2013).

6. Current Limitations and Future Directions

6.1 Low-Heritability and Novel Traits

Genomic prediction accuracy for traits with low heritability ($h^2 < 0.10$), including many fertility, health, and longevity traits, remains modest even with large reference populations, because the signal-to-noise ratio in phenotypic records is inherently poor. Novel recording technologies — including automated milking system data, pedometers for activity-based fertility detection, infrared spectroscopy of milk composition, and genomic typing for infectious disease resistance — are expanding the phenotypic landscape available for genomic analysis but require careful validation before incorporation into official selection indices (König & Swalve, 2009).

6.2 Whole-Genome Sequence Based Selection

The ultimate frontier of GS is the use of complete WGS data (~27 million SNPs in cattle) as the predictor set, with the expectation that causative variants will be directly included rather than imperfectly tagged by LD proxies. Early results from the 1000 Bull Genomes Project indicate that sequence-based prediction provides modest but significant accuracy gains for traits influenced by large-effect QTL, particularly when bioinformatic annotation is used to pre-select variants in functional genomic regions (Koufariotis et al., 2014; van Binsbergen et al., 2014). Widespread sequence-based GS awaits further reduction in sequencing costs and development of robust variant prioritisation methods.

6.3 Genomic Selection in Developing Countries

The benefits of GS have so far accrued disproportionately to dairy industries in high-income countries with large reference populations and well-developed genotyping infrastructure. Extending GS to emerging dairy sectors in South Asia, East Africa, and Latin America is a priority for global food security. Strategies being explored include imputation from low-density chips genotyped in target populations to reference panels from phenotypically similar breeds, across-country reference population sharing, and genotype-based parentage verification to improve pedigree accuracy as a prerequisite for reliable genomic predictions in these environments.

7. Conclusions

Genomic selection has irrevocably altered the landscape of dairy cattle breeding. Within six years of commercial implementation, it has doubled annual genetic gain, slashed progeny testing costs, and democratised elite genetics by making accurate selection possible in animals with no performance data. As reference populations continue to grow, genotyping costs decline, and WGS-based methods mature, the technology will become increasingly accessible to minor breeds and developing-country programmes. The integration of genomic selection with precision recording technologies, functional genomics, and advanced reproductive biotechnologies represents the most powerful strategy for meeting global dairy demand sustainably in the coming decades.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., & Lawlor, T.J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93: 743–752.

Berry, D.P., & Kearney, J.F. (2013). Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. Animal 5: 1162–1169.

Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W., & Veerkamp, R.F. (2008). Accuracy of genomic selection using different methods to define haplotypes. Genetics 178: 553–561.

Christensen, O.F., & Lund, M.S. (2010). Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42: 2.

Cole, J.B., & VanRaden, P.M. (2011). Use of haplotypes to estimate Mendelian sampling effects and selection limits. J. Anim. Breed. Genet. 128: 446–455.

Colleau, J.J., Tual, K., Palhière, I., et al. (2011). An indirect approach for across-breed genomic prediction. Genet. Sel. Evol. 43: 28.

Daetwyler, H.D., Villanueva, B., & Woolliams, J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3: e3395.

Daetwyler, H.D., Capitan, A., Pausch, H., et al. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. 46: 858–865.

Erbe, M., Hayes, B.J., Matukumalli, L.K., et al. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density SNP panels. J. Dairy Sci. 95: 4114–4129.

Goddard, M.E., Hayes, B.J., & Meuwissen, T.H.E. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed. Genet. 128: 409–421.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., & Goddard, M.E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92: 433–443.

König, S., & Swalve, H.H. (2009). Application of selection index calculations to determine selection strategies in genomic breeding programs. J. Dairy Sci. 92: 5292–5303.

Lund, M.S., de Roos, A.P.W., de Vries, A.G., et al. (2011). A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43: 43.

Meuwissen, T.H.E., Hayes, B.J., & Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Mulder, H.A., Calus, M.P.L., Druet, T., & Schrooten, C. (2012). Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. J. Dairy Sci. 95: 876–889.

Schaeffer, L.R. (2006). Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123: 218–223.

Schefers, J.M., & Weigel, K.A. (2012). Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Anim. Front. 2: 4–9.

Thomasen, J.R., Guldbrandtsen, B., Su, G., Brøndum, R.F., & Lund, M.S. (2012). Reliabilities of genomic estimated breeding values in Danish Jersey. Animal 6: 789–796.

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., et al. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92: 16–24.

Wiggans, G.R., VanRaden, P.M., & Cooper, T.A. (2011). The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. 94: 3202–3211.