THESES OF THE DOCTORAL (PhD) DISSERTATION

LÁSZLÓ BOGNÁR

MOSONMAGYARÓVÁR

2025

THESES OF THE DOCTORAL (PhD) DISSERTATION

SZÉCHENYI ISTVÁN UNIVERSITY
ALBERT KÁZMÉR FACULTY OF MOSONMAGYARÓVÁR

WITTMANN ANTAL MULTIDISCIPLINARY DOCTORAL SCHOOL OF PLANT, ANIMAL, AND FOOD SCIENCES

UJHELYI IMRE ANIMAL SCIENCE DOCTORAL PROGRAM

HEAD OF THE DOCTORAL SCHOOL:

DR. LÁSZLÓ VARGA, DOCTOR OF THE HUNGARIAN ACADEMY OF SCIENCES (DSC),
UNIVERSITY PROFESSOR

HEAD OF THE DOCTORAL PROGRAM:

DR. FERENC SZABÓ, DOCTOR OF THE HUNGARIAN ACADEMY OF SCIENCES (DSC),
PROFESSOR EMERITUS

SCIENTIFIC SUPERVISORS:

DR. FERENC SZABÓ, DOCTOR OF THE HUNGARIAN ACADEMY OF SCIENCES (DSC),
PROFESSOR EMERITUS
DR. ISTVÁN ANTON, DOCTOR OF THE HUNGARIAN ACADEMY OF SCIENCES (DSC),
PROFESSOR EMERITUS

RESEARCH TOPIC TITLE:

PRACTICAL EXPERIENCES IN THE IMPLEMENTATION OF GENOMIC BREEDING VALUE ESTIMATION AND GENOMIC SELECTION IN THE HUNGARIAN HOLSTEIN-FRIESIAN POPULATION

AUTHOR:

LÁSZLÓ BOGNÁR

MOSONMAGYARÓVÁR

2025

TABLE OF CONTENT

1. INTRODUCTION AND OBJECTIVES

The Holstein-Friesian (HF) is the most widespread dairy cattle breed worldwide, known for its outstanding milk production capacity. The fundamental principles of breeding selection, genetic improvement, and selection strategies have remained unchanged over the years; however, the accuracy of the information used for decision-making has significantly improved. In the past, selection was primarily based on phenotypic performance, but to enhance the precision of genetic evaluations, genotyping and genomic breeding value estimation have become increasingly important. The emergence of genomic selection has revolutionized cattle breeding, relying on single nucleotide polymorphisms (SNPs) as key genetic markers. Proper statistical methods allow the identification of SNPs associated with milk production, functional conformation, and other economically significant traits. Genotyping enables early and highly reliable prediction of individual performance, which is essential for improving breeding efficiency.

The accuracy of genomic breeding value estimation depends on the availability and continuous updating of the reference population. A sufficiently large and genetically diverse reference population allows for the reliable application of SNP-based genetic evaluations. In dairy cattle breeding, the widely used Illumina 50K chip, which examines approximately 50,000 SNPs, provides high-resolution genetic information, enabling more accurate breeding value estimation. One of the major advantages of genomic breeding value estimation and selection is the reduction of the generation interval, allowing reliable breeding decisions to be made at an early stage of an animal's life.

The objective of this study is to analyse and validate genomic breeding value estimation methods in the Holstein-Friesian breed, with a particular focus on the genetic background of milk production traits and the practical implementation of genomic selection.

Based on the above considerations, the following research objectives were set:

1. Implementation and validation of genomic breeding value estimation in the Hungarian Holstein-Friesian (HF) population at the population level.
2. Estimation of production traits in first-lactation HF cows using three different methods: pedigree-based BLUP, conventional BLUP, and genomic BLUP.
3. Evaluation of accuracy, predictive ability, and reliability of the three BLUP methods in first-lactation cows born and raised under identical herd, seasonal, and environmental conditions.
4. Assessment of the applicability of different estimation procedures in modern precision breeding programs.

5. Identification of SNP markers, determined using SNP-chip technology, that contribute to the formation of breeding values (BV) for two or three traits simultaneously.
6. Determination of regression coefficients between the selected SNP markers and breeding values, analysing the direction and magnitude of their effects.

2. MATERIALS AND METHODS

2.1. Genomic Analysis and Genomic Breeding Value Estimation Method

The research was conducted within the HUNGENOM project, utilizing Illumina EuroG MD BeadChip technology. Through international collaboration, the laboratory processing the Hungarian samples had access to a large-scale SNP reference database, which served as the foundation for breeding value estimation.

The study analysed phenotypic and genotypic data from 2,963 Holstein-Friesian cows, provided by the Holstein-Friesian Breeders' Association (HFTE). Genotyping was performed using the EuroG_MDv4 microarray (EuroGenomics, Amsterdam), and only samples with a call rate above 0.95 were included, resulting in a final dataset of 59,151 SNPs.

Breeding value estimation was conducted using a model developed by CRV (Arnhem, Netherlands), which incorporated 40,947 SNPs, excluding those located on the X chromosome. The genomic evaluation model employed a Bayesian multi-QTL approach, which directly assesses the effects of SNPs without considering haplotypes or pedigree information. Although the method can be applied to multiple traits simultaneously, routine genomic evaluations are performed trait-by-trait, meaning m = 1. The model for M traits is as follows:

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{u}_i + \sum_{j=1}^{40947} \boldsymbol{z}_{ij}\boldsymbol{q}_j\boldsymbol{v}_j + \boldsymbol{e}_i$$

Where $\mathbf{y}_i$ = vector of phenotypes (deregressed proofs) of bull i; $\boldsymbol{\mu}$ = vector of fixed trait means; $\mathbf{u}_i$ = vector of random polygenic effects of bull i; $\mathbf{q}_j$ = vector of random non-scaled SNP effects for SNP j with alleles 0, 1, and 2, where SNP allele 0 corresponds to a missing genotype; $\mathbf{v}_j$ = random scaling vector for SNP j; $\mathbf{z}_{ij}$ = design vector for bull i and SNP j - $\mathbf{z}_{ij}$ = [0 2 0], [0 1 1], [0 0 2] or [2 0 0] for homozygous (AA), heterozygous (AB), homozygous (BB), or non-genotyped (00) bulls at SNP j, respectively; $\mathbf{e}_i$ = vector of residuals of bull i.

In the Genome-Wide Association Study (GWAS), Holstein-Friesian cows were categorized based on their breeding values for milk yield (BV_{Milk}), fat yield (BV_{Fat}), and protein yield ($BV_{Protein}$). The high and low groups were defined using the following thresholds: for BV_{Milk}, values above 1,465 were classified as high, while values below 328 were categorized as low; for BV_{Fat}, the high group included values >65, while the low group had values <19; and for $BV_{Protein}$, the high group was defined as >51, and the low as <21.

To identify associations between breeding values and SNPs, we applied three independent approaches: genetic differentiation analysis (Fst_marker), linear regression, and haplotype association tests, conducted using the SNP and Variation Suite (SVS) software. In the haplotype association test, we set the analysis for 5 markers, followed by a chi-square test. Haplotype reconstruction was performed using the Expectation-Maximization (EM) algorithm, with a maximum of 50 iterations and a convergence tolerance of 0.0001.

Based on Manhattan plot evaluations, the threshold values for Fst_marker and -log10(p) were determined as follows: 0.06; 8; and 8 for milk yield, 0.06; 9; and 9 for fat yield, and 0.08; 9; and 9 for protein yield. SNPs exceeding these thresholds were identified for each trait, and those associated with at least two traits were further analysed. A total of 74 SNPs were identified, with a False Discovery Rate (FDR) ranging between 1.3×10^{-21} and 6.0×10^{-6}. The results of Fst_marker, linear regression, and haplotype association tests were converted into unified indices by rescaling values between 0 and 1 and averaging them.

The analysis revealed that: 5 SNPs were associated with both milk and fat yield breeding values, 44 SNPs were linked to milk and protein yield breeding values, 16 SNPs influenced both fat and protein yield breeding values, 9 SNPs showed significant associations with all three traits (milk, fat, and protein yield).

To identify genes located near these SNPs, we used the Bos taurus genome assembly ARS-UCD1.2, analysing regions ±1 megabase pairs (Mbp) around the associated loci (Appendix 1. of the Doctoral Dissertation). If no gene was found within this range, the search window was expanded to ±3 Mbp.

2.2. Comparison of BLUP Breeding Value Estimation Methods

The Hungarian dairy sector is dominated by intensively managed Holstein-Friesian herds, with an average of 453 milk-recorded and herd-book registered cows per farm. Only cows with both pedigree-based and genomic breeding values ($BV_{Pedigree}$, $BV_{Genomic}$) were included in the study. The BLUP AM breeding value estimation was conducted after calving, completion of the first lactation, and official conformation scoring of previously genotyped female animals.

A total of 1,616,549 females received traditional breeding values, of which 23,561 cows also underwent genomic breeding value estimation. Additional filtering criteria included farm, birth year, season, and age, resulting in a final dataset of 190 cows from the same herd, age group, and lactation stage. The genotyping was followed by genomic breeding value estimation, jointly conducted by CRV (Netherlands) and HFTE.

The cows in the study had an average 305-day standard lactation yield of: 10,910 kg milk (MY), 398 kg fat (FY), 365 kg protein (PY). The study population was kept under uniform conditions, housed in free-stall barns with cubicles, and fed a Total Mixed Ration (TMR) diet. The feed primarily consisted of corn silage, concentrated feed, and additional supplements. The standardized housing and feeding conditions were crucial in minimizing environmental variation, ensuring that production traits and breeding values accurately reflected genetic differences.

Comparison of BLUP Breeding Value Estimation Methods:

Pedigree-based breeding value **($BV_{Pedigree}$):** It was calculated as the simple arithmetic mean of the traditional breeding values ($BV_{Traditional}$) of the dam ($BLUP_{dam}$) and the sire ($BLUP_{sire}$) using the following equation:

$$BV_{Pedigree} = \frac{BV_{BLUP_{dam}} + BV_{BLUP_{sire}}}{2}$$

Traditional BLUP breeding value **(BV_{BLUP}):** This method utilized two matrices: a database matrix and a pedigree matrix. The pedigree matrix included full siblings, half-siblings, sires, dams, and grandparents. The BLUP models accounted for maternal genetic effects and permanent environmental effects as random factors. The model structure was defined as follows:

$$\mathbf{y} = X_{\mathbf{b}} + Z_{\mathbf{a}} + W_{\mathbf{pe}} + \mathbf{e}$$

Where "**y**" is the vector of observations; "**b**" is the vector of fixed effects; "**a**" is the vector of random animal effects; "**pe**" is the random vector of permanent environmental effects; "**e**" is the vector of random residual effects; X, Z and W are the incidence matrices relating records to fixed, animal and random permanent environmental effects, respectively.

Genomic breeding value **($BV_{Genomic}$):** The methodology for genomic breeding value estimation was detailed in **Section 2.1** of the study.

The normality of production traits was assessed using the Kolmogorov-Smirnov test, while the homogeneity of variances was evaluated with the Levene test. To analyse relationships between traits, a multifactorial analysis of variance (ANOVA) was applied. The associations between breeding values and production data were determined using Pearson's correlation coefficient and Spearman's rank correlation. Data preparation and organization were performed using Microsoft Excel 2019 and Word 2019, while statistical analyses were conducted with the SPSS 27.0 statistical software package.

3. RESULTS AND DISCUSSION

3.1. Results of SNP Analysis

A total of 5 SNPs were associated with milk yield (MY) and fat yield (FY) on chromosomes 9, 18, and 19. Additionally, 44 SNPs were linked to MY and protein yield (PY) across chromosomes 1-6, 11, 13-15, 18, 19, 24, 28, and X. For FY and PY, 16 SNPs were identified on chromosomes 3, 11, 19, 22, and X. Furthermore, 9 SNPs were associated with all three production traits (MY, FY, and PY) and were located on chromosomes 2, 5, 28, and X. The maximum values observed were 0.17 for Fst_marker, -log10(p) = 24.9 for linear regression, and -log10(p) = 26.4 for haplotype association analysis. The most significant SNPs (average >0.8) were found on chromosomes 2, 11, 19, 28, and X for MY, chromosomes 3, 22, 28, and X for FY, and chromosomes 1 and 28 for PY. Among the 9 SNPs linked to all three production traits, 7 were classified as having the highest impact (average >0.8), while 3 SNPs were located within 1.18 million base pairs on chromosome 28 (Appendix 2. of the Doctoral Dissertation).

3.2. Results of the Comparison of BLUP Breeding Value Estimation Methods

The standard lactation performance of the examined population, consisting of 190 cows, was favourable and aligned with the national average: 10,910.50 kg milk yield (MY), 397.86 kg fat yield (FY), and 365.33 kg protein yield (PY). The homogeneity of the population was confirmed by the coefficient of variation (CV%) values remaining below 15%. Significant differences were observed between the different breeding values (BV), particularly in the minimum and maximum values, while the standard errors (SE) relative to the mean values remained below 5%. The Pedigree BV ($BV_{Pedigree}$) consistently showed higher values than the other two BV estimates, which can be attributed to the five-year rolling base change between the parental and offspring generations. Based on the mean breeding values, for milk yield (MY), Traditional BLUP BV ($BV_{Traditional}$) was 739.16 kg, Genomic BLUP BV ($BV_{Genomic}$) was 718.11 kg, and Pedigree BLUP BV ($BV_{Pedigree}$) was 1,379.57 kg. For fat yield (FY), $BV_{Traditional}$ was 34.47 kg, $BV_{Genomic}$ was 33.44 kg, and $BV_{Pedigree}$ was 55.58 kg. For protein yield (PY), $BV_{Traditional}$ was 27.01 kg, $BV_{Genomic}$ was 26.05 kg, and $BV_{Pedigree}$ was 49.33 kg. The standard errors (SE) were consistently low, indicating the stability of the estimates.

The correlation coefficients between phenotypic performance and the breeding values estimated using different BLUP methods ($BV_{Traditional}$, $BV_{Genomic}$, $BV_{Pedigree}$) were positive, with most being statistically significant. The strongest relationship with phenotypic performance was observed for $BV_{Traditional}$ (r_{gp} = 0.61–0.70), followed by $BV_{Genomic}$ (r_{gp} = 0.31–0.48), while the weakest correlation was found between $BV_{Pedigree}$ and phenotype (r_{gp} = 0.15–0.24).

A strong genetic correlation was observed between Traditional BLUP Breeding Value ($BV_{Traditional}$) and Genomic BLUP Breeding Value ($BV_{Genomic}$) (r_g = 0.66–0.67), whereas the correlation between $BV_{Traditional}$ and Pedigree Breeding Value ($BV_{Pedigree}$) was moderate to strong (r_g = 0.56–0.66). The relationship between BVGenomic and BVPedigree was slightly weaker, showing a moderate-to-strong correlation (r_g = 0.43–0.56).

Rank correlations further confirmed the genetic associations, with $BV_{Traditional}$ and $BV_{Genomic}$ showing a strong connection (r_{rank} = 0.65–0.66). The rank correlations between $BV_{Pedigree}$ and the other two BVs (r_{rank} = 0.12–0.57) indicated moderate associations, particularly between $BV_{Genomic}$ and $BV_{Pedigree}$ (r_{rank} = 0.40–0.56).

The regression relationships between breeding values and phenotypic traits were also examined. For fat yield (FY), all regression coefficients were significant ($p < 0.01$), except for $BV_{Pedigree}$. The regression slope (b) was positive in all cases, with the highest values recorded for $BV_{Traditional}$ (b = 0.17–0.21), followed by $BV_{Genomic}$ (b = 0.10–0.15), while $BV_{Pedigree}$ had the lowest values (b = 0.04–0.07).

Regarding prediction accuracy, $BV_{Traditional}$ was the most precise for all phenotypic traits (R^2 = 0.37–0.48), followed by $BV_{Genomic}$ (R^2 = 0.09–0.23), while $BV_{Pedigree}$ had the lowest accuracy (R^2 = 0.02–0.06). The strongest relationships were observed for milk yield (MY) and fat yield (FY), while protein yield (PY) showed a weaker association.

The results confirm that traditional BLUP-based breeding values provide the most reliable predictions, while genomic BLUP demonstrates a moderate but still significant association. Pedigree-based breeding values were the least accurate compared to the other two methods. These findings highlight the varying reliability of different breeding value estimation methods, emphasizing the superior accuracy of traditional BLUP models, followed by genomic BLUP, while pedigree-based estimates showed the highest uncertainty.

Although genomic breeding values did not always surpass traditional BLUP in accuracy, their importance is particularly evident for hard-to-measure/estimate traits, such as longevity, metabolic traits, and health indicators. The results support the role of genomic selection in early breeding stock selection, which can reduce maintenance costs while improving genetic progress.

Traditional breeding values remain a critical reference, ensuring the control and credibility of genomic evaluations. Comparative studies contribute to refining breeding value estimations, thereby enhancing the precision of breeding decisions and securing long-term economic benefits for the dairy industry. The advantages of genomic BLUP were less pronounced in small populations or under uniform environmental conditions, but it offers significant benefits for traits that are difficult to estimate. Early genomic selection enhances breeding efficiency, reduces maintenance costs, and improves herd genetic quality. Meanwhile, traditional breeding value estimation continues to serve as a control mechanism, ensuring the reliability of genomic predictions.

4. NEW SCIENTIFIC FINDINGS AND THEIR APPLICATIONS

4.1. New Scientific Findings

1. I was the first to apply the Pearson correlation between phenotype and genotype to characterize the reliability of different breeding value estimation (BVE) methods. Based on the comparison of the three BLUP methods, I found that the conventional BLUP provided the most reliable estimates (r = 0.61–0.70), followed by the genomic BLUP (r = 0.31–0.48), while the pedigree-based BLUP proved to be the least reliable (r = 0.15–0.24).
2. I demonstrated that the 305-day milk, fat, and protein yields of Holstein-Friesian cows born in the same year and season, and kept under uniform environmental, nutritional, and management conditions, are suitable for validating breeding values estimated for production traits.
3. For the first time in the Holstein-Friesian breed, I identified SNPs involved in the development of two or three breeding values related to milk production, using three independent methods: genetic distance (Fst_marker), linear regression, and haplotype association analysis.
4. Based on the regression analysis of the 74 identified SNPs, I found that most markers exerted effects in the same direction on the evaluated breeding values. However, I also identified SNPs (BTB-00219372 and BovineHD3000027615) that require careful consideration, as they may increase one breeding value while decreasing one or both of the others.
5. I identified four genes (EFCAB10, GLOD5, NONO, and TMEM70) that have not previously been studied in cattle or associated with milk production traits.

4.2. Practical Applications of the Results

Responsible Holstein-Friesian breeding requires a novel approach that emphasizes the combined evaluation of multiple factors, including breeding values, in the interest of sustainability and regenerative development. Maintaining the balance between genotype and environment can significantly improve the sustainability and adaptability of milk production based on the Holstein-Friesian breed. The application of genomic information and genomic selection systems enables breeders to make informed decisions that promote healthier cows with improved reproductive performance, increased efficiency, and higher milk yields with reduced feed intake — thus reducing the negative environmental impact of dairy production.

The SNPs identified in the course of this study are located near genes that have not previously been examined in cattle and may serve as potential targets for new practical applications related to milk production traits. Markers associated with two or three breeding values can be effectively applied in selection, thereby accelerating genetic progress and supporting breeders, geneticists, and dairy farmers in achieving their breeding and selection goals. Markers with consistent directional effects facilitate sire selection, while those with antagonistic effects require increased caution, as they may improve one breeding value while simultaneously decreasing one or even two others.

The true value of the results lies in their applicability for validating breeding values—especially through the direct comparison of different estimation methods with actual performance data under standardized environmental conditions. This provides a new perspective for the practical implementation of genomic selection in Holstein-Friesian breeding.

The findings of this study offer tangible economic benefits for domestic Holstein-Friesian breeders. Early sire selection and the implementation of efficient breeding programs—both at the national and farm level—can help establish optimal herd sizes, thereby increasing the profitability of milk production while significantly reducing the environmental footprint.

5. PUBLICATIONS RELATED TO THE DISSERTATION

5.1. Peer-Reviewed Article Published in a Foreign-Language Journal

Bognár, L. – Kőrösi, Zs. J. – Anton, I. – Bene, Sz. - Szabó, F. (2024): Different Breeding Values Under Uniform Environmental Condition for Milk Production Yield Traits in Holstein-Friesian Cows. *Animals* 15 (51), 11. https://doi.org/10.3390/ani15010051 Q1, IF=2,7

Bognár, L. – Kőrösi, Zs. J. – Bene, Sz. A. – Szabó, F. – Anton, I. - Zsolnai, A. (2024): Simultaneous Effects of Single-Nucleotide Polymorphisms on the Estimated Breeding Value for Milk, Fat, and Protein Yield of Holstein-Friesian Cows in Hungary. *Animals* 14 (23), 3518. https://doi.org/10.3390/ani14233518 Q1, IF=2,7

Bene, Sz. – Kőrösi, Zs. J. – **Bognár, L.** – Polgár, J. P. – Szabó, F. (2024): Population Genetic Features of Calving Interval of Holstein-Friesian Cows Bred in Hungary. *Animals* 14 (17), 2513. https://doi.org/10.3390/ani14172513 Q1, IF=2,7

Bognár, L. – Szabó, F. (2023): Management of "Modern" Holstein Cows Focusing on Sustainability and Resilience – Review of Recent Achievements. *Chemical Engineering Transactions* 107, pp. 169–174. Q3

5.2. Peer-Reviewed Article Published in a Hungarian-Language Journal

Bognár L. - Kőrösi Zs. J. - Bene Sz. - Szabó F. (2024): A modern holstein-fríz tehén tenyésztésének és tartásának néhány fenntarthatósági szempontja (Irodalmi áttekintés) (Some sustainability aspects for the breeding and management of modern Holstein cow. [Literature review]) *Állattenyésztés és Takarmányozás* 73 (2), 109–119. old.

Bognár L. (2021): A hazai genomikai tenyészértékbecslés gyakorlati tapasztalatai a holstein-fríz fajta tenyésztési programjában (Practical Experiences of the Hungarian Genomic Breeding Value Estimation in the Breeding Program of the Holstein-Friesian Genetic Programme). *Állattenyésztés és Takarmányozás*, 70 (3), 250–269. old.

Bognár L. (2013): Új irányzatok a tejtermelő szarvasmarha-tenyésztésben. A genomikus tenyészértékbecslés. (New Trends in Dairy Cattle Breeding. Genomic Breeding Value Estimation). *Állattenyésztés és Takarmányozás*, 62 (4), 367–373. old.