J. Dairy Sci. TBC
https://doi.org/10.3168/jds.2026-28483

Effects of ROH-based versus GRM-based future inbreeding penalties on genetic gain, variance components, and inbreeding depression in dairy cattle

C. Maltecca,†#††* J. Jiang,† J. B. Cole,§†# C. F. Baes,# L. Ma,†† and F. Tiezzi‡‡
†Department of Animal Science, North Carolina State University, Raleigh, NC 27695, USA
#Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON N1G 2W1, Canada
††Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742, USA
§Council on Dairy Cattle Breeding, Bowie, MD, USA;
&Department of Animal Sciences, Donald Henry Barron Reproductive and Perinatal Biology Research Program, and the Genetics Institute, University of Florida, Gainesville, FL, USA;
‡‡Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, 50144 Florence, Italy

ABSTRACT

Managing inbreeding while maintaining genetic gain is a critical challenge in dairy cattle breeding programs. Current North American genetic evaluations adjust breeding values for expected future inbreeding using genomic relationship matrix (GRM) approaches, but runs of homozygosity (ROH)-based methods may offer advantages for detecting recent autozygosity. This study compared ROH-based versus GRM-based inbreeding penalties across 20 generations of simulated genomic selection. Using AlphaSimR, we simulated dairy cattle populations under 3 genetic scenarios: additive only, low non-additive effects ($\sigma^2D/\sigma^2A = 0.10$, inbreeding depression = $0.30\sigma G$ per 10% ΔF), and high non-additive effects ($\sigma^2D/\sigma^2A = 0.50$, inbreeding depression = $1.20\sigma G$ per 10% ΔF). Each penalty method was tested at 3 intensities (0.5 × , 1 × , 2 × the inbreeding depression parameter) with 10 replicates per scenario-treatment combination. GRM-based penalties achieved lower inbreeding accumulation than ROH-based penalties, with differences of 0.13 to 0.22 in genomic F at matched penalty intensities. However, this enhanced inbreeding control was associated with reduced genetic gain, with GRM-based selection achieving 0.6 to 1.7 genetic standard deviations less gain than ROH-based selection. GRM-based selection retained more additive variance (66–107%) compared with ROH-based selection (63–74%), with values exceeding 100% indicating partial recovery of variance depleted during burn-in. Dominance variance retention was also greater for GRM-based selection (105–114% vs. 93–98%), reflecting maintained heterozygosity. Interestingly, GRM-based selection accumulated more inbreeding depression than ROH-based approaches despite lower inbreeding levels, potentially reflecting more effective purging under the stronger directional selection maintained by ROH-based methods. Results illustrate trade-offs between inbreeding control and genetic gain, with optimal strategy depending on breeding program objectives and trait genetic architecture.

INTRODUCTION

Maintaining genetic diversity in dairy cattle populations is essential for preserving adaptability to future breeding goals and avoiding fitness declines (Lozada-Soto et al., 2022). Genomic selection has substantially accelerated genetic gain, with annual progress increasing by 50–192% in various populations (Doublet et al., 2019; Guinan et al., 2023). This acceleration has been accompanied by elevated inbreeding accumulation: post-genomic selection inbreeding rates of 1.19–2.06% per generation have been documented in North American Holsteins (Makanjuola et al., 2020a), exceeding the FAO-recommended threshold of 1% per generation (FAO, 2013), and effective population sizes for Holstein sires have declined to 13–18 in the genomic era (Lozada-Soto et al., 2022). These trends raise concerns about long-term breeding program sustainability and the accumulation of inbreeding depression.

Inbreeding depression — a reduction in phenotypic performance due to increased homozygosity at loci exhibiting directional dominance (Charlesworth and Willis, 2009) — has been documented across production, fertility, and health traits in dairy cattle. For production traits, a 1% increase in inbreeding coefficient is associated with reductions of approximately 20–60 kg in milk yield, depending on the inbreeding measure used (Doekes et al., 2019; Makanjuola et al., 2020b). Fertility and health traits, being more closely related to fitness, often exhibit proportionally greater depression effects (Howard et al.,

Received February 18, 2026.
Accepted April 1, 2026.
*Corresponding author: cmaltec@ncsu.edu

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-26. Nonstandard abbreviations are available in the Notes.

2017). Dominance variance contributes substantially to genetic variation in dairy cattle, accounting for 5–7% of total phenotypic variance for yield traits and up to 50% of genetic variance for some functional traits (Sun et al., 2014; Ertl et al., 2014), underscoring the importance of understanding how inbreeding management strategies affect dominance variance dynamics.

Several approaches have been proposed to manage inbreeding in dairy cattle breeding programs. Optimal contribution selection (OCS) maximizes genetic gain while constraining the rate of inbreeding through optimization of parental contributions (Meuwissen, 1997), but practical implementation in commercial dairy breeding has proven difficult given the dispersed nature of breeding decisions across AI organizations and farms (Weigel and Lin, 2002; Cole, 2023). As an alternative, North American genetic evaluation systems have developed measures to predict the expected inbreeding of future progeny. In the United States, VanRaden and Smith (1999) introduced expected future inbreeding (EFI), which estimates an animal's average pedigree relationship to a recent female cohort; this measure is formally incorporated into CDCB genetic evaluations as a deduction from published PTAs, applied as b(EFI − EFIbase) where b reflects trait-specific inbreeding depression estimates (approximately $40 per 1% EFI for Net Merit). More recently, genomic future inbreeding (GFI) was developed using genomic rather than pedigree relationships to the same female cohort (VanRaden et al., 2011); GFI is published alongside PTAs as an information metric and correlates 0.83–0.84 with pedigree-based EFI in Holstein populations (VanRaden et al., 2011). Similarly, Van Doormaal et al. (2003) developed the Relationship Value (R-Value) for the Canadian Dairy Network, which is published as producer guidance to identify outcross bulls; unlike EFI in the US, the R-Value is not applied as an automatic penalty on published LPI or Pro$ values. To our knowledge, no other major national dairy cattle evaluation currently adjusts published total merit indexes for expected future inbreeding in this manner: the Dutch NVI, Nordic NTM, German RZG and RZ€, French ISU, UK PLI, and Australian BPI all manage inbreeding through optimal contribution selection constraints, mating tools, or trait-model covariates. Italy's ANAFIBJ is developing a ROH-based genomic expected future inbreeding measure (Gefi) with direct comparisons to CDCB GFI (van Kaam et al., 2024), representing the most substantive development toward equivalent implementation outside North America. Limitations of EFI adjustments have been noted: elite bulls that drive genetic change are often extreme outliers for which PTA adjustments may not cause substantial re-ranking (Cole, 2023), and the Holstein base population EFI rising from 7.5% to 9.4% between 2015 and 2020 (CDCB, 2025) indicates that the population is becoming uniformly more related, reducing the discriminatory power of individual future inbreeding penalties.

The choice of inbreeding coefficient calculation method affects both the accuracy of estimation and the effectiveness of management strategies. Genomic relationship matrix (GRM)-based inbreeding coefficients reflect genome-wide departure from Hardy-Weinberg expectations relative to a base population (VanRaden, 2008). Runs of homozygosity (ROH)-based inbreeding coefficients (FROH) identify contiguous stretches of homozygous genotypes likely originating from a common ancestor (McQuillan et al., 2008) and show higher correlations with homozygous mutation load than marker-by-marker approaches (Keller et al., 2011). ROH length also carries temporal information: segments greater than 16 Mb reflect inbreeding within approximately 3 generations (Curik et al., 2014), which is relevant because recent inbreeding causes more severe depression than ancient inbreeding (Doekes et al., 2019; Ablondi et al., 2023). GRM-based approaches, by contrast, offer computational advantages and direct integration with genomic prediction models. Despite these established differences, the relative effectiveness of ROH-based versus GRM-based measures for long-term inbreeding management has not been thoroughly evaluated.

Several knowledge gaps remain. The relative effectiveness of ROH-based versus GRM-based expected progeny inbreeding penalties for controlling inbreeding accumulation over multiple generations of selection has not been compared through simulation. The effects of these approaches on inbreeding depression dynamics and dominance variance are poorly understood: selection might create homozygosity at QTL where dominance effects are expressed, increasing depression even when genome-wide measures are controlled. No study has examined how genetic architecture — specifically the proportion of dominance variance relative to additive variance — affects the relative performance of these strategies. Finally, the potential for trait architecture to shift under sustained genomic selection, with dominance effects becoming relatively more important as additive variance is eroded by the Bulmer effect (Wientjes et al., 2022), has not been investigated in the context of inbreeding management.

Through a simulation framework, the objectives of this study were to: compare the effectiveness of ROH-based versus GRM-based expected progeny inbreeding penalties for controlling inbreeding accumulation over 20 generations of simulated dairy cattle genomic selection; evaluate the effects of these penalty approaches on genetic gain, additive and dominance variance components, and inbreeding depression across genetic scenarios with different levels of non-additive variance; and assess how genetic architecture, specifically the ratio of dominance

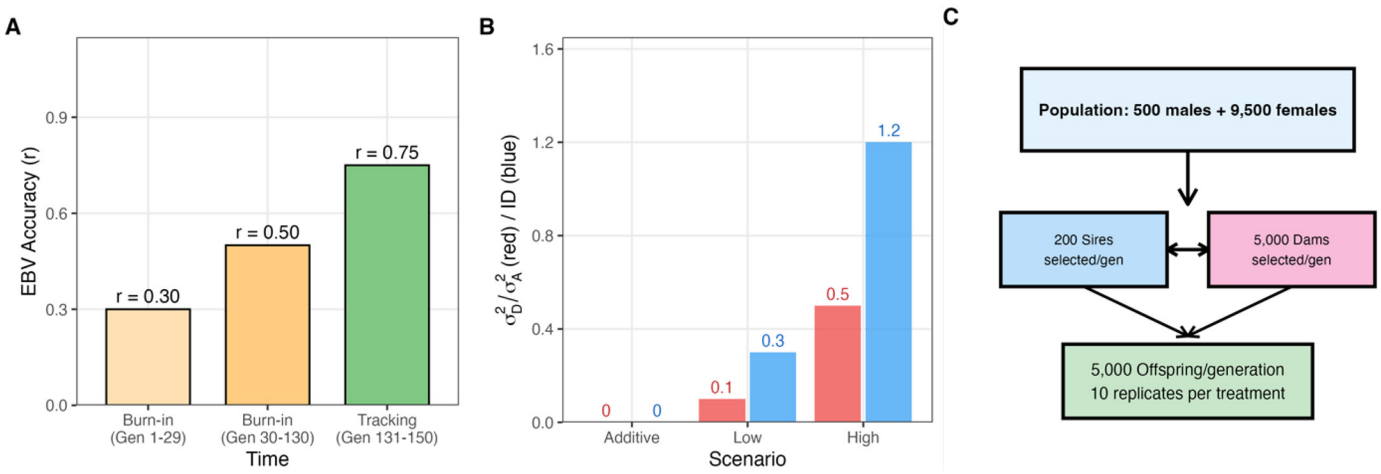

Figure 1. Experimental design. (A) EBV accuracy progression across burn-in and tracking phases. (B) Three genetic scenarios varying in dominance variance ratio (VarD/VarA) and inbreeding depression severity. (C) Population structure and selection methods compared during the tracking phase. Ten replicates per scenario-method combination.

to additive variance, affects the relative performance of these inbreeding management strategies.

MATERIALS AND METHODS

All simulations were performed with AlphaSimR version 1.5.3 (Gaynor et al., 2021) in R (v 4.3.1). The simulation framework was designed to mimic a commercial dairy cattle breeding program operating under genomic selection and followed the parameterization from Lozada-Soto et al. (2024). The overall experimental design is illustrated in Figure 1.

Genome and Population Simulation

A diploid genome consisting of 6 chromosomes was simulated, with each chromosome containing 2,000 biallelic single nucleotide polymorphism (SNP) markers and 2,000 quantitative trait loci (QTL), yielding a total of 12,000 markers and 12,000 QTL genome-wide. This was done to approximate a medium-density SNP array commonly used in livestock genomic selection. Markers and QTL were uniformly distributed along chromosomes with positions sampled from a uniform distribution. Founder population initial allele frequencies were sampled from a β distribution $B(\alpha = 1.0, \beta = 1.0)$ to approximate a uniform distribution of allele frequencies reflecting standing variation in outbred livestock populations before selection.

True Breeding Value selection (**TBV**) served as a positive control representing the theoretical maximum genetic gain achievable with perfect information. Under TBV selection, individuals were ranked and selected directly on their additive breeding value without estimation error. The breeding value was calculated as $TBV_i = \sum_j \alpha_j x_{ij}$ where x_{ij} is the gene content (centered on the mean) for individual I at QTL j.

Genetic Scenarios

Three genetic scenarios were simulated to represent a spectrum of non-additive genetic architectures commonly observed in livestock populations. The *Additive Only* scenario served as a baseline. Dominance variance was set to zero ($\sigma^2D = \sigma^2D/\sigma^2A = 0$) and consequently no inbreeding depression (**ID**) occurred ($ID = 0$).

In the *Low Non-Additive* scenario dominance variance was set with a ratio of $\sigma^2D/\sigma^2A = 0.10$ ($\sigma^2D = 10$) and an inbreeding depression parameter of $ID = 3$, corresponding to a loss of 0.30 genetic standard deviations (σG) per 10% increase in the genomic inbreeding coefficient. This scenario represents production traits exhibiting modest directional dominance, consistent with estimates for production and growth rate in cattle (Leroy, 2014).

The *High Non-Additive* scenario was created with a dominance variance ratio of $\sigma^2D/\sigma^2A = 0.50$ ($\sigma^2D = 50$) and an inbreeding depression parameter of $ID = 12$, corresponding to a loss of 1.20 σG per 10% increase in inbreeding, to mimic fitness-related traits with substantial inbreeding depression exhibiting stronger directional dominance.

We defined *ID* as the expected reduction in mean phenotypic value per unit increase in the genomic inbreeding coefficient, reflecting the cumulative effect of increased homozygosity at loci exhibiting directional dominance so that magnitude of inbreeding depression scaled appropriately with the level of dominance variance in each scenario.

A base population of 10,000 individuals (500 males, 9,500 females) was generated through 10 generations of random mating to establish linkage disequilibrium patterns and minor allele frequency distributions consistent with populations that have experienced historical genetic drift but little selection. Each generation, 200 sires and 5,000 dams were selected from 500 male and 9,500 female candidates based on the criterion specific to each experimental treatment, and mated at approximately 25 matings per sire to produce 5,000 offspring. This ratio reflects the reproductive asymmetry of dairy cattle breeding programs, where artificial insemination enables high selection intensity on the male pathway while female selection intensity is constrained by reproductive capacity.

Burn-in and Tracking Phases

Simulations comprised 2 phases. The burn-in phase (generations 1–130) subjected all replicates to identical EBV selection to establish realistic patterns of inbreeding accumulation and genetic variance depletion before the comparison of selection methods, creating common starting conditions across experimental treatments. EBV accuracy was varied across the burn-in phase to mimic the historical evolution of genetic evaluation methods in dairy cattle populations, with r = 0.30 during generations 1–29 and r = 0.50 during generations 30–130.

This burn-in period rather than providing a historically accurate reconstruction of population selection history was designed to ensure that genomic inbreeding had accumulated to levels typical of modern intensively selected livestock populations (mean FG ≈0.55 at burn-in end, relative to the founder base population), that additive genetic variance had been substantially depleted through directional selection (from the founder value of 100 to ~40), that realistic patterns of linkage disequilibrium had developed, and that all replicates shared a common genetic history before experimental treatments were applied.

The tracking phase from generation 130–150, was a 20-generation period during which alternative selection methods were compared. During this phase, EBV accuracy was increased to r = 0.75, mimicking a modern genomic selection scenario. The different selection methods were applied beginning at generation 131, and all genetic parameters were recorded at each generation for subsequent analysis. For clarity of discussion, generation numbering in results refer to the tracking phase, so that "Generation 1" refers to generation 131 of the full simulation timeline and "Generation 20" refers to generation 150.

Selection Methods

Two control methods and 2 penalized selection methods were evaluated, with penalized methods tested at 3 penalty intensities. Under TBV selection (positive control), individuals were ranked directly on true genetic merit without estimation error, providing an upper bound on selection response equivalent to EBV accuracy of unity. Under EBV selection (standard practice control), individuals were ranked on estimated breeding values generated as $EBV_i = TBV_i + \varepsilon_i$, where ε_i is normally distributed with mean zero and variance calibrated to achieve $r = cor(\text{TBV, EBV}) \approx 0.75$. A uniform accuracy was applied to both sexes, rather than the lower female accuracy typical of progeny-based evaluation, to isolate the effects of inbreeding penalty methods from sex-specific accuracy differences and to reflect the reduced accuracy gap achievable under genomic selection, where direct genotyping provides information independent of progeny performance.

Inbreeding Penalty Indices

The GRM-based penalty method (**GFI**) incorporated expected progeny inbreeding into the selection criterion by constructing a penalized index that combined EBV with the expected inbreeding of future offspring. Selection was performed on $GFI_i = EBV_i - \lambda GRM \times Fprog_i$, where EBV_i is the estimated breeding value for selection candidate i, $Fprog_i$ is the expected genomic inbreeding coefficient of progeny if candidate i were mated to the current breeding female population, and λ_{GRM} is the penalty weight. Expected progeny inbreeding was calculated as half the average genomic relationship between each selection candidate and all active dams: $Fprog_i = \left(\frac{1}{2}\right) \times \left(\frac{1}{n_{dams}}\right) \times £_j G_{ij}$, where G_{ij} is the genomic relationship between candidate i and dam j. The genomic relationship matrix **G** was constructed following VanRaden (2008) Method 1 using base population allele frequencies to ensure compatibility with pedigree-based inbreeding scales.

The ROH-based penalty method (**ROFI**) similarly incorporated expected progeny inbreeding into selection but utilized runs of homozygosity to quantify autozygosity. Selection was performed on $ROFI_i = EBV_i - \lambda ROH \times Fprog_i$, where expected progeny inbreeding was approximated as $Fprog_i = \left(F_{ROH_i} + F_{ROH_{dams}}\right)/2$, with F_{ROH_i} being the ROH-based inbreeding coefficient of candidate i and $F_{ROH_{dams}}$ being the mean ROH-based inbreeding coefficient across all active dams.

An exact ROH-based measure of expected progeny inbreeding would require detecting shared haplotype segments between each candidate and all potential mates to construct a segment-based relationship matrix, with expected progeny inbreeding calculated as half the average relationship to the dam population, analogous to the GRM method. However, pairwise haplotype comparisons scale quadratically with population size, making this approach computationally challenging at breeding program scale. The approximation used here exploits the correlation between individual autozygosity and average relatedness to the population, requiring only individual-level ROH calculations. Both ROFI and GFI indexes were implemented in custom C++ code.

ROH Detection. Runs of homozygosity were identified using a sliding window approach implemented in custom C++ functions. Genotypes were converted to a binary homozygosity indicator (1 for homozygous, 0 for heterozygous), and consecutive homozygous segments were identified and merged into ROH for each individual and chromosome. A segment was classified as an ROH if it contained a minimum of 70 consecutive homozygous SNPs with no heterozygous genotypes permitted, corresponding to a minimum physical length of approximately 3.5 Mb given the marker density of 20 SNPs/Mb. This threshold was chosen to ensure reliable detection of true autozygous segments while accommodating the moderate marker density of the simulated array. Higher thresholds capture more recent inbreeding but yield fewer and shorter detectable segments, reducing discriminatory power; more stringent thresholds targeting recent inbreeding (e.g., > 16 Mb segments reflecting ~3 generations) would require 320+ consecutive homozygous SNPs, substantially reducing detection power at the marker density used here.

The ROH-based inbreeding coefficient for each individual was calculated as $F_{ROH_i} = £_k L_k / L_{genome}$, where L_k represents the physical length of the k-th ROH segment and Lgenome represents the total length of the autosomal genome covered by markers (600 Mb across 6 chromosomes).

Penalty weights (λ) were scaled relative to the inbreeding depression parameter (ID) of each scenario, so that a '1 × penalty' carries equivalent economic meaning across scenarios with different levels of non-additive effects. Three intensities were applied: 0.5 × , 1 × , and 2 × ID. In the Low Non-Additive scenario (ID = 3), λ = 1.5, 3, and 6; in the High Non-Additive scenario (ID = 12), λ = 6, 12, and 24. In the Additive Only scenario (ID = 0), fixed values of λ = 0, 1, and 2 were used for experimental consistency. For each candidate, the penalized index was computed as EBV minus the product of λ and the expected progeny inbreeding coefficient, calculated as described separately for each method.

Performance Metrics. Different parameters were recorded at each generation during the 20-generation tracking phase to characterize the genetic status and selection outcomes of each population. Three measures of inbreeding were recorded: the genomic inbreeding coefficient (**FG**) derived from GRM diagonal elements as described above, the pedigree-based inbreeding coefficient (**FP**) calculated from the pedigree relationship matrix, and the ROH-based inbreeding coefficient (**FROH**) calculated as described above.

All variance components were recorded at every generation throughout the 20-generation tracking phase, including additive genetic variance (**VarA**) calculated as the genic variance from QTL additive effects given current allele frequencies, and dominance genetic variance (**VarD**) calculated from QTL dominance deviations and heterozygosity levels. Cumulative genetic gain (**ΔG**) was recorded as the difference in mean true breeding value relative to the founder population mean, expressed in the original trait units. Changes in genome-wide homozygosity (**ΔHom**) were also monitored. Cumulative inbreeding depression was calculated at each generation as the reduction in mean phenotypic value attributable to the increase in homozygosity at QTL with directional dominance effects, relative to the expected phenotypic value under Hardy-Weinberg equilibrium.

The 3 genetic scenarios (*Additive Only, Low Non-Additive, High Non-Additive*) were crossed with 6 penalized treatments (2 methods × 3 intensities) and 2 control methods (TBV, EBV), yielding 8 combinations per scenario. Each of the 24 scenario-by-treatment combinations was replicated 10 times using independent random number seeds, producing a total of 240 simulation runs.

Changes in genetic parameters over the 20-generation tracking phase were calculated as the difference between values at generation 20 and generation 1 of tracking. The primary response variables included change in genomic inbreeding ($\mathbf{\Delta F} = FG_{Gen20} - FG_{Gen1}$), genetic gain expressed in genetic standard deviation units ($\mathbf{\Delta G} = gain_{Gen20} - gain_{Gen1}$), change in additive genetic variance ($\mathbf{\Delta\sigma^2 A} = VarA_{Gen20} - VarA_{Gen1}$), change in dominance genetic variance ($\mathbf{\Delta\sigma^2 D} = VarD_{Gen20} - VarD_{Gen1}$), and accumulated inbreeding depression ($\mathbf{\Delta ID} = cID_{Gen20} - cID_{Gen1}$), where **cID** is the cumulative ID.

Variance retention was calculated as the proportion of starting variance maintained through the tracking phase: Variance Retained (%) = (σ^2Gen20 / σ^2Gen1) × 100, where the denominator refers to variance at generation 1 of tracking, not the original founder variance of 100. By the start of tracking, burn-in selection had already depleted additive genetic variance from the founder value of 100 to approximately 30–45 depending on the scenar-

io. Consequently, retention values exceeding 100% do not indicate creation of new genetic variance but rather partial recovery of variance depleted during burn-in, occurring when reduced selection intensity under penalized selection allows alleles approaching fixation to stabilize at intermediate frequencies.

The rate of inbreeding per generation during tracking was calculated as $\mathbf{\Delta F_{per\ gen}} = (F_{Gen20} - F_{Gen1}) / 19$, where the denominator of 19 reflects the number of generation intervals between generation 1 and generation 20. This rate was calculated for the genomic inbreeding coefficient (FG), which is the primary inbreeding measure reported throughout. Negative values indicate net reduction in genomic inbreeding during the tracking phase, which can occur when penalty methods reduce the average relatedness among selected parents below the level established during burn-in.

Statistical Analysis.

The 3 genetic scenarios were crossed with 6 penalized treatments (2 methods × 3 intensities) and 2 control methods, yielding 8 combinations per scenario. Each of the 24 scenario-by-treatment combinations was replicated 10 times using independent random number seeds (240 simulation runs total). Summary statistics were computed for each unique combination of scenario, method, and penalty level using the dplyr package (Wickham et al., 2023). Results are presented as mean ± standard deviation across 10 replicates. Visualization used ggplot2 (Wickham, 2016) with multi-panel figures composed using cowplot (Wilke, 2020). In figures, error bars represent ± 1 SE; in Figure 3, shaded ribbons represent ± 2 SE to enhance visibility given the scale of the panels. All analysis code and simulation scripts are available at [https://github.com/cmaltec/ROH-GRM-penalty-simulation].

RESULTS

Simulations compared GRM-based and ROH-based inbreeding penalties across 2 genetic scenarios differing in dominance variance and inbreeding depression severity (Table 1; Figure 1). The Low scenario ($\sigma^2_D/\sigma^2_A = 0.1$; ID = $0.3\sigma_G$ per 10% ΔF) represented mild non-additive effects, while the High scenario ($\sigma^2_D/\sigma^2_A = 0.5$; ID = $1.2\sigma_G$ per 10% ΔF) represented substantial dominance effects typical of fitness-related traits. Results are reported as changes (Δ) over the 20-generation tracking period (generations 131–150), presented as means ± standard deviations across 10 simulation replicates.

Control Selection Methods

Selection on estimated breeding values (EBV) without inbreeding penalties served as the baseline representing

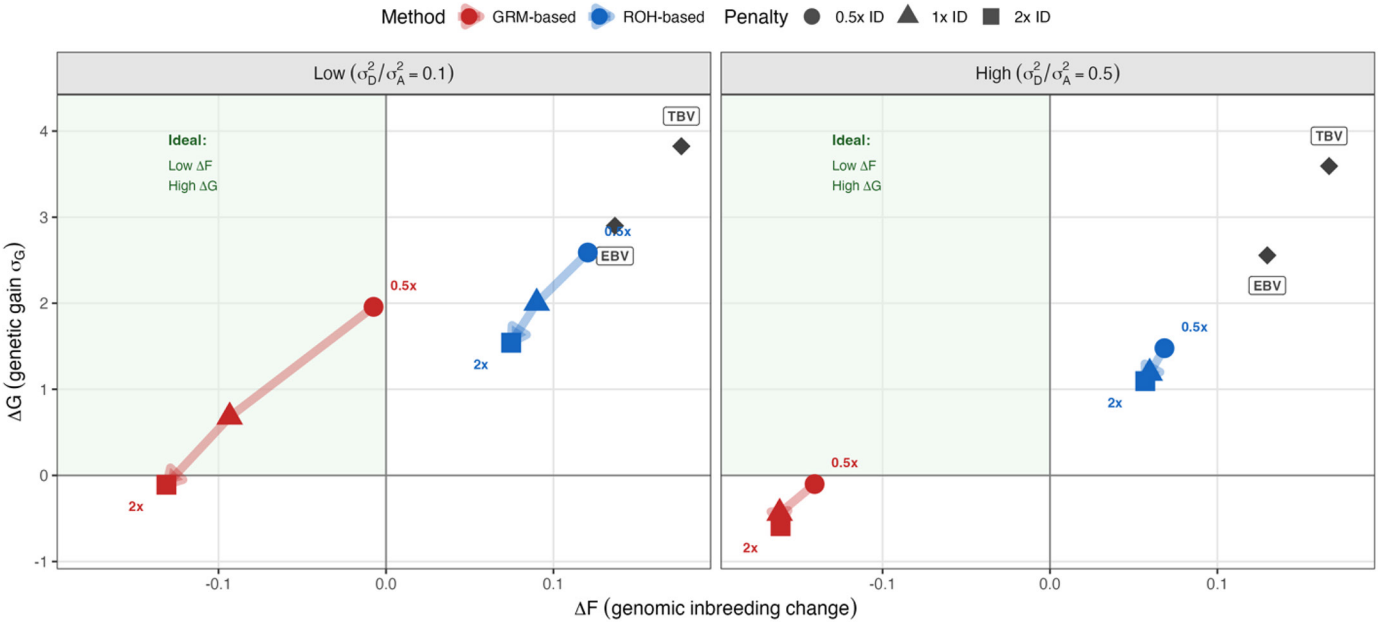

Figure 2. Trade-off between genetic gain (ΔG) and inbreeding accumulation (ΔF) over 20 generations. Green shading indicates the ideal region (reduced inbreeding with positive gain). GRM-based (red) and ROH-based (blue) penalties shown at 3 intensities (0.5x, 1x, 2x the inbreeding depression parameter); arrows indicate direction of increasing penalty. Controls: TBV = true breeding value selection, EBV = estimated breeding value selection. Left: Low scenario (VarD/VarA = 0.1, ID = 0.3 σ 10% ΔF). Right: High scenario (VarD/VarA = 0.5, ID = 1.2 σ 10% ΔF). Strong penalties in the High scenario sacrifice genetic gain to control inbreeding.

Table 1. Genetic scenario parameters. VarA = additive genetic variance (fixed at 100); VarD = dominance variance. ID = inbreeding depression parameter; ID (σ10%dF) = depression in genetic standard deviations per 10 percentage point increase in genomic F. Genome: 6 chromosomes x 2,000 markers = 12,000 SNPs; 12,000 QTL. Population: 10,000 founders; 500 males + 9,500 females; 200 sires x 5,000 dams = 5,000 offspring/generation. Burn-in: 130 generations with two accuracy phases (r = 0.30 for gen 1–29, r = 0.50 for gen 30–130). Tracking: 20 generations (131–150, r = 0.75) with experimental treatments; 10 replicates per combination

Scenario	VarA	VarD	VarD/VarA	ID	ID (σ10%ΔF)	Penalty weights	Penalty/ID
Additive Only	100	0	0	0	0	0, 1, 2	—
Low Non-Additive	100	10	0.10	3	0.30	1, 3, 6	0.5x, 1x, 2x
High Non-Additive	100	50	0.50	12	1.20	6, 12, 24	0.5x, 1x, 2x

current genomic selection practice (Table 2). Over the 20-generation tracking period, EBV selection accumulated $\Delta F = 0.137 \pm 0.012$ in the Low scenario and $\Delta F = 0.130 \pm 0.011$ in the High scenario. Genetic gain was $\Delta G = 2.90 \pm 0.09\ \sigma_G$ and $2.56 \pm 0.12\ \sigma_G$ in the Low and High scenarios, respectively, with the lower gain in the High scenario reflecting the impact of inbreeding depression. EBV selection retained approximately 65% of additive genetic variance present at tracking start.

Selection on true breeding values (TBV) served as a positive control representing the theoretical maximum achievable with perfect information. TBV selection achieved higher genetic gains than EBV ($\Delta G = 3.82\ \sigma_G$ in Low; $3.59\ \sigma_G$ in High) but also accumulated more inbreeding ($\Delta F = 0.176$ in Low; 0.167 in High), consis-

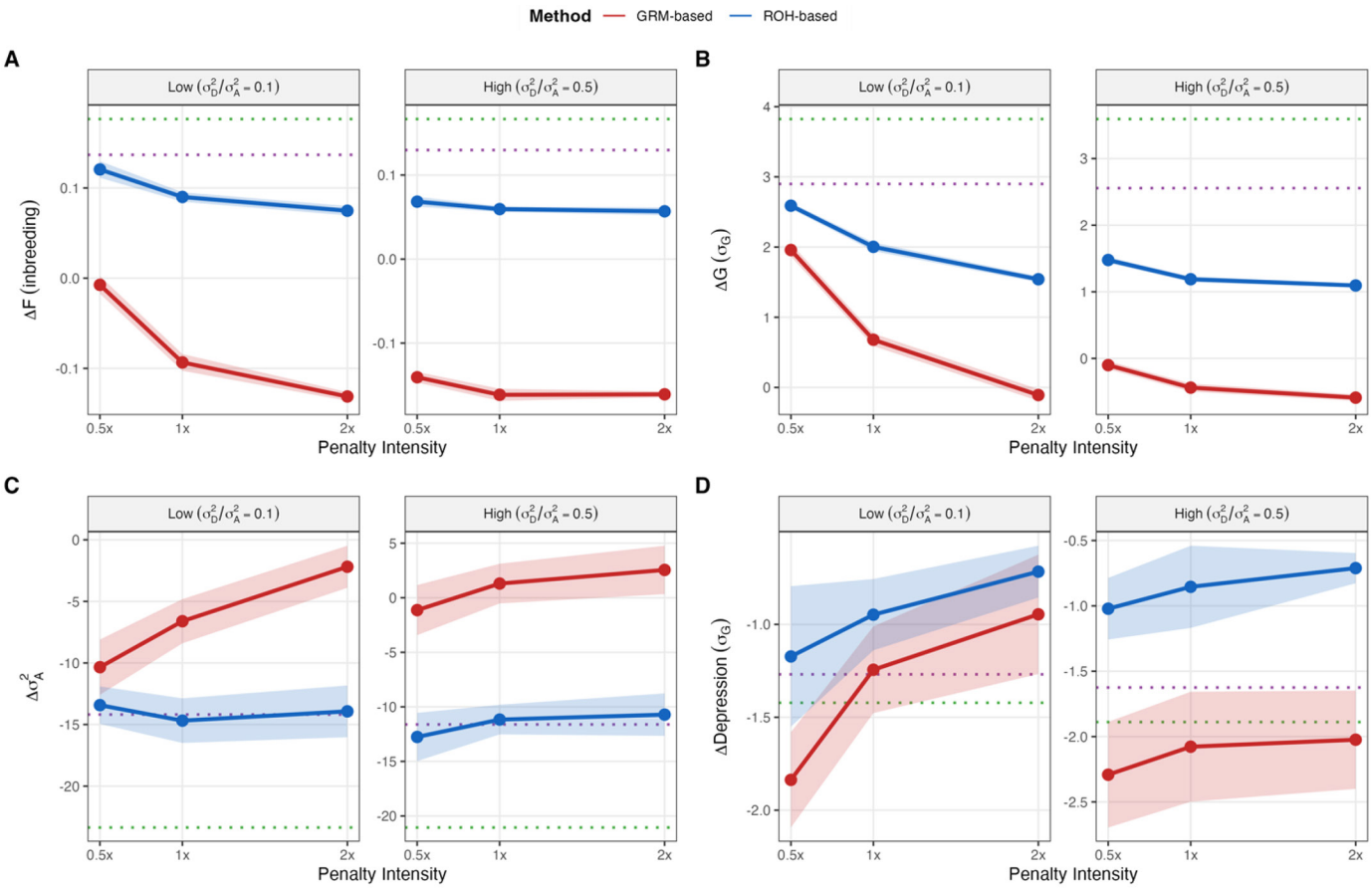

Figure 3. Effect of penalty intensity on selection outcomes over 20 generations. **(A)** Change in genomic inbreeding (DeltaF); negative values indicate net reduction below starting levels. **(B)** Genetic gain (DeltaG) in genetic standard deviation units. **(C)** Change in additive genetic variance (DeltaVarA). **(D)** Change in cumulative inbreeding depression in sigma units; more negative values indicate greater depression accumulated. GRM-based penalties (red) and ROH-based penalties (blue) compared at 0.5x, 1x, and 2x the inbreeding depression parameter. Left facets: Low scenario (VarD/VarA = 0.1); Right facets: High scenario (VarD/VarA = 0.5). Dotted horizontal lines show control values (TBV, EBV). Shaded ribbons: +/−2 SE (n = 10 replicates).

Table 2. Performance of control selection methods. Values are mean +/− SD (n = 10 replicates). TBV = True Breeding Value. EBV = Estimated Breeding Values. ID expressed as genetic standard deviations lost per 10% increase in genomic F

Scenario	Method	Role	ΔF	ΔG (σ)	ΔDepression (σ)
Additive (ID = 0)	TBV	Positive control (theoretical maximum)	0.178 +/− 0.020	4.01 +/− 0.13	0.00 +/− 0.00
	EBV	Baseline (current practice)	0.145 +/− 0.014	3.03 +/− 0.15	0.00 +/− 0.00
Low (ID = 0.3σ10%dF)	TBV	Positive control (theoretical maximum)	0.176 +/− 0.015	3.82 +/− 0.14	−1.42 +/− 0.59
	EBV	Baseline (current practice)	0.137 +/− 0.012	2.90 +/− 0.09	−1.27 +/− 0.50
High (ID = 1.2σ10%dF)	TBV	Positive control (theoretical maximum)	0.167 +/− 0.011	3.59 +/− 0.08	−1.89 +/− 0.55
	EBV	Baseline (current practice)	0.130 +/− 0.011	2.56 +/− 0.12	−1.62 +/− 0.55

tent with stronger selection pressure on the true genetic signal. TBV selection retained only 42–45% of additive variance, reflecting more aggressive allele frequency changes.

Trade-off Between Genetic Gain and Inbreeding Control

The trade-off between genetic gain and inbreeding control differed between GRM-based and ROH-based approaches (Figure 2; Table 3). GRM-based penalties generally achieved lower inbreeding accumulation than ROH-based penalties across penalty intensities tested. In the Low scenario, GRM penalties at 0.3 × , 1 × , and 2 × the inbreeding depression parameter achieved ΔF = −0.007, −0.093, and −0.131, respectively, while ROH penalties at the same intensities achieved ΔF = +0.120, +0.090, and +0.075. The negative ΔF values for GRM indicate net reduction in inbreeding below starting levels.

Similar patterns emerged in the High scenario, where GRM penalties achieved ΔF ranging from −0.141 to −0.161 compared with ΔF = +0.057 to +0.068 for ROH penalties. Across all comparisons, GRM-based selection achieved ΔF approximately 0.13 to 0.22 lower than ROH-based selection at matched penalty intensities (Table 4).

However, the greater inbreeding control of GRM-based methods was associated with lower genetic gain. ROH-based selection achieved higher genetic gain than GRM-based selection at matched penalty intensities, with differences ranging from 0.6 to 1.7 σ_G (Table 4). In the Low scenario at the 1 × penalty level, ROH achieved ΔG = 2.00 σ_G compared with 0.68 σ_G for GRM. In the High scenario, GRM penalties resulted in negative genetic gain (ΔG = −0.10 to −0.59 σ_G), indicating limited net genetic progress under these conditions.

Effect of Penalty Intensity

Increasing penalty intensity produced different responses between methods (Figure 3). For GRM-based selection, stronger penalties progressively reduced inbreeding accumulation, with ΔF declining from −0.007 at 0.3 × to −0.131 at 2 × in the Low scenario, with corresponding reductions in genetic gain from 1.96 σ_G to −0.11 σ_G (Figure 3A, B). Changes in additive variance ($\Delta\sigma^2_A$) followed a similar pattern, with GRM penalties preserving more variance than EBV selection across intensities (Figure 3C).

ROH-based penalties showed a flatter response to increasing intensity. In the Low scenario, ΔF changed from +0.120 to +0.075 across the penalty range, while genetic gain decreased from 2.59 to 1.54 σ_G. At the strongest penalty level, ROH-based selection maintained positive inbreeding accumulation while achieving approximately half the genetic gain of EBV selection.

Comparison Across Dominance Scenarios

The 2 methods responded differently to increasing dominance variance and inbreeding depression severity (Table 3). GRM-based selection showed greater sensitivity to scenario: at the 1 × penalty level, ΔF shifted from −0.093 in Low to −0.161 in High, while ΔG shifted from +0.68 to −0.44 σ_G. ROH-based selection showed more consistent behavior across scenarios, with ΔF changing from +0.090 to +0.059 and ΔG from +2.00 to +1.19 σ_G at the same penalty level.

The difference in depression accumulation between methods was larger in the High scenario. At 1 × penalty, GRM accumulated 2.4 × more depression than ROH in the High scenario (−2.08 vs. −0.85 σ_G) compared with 1.3 × more in the Low scenario (−1.24 vs. −0.95 σ_G). Variance retention patterns also differed: GRM achieved > 100% additive variance retention only in the High scenario, suggesting that the stronger penalties applied in response to greater inbreeding depression were sufficient to recover variance lost during burn-in.

Genetic Variance Retention

GRM-based selection retained more additive genetic variance than ROH-based selection (Figure 4A; Table 3). In the Low scenario, GRM penalties retained 76–95% of starting additive variance compared with 66–69% for ROH penalties. In the High scenario, GRM penalties retained 98–107% compared with 69–74% for ROH. Val-

Table 3. Changes in genetic parameters over 20 generations of selection. Values are mean +/− SD (n = 10 replicates). Δ*F* = change in genomic inbreeding; ΔG = genetic gain (sigma). VarA0 = additive variance at tracking start (after 130-gen burn-in; founder value = 100). VarA Ret, VarD Ret = variance retained as % of tracking start value. Retention > 100% indicates recovery of variance depleted during burn-in via maintained heterozygosity. ΔF/gen = rate of inbreeding per generation; negative = net reduction

Scenario	Method	ΔF	ΔG (σ)	VarA0	VarA Ret	VarD Ret	ΔF/gen
Additive (VarD/VarA = 0, ID = 0)	EBV	0.145 +/− 0.014	3.03 +/− 0.15	40.9	67%	—	0.008
	GRM (0.5x)	0.084 +/− 0.010	2.73 +/− 0.13	41.4	66%	—	0.004
	GRM (1x)	0.033 +/− 0.020	2.36 +/− 0.13	41.4	71%	—	0.002
	GRM (2x)	−0.052 +/− 0.019	1.32 +/− 0.16	40.8	78%	—	−0.003
	ROH (0.5x)	0.135 +/− 0.014	2.92 +/− 0.12	41.4	65%	—	0.007
	ROH (1x)	0.124 +/− 0.012	2.73 +/− 0.14	41.4	63%	—	0.007
	ROH (2x)	0.104 +/− 0.008	2.32 +/− 0.10	40.8	65%	—	0.005
	TBV	0.178 +/− 0.020	4.01 +/− 0.13	42.2	42%	—	0.009
Low (VarD/VarA = 0.1, ID = 0.3σ10%ΔF)	EBV	0.137 +/− 0.012	2.90 +/− 0.09	42.1	66%	91%	0.007
	GRM (0.5x)	−0.007 +/− 0.016	1.96 +/− 0.15	43.4	77%	105%	−0.000
	GRM (1x)	−0.093 +/− 0.015	0.68 +/− 0.14	43.3	85%	111%	−0.005
	GRM (2x)	−0.131 +/− 0.008	−0.11 +/− 0.13	43.9	95%	114%	−0.007
	ROH (0.5x)	0.121 +/− 0.015	2.59 +/− 0.07	43.4	69%	93%	0.006
	ROH (1x)	0.090 +/− 0.008	2.00 +/− 0.10	43.3	66%	93%	0.005
	ROH (2x)	0.075 +/− 0.008	1.54 +/− 0.07	43.9	69%	95%	0.004
	TBV	0.176 +/− 0.015	3.82 +/− 0.14	40.4	42%	91%	0.009
High (VarD/VarA = 0.5, ID = 1.2σ10%ΔF)	EBV	0.130 +/− 0.011	2.56 +/− 0.12	32.0	64%	96%	0.007
	GRM (0.5x)	−0.141 +/− 0.010	−0.10 +/− 0.10	41.1	98%	109%	−0.007
	GRM (1x)	−0.161 +/− 0.011	−0.44 +/− 0.10	40.2	104%	110%	−0.008
	GRM (2x)	−0.161 +/− 0.006	−0.59 +/− 0.08	40.8	107%	108%	−0.008
	ROH (0.5x)	0.068 +/− 0.009	1.48 +/− 0.07	41.1	69%	98%	0.004
	ROH (1x)	0.060 +/− 0.005	1.19 +/− 0.07	40.2	72%	97%	0.003
	ROH (2x)	0.057 +/− 0.007	1.09 +/− 0.04	40.8	74%	97%	0.003
	TBV	0.167 +/− 0.011	3.59 +/− 0.08	37.9	45%	97%	0.009

ues exceeding 100% indicate partial recovery of variance depleted during the 130-generation burn-in phase.

Dominance variance showed similar patterns (Figure 4B). GRM-based selection retained more dominance variance than ROH selection in both scenarios, consistent with the frequency-dependent nature of dominance variance ($\sigma^2_D = \Sigma(2pq)^2d^2$), where maintaining allele frequencies closer to intermediate values preserves more dominance variance.

The rate of inbreeding accumulation per generation (Figure 4C) differed between approaches. GRM-based methods achieved negative rates (approximately −0.5% to −0.8% per generation), while ROH-based methods maintained positive rates (approximately +0.3% to +0.6% per generation), yet lower than EBV selection (+0.7% per generation).

DISCUSSION

This study aimed to fill a gap in the literature by comparing GRM-based and ROH-based inbreeding penalties when used directly in selection indices. Previous work has explored these approaches within optimal contribution selection or genomic mating frameworks, but to our knowledge, no simulation study has tested them as expected progeny inbreeding penalties across multiple generations of selection.

Differential performance of GRM-based and ROH-based penalties

Our results suggest that GRM-based and ROH-based penalties may work differently rather than being interchangeable. At similar penalty intensities, GRM-based

Table 4. Direct comparison of GRM vs ROH at equivalent penalty intensities. Diff = GRM value minus ROH value. Negative ΔF difference indicates GRM achieved lower inbreeding than ROH. Positive VarA retained difference indicates GRM preserved more genetic variance

		ΔF (inbreeding)			ΔG (gain, sigma)			VarA retained (%)		
Scenario	Penalty	GRM	ROH	Diff	GRM	ROH	Diff	GRM	ROH	Diff
High (ID = 1.2σ10%ΔF)	0.5x	−0.141	0.068	−0.209	−0.10	1.48	−1.58	98	69	+28
	1x	−0.161	0.060	−0.221	−0.44	1.19	−1.63	104	72	+31
	2x	−0.161	0.057	−0.218	−0.59	1.09	−1.68	107	74	+33
Low (ID = 0.3σ10%ΔF)	0.5x	−0.007	0.121	−0.128	1.96	2.59	−0.63	77	69	+7
	1x	−0.093	0.090	−0.183	0.68	2.00	−1.32	85	66	+19
	2x	−0.131	0.075	−0.206	−0.11	1.54	−1.65	95	69	+27

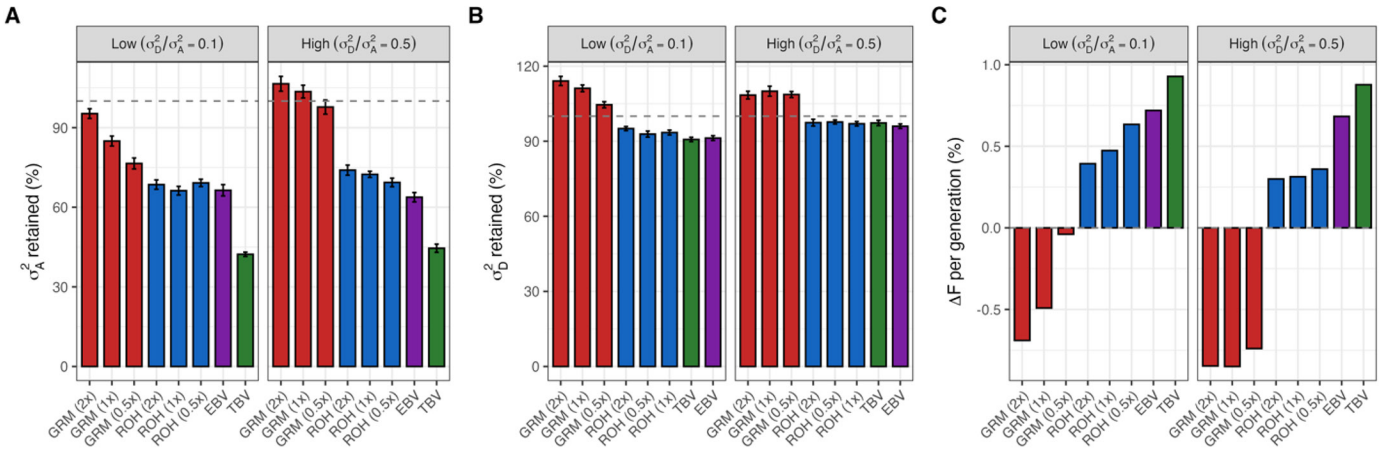

Figure 4. Long-term selection sustainability. **(A)** Proportion of additive genetic variance (VarA) retained after 20 generations relative to tracking start. Dashed line at 100% indicates no net change. Note: tracking begins after 130 generations of EBV burn-in, during which VarA declined from founder value (100) to ~30–45 depending on scenario. Retention > 100% indicates partial recovery of variance lost during burn-in. **(B)** Proportion of dominance variance (VarD) retained. Values exceeding 100% reflect maintained heterozygosity: since VarD depends on $(2pq)^2$, reduced selection intensity keeps allele frequencies closer to intermediate values. **(C)** Rate of inbreeding accumulation per generation; negative values indicate net inbreeding reduction. Methods compared across Low (VarD/VarA = 0.1) and High (VarD/VarA = 0.5) scenarios. Error bars: +/−1 SE (n = 10 replicates).

selection tended to achieve lower genomic inbreeding (ΔF roughly 0.13 to 0.22 lower than ROH across scenarios), and in some cases produced negative ΔF values, pointing to a net reduction in inbreeding over 20 generations. ROH-based selection, on the other hand, appeared to maintain higher genetic gain (around 1.0 to 2.0 σG more than GRM) while slowing rather than reversing inbreeding accumulation. These patterns likely reflect what each metric captures: GRM measures genome-wide allelic similarity through departures from Hardy-Weinberg expectations, while ROH identifies contiguous stretches of homozygosity arising from recent common ancestry (VanRaden, 2008; McQuillan et al., 2008).

Comparison with prior studies

These findings appear to contrast with those of Zhao et al. (2023), who reported in a simulation study using Large White pigs, that ROH-based genomic mating in Large White pigs outperformed SNP-based approaches for genetic gain, though their largest reported differences (0.9–2.6% higher ΔG, 13–83% lower ΔF) were relative to positive assortative mating. This difference might stem from how the methods were applied: their study used ROH-based genomic relationship matrices to optimize mate allocation after selection, while we applied ROH-based penalties directly during candidate ranking. Genomic mating can exploit complementarity between selected parents in ways that may not be available when ROH information is used only for penalizing candidates. Their shorter time frame (5 generations versus our 20) may also explain why the longer-term trade-offs we observed did not appear in their results.

Alemu et al. (2021) found that FUNI, the inbreeding coefficient based on the correlation between uniting gametes (Yang et al., 2011) and FGRM show stronger correlations than FROH with the burden of rare deleterious alleles (minor allele frequency <0.15), while FROH provides a more accurate measure of total genome-wide autozygosity and captures autozygosity at loci where derived alleles segregate at low-to-moderate frequencies.

Forutan et al. (2018) showed in gene-dropping simulations of North American Holsteins that ROH with minimum window sizes of 20–50 SNPs came closest to true inbreeding, while GRM-based estimates varied with assumptions about base allele frequencies. Together, these findings suggest that GRM-based penalties may be more effective at limiting rare deleterious alleles across the genome, while ROH-based penalties focus on avoiding long homozygous segments from recent shared ancestry.

Role of genetic architecture and dominance variance

The size of the difference between methods depended heavily on the genetic scenario. In the purely additive scenario, GRM and ROH penalties produced nearly the same outcomes across all metrics, which suggests that dominance and inbreeding depression may be necessary for the methods to perform differently. Differences were roughly twice as large in the High dominance scenario ($\sigma D^2/\sigma A^2$ = 0.5, inbreeding depression = 1.2σG per 10% ΔF) compared with the Low scenario ($\sigma D^2/\sigma A^2$ = 0.1, inbreeding depression = 0.3σG per 10% ΔF).

The dominance variance ratios we simulated fall within ranges reported for dairy cattle. Sun et al. (2014) estimated dominance variance at about 5% of phenotypic variance in US Holsteins and 7% in Jerseys for yield traits, with other traits showing little dominance contribution. Ertl et al. (2014) found dominance variance ranging from 3.3% to 50.5% of total genetic variance across production and functional traits in Fleckvieh cattle. More recently, Schneider et al. (2023) reported that while dominance contributes relatively little to phenotypic variance (ranging from 1.8% for ovarian cysts to 7.8% for milk yield), it can account for a substantial share of genetic variance for health traits specifically, reaching 55.1% for mastitis and 23.3% for ovarian cysts. These estimates suggest both our scenarios have practical relevance: the Low scenario ($\sigma D^2/\sigma A^2 = 0.1$) may approximate production traits, while the High scenario ($\sigma D^2/\sigma A^2 = 0.5$) may better reflect health and fitness traits where non-additive effects seem more important.

Genetic variance conservation under contrasting penalty approaches

GRM-based selection appeared to conserve genetic variance better, retaining roughly 8–13% more additive variance than ROH at similar penalty intensities. This difference was more noticeable in the High dominance scenario, where GRM achieved over 100% variance retention relative to the start of tracking, suggesting it may have recovered some diversity lost during the burn-in period. This pattern could reflect the genome-wide nature of GRM penalties, which may spread selection pressure more evenly across the genome compared with ROH penalties that act on specific homozygous segments.

These observations fit with what is known about long-term variance changes under genomic selection. Wientjes et al. (2022) found that GBLUP retained less than 20% of initial additive genetic variance after 50 generations under purely additive models, similar to pedigree-based BLUP. However, the mechanisms differed: GBLUP kept more loci segregating (only 40–42% lost) but at lower minor allele frequencies, while pedigree BLUP lost more loci (~50%) but maintained higher frequencies at remaining ones. Our results hint that GRM-based inbreeding penalties might further help retain loci by working against the drift toward fixation that comes with strong selection.

The Bulmer effect, which reduces genetic variance through negative covariance between allelic effects caused by selection, has been measured by Van Grevenhof et al. (2012), who found that at 5% selection intensity, response drops by 27% regardless of initial accuracy. This reduction appears the same for genomic and pedigree selection. In our simulations, TBV-based selection (the most intense) depleted additive variance fastest, while GRM-based penalties at higher intensities showed the greatest variance preservation. The combination of reduced selection intensity and genome-wide inbreeding control under GRM penalties may buffer against both the Bulmer effect and longer-term variance erosion.

Dominance variance behaved differently from additive variance. Under ROH-based selection, dominance variance declined steadily, consistent with increased homozygosity reducing heterozygote frequency and thus dominance expression. Under GRM-based selection, dominance variance held steady or even increased, especially in the High scenario. This pattern makes sense given that dominance variance ($\sigma D^2 = \Sigma(2pq)^2 d^2$) is largest when allele frequencies are intermediate. The stronger inbreeding control from GRM-based penalties may keep more loci at intermediate frequencies, preserving conditions for dominance variance.

Inbreeding depression dynamics and the role of purging

An unexpected pattern emerged for inbreeding depression (Figure 3D). GRM-based selection accumulated more inbreeding depression than ROH-based selection at matched penalty intensities, despite achieving lower inbreeding levels. In the Low scenario at the 1 × penalty level, GRM accumulated ΔDepression = −1.24 σ_G compared with −0.95 σ_G for ROH, despite GRM having lower inbreeding (ΔF = −0.093 vs. +0.090). In the High scenario, GRM accumulated ΔDepression = −2.08 σ_G at 1 × compared with −0.85 σ_G for ROH.

This pattern may reflect differences in selection intensity and purging dynamics between methods. ROH-based selection maintains stronger directional selection pressure, which could allow more effective purging of deleterious variants responsible for inbreeding depression. GRM-based selection, by constraining selection more heavily to control inbreeding, may retain more segregating deleterious recessives.

One possible explanation is that purging of deleterious recessive alleles may be more effective under ROH-based selection. The higher selection intensity permitted by moderate inbreeding control could allow stronger selection against individuals expressing recessive genetic load. When deleterious alleles become homozygous and reduce performance, selection removes them more effectively. In contrast, the strict inbreeding control from GRM-based penalties maintains heterozygosity, which may shelter deleterious recessive alleles from selection in carrier individuals.

This idea aligns with studies distinguishing recent from ancient inbreeding effects in dairy cattle. Doekes et al. (2019) found that in Dutch Holstein-Friesians, new

inbreeding (FNEW) caused −2.42 kg fat yield per 1% increase, while ancient inbreeding (FANC) had essentially no effect (+0.03 kg). Makanjuola et al. (2020) reported similar patterns in Canadian Holsteins: overall genomic inbreeding (FROH) caused roughly −40 to −49 kg milk yield per 1% increase, with longer ROH (>4 Mb) showing significant unfavorable effects, while shorter ROH (<4 Mb, representing ancient inbreeding) showed nonsignificant effects. This fading impact of older inbreeding is consistent with historical purging having removed the most harmful alleles over time.

The theoretical framework developed by García-Dorado (2012) predicts that genetic purging becomes effective when the product of effective population size and the purging coefficient (Nd) equals or exceeds 1. This threshold translates empirically to populations with Ne ≥10, where purging can substantially reduce inbreeding depression across a range of deleterious mutations (Pérez-Pereira et al., 2022).

Gulisija and Crow (2007) estimated about 17% reduction of total genetic load from purging in US Jersey cattle with deep pedigrees, mainly affecting recessive alleles with large fitness effects. Our simulation setup, with moderate effective population sizes and defined inbreeding depression, may have created conditions where purging could occur under the higher inbreeding rates allowed by ROH-based selection.

Practical evidence supports the idea that targeted selection against deleterious alleles can work. The Council on Dairy Cattle Breeding (CDCB) now tracks 27 recessive conditions (Cole et al., 2022), and carrier frequencies for many haplotypes have decreased substantially since genomic testing became widespread. Economic losses from defect haplotypes fell from $11 million (Cole et al., 2016) to $4.1 million through genomic management (Cole et al., 2025). This shows that selecting against carriers of harmful alleles, while accepting some increase in homozygosity elsewhere, can reduce genetic load without causing severe inbreeding depression.

Practical implications for breeding program design

These results suggest that choosing a penalty method should depend on breeding program goals. GRM-based penalties may be more appropriate when long-term genetic diversity is the priority, such as in populations with small effective sizes, conservation programs, or when maintaining future selection response matters most. Current estimates put Holstein Ne at 43–102 across populations (Makanjuola et al., 2020; Doekes et al., 2018), close to or below the FAO threshold of 50 where populations may start losing long-term fitness. For such populations, the better inbreeding control and variance retention from GRM-based penalties could outweigh the short-term gain advantage of ROH-based approaches.

ROH-based penalties might work better when short-term genetic gain is the focus and moderate inbreeding is acceptable, as in commercial populations with regular outside genetics or for traits with minimal inbreeding depression. The lower depression observed under ROH-based selection, possibly from enhanced purging, could be particularly useful when breeding goals include fitness traits where deleterious recessive alleles are common.

Cole (2024) described the core challenge in dairy breeding: demand for elite genetics conflicts with managing inbreeding. Current EFI adjustments in US genetic evaluations account for inbreeding depression by reducing breeding values proportionally to expected future inbreeding (approximately $40 per 1% EFI for Net Merit), though these adjustments reflect population-average relationships rather than mate-specific inbreeding risk. As the collective dairy herd becomes more genomically homogeneous, the ability of EFI adjustments to meaningfully rerank individuals diminishes: between 2015 and 2020, the Holstein base population's EFI rose from 7.5% to 9.4%, compressing the variance in individual EFI values and reducing the discriminatory power of inbreeding adjustments in genetic evaluations (CDCB, 2025).

Limitations

Several limitations should be kept in mind. First, the simulations assumed complete linkage equilibrium between markers and QTL, which may not hold when strong selection creates marker-QTL associations. Henryon et al. (2019) found that genomic optimal contribution selection can inadvertently penalize beneficial allele frequency changes at markers linked to QTL, achieving 1.5–5.7% less true genetic gain than pedigree-based OCS at the same inbreeding rate. Similar effects might influence penalty-based approaches and could partly explain the gain reduction we observed under GRM-based selection.

Second, the ROH-based expected progeny inbreeding used a heuristic approximation rather than a theoretically exact segment-based relationship matrix. This creates an asymmetry in how the 2 methods estimate future inbreeding: the GRM method directly calculates genomic relationships between each candidate and the dam population, while the ROH method averages parental autozygosity without accounting for whether homozygous segments overlap between potential mates. Two candidates with identical FROH values but autozygosity in different genomic regions would receive the same penalty despite potentially producing offspring with different inbreeding levels. This limitation may have disadvantaged the ROH method in our comparison, as it cannot exploit mate com-

plementarity the way a true segment-based coancestry matrix would. The relative underperformance of ROH-based penalties for inbreeding control could therefore be partially attributable to this methodological constraint rather than an inherent limitation of ROH as an inbreeding metric. Future work implementing computationally efficient segment-based relationship matrices, perhaps using approximate haplotype matching or dimensionality reduction techniques, could help determine whether ROH-based penalties can achieve inbreeding control comparable to GRM when both methods use equivalent relationship information.

Third, the ROH detection threshold (1 Mb) was chosen to balance sensitivity and specificity but may not be optimal for all situations. Curik et al. (2014) showed that different ROH lengths correspond to different inbreeding ages: FROH > 16Mb reflects about 3 generations ago, FROH > 4Mb about 12.5 generations, and FROH > 1Mb about 50 generations. Our 1 Mb threshold captures both recent and older inbreeding, which may not suit all objectives. Forutan et al. (2018) found that ROH with 20–50 SNP minimum windows came closest to true inbreeding, suggesting that parameter tuning could improve ROH-based penalty performance.

Fourth, we did not include epistasis, genotype-by-environment interactions, or the interplay between penalty methods and mate allocation. Wientjes et al. (2022) found that including epistasis in models led to about 24% variance retention versus less than 20% for additive-only models, because epistatic effects can reverse selection direction at individual loci across generations. Real dairy traits likely involve some epistasis, which could change how the penalty methods compare.

Fifth, the depression paradox we observed needs validation with real data, as the purging explanation remains a hypothesis. The relationship between selection intensity and purging has competing effects: higher intensity can help purging by increasing homozygote frequency, but can also hurt it by increasing drift that fixes deleterious alleles before selection acts. Grossen et al. (2020) found in Alpine ibex that severe bottlenecks purged highly deleterious mutations but allowed mildly deleterious ones to accumulate, showing that mutation effect size matters for purging success. Our simulations used uniform inbreeding depression effects, which may not capture this complexity.

Sixth, One practical consideration not fully captured in our simulation is that GFI and EFI values are expressed relative to a rolling reference population, so individual bulls' PTA adjustments will shift between evaluation rounds as the base population becomes more uniformly related — not because the measurement is erratic but because the reference is changing; this is an intrinsic property of relative indexing against a dynamic population that breeders and program managers need to account for when comparing or interpreting proof changes across evaluation rounds (CDCB, 2025).

Finally, while designed to approximate dairy cattle breeding, our simulated population structure is a simplification. Real breeding decisions are spread across thousands of farms and multiple AI organizations, making coordinated inbreeding management difficult (Cole, 2024). The penalty approaches we evaluated assume centralized selection decisions, and how well they would work in more realistic decentralized settings remains unclear.

CONCLUSIONS

This study suggests that GRM-based and ROH-based inbreeding penalties offer complementary approaches to managing genetic diversity in genomic selection programs, with different trade-offs between genetic gain and inbreeding control. GRM-based penalties tended to achieve better inbreeding control and genetic variance retention but at the cost of lower genetic gain, while ROH-based penalties appeared to maintain higher genetic gain with less cumulative inbreeding depression despite higher autozygosity accumulation.

Several findings may have practical relevance: GRM-based penalties can achieve negative inbreeding rates, suggesting potential to recover genetic diversity lost during historical selection, which could be valuable for populations approaching critical effective population sizes; ROH-based penalties allow higher selection intensity and possibly more effective purging of deleterious recessive alleles, which may be useful when breeding for fitness traits; method differences scale with dominance variance, suggesting that trait-specific penalty selection might improve outcomes when indices span traits with different genetic architectures; and neither method appears universally better, highlighting the importance of matching penalty choice to specific breeding goals.

Future work could explore optimal penalty calibration across populations with different effective sizes, examine how penalty-based selection interacts with mate allocation strategies as used in genomic mating programs, and test the purging hypothesis with real breeding data to clarify the biological basis of these findings. Looking further ahead, if site-directed genome editing is used to introduce targeted genetic variation or to eliminate recessive defects, the relationship between ROH-based and GRM-based inbreeding metrics and functional genetic load may diverge in ways that current simulation frameworks — grounded in classical mutation-selection balance — cannot fully capture, warranting reconsideration of how future inbreeding is quantified as these technologies approach commercial application. As dairy cattle populations continue experiencing elevated in-

breeding rates following genomic selection, developing and evaluating effective inbreeding management tools remains important for sustainable genetic improvement.

NOTES

Christian Maltecca, and Francesco Tiezzi acknowledge funding from the National Italian Holstein Brown and Jersey breeders' association ANAFIBJ (Cremona, ITA).

REFERENCES

Ablondi, M., A. Summer, G. Stocco, R. Finocchiaro, J. T. van Kaam, M. Cassandro, C. Dadousis, A. Sabbioni, and C. Cipolat-Gotet. 2023. The role of inbreeding depression on productive performance in the Italian Holstein breed. J. Anim. Sci. 101:skad382. https://doi.org/10.1093/jas/skad382.

Alemu, S. W., N. K. Kadri, C. Harland, P. Faux, C. Charlier, A. Caballero, and T. Druet. 2021. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. Heredity 126:410–423. https://doi.org/10.1038/s41437-020-00383-9.

Baes, C. F., B. O. Makanjuola, F. Miglior, G. Marras, J. T. Howard, A. Fleming, and C. Maltecca. 2019. Symposium review: The genomic architecture of inbreeding: How homozygosity affects health and performance. J. Dairy Sci. 102:2807–2817. https://doi.org/10.3168/jds.2018-15520.

CDCB. 2025. What Role Does Inbreeding Play in Genetic Base Changes? Council on Dairy Cattle Breeding. https://uscdcb.com/what-role-does-inbreeding-play-in-genetic-base-changes/

Charlesworth, D., and J. H. Willis. 2009. The genetics of inbreeding depression. Nat. Rev. Genet. 10:783–796. https://doi.org/10.1038/nrg2664.

Cole, J. B. 2024. Perspective: Can we actually do anything about inbreeding? J. Dairy Sci. 107:643–648. https://doi.org/10.3168/jds.2023-23958.

Cole, J. B., C. F. Baes, S. A. E. Eaglen, T. J. Lawlor, C. Maltecca, M. S. Ortega, and P. M. VanRaden. 2025. Invited review: Management of genetic defects in dairy cattle populations. J. Dairy Sci. 108:3045–3067. https://doi.org/10.3168/jds.2024-26035.

Cole, J. B., D. J. Null, and P. M. VanRaden. 2016. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. J. Dairy Sci. 99:7274–7288. https://doi.org/10.3168/jds.2015-10777.

Cole, J. B., P. M. VanRaden, D. J. Null, J. L. Hutchison, and S. M. Hubbard. 2022. Haplotype tests for economically important traits of dairy cattle. AIP Research Report Genomic5 (12–20). Available at: https://www.ars.usda.gov/ARSUserFiles/80420530/Publications/ARR/Haplotype_tests_ARR-Genomic5.pdf

Curik, I., M. Ferenčaković, and J. Sölkner. 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci. 166:26–34. https://doi.org/10.1016/j.livsci.2014.05.034.

Doekes, H. P., R. F. Veerkamp, P. Bijma, G. de Jong, S. J. Hiemstra, and J. J. Windig. 2019. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein-Friesian dairy cattle. Genet. Sel. Evol. 51:54. https://doi.org/10.1186/s12711-019-0497-z.

Doekes, H. P., R. F. Veerkamp, P. Bijma, S. J. Hiemstra, and J. J. Windig. 2018. Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein-Friesian breeding program from 1986 to 2015. Genet. Sel. Evol. 50:15. https://doi.org/10.1186/s12711-018-0385-y.

Doublet, A. C., P. Croiseau, S. Fritz, A. Michenet, C. Hozé, C. Danchin-Burge, D. Laloë, and G. Restoux. 2019. The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. Genet. Sel. Evol. 51:52. https://doi.org/10.1186/s12711-019-0495-1.

Ertl, J., A. Legarra, Z. G. Vitezica, L. Varona, C. Edel, R. Emmerling, and K. U. Götz. 2014. Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. Genet. Sel. Evol. 46:40. https://doi.org/10.1186/1297-9686-46-40.

FAO. (2013). In vivo conservation of animal genetic resources. FAO Animal Production and Health Guidelines, No. 14. Rome: Food and Agriculture Organization of the United Nations.

Forutan, M., S. Ansari Mahyari, C. Baes, N. Melzer, F. S. Schenkel, and M. Sargolzaei. 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genomics 19:98. https://doi.org/10.1186/s12864-018-4453-z.

García-Dorado, A. 2012. Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. Genetics 190:1461–1476. https://doi.org/10.1534/genetics.111.135541.

Gaynor, R. C., G. Gorjanc, and J. M. Hickey. 2021. AlphaSimR: an R package for breeding program simulations. G3 (Bethesda) 11:jkaa017. https://doi.org/10.1093/g3journal/jkaa017.

Grossen, C., F. Guillaume, L. F. Keller, and D. Croll. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nat. Commun. 11:1001. https://doi.org/10.1038/s41467-020-14803-1.

Guinan, F. L., G. Wiggans, D. Norman, J. Dürr, J. Cole, C. Van Tassell, I. Misztal, and D. Lourenco. 2023. Changes in genetic trends in US dairy cattle since the implementation of genomic selection. J. Dairy Sci. 106:1110–1129. https://doi.org/10.3168/jds.2022-22205.

Gulisija, D., and J. F. Crow. 2007. Inferring purging from pedigree data. Evolution 61:1043–1051. https://doi.org/10.1111/j.1558-5646.2007.00088.x.

Henryon, M., H. Liu, P. Berg, G. Su, H. M. Nielsen, G. T. Gebregiwergis, and A. C. Sørensen. 2019. Pedigree relationships to control inbreeding in optimum-contribution selection realise more genetic gain than genomic relationships. Genet. Sel. Evol. 51:39. https://doi.org/10.1186/s12711-019-0475-5.

Howard, J. T., J. E. Pryce, C. Baes, and C. Maltecca. 2017. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J. Dairy Sci. 100:6009–6024. https://doi.org/10.3168/jds.2017-12787.

Keller, M. C., P. M. Visscher, and M. E. Goddard. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics 189:237–249. https://doi.org/10.1534/genetics.111.130922.

Leroy, G. 2014. Inbreeding depression in livestock species: review and meta-analysis. Anim. Genet. 45:618–628. https://doi.org/10.1111/age.12178.

Lozada-Soto, E. A., C. Maltecca, J. Jiang, J. B. Cole, P. M. VanRaden, and F. Tiezzi. 2024. Effects of germplasm exchange strategies on genetic gain, homozygosity, and genetic diversity in dairy stud populations: A simulation study. J. Dairy Sci. 107:11149–11163. https://doi.org/10.3168/jds.2024-24992.

Lozada-Soto, E. A., F. Tiezzi, J. Jiang, J. B. Cole, P. M. VanRaden, and C. Maltecca. 2022. Genomic characterization of autozygosity and recent inbreeding trends in all major breeds of US dairy cattle. J. Dairy Sci. 105:8956–8971. https://doi.org/10.3168/jds.2022-22116.

Makanjuola, B. O., C. Maltecca, F. Miglior, F. S. Schenkel, and C. F. Baes. 2020. Effect of recent and ancient inbreeding on production and fertility traits in Canadian Holsteins. BMC Genomics 21:605. https://doi.org/10.1186/s12864-020-07031-w.

Makanjuola, B. O., F. Miglior, E. A. Abdalla, C. Maltecca, F. S. Schenkel, and C. F. Baes. 2020. Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. J. Dairy Sci. 103:5183–5199. https://doi.org/10.3168/jds.2019-18013.

McQuillan, R., A. L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. Macleod, S. M. Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S. H. Wild, M. G. Dunlop, A. F. Wright, H. Campbell, and J. F. Wilson. 2008. Runs of homozygosity in European populations. Am. J. Hum. Genet. 83:359–372. https://doi.org/10.1016/j.ajhg.2008.08.007.

Meuwissen, T. H. 1997. Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75:934–940. https://doi.org/10.2527/1997.754934x.

Pérez-Pereira, N., A. Caballero, and A. García-Dorado. 2022. Reviewing the consequences of genetic purging on the success of rescue

programs. Conserv. Genet. 23:1–17. https://doi.org/10.1007/s10592-021-01405-7.
Schneider, H., J. Heise, J. Tetens, G. Thaller, R. Wellmann, and J. Bennewitz. 2023. Genomic dominance variance analysis of health and milk production traits in German Holstein cattle. J. Anim. Breed. Genet. 140:390–399. https://doi.org/10.1111/jbg.12765.
Sun, C., P. M. VanRaden, J. B. Cole, and J. R. O'Connell. 2014. Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects. PLoS One 9:e103934. https://doi.org/10.1371/journal.pone.0103934.
Tenhunen, S., J. R. Thomasen, L. P. Sørensen, P. Berg, and M. Kargo. 2024. Genomic analysis of inbreeding and coancestry in Nordic Jersey and Holstein dairy cattle populations. J. Dairy Sci. 107:5897–5912. https://doi.org/10.3168/jds.2023-24553.
Van Doormaal, B., C. Jaton, H. Sweett, P. G. Sullivan, A. Fleming, F. Miglior, and C. M. Richardson. Modernizing Canada's Lifetime Performance Index (LPI). Interbull Bulletin. 2025;61. Published November 17, 2025. Available at: https://journal.interbull.org/index.php/ib/article/view/1988
Van Doormaal, B., F. Miglior, G. Kistemaker, and P. Brand. 2003. Genetic diversification of the Holstein breed in Canada and internationally. Interbull Bull. 30:93–97.
Van Grevenhof, E. M., J. A. Van Arendonk, and P. Bijma. 2012. Response to genomic selection: the Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. Genet. Sel. Evol. 44:26. https://doi.org/10.1186/1297-9686-44-26.
van Kaam, J. B. C. H. M., C. Dadousis, F. Tiezzi, and M. Cassandro. Next-Level Genomic Selection: Mitigating Inbreeding. Interbull Bulletin. 2024;60. Published September 4, 2024. Available at: https://journal.interbull.org/index.php/ib/article/view/1916
VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.
VanRaden, P. M., K. M. Olson, G. R. Wiggans, J. B. Cole, and M. E. Tooker. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 94:5673–5682. https://doi.org/10.3168/jds.2011-4500.
VanRaden, P. M., and L. A. Smith. 1999. Selection and mating considering expected inbreeding of future progeny. J. Dairy Sci. 82:2771–2778. https://doi.org/10.3168/jds.S0022-0302(99)75534-7.
Weigel, K. A., and S. W. Lin. 2002. Controlling inbreeding by constraining the average relationship between parents of young bulls entering AI progeny test programs. J. Dairy Sci. 85:2376–2383. https://doi.org/10.3168/jds.S0022-0302(02)74318-X.
Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.2.
Wientjes, Y. C. J., P. Bijma, M. P. L. Calus, B. J. Zwaan, Z. G. Vitezica, and J. van den Heuvel. 2022. The long-term effects of genomic selection: 1. Response to selection, additive genetic variance, and genetic architecture. Genet. Sel. Evol. 54:19. https://doi.org/10.1186/s12711-022-00709-7.
Wilke, C.O. (2020). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1.
Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. 56:330–338. https://doi.org/10.1086/279872.
Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
Zhao, F., P. Zhang, X. Wang, D. Akdemir, D. Garrick, J. He, and L. Wang. 2023. Genetic gain and inbreeding from simulation of different genomic mating schemes for pig improvement. J. Anim. Sci. Biotechnol. 14:87. https://doi.org/10.1186/s40104-023-00872-x.

ORCIDS

C. Maltecca, https://orcid.org/0000-0002-9996-4680
J. Jiang, https://orcid.org/0000-0001-6890-7539
J. B. Cole, https://orcid.org/0000-0003-1242-4401
C. F. Baes, https://orcid.org/0000-0001-6614-8890
L. Ma, https://orcid.org/0000-0003-1038-1081
F. Tiezzi https://orcid.org/0000-0002-4358-9236