J. Dairy Sci. 108:11254–11268
https://doi.org/10.3168/jds.2025-26738
Published by Elsevier Inc. on behalf of the American Dairy Science Association®.

Genomic partition of inbreeding depression in production traits of US Jersey cattle using functional annotations

Chang Xu,[1]* Christian Maltecca,[1] Francesco Tiezzi,[1,2] John B. Cole,[1,3,4] Sajjad Toghiani,[5] Li Ma,[6] and Jicai Jiang[1]
[1]Department of Animal Science, North Carolina State University, Raleigh, NC 27607
[2]Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, 50144 Firenze, Italy
[3]Council on Dairy Cattle Breeding, Bowie, MD 20716
[4]Department of Animal Sciences, Donald Henry Barron Reproductive and Perinatal Biology Research Program, and the Genetics Institute, University of Florida, Gainesville, FL 32611-0910
[5]USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350
[6]Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742

ABSTRACT

Inbreeding depression (InD) refers to the mean reduction in trait values due to inbreeding, with detrimental effects on survival, production, and reproduction traits that have been observed in many natural and domesticated populations. Despite efforts to measure how much reduction in the traits of interest was caused by InD, the genetic and molecular basis of these declines remains unclear, particularly in dairy cattle. In this research, we used a linear mixed model to partition the InD of 3 production traits in 245,517 genotyped Jersey cows from the Council on Dairy Cattle Breeding (Bowie, MD) database. We mapped 9,532,696 imputed sequence variants into 5 functional annotation categories (i.e., intron, promoter, genomic evolutionary rate profiling [GERP] constrained elements, coding sequence, untranslated regions [UTR], and remaining). We estimated the effects of InD attributed to each functional annotation category by a mixed model method accounting for additive effects and relatedness through a genomic relationship matrix. The InD for milk yield was significantly enriched for promoter regions (enrichment ratio [R_k] = 20.11, SE = 6.44), UTR regions (R_k = 57.96, SE = 16.62) and GERP regions (R_k = 35.91, SE = 7.00). The enrichment ratio R_k represents the disproportionate effect that annotation-specific homozygosity has on the trait mean compared with the magnitude of InD on the whole-genome level. Similarly, protein yield showed significant enrichment of InD for promoter regions (R_k = 15.25, SE = 5.45), UTR regions (R_k = 46.44, SE = 14.07), and GERP regions (R_k = 32.73, SE = 5.92), whereas fat yield showed significant enrichment of InD for UTR regions (R_k = 40.20, SE = 12.77) and GERP regions (R_k = 28.72, SE = 5.34). Our results indicate that certain functional annotations in dairy cattle genome are disproportionally responsible for the detrimental effects of inbreeding, which could be vulnerable to deleterious mutations. This research can help better elucidate the genetic and molecular basis of InD in dairy cattle genome and potentially guide breeding strategies for genomic selection.

Key words: inbreeding depression, dairy cattle, functional annotation

INTRODUCTION

Inbreeding depression (**InD**), the reduction in trait performance due to inbreeding, presents a notable challenge across natural and managed populations (Charlesworth and Charlesworth, 1987). This issue is particularly pressing in dairy cattle, as inbreeding negatively affects important traits related to longevity, production, survival, and reproduction (Fuerst-Waltl and Fuerst, 2012; Mugambe et al., 2024). Despite efforts to measure the extent of trait reduction caused by InD, the genetic and molecular basis contributing to these declines are still not well understood. Research focused on the genetic mechanisms behind InD in livestock population is limited (Ferenčaković et al., 2017), hindering the development of management strategies in breeding programs (Cole, 2024).

Historically, the understanding and quantification of InD have predominantly relied on pedigree-based inbreeding coefficients, which track relatedness over generations (Cassell et al., 2003). However, pedigree inbreeding coefficients ($\boldsymbol{F_{Ped}}$) often face challenges due to incomplete data and their inability to capture Mendelian sampling variability. To address these limitations, researchers have recently turned to genomic inbreeding

Received April 10, 2025.
Accepted July 28, 2025.
*Corresponding author: cxu33@ncsu.edu

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-25. Nonstandard abbreviations are available in the Notes.

coefficients, which use data derived from SNPs (VanRaden, 2008) or runs of homozygosity (**ROH**; Purfield et al., 2012). These genomic measures provide a more accurate representation of actual inbreeding levels in individuals. However, genomic methods still have shortcomings, particularly in distinguishing between alleles that are identical by descent and those that are identical by state (Caballero et al., 2021). Further initiatives to refine the estimation of inbreeding measures have emphasized the importance of distinguishing between recent and ancient inbreeding events (Sumreddee et al., 2020, 2021). Specifically, new methodologies have emerged for classifying ROH by their length, as this classification correlates with the recent inbreeding occurrences (Doekes et al., 2019).

Recent progress in the assessment of inbreeding provided a foundation for exploring the impacts of InD within specific population, where the unique interplay of genetic architecture and population dynamics shapes inbreeding patterns. However, substantial gaps remain in understanding the genetic and molecular basis of InD in dairy cattle. Recent studies focusing on the refinement of heritability partitioning through functional annotation have established frameworks for studying the genetic basis of complex traits in both human and livestock populations (Edwards et al., 2015; Finucane et al., 2015; Schneider et al., 2024). In human studies, for instance, stratified linkage disequilibrium (**LD**) score regression has been applied for partitioning heritability by using summary statistics derived from GWAS (Gazal et al., 2019). However, dairy cattle require alternative strategies and methodologies due to their distinct population history and genetic architectures, such as small effective population sizes, higher levels of relatedness, and long LD blocks (Jiang, 2024; Yuan et al., 2024). Certain genomic regions may have a disproportionate impact on InD. In dairy cattle populations, the extensive LD blocks can obscure the effects of individual variants, complicating efforts to partition genetic contributions to InD throughout the genome (Qanbari, 2020).

One of the hypothesized genetic bases of InD states that dominance effects, which represent the differences in phenotypic value between heterozygotes and homozygotes at specific loci, contribute to InD by revealing recessive deleterious alleles that are normally masked in heterozygous individuals as homozygosity increases through inbreeding (Charmantier et al., 2014). Across causal loci of traits under selection, directional dominance arises when heterozygotes systematically deviate from the midpoint of the 2 homozygotes, which means heterozygosity confers a net performance advantage and its loss under inbreeding produces inbreeding depression (Maltecca et al., 2020). An enrichment of InD within certain functional classes might indicate that those genomic regions harbor variants with stronger directional dominance effects. However, particularly with moderate sample sizes, identifying dominance effects can be difficult and challenging using current existing methods due to insufficient statistical power (Boysen et al., 2013; Sun et al., 2014). Statistical power plays a vital role in identifying and detecting the subtle negative impact of deleterious alleles, particularly in regions with low variant density. Without adequate statistical power and appropriate methods, there is a risk of false positives or the inability to differentiate between causal and noncausal variants in regions associated with InD (Li et al., 2024).

Substantial advancements has been achieved in understanding the genetic basis of InD in humans and model organisms (Charlesworth and Willis, 2009). However, replicating these advancements in livestock and natural populations has been challenging due to their distinct population histories and evolutionary implications. To address these limitations, we propose leveraging established methodologies from human genetics. For instance, Yengo et al. (2021) effectively partitioned the average effect of InD across 11 traits into 8 genomic regions, revealing an enrichment in regions characterized by high recombination rates. On the other hand, one study on ROH in cattle populations has similarly highlighted region-specific homozygosity, which could allow InD to be broken down into genomic regions (Howard et al., 2015). However, genome-wide analyses have demonstrated that InD effects are uniformly distributed in certain cattle breeds, such as the Dutch Holstein Friesian, indicating variation even within closely related populations (Doekes et al., 2020). These findings indicate that further investigation into region-specific InD in dairy cattle could provide valuable insights for managing and mitigating InD while simultaneously promoting genetic improvement.

Since the implementation of official genomic evaluations for US cattle in 2009 (Wiggans et al., 2011), inbreeding rates have increased substantially, especially within Jersey and Holstein breeds. This trend persists despite earlier forecasts suggesting that enhanced selection techniques would limit inbreeding levels (Daetwyler et al., 2007). This acceleration is largely attributed to the increased use of young bulls and reduction in generation intervals (Guinan et al., 2023). The extensive genomic and phenotypic datasets available from sources such as the Council on Dairy Cattle Breeding (**CDCB**; Bowie, MD) offer a valuable opportunity to investigate the impact of inbreeding on a broad range of production, reproductive, and fitness traits in dairy cattle. For example, in US Jersey cattle, the observed increase in average inbreeding coefficient is associated with reduced productivity and survival, along with in-

Table 1. Number, mean, SD, and median of yield deviation records for milk, fat, and protein

Trait	Yield deviation records	Mean	SD	Median
Milk	242,103	−218.14	2,257.49	−201.054
Fat	239,063	136.17	82.25	135.86
Protein	239,501	70.26	60.49	70.64

creased replacement costs when inbreeding surpasses critical thresholds (Lozada-Soto et al., 2022).

The primary objective of this research is to evaluate whether genomic functional annotations contribute disproportionately to InD in key production traits within a US Jersey population. By partitioning the effects of InD through functional annotations, we aim to assess the effectiveness of our current methodology for analyzing the dairy cattle genome, and to identify improved models that facilitate more accurate partitioning. Additionally, our research focuses on exploring how different genomic parameters, including variant density, LD levels, and minor allele frequency (**MAF**), affect the partitioning of InD.

MATERIALS AND METHODS

Phenotypic Data

Production traits such as milk, fat, and protein yields were the focus of the investigation. Large-scale phenotypic records on these economically important traits revealed a substantial negative correlation between inbreeding levels and production traits, establishing Jersey as a suitable model for dissecting and partitioning InD effects (Pryce et al., 2014). Furthermore, the limited contribution of dominance effects to phenotypic variance of reproductive traits in Jersey cattle restricts their applicability in the analysis of InD.

Adjusted lactation yield data for these 3 production traits were obtained from CDCB for 248,488 Jersey cows. These yield deviation (**YD**) records, reported as weighted average yields in pounds, were adjusted by CDCB to account for management group, permanent environmental effects, and herd-sire interactions. This adjustment process enhances unbiased assessment by mitigating the influence of nongenetic factors or systematic effects (Wiggans and VanRaden, 1989).

To identify and remove potential outliers, the absolute z-scores of the YD records were calculated. Observations with absolute z-scores exceeding 2.5 were excluded, resulting in a dataset of 241,918 Jersey cows for subsequent analyses. The summary statistics for the refined dataset are presented in Table 1. To ensure uniformity across traits, the data were standardized by centering the values through mean subtraction and scaling them by dividing by the SD.

Genomic Data

A variety of low- to medium-density SNP panels was applied in the genotyping process. The genotypic dataset included 78,964 imputed SNPs that were incorporated into genomic evaluations for US dairy cattle (Déru et al., 2024). In addition, sequence imputation was conducted using reference panels of run 8 and run 9 from the 1000 Bull Genomes Project (Hayes and Daetwyler, 2019). A total number of 15,758,692 imputed whole-genome variants was obtained from the imputation. Both SNPs and short insertions and deletions were included. The quality control measures implemented for the imputed data using PLINK 2 (Chang et al., 2015) involved the exclusion of variants based on 2 criteria: (1) MAF of less than 1% and (2) a Hardy–Weinberg equilibrium (**HWE**) *P*-value lower than 10^{-6}. The quality of imputation was assessed using an information metric (**INFO**) score, which ranges from 0 to 1, with values approaching 1 indicating a high level of confidence in the imputed variants (Zheng et al., 2015). With IMPUTE (version 2; Howie et al., 2009), the INFO score is calculated by comparing the variance of imputed genotype dosages to the theoretical maximum variance under Hardy–Weinberg equilibrium (Stahl et al., 2025). It estimates the proportion of statistical information retained after imputation and is conceptually related to Fisher information used in the score test framework (Balakrishnan et al., 2007). In this study, only variants with an INFO score exceeding 0.3 were retained, resulting in a total of 9,532,696 variants used for the calculation of inbreeding coefficients.

Inbreeding Coefficients

Our basic assumption is that each variant has an equal and small contribution to InD, with an expectation of $E\left(\beta_j\right)=\frac{b}{M}$, where β_j is the contribution of SNP j to InD, M represents the total number of variants, and b denotes the genome-wide InD (Yengo et al., 2021). Under this assumption, the genomic measure of inbreeding (F_{UNI}), which is defined as the correlation between the parents' uniting gametes (Wright, 1922), is a more suitable estimator for interpreting InD. The calculation of F_{UNI} was performed using PLINK 1.9 (Chang et al., 2015) with the command *--ibc*, following the formula outlined by Yang et al. (2011):

$$F_{UNI}=\frac{1}{M}\sum_{i=1}^{M}\frac{x_i^2-\left(1+2p_i\right)x_i+2p_i^2}{2p_i\left(1-p_i\right)},$$

where x_i indicates the number of alternative alleles (0, 1, or 2) at the ith variant and p_i is the frequency of the alternative allele.

Functional Annotations

The study incorporated 5 annotation categories into its analysis. Annotations for intron, promoter, genomic evolutionary rate profiling (**GERP**) constrained elements, coding sequence (**CDS**), and untranslated regions (**UTR**) were extracted from the ARS-UCD1.2 cattle reference genome (Ensembl release 109; Rosen et al., 2020). Sequence variants were aligned with these 5 functional annotations as well as the remaining genomic regions. The intron category includes all noncoding sequences that are transcribed into precursor mRNA, with ~32.4% of the variants (totaling 3,088,751) classified within this category. The promoter category refers to DNA sequences that initiate gene transcription located within 2 kb upstream of the transcription start site, with ~1.3% of the variants (totaling 121,762) mapped to promoters. Constrained elements represent highly conserved regions identified within a multiple sequence alignment through GERP, which employs a permutation-based scoring method. These regions show higher levels of sequence conservation than expected by random chance, suggesting their potential functional importance within the genome (Huber et al., 2020). Approximately 1.2% of the total variants (totaling 115,275) were found within GERP constrained elements. The CDS category includes DNA sequences that correspond to the sequence of amino acids in proteins, with ~0.7% of the variants (totaling 70,838) mapped to CDS regions. The UTR category included regions of mRNA, specifically the 5′ and 3′ UTR, which are generally not translated into proteins; ~0.5% of the variants (totaling 44,268) were identified in UTR. The annotation-level average inbreeding coefficient, denoted as $\bar{F}_k$, was calculated based on the variants in each category k, resulting in a total of 6 inbreeding coefficients for each individual, which includes 5 annotation categories and one for the whole genome.

Models for Partitioning Inbreeding Depression

Our initial focus was on a linear model, assuming all variants contribute to InD with small and minor effects. This model can be represented as

$$y = b_0 + \sum_{j=1}^{M} \beta_j F_j + e,$$

referred to as the infinitesimal SNP-based InD model (**ISIM**), where y represents the YD of a quantitative trait affected by InD; b_0 is the model intercept, indicating the mean YD of the quantitative trait; M denotes the total number of variants included; β_j represents the effect size of variant j on InD; F_j denotes the per-SNP inbreeding coefficient for variant j, quantifying the deviation of an individual's genotype from Hardy–Weinberg expectations at that locus, and was computed as F_{UNI}; and e is the residual term that captures all the other effects. In ISIM, each SNP effect $\hat{\beta}_j$ can be written as $\hat{\beta}_j = b + \hat{b}_{k(j)}$, where $\hat{b}_{k(j)}$ is the SNP effect on InD specific to functional annotation k to which SNP j belongs. The ISIM can be further expressed as

$$y = b_0 + \sum_{j=1}^{M} \left(b + \hat{b}_{k(j)}\right) F_j + e,$$

$$y = b_0 + b\sum_{j=1}^{M} F_j + \sum_{k=1}^{K} \sum_{j \in k} \hat{b}_k F_j + e.$$

Moreover, the whole-genome inbreeding coefficient is the average of F_j:

$$\hat{F}_g = \frac{1}{M} \sum_{j=1}^{M} F_j,$$

which is based on the correlation between uniting gametes, where g is whole genome. Similarly, the average inbreeding coefficient for functional annotation k is

$$\hat{\bar{F}}_k = \frac{1}{M_k} \sum_{j \in k} F_j,$$

where M_k is the number of SNPs in functional annotation k. Hence, $\sum_{j=1}^{M} F_j = M\hat{F}_g$ and $\sum_{j \in k} F_j = M_k \hat{\bar{F}}_k$.

By substituting $\sum_{j=1}^{M} F_j$ and $\sum_{j \in k} F_j$ into ISIM as follows:

$$y = b_0 + b\sum_{j=1}^{M} F_j + \sum_{k=1}^{K} \sum_{j \in k} \hat{b}_k F_j + e,$$

$$y = b_0 + bM\hat{F}_g + \sum_{k=1}^{K} \hat{b}_k M_k \hat{\bar{F}}_k + e,$$

and writing $b' = bM$ and $\hat{b}_k = \hat{b}_k M_k$, we can get

$$y = b_0 + b'\hat{F}_g + \sum_{k=1}^{K} \hat{b}_k \hat{\bar{F}}_k + e.$$

Hence, the effect of InD was dissected by assuming that variants within certain functional annotation categories make disproportional contributions:

$$y = b_0 + bF_g + \sum_{k=1}^{K} b_k \bar{F}_k + e,$$

referred to as the functional annotation partition model (**FAPM**). In the FAPM, b is the overall contribution of the whole genome-wide variants to InD; F_g is the average inbreeding coefficient across all M variants; K is the total number of functional annotations; b_k is the contribution of variants within annotation k to InD; $\bar{F}_k$, the annotation-level average inbreeding coefficient, represents the average inbreeding coefficient across all variants in annotation k; and e is the residual term. This model, akin to the one used by Yengo et al. (2021) to partition InD in human population, is included here for comparison purposes. Contrary to the to human population, dairy cattle populations have a substantially smaller effective population size due to the selective breeding practices and the use of elite sires through artificial insemination. Consequently, individuals within the dairy population are more closely related, leading to a higher probability of sharing larger segments of their genome, particularly haplotype blocks. This shared ancestry can complicate the process of accurately partitioning the effects of InD. Hence, a linear mixed model was employed, as follows:

$$y = b_0 + bF_g + \sum_{k=1}^{K} b_k \bar{F}_k + g + e,$$

referred to as the structured InD partitioning mixed model (**SIPMM**), which is a genomic restricted maximum likelihood (GREML) framework designed to partition the effects of InD while considering the population structure of dairy cattle. In this context, g represents genomic additive genetic effect, which follows a normal distribution, $g \sim N\left(0, \mathbf{G}\sigma_a^2\right)$. The genomic relationship matrix (**GRM**), denoted as **G**, is computed using the formula $\mathbf{G} = \frac{\mathbf{ZZ'}}{M_g}$, where **Z** is the matrix of standardized genotypes (VanRaden, 2008) and M_g indicates the total number of SNPs used for calculating **G**.

All linear models were fitted using R 4.3.2 (R Core Team, 2023), except for the SIPMM model, which was employed through the SLEMM software (Cheng et al., 2023). Whole-genome InD was estimated using both FAPM and SIPMM, omitting the annotation partitioning term $\sum_{k=1}^{K} b_k \bar{F}_k$.

Inbreeding Depression Enrichment Ratios

The concept of InD enrichment refers to the disproportionate impact of homozygosity specific to certain annotation on the trait mean in contrast to the overall impact of InD across the whole genome. When SNPs are annotated into multiple functional annotations, the SNP effect estimates can be confounded and inflated. Hence, we need to adjust the raw estimates to account for the overlapping. We first constructed an $N \times K$ binary annotation matrix **A**, where N is the total number of SNPs, K is the number of functional annotations, and $A_{ij} = 1$ if the ith SNP is in the jth annotation. A cross-product matrix **X** was used to quantify overlaps among annotations: $\mathbf{X} = \mathbf{A}^T\mathbf{A}$, where the diagonal elements x_{ij} indicated the number of variants in annotation j, off-diagonal elements x_{jk} indicated the number of variants in both annotations j and k, and T is the matrix transpose. Based on the key assumption in ISIM, $\hat{\beta}_j = b + \hat{b}_{k(j)}$, $b_j = b_k$ if SNP j belongs to annotation k, and b_j is the sum of the corresponding b_k if SNP j belongs to multiple annotations. Hence, we used estimates $\hat{b}_k$ from FAPM or SIPMM to infer SNP-specific estimates $\hat{b}_j$. A per-SNP InD estimate of a functional annotation category was obtained by summing up $\hat{b}_j$ across all SNPs in the category and dividing the sum by $M_k\left(\bar{b}_k = \frac{S_k}{M_k}\right)$, where $S_k = \sum_{j \in SNPs\ in\ k} \hat{b}_j$ and M_k is the number of SNPs in that category. The whole-genome per-SNP InD estimate was calculated by summing up $\hat{b}_j$ across all SNPs in the genome and dividing the sum by $M\left(\bar{b}_0 = \frac{S_0}{M}\right)$, where $S_0 = \sum_{j=1}^{M} \hat{b}_j$ and M is the total number of SNPs. The enrichment ratio R_k for each annotation k is defined as the ratio of the per-SNP InD estimate of annotation k to the whole-genome per-SNP InD estimate $\left(R_k = \frac{\bar{b}_k}{\bar{b}_0}\right)$.

The SE of R_k was approximated using the delta method (Yang et al., 2015). First, we computed the gradient of the transformation matrix $\nabla\mathbf{G}(\mathbf{X})$ of S_k/S_0. Second, we approximated the SE of R_k by the following formula:

$$\mathrm{SE}[R_k] \approx \sqrt{\nabla\mathbf{G}(\mathbf{X})\, Cov(\mathbf{X})\, \nabla\mathbf{G}(\mathbf{X})^T},$$

where **X** is a vector of the adjusted model estimates; the transformation function **G** converts **X** into a set of ratios; $\nabla\mathbf{G}(\mathbf{X})$ is the gradient of the transformation function; and $Cov(\mathbf{X})$ is the variance-covariance matrix of **X**. Last, we divided SE[R_k] by M_k/M to match the normalization in the enrichment ratio.

The average enrichment of InD across traits and its SE were calculated as follows:

$$\overline{R_k} = \frac{1}{T}\sum_{1}^{T} R_k, \mathrm{SE}\left[\overline{R_k}\right] = \sqrt{\frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} r_{t,t'}\left[Var\left(R_{k,t}\right)Var\left(R_{k,t'}\right)\right]^{\frac{1}{2}}},$$

where T is the total number of traits and $Var(R_{k,t})$ is the sampling variance of R_k for trait t. Similarly, $Var(R_{k,t'})$ is the sampling variance of R_k for trait t' and $r_{t,t'}$ is the genetic correlation between traits t and t'. When no genetic correlation is considered ($t = t'$ and $r_{t,t'} = 1$), only variance terms remain, and the equation becomes $\mathrm{SE}\left[R_k\right] = \sqrt{\frac{1}{T^2}\sum_{t=1}^{T} Var\left(R_{k,t}\right)}$.

Impact of Genomic Parameters on Inbreeding Depression

A sensitivity analysis was performed to investigate the impact of various genomic parameters used in the InD partitioning analysis on InD enrichment. The imputed sequence variants were filtered based on the following criteria: (1) MAF classified into ranges of 0.01 to 0.1, 0.1 to 0.2, 0.2 to 0.3, 0.3 to 0.4, and 0.4 to 0.5; (2) LD-pruned variants with r^2 <0.6 or r^2 <0.3 (where r^2 is the squared correlation coefficient between allelic states at 2 loci); and (3) a random sampling of 250k or 700k variants that were evenly distributed along the genome. Linkage disequilibrium pruning was performed with the PLINK 2 command *--indep-pairwise*, which specified a window size of 50 variants and a step size of 5 variants. Random sampling was carried out using the PLINK 2 command *--thin-count*. The numbers of variants remained after each filtering procedure are shown in Supplemental Table S1 (see Notes). Inbreeding coefficients F_{UNI} were calculated using the filtered variants and subsequently incorporated into FAPM and SIPMM models, along with their corresponding null models.

The code used for data processing and statistical analysis is available at https://github.com/cxu33/Genomic-Partition-of-Inbreeding-Depression-Using-Functional-Annotations.git.

RESULTS

Inbreeding Coefficients and Inbreeding Depression Estimates

The summary statistics for the whole-genome F_g and the annotation-level inbreeding coefficients across each functional annotation class, derived from sequence variants, are presented in Supplemental Table S2 (see Notes). The whole genome and annotation-level inbreeding coefficients, calculated as F_{UNI}, exhibited a range from approximately −0.1 to 0.5, with a mean value near −0.003 and an SD of ~0.3.

When the FAPM was applied to the whole-genome inbreeding coefficients, significant positive effects of whole-genome InD on production traits were identified. The estimated regression coefficient $\hat{b}$ quantifies the change in trait value corresponding to a one-unit increase in the inbreeding coefficient. Specifically, the coefficient was ~0.67 (SE = 0.10, $P < 0.001$) for milk yield, 1.84 (SE = 0.10, $P < 0.001$) for fat yield, and 1.87 (SE = 0.10, $P < 0.001$) for protein yield. These unexpectedly positive coefficients likely stem from selective pressure acting on the production traits. As a result, the influence of inbreeding might give the impression of enhancing production traits within the model, potentially masking the usual negative biological consequences associated with heightened levels of inbreeding. Upon the adjustment for population structure, significant negative effects of whole-genome inbreeding on all production traits were identified (milk yield: $\hat{b} = -1.87$, SE = 0.087, $P < 0.001$; fat yield: $\hat{b} = -2.52$, SE = 0.094, $P < 0.001$; protein yield: $\hat{b} = -2.25$, SE = 0.091, $P < 0.001$). These negative coefficients suggest that once population structure was considered, inbreeding had a detrimental impact on production traits, aligning with the expected deleterious effects of inbreeding on productivity. The FAPM approach is analogous to the model proposed by Yengo et al. (2021). Their model exhibited robust performance in human populations, likely due to the relatively lower levels of inbreeding compared to dairy cattle populations.

Enrichment of Inbreeding Depression Within Functional Annotations

Inbreeding depression refers to the reduction in fitness or performance observed in offspring resulting from the mating of closely related individuals. Enrichment of InD within functional annotations indicates that inbreeding within these regions, associated with particular biological roles, contributes more to the adverse impacts than would typically be anticipated. Conversely, a depletion of InD enrichment suggests that in specific areas, the negative consequences are less pronounced than expected. Our study examined the impact of InD on milk, fat, and protein yields using FAPM and SIPMM models, with a particular emphasis on whole-genome InD (F_g) and specific functional annotations. The findings derived from the FAPM model are shown in the supplemental material (see Notes).

The application of SIPMM revealed notable adverse impacts of inbreeding on milk yield across various genomic regions, specifically in promoter ($\hat{b}_k = -0.40$, SE = 0.15, $P = 0.009$), UTR ($\hat{b}_k = -0.45$, SE = 0.14, $P = 0.001$),

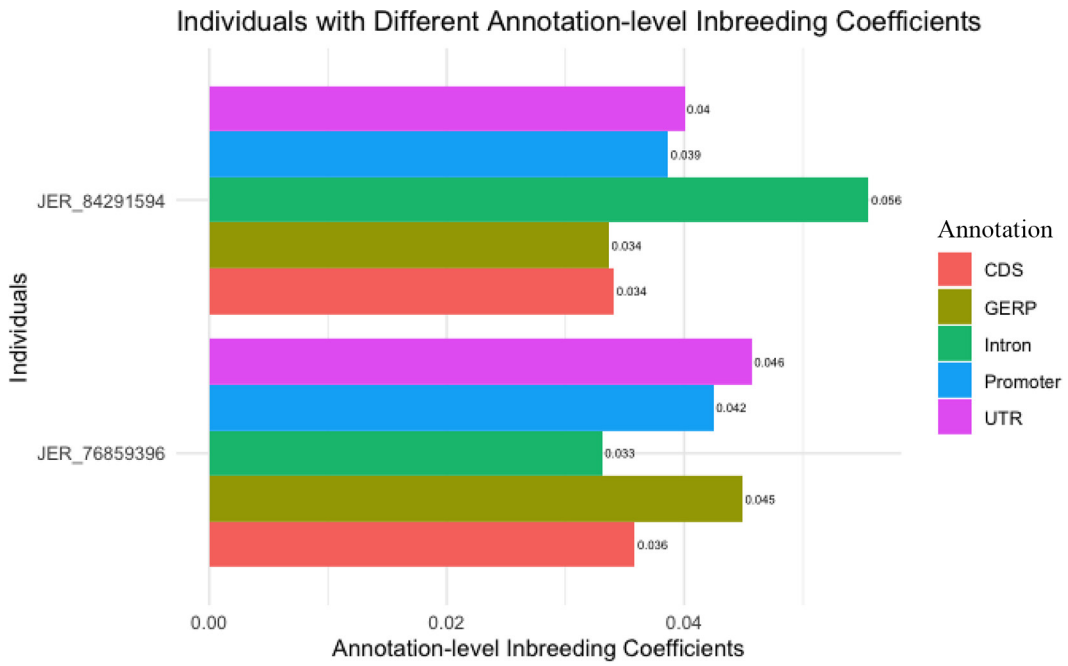

Figure 1. Comparisons of annotation-level inbreeding coefficients among different functional annotations between 2 individuals from the population.

and GERP ($\hat{b}_k$ = −0.73, SE = 0.14, P < 0.001) regions. Significant enrichment of InD was observed in promoter (R_k = 20.11, SE = 6.44), UTR (R_k = 57.96, SE = 16.62), and GERP (R_k = 35.91, SE = 7.00). For fat yield, significant negative effects of inbreeding were identified for the whole-genome ($\hat{b}$ = −0.63, SE = 0.13, P < 0.001), as well as in UTR ($\hat{b}_k$ = −0.43, SE = 0.15, P = 0.004) and GERP ($\hat{b}_k$ = −0.78, SE = 0.15, P < 0.001) regions. Enrichment of InD was significant in UTR (R_k = 40.20, SE = 12.77) and GERP (R_k = 28.72, SE = 5.34). For protein yield, significant negative effects were noted in the promoter ($\hat{b}_k$ = −0.36, SE = 0.16, P = 0.02), UTR ($\hat{b}_k$ = −0.44, SE = 0.14, P = 0.003), and GERP ($\hat{b}_k$ = −0.80, SE = 0.15, P<0.001) regions, with InD enrichment being significant in promoter (R_k = 15.25, SE = 5.45), UTR (R_k = 46.44, SE = 14.07), and GERP (R_k = 32.73, SE = 5.92). The average InD enrichment level in promoter was 17.68, with an SE of 2.71, computed across milk and protein yields with a consideration of their genetic correlation. The average InD enrichment levels and SE in UTR (R_k = 48.20, SE = 6.55), and GERP (R_k = 32.45, SE = 2.75) were computed across milk, fat, and protein yields accounting their genetic correlations.

In dairy cattle breeding, the selection process for reproductive animals plays a pivotal role in controlling and managing inbreeding levels. Our study demonstrates that the incorporation of functional annotation information could improve the differentiation of individuals based on their annotation-level inbreeding coefficients and the enrichment of InD within specific functional annotations. Figure 1 illustrates the annotation-level inbreeding coefficients across 5 functional annotations for 2 Jersey cows (JER_76858386 and JER_84291594). For instance, to mitigate the impact of InD on production traits, a selection decision may favor JER_84291594, which shows lower excessive homozygosity in the UTR and GERP regions.

Impact of Genomic Parameters on Inbreeding Depression

Impact of Variant Density on Inbreeding Depression. Table 2 presents the estimated F_g and annotation-specific InD, along with their significance levels derived from SIPMM for all the production traits using different numbers of variants. Figure 2 illustrates the annotation-specific InD enrichment and SE for each trait, as well as the average InD enrichment across the production traits. Our analysis indicates that an increase in the density of SNPs facilitates the detection of significant InD enrichment in GERP.

The initial genomic parameter investigated was the variant density used in InD partitioning analysis to determine whether varying variant density impact on InD enrichment. A random selection of ~250,000 and 700,000 uniformly distributed variants was used to establish a gradient in variant quantity. The whole genome

Table 2. Significant ($P < 0.05$) annotation-level InD estimates $\left(\hat{b}_k\right)$ and P-values from the structured InD partitioning mixed model using imputed 79k chip, randomly sampled evenly distributed 250k, randomly sampled evenly distributed 700k, and sequence variants

	Imputed 79k chip variant		Randomly sampled 250k variant		Randomly sampled 700k variant		Imputed sequence variant	
Annotation	$\hat{b}_k$	P-value	$\hat{b}_k$	P-value	$\hat{b}_k$	P-value	$\hat{b}_k$	P-value
Milk								
F_g[1]	−0.841	0.000						
Intron	−0.665	0.000	−1.132	0.000	−0.832	0.001		
Promoter	−0.116	0.022	−0.243	0.018	−0.293	0.023	−0.400	0.009
GERP			−0.422	0.000	−0.722	0.000	−0.730	0.000
CDS	−0.118	0.015	−0.261	0.001	−0.246	0.033		
UTR	−0.108	0.004	−0.285	0.000	−0.434	0.000	−0.448	0.001
Fat								
F_g	−0.916	0.000					−0.633	0.000
Intron	−0.804	0.000	−1.209	0.000	−0.758	0.007		
Promoter	−0.145	0.008	−0.260	0.019	−0.428	0.002		
GERP			−0.402	0.001	−0.844	0.000	−0.778	0.000
CDS	−0.157	0.003	−0.373	0.000				
UTR	−0.082	0.039	−0.228	0.003	−0.370	0.001	−0.429	0.004
Protein								
F_g	−1.020	0.000						
Intron	−0.648	0.001	−1.082	0.000	−0.830	0.002		
Promoter	−0.156	0.003	−0.304	0.004	−0.300	0.026	−0.362	0.023
GERP			−0.402	0.000	−0.747	0.000	−0.798	0.000
CDS	−0.163	0.001	−0.345	0.000				
UTR	−0.082	0.033	−0.268	0.000	−0.432	0.000	−0.436	0.003

[1]F_g = whole-genome inbreeding coefficient (average inbreeding coefficient across M variants).

InD estimates derived from SIPMM, based on imputed 79k SNP chip (milk yield: $\hat{b} = -1.80$, SE = 0.053, $P < 0.001$; fat yield: $\hat{b} = -2.03$, SE = 0.057, $P < 0.001$; protein yield: $\hat{b} = -1.99$, SE = 0.055, $P < 0.001$), were consistent with those obtained from imputed whole-genome sequencing data. Upon increasing the number of variants to 250k, a slight increase in InD estimates was observed for all the production traits from SIPMM compared to the estimates derived from imputed 79k SNP chip (milk yield: $\hat{b} = -1.92$, SE = 0.073, $P < 0.001$; fat yield: $\hat{b} = -2.43$, SE = 0.078, $P < 0.001$; protein yield: $\hat{b} = -2.22$, SE = 0.076, $P < 0.001$). When the analysis was conducted using randomly selected evenly distributed 700k variants, a slightly greater increase in InD estimates was found for all the production traits from SIPMM in comparison to the 250k estimates (milk yield: $\hat{b} = -2.21$, SE = 0.082, $P < 0.001$; fat yield: $\hat{b} = -2.65$, SE = 0.088, $P < 0.001$; protein yield: $\hat{b} = -2.47$, SE = 0.085, $P < 0.001$). In summary, the linear mixed model shows minimal sensitivity to the variations in SNP densities when it comes to estimating InD.

Impact of LD on Inbreeding Depression. The estimated F_g values along with the annotation-specific InD and their significance levels derived from SIPMM for all 3 production traits characterized by varying levels of LD are presented in Table 3. Figure 3 illustrates the annotation-specific InD enrichment and SE for each trait as well as the average InD enrichment across 3 production traits. The InD enrichment estimated from randomly selected 60k variants was found to be comparable to that of the imputed 79k SNP chip with the exception of a notable absence of InD enrichment in UTR. This discrepancy may be attributed to the limited representation of variants in the UTR region during the random sampling. Additionally, the InD enrichment estimates for CDS were found to be higher when using the randomly selected 60k variants compared to the imputed 79k SNP chip. It was observed that varying the r^2 parameters did not considerably influence the estimated InD enrichment levels. However, LD pruning indicated a low-level significant InD enrichment in the intronic region when compared to the unpruned sequencing data.

In order to assess the impact of LD on the detection and quantification of within-annotation InD enrichment, a partitioning analysis was performed with LD-pruned sequence variants. A gradient of LD levels was created by varying the LD parameter r^2. The LD pruning technique effectively removes pairs of highly correlated variants, leaving a set of independent variants for further testing. Although this approach is widely adopted, subsequent analyses revealed that some degree of residual correlation remained among markers, indicating that our selected pruning parameters may not have been sufficiently rigorous. Consequently, our dataset may still contain some moderate levels of LD. However, because many published genome-wide studies use similar or identical pruning parameters, our approach aligns with established practices (Meyermans et al., 2020).

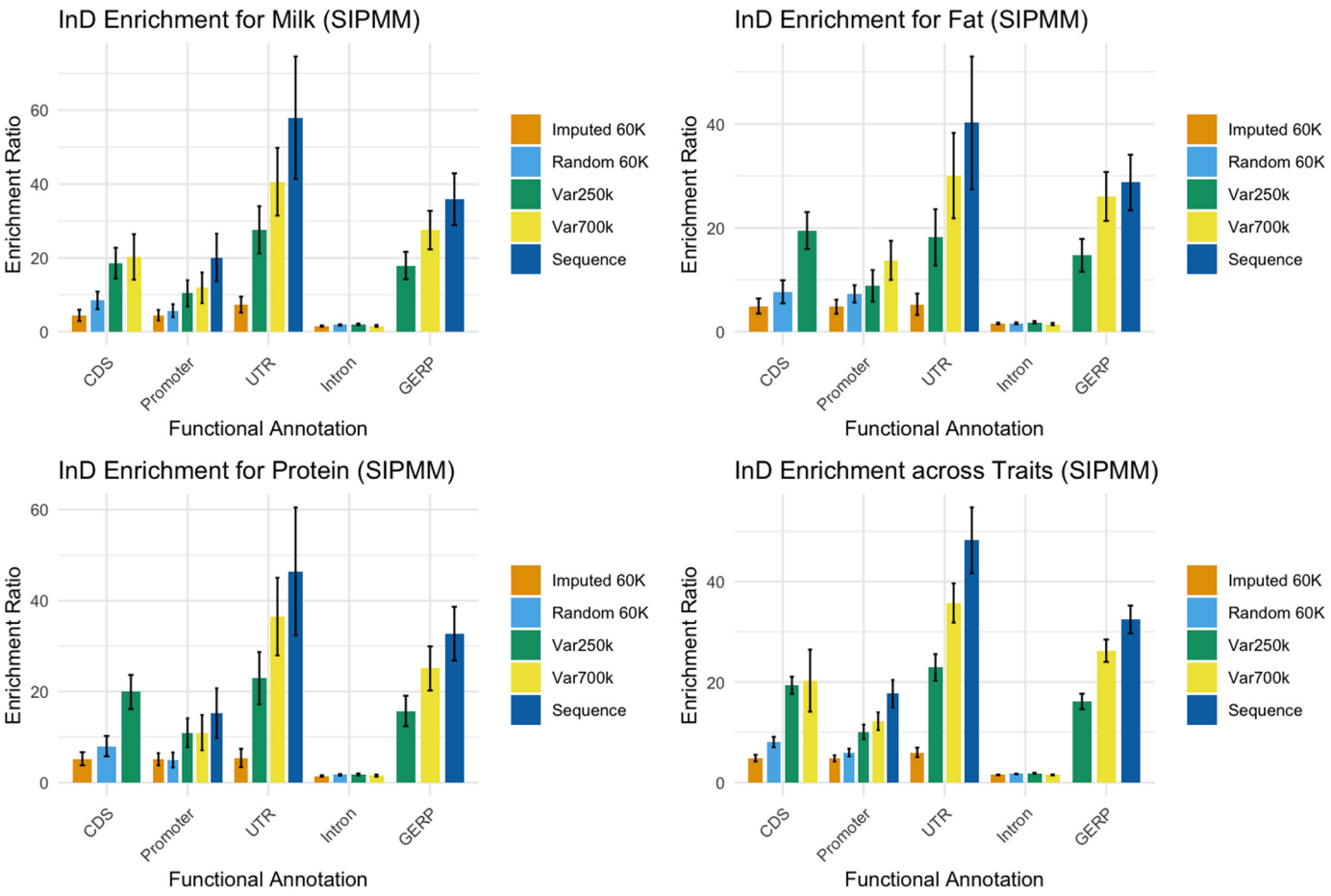

Figure 2. Significant ($P < 0.05$) InD enrichment ratios (enrichment SE) among different variant density groups (imputed 79k chip, 250k, 700k, sequence) for each trait and average InD enrichment ratios (enrichment SE) across traits. Var = variant density.

A random sample of ~60k variants was obtained from sequence variants that are uniformly distributed across the genome. This sampling served as a comparison to the imputed 79k SNP chip, which exhibits strong associations with quantitative traits.

Impact of MAF on Inbreeding Depression. The estimates of InD, along with their significance levels across all MAF groups for the 3 traits, are presented in Table 4. Figure 4 illustrates the comparison of InD enrichment across the various MAF groups. The MAF of variants used in InD detection may influence the enrichment of InD within particular functional regions. To explore these effects on InD enrichment, a grouping methodology was employed. Variants with MAF ranging from 0.4 to 0.5 were categorized as group g1, and those with MAF between 0.01 and 0.1 were classified as group g5. Intermediate MAF were represented in groups g2, g3, and g4, corresponding to ranges of 0.3 to 0.4, 0.2 to 0.3, and 0.1 to 0.2, respectively. The group with the rarest alleles exhibited less significant InD enrichment, likely due to the limited statistical power associated with a smaller sample size. In contrast, groups g3 and g4, which encompass intermediate MAF from 0.1 to 0.3, demonstrated significant InD enrichment in CDS, whereas both the common alleles (MAF from 0.3 to 0.5) and the rare alleles did not show such enrichment.

DISCUSSION

The challenge of partitioning InD in dairy cattle populations is compounded by the need to adapt estimation methods to specific genomic characteristics. The influences of selection, purging, and genetic drift have resulted in alterations of allele frequencies in dairy cattle that diverge from those in human populations, where most existing frameworks for partitioning InD derived (Pemberton et al., 2012; Yengo et al., 2021). In particular, the strong directional selection for production-related

Table 3. Significant ($P < 0.05$) annotation-level InD estimates $\left(\hat{b}_k\right)$ and P-values from the structured InD partitioning mixed model using imputed 79k chip, randomly sampled evenly distributed 60k, LD-pruned ($r^2 < 0.3$), LD-pruned ($r^2 < 0.6$), and sequence variants

	Imputed 79k chip variant		Randomly sampled 60k variant		$r^2 < 0.3$		$r^2 < 0.6$		Imputed sequence variant	
Annotation	$\hat{b}_k$	*P*-value	$\hat{b}_k$	*P*-value	$\hat{b}_k$	*P*-value	$\hat{b}_k$	*P*-value	$\hat{b}_k$	*P*-value
Milk										
F_g[1]	−0.841	0.000	−0.784	0.000						
Intron	−0.665	0.000	−0.722	0.000	−0.936	0.000	−0.888	0.000		
Promoter	−0.116	0.022	−0.103	0.011	−0.365	0.001	−0.469	0.000	−0.400	0.009
GERP					−0.293	0.013	−0.481	0.000	−0.730	0.000
CDS	−0.118	0.015	−0.099	0.002						
UTR	−0.108	0.004			−0.194	0.015	−0.233	0.018	−0.448	0.001
Fat										
F_g	−0.916	0.000	−1.097	0.000					−0.633	0.000
Intron	−0.804	0.000	−0.612	0.000	−1.029	0.000	−1.009	0.000		
Promoter	−0.145	0.008	−0.160	0.000	−0.442	0.000	−0.508	0.000		
GERP					−0.270	0.033	−0.373	0.009	−0.778	0.000
CDS	−0.157	0.003	−0.104	0.002						
UTR	−0.082	0.039							−0.429	0.004
Protein										
F_g	−1.020	0.000	−1.028	0.000						
Intron	−0.648	0.001	−0.686	0.000	−0.808	0.001	−0.829	0.001		
Promoter	−0.156	0.003	−0.096	0.022	−0.425	0.000	−0.453	0.001	−0.362	0.023
GERP					−0.254	0.038	−0.449	0.001	−0.798	0.000
CDS	−0.163	0.001	−0.105	0.001						
UTR	−0.082	0.033			−0.267	0.001	−0.278	0.007	−0.436	0.003

[1]F_g = whole-genome inbreeding coefficient (average inbreeding coefficient across M variants).

traits, such as milk yield, fat yield, and protein yield, has uniquely shaped allele frequencies in dairy cattle, a phenomenon not observed in humans (Qanbari, 2020).

By applying SIPMM to the whole genome and calculating annotation-specific inbreeding coefficients using sequence data, a notable enrichment of InD was observed within promoter regions, UTR, and GERP across 3 production traits. It has been established that deleterious variants are frequently enriched in functionally critical regions, including promoters and UTR (Gazal et al., 2018). Homozygous deleterious mutations in these regions can disrupt gene function, leading to decreased fitness. Finucane et al. (2015) found that heritability enrichment in conserved regions among mammals is considerably greater across numerous traits than in coding regions, highlighting the biological relevance of conserved regions, despite many of their functions remaining poorly characterized. Their results regarding heritability enrichment align with our findings of InD enrichment in GERP and the absence of InD enrichment in CDS.

The complexity of understanding variant causality in noncoding regions is notable. Innovative approaches and methodologies are being developed to assess the effects of variants on gene functionality, such as mutations located in the 3′-UTR of *TRIM14*, which interfere with miRNA binding sites and consequently affect gene regulation (Griesemer et al., 2021). Schneider et al. (2024) have demonstrated the critical role of UTR in shaping milk and health traits through their impacts on mRNA stability, localization, and translation efficiency. However, other research has indicated a lack of significant association between UTR regions and milk yield traits in dairy cattle (Koufariotis et al., 2014). Given their evolutionary conservation and regulatory importance in mammals, UTR regions may harbor deleterious mutations that reduce fitness when present in a homozygous state, thereby contributing to InD even if these mutations are not directly associated with phenotypic traits (Chatterjee and Pal, 2009).

Higher inbreeding is not necessarily associated with higher mutational load in genomic areas characterized by elevated GERP scores, typically influenced by strong purifying selection (Wootton et al., 2023). Mutations occurring in these highly conserved regions, marked by high GERP scores, are likely to result in more severe detrimental effects.

The findings from our sensitivity analysis indicated that whole-genome InD estimates derived from SIPMM exhibit only minor variations as variant density increases. Nevertheless, the observed trends in InD enrichment across functional annotation categories reveal a tendency for enrichment levels to rise with the utilization of denser variant datasets. It is crucial to acknowledge that the integration of genomic data could enhance these estimates, thereby improving the accuracy and biological significance of InD partitioning. Whole-genome sequencing is favored over chip data, as the currently used low- and medium-density variant chips primarily capture com-

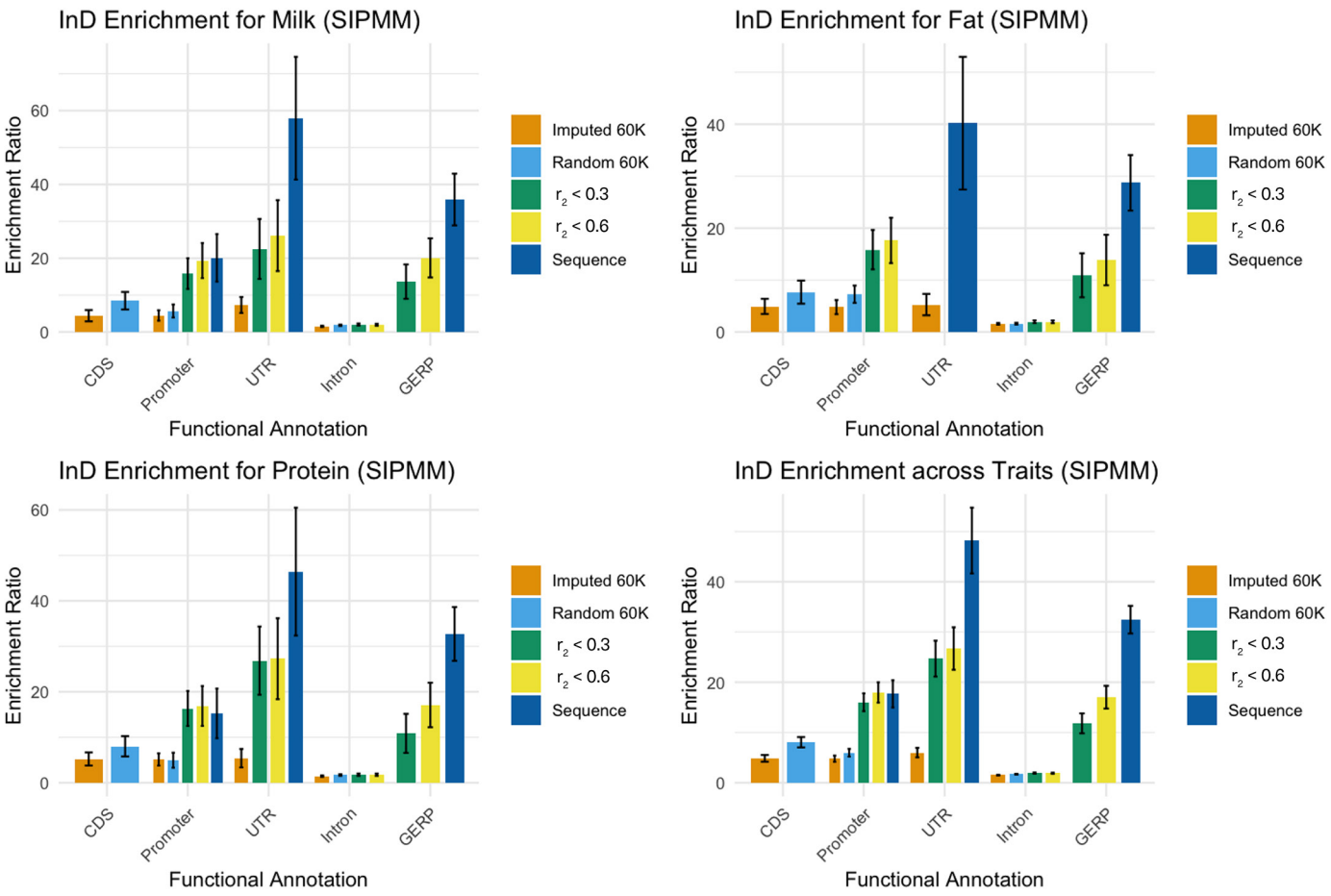

Figure 3. Significant ($P < 0.05$) InD enrichment (enrichment SE) among different LD levels ($r^2 < 0.3$, $r^2 < 0.6$) for each trait and average InD enrichment ratios (enrichment SE) across traits.

mon variants, which limits the coverage of functional genomic regions, particularly regulatory elements. This is critical, as InD may result from rare deleterious mutations that become homozygous. The lower level of InD enrichment observed may reflect the limited number of variants available for detection.

The presence of unpruned data can obscure the contributions of variants to InD due to the noise introduced by recombination (Yengo et al., 2017), which ultimately weakens the ability to identify InD enrichment. Conversely, although pruning decreases the number of markers, already constrained in specific functional annotation areas like UTR, it may further hinder the detection of significant InD enrichment in these regions. Future research could benefit from utilizing a larger population with an adequate number of sequenced individuals, such as those found in Holstein populations, to enhance InD estimates.

Interestingly, InD enrichment was detected in the CDS and intron categories when using imputed 79k SNP chip, whereas this was not the case with whole-genome sequencing data. This discrepancy could suggest that common alleles in CDS regions may not substantially contribute to InD. Natural selection often eliminates deleterious mutations in protein-coding regions to preserve the fitness of organisms, which results in a diminished effect of CDS alleles on InD. Conversely, functionally important regions outside of coding sequences may harbor rare or intermediate alleles with deleterious effects that contribute more significantly to InD. In the study of natural selection on deleterious alleles within coding and noncoding regions among 2 passerine bird species, Corcoran et al. (2017) found that purifying selection acts more strongly on coding regions due to the critical role these regions play in protein function. Consequently, deleterious mutations in coding sequences were more likely to be purged from the population, leading to a decrease in their overall frequency. Furthermore, the presence of common alleles and the intricate interactions among them can obscure signals of InD in whole-genome datasets, underscoring the necessity for meticulous filtering and analysis. Therefore, it is crucial to exercise caution when interpreting findings from different genomic data

Table 4. Significant ($P < 0.05$) annotation-level InD estimates $\left(\hat{b}_k\right)$ and P-values from the structured InD partitioning mixed model using variants grouped based on MAF (g1: MAF 0.4–0.5, g2: MAF 0.3–0.4, g3: MAF 0.2–0.3, g4: MAF 0.1–0.2, g5: MAF 0.01–0.1)

	g1		g2		g3		g4		g5	
Annotation	$\hat{b}_k$	P-value	$\hat{b}_k$	P-value	$\hat{b}_k$	P-value	$\hat{b}_k$	P-value	$\hat{b}_k$	P-value
Milk										
F_g[1]			−0.277	0.034	−0.302	0.027				
Intron	−0.405	0.000			−0.275	0.019				
Promoter			−0.346	0.000	−0.160	0.036				
GERP	−0.510	0.000	−0.529	0.000	−0.435	0.000	−0.512	0.000	−0.741	0.000
CDS					−0.158	0.019	−0.325	0.000		
UTR	−0.185	0.002			−0.159	0.008	−0.149	0.049		
Fat										
F_g			−0.447	0.002	−0.466	0.002	−0.486	0.002		
Intron	−0.370	0.001			−0.435	0.001				
Promoter	−0.190	0.009	−0.368	0.000						
GERP	−0.556	0.000	−0.545	0.000	−0.375	0.001	−0.506	0.000	−1.179	0.000
CDS					−0.212	0.004	−0.343	0.000		
UTR			−0.138	0.032			−0.206	0.012	−0.295	0.043
Protein										
F_g			−0.433	0.001	−0.326	0.022				
Intron	−0.380	0.000			−0.371	0.002				
Promoter			−0.388	0.000	−0.191	0.016				
GERP	−0.532	0.000	−0.517	0.000	−0.492	0.000	−0.596	0.000	−0.900	0.000
CDS					−0.153	0.029	−0.334	0.000		
UTR	−0.147	0.020	−0.133	0.032			−0.208	0.008		

[1]F_g = whole-genome inbreeding coefficient (average inbreeding coefficient across M variants).

sources, as differences in marker types and allele frequencies may lead to distinct insights into the genetic basis of InD.

CONCLUSIONS

Our study detected and quantified significant enrichment of InD within functional genomic regions of US Jersey cattle. Linear mixed models, which incorporate a genomic relationship matrix, proved more effective for partitioning and quantifying enrichment by accounting for relatedness among individuals. Notably, promoter, UTR, and GERP-conserved regions showed strong InD enrichment, underscoring their importance in the molecular basis of inbreeding. The extent of enrichment was influenced by marker density, linkage disequilibrium, and minor allele frequency; higher marker density and LD pruning improved precision, emphasizing the need for careful marker selection. These findings provide a foundation for tailoring inbreeding management strategies to mitigate negative effects in key regulatory regions while supporting genetic progress. Future work should validate these results across additional traits, integrate functional data such as transcriptomics or epigenomics, and explore comparative analyses across breeds and species to refine region-specific strategies and enhance genomic selection in dairy cattle.

NOTES

This work is supported by the Agriculture and Food Research Initiative (AFRI) Foundational and Applied Science Program, project award no. 2023-67015-39260, and the Research Capacity Fund (HATCH), project award no. 7008128, from the U.S. Department of Agriculture's National Institute of Food and Agriculture. Sajjad Toghiani and Paul VanRaden were supported by appropriated project 8042-31000-002-00-D, "Improving Dairy Animals by Increasing Accuracy of Genomic Prediction, Evaluating New Traits, and Redefining Selection Goals" of the USDA Agricultural Research Service (Beltsville, MD). We extend our sincere gratitude to the Council on Dairy Cattle Breeding (Bowie, MD) for granting access to their dairy cattle genotype database and for their invaluable scientific and technical support. We also express our appreciation to Li Ma's group at the University of Maryland (College Park, MD) for their essential scientific insights. We acknowledge the 1000 Bull Genomes Project (Agriculture Victoria; https://www.ebi.ac.uk/ena/browser/view/PRJEB42783; https://www.ebi.ac.uk/ena/browser/view/PRJEB56689) for providing the reference panels. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of

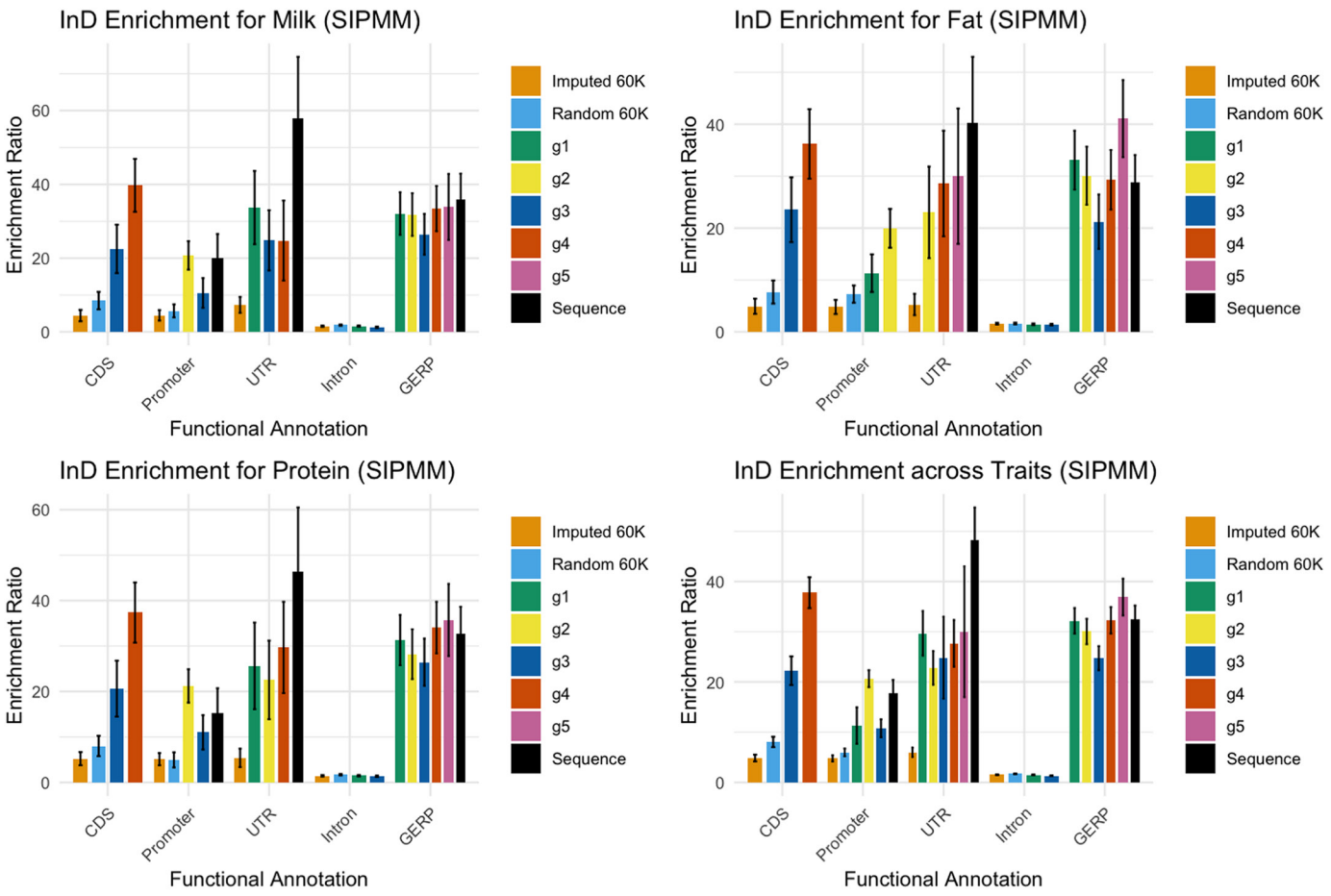

Figure 4. Significant ($P < 0.05$) InD enrichment (enrichment SE) among different MAF groups (g1: MAF 0.4–0.5, g2: MAF 0.3–0.4, g3: MAF 0.2–0.3, g4: MAF 0.1–0.2, g5: MAF 0.01–0.1) for each trait and average InD enrichment ratios (enrichment SE) across traits.

Agriculture. The USDA is an equal opportunity provider and employer. Supplemental material for this article is available at https://doi.org/10.5281/zenodo.16895796 . No human or animal subjects were used, so this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board. The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: CDCB = Council on Dairy Cattle Breeding; CDS = coding sequence; FAPM = functional annotation partition model; F_{Ped} = pedigree inbreeding coefficient; GERP = genomic evolutionary rate profiling; GRM = genomic relationship matrix; HWE = Hardy–Weinberg equilibrium; InD = inbreeding depression; INFO = information metric; ISIM = infinitesimal SNP-based InD model; LD = linkage disequilibrium; MAF = minor allele frequency; R_k = enrichment ratio; ROH = runs of homozygosity; SIPMM = structured InD partitioning mixed model; UTR = untranslated region; Var = variant; YD = yield deviation.

REFERENCES

Balakrishnan, N., N. Kannan, and H. N. Nagaraja. 2007. Advances in Ranking and Selection, Multiple Comparisons, and Reliability: Methodology and Applications. Springer Science & Business Media.

Boysen, T.-J., C. Heuer, J. Tetens, F. Reinhardt, and G. Thaller. 2013. Novel use of derived genotype probabilities to discover significant dominance effects for milk production traits in dairy cattle. Genetics 193:431–442. https://doi.org/10.1534/genetics.112.144535.

Caballero, A., B. Villanueva, and T. Druet. 2021. On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. Evol. Appl. 14:416–428. https://doi.org/10.1111/eva.13126.

Cassell, B. G., V. Adamec, and R. E. Pearson. 2003. Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. J. Dairy Sci. 86:2967–2976. https://doi.org/10.3168/jds.S0022-0302(03)73894-6.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 4:s13742-015-0047-8. https://doi.org/10.1186/s13742-015-0047-8.

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. 18:237–268. https://doi.org/10.1146/annurev.es.18.110187.001321.

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

Charmantier, A., D. Garant, and L. E. B. Kruuk. 2014. Quantitative Genetics in the Wild. Oxford University Press.

Chatterjee, S., and J. K. Pal. 2009. Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. Biol. Cell 101:251–262. https://doi.org/10.1042/BC20080104.

Cheng, J., C. Maltecca, P. M. VanRaden, J. R. O'Connell, L. Ma, and J. Jiang. 2023. SLEMM: million-scale genomic predictions with window-based SNP weighting. Bioinformatics 39:btad127. https://doi.org/10.1093/bioinformatics/btad127.

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.

Corcoran, P., T. I. Gossmann, H. J. BartonThe Great Tit HapMap Consortium, J. Slate, and K. Zeng. 2017. Determinants of the efficacy of natural selection on coding and noncoding variability in two passerine species. Genome Biol. Evol. 9:2987–3007. https://doi.org/10.1093/gbe/evx213.

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. J. Anim. Breed. Genet. 124:369–376. https://doi.org/10.1111/j.1439-0388.2007.00693.x.

Déru, V., F. Tiezzi, P. M. VanRaden, E. A. Lozada-Soto, S. Toghiani, and C. Maltecca. 2024. Imputation accuracy from low- to medium-density SNP chips for US crossbred dairy cattle. J. Dairy Sci. 107:398–411. https://doi.org/10.3168/jds.2023-23250.

Doekes, H. P., P. Bijma, R. F. Veerkamp, G. de Jong, Y. C. J. Wientjes, and J. J. Windig. 2020. Inbreeding depression across the genome of Dutch Holstein Friesian dairy cattle. Genet. Sel. Evol. 52:64. https://doi.org/10.1186/s12711-020-00583-1.

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.

Edwards, S. M., B. Thomsen, P. Madsen, and P. Sørensen. 2015. Partitioning of genomic variance reveals biological pathways associated with udder health and milk production traits in dairy cattle. Genet. Sel. Evol. 47:60. https://doi.org/10.1186/s12711-015-0132-6.

Ferenčaković, M., J. Sölkner, M. Kapš, and I. Curik. 2017. Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population. J. Dairy Sci. 100:4721–4730. https://doi.org/10.3168/jds.2016-12164.

Finucane, H. K., B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, and F. R. Day. ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, The RACI Consortium, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, and A. L. Price. 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47:1228–1235. https://doi.org/10.1038/ng.3404.

Fuerst-Waltl, B., and C. Fuerst. 2012. Effect of inbreeding depression on survival of Austrian Brown Swiss calves and heifers. J. Dairy Sci. 95:6086–6092. https://doi.org/10.3168/jds.2011-4684.

Gazal, S., P.-R. Loh, H. K. Finucane, A. Ganna, A. Schoech, S. Sunyaev, and A. L. Price. 2018. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. Nat. Genet. 50:1600–1607. https://doi.org/10.1038/s41588-018-0231-8.

Gazal, S., C. Marquez-Luna, H. K. Finucane, and A. L. Price. 2019. Reconciling S-LDSC and LDAK functional enrichment estimates. Nat. Genet. 51:1202–1204. https://doi.org/10.1038/s41588-019-0464-1.

Griesemer, D., J. R. Xue, S. K. Reilly, J. C. Ulirsch, K. Kukreja, J. R. Davis, M. Kanai, D. K. Yang, J. C. Butts, M. H. Guney, J. Luban, S. B. Montgomery, H. K. Finucane, C. D. Novina, R. Tewhey, and P. C. Sabeti. 2021. Genome-wide functional screen of 3′UTR variants uncovers causal variants for human disease and evolution. Cell 184:5247–5260.e19. https://doi.org/10.1016/j.cell.2021.08.025.

Guinan, F. L., G. R. Wiggans, H. D. Norman, J. W. Dürr, J. B. Cole, C. P. 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.

Hayes, B. J., and H. D. Daetwyler. 2019. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. Annu. Rev. Anim. Biosci. 7:89–102. https://doi.org/10.1146/annurev-animal-020518-115024.

Howard, J. T., M. Haile-Mariam, J. E. Pryce, and C. Maltecca. 2015. Investigation of regions impacting inbreeding depression and their association with the additive genetic effect for United States and Australia Jersey dairy cattle. BMC Genomics 16:813. https://doi.org/10.1186/s12864-015-2001-7.

Howie, B. N., P. Donnelly, and J. Marchini. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5:e1000529.

Huber, C. D., B. Y. Kim, and K. E. Lohmueller. 2020. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. PLoS Genet. 16:e1008827. https://doi.org/10.1371/journal.pgen.1008827.

Jiang, J. 2024. MPH: Fast REML for large-scale genome partitioning of quantitative genetic variation. Bioinformatics 40:btae298. https://doi.org/10.1093/bioinformatics/btae298.

Koufariotis, L., Y.-P. P. Chen, S. Bolormaa, and B. J. Hayes. 2014. Regulatory and coding genome regions are enriched for trait associated variants in dairy and beef cattle. BMC Genomics 15:436. https://doi.org/10.1186/1471-2164-15-436.

Li, G., J. Wu, and X. Wang. 2024. Predicting functional UTR variants by integrating region-specific features. Brief. Bioinform. 25:bbae248. https://doi.org/10.1093/bib/bbae248.

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.

Maltecca, C., F. Tiezzi, J. B. Cole, and C. Baes. 2020. Symposium review: Exploiting homozygosity in the era of genomics—Selection, inbreeding, and mating programs. J. Dairy Sci. 103:5302–5313. https://doi.org/10.3168/jds.2019-17846.

Meyermans, R., W. Gorssen, N. Buys, and S. Janssens. 2020. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. BMC Genomics 21:94. https://doi.org/10.1186/s12864-020-6463-x.

Mugambe, J., R. H. Ahmed, G. Thaller, and C. Schmidtmann. 2024. Impact of inbreeding on production, fertility, and health traits in German Holstein dairy cattle utilizing various inbreeding estimators. J. Dairy Sci. 107:4714–4725. https://doi.org/10.3168/jds.2023-23728.

Pemberton, T. J., D. Absher, M. W. Feldman, R. M. Myers, N. A. Rosenberg, and J. Z. Li. 2012. Genomic Patterns of Homozygosity in Worldwide Human Populations. Am. J. Hum. Genet. 91:275–292. https://doi.org/10.1016/j.ajhg.2012.06.014.

Pryce, J. E., M. Haile-Mariam, M. E. Goddard, and B. J. Hayes. 2014. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genet. Sel. Evol. 46:71. https://doi.org/10.1186/s12711-014-0071-7.

Purfield, D. C., D. P. Berry, S. McParland, and D. G. Bradley. 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13:70. https://doi.org/10.1186/1471-2156-13-70.

Qanbari, S. 2020. On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. Front. Genet. 10:1304. https://doi.org/10.3389/fgene.2019.01304.

R Core Team. 2023. R: A Language and Environment for Statistical Computing [Computer software]. R Foundation for Statistical Computing. Accessed Apr. 10, 2025. https://www.R-project.org/.

Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDaneld, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schultheiss, S. G. Schroeder, A. M. Phillippy, J. B. Cole, C. P. Van Tassell, G. Liu, T. P. L. Smith, and J. F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience 9:giaa021. https://doi.org/10.1093/gigascience/giaa021.

Schneider, H., A.-M. Krizanac, C. Falker-Gieske, J. Heise, J. Tetens, G. Thaller, and J. Bennewitz. 2024. Genomic dissection of the correlation between milk yield and various health traits using functional and evolutionary information about imputed sequence variants of 34,497 German Holstein cows. BMC Genomics 25:265. https://doi.org/10.1186/s12864-024-10115-6.

Stahl, K., S. Papiol, M. Budde, M. Heilbronner, M. Oraki Kohshour, P. Falkai, T. G. Schulze, U. Heilbronner, and H. Bickeböller. 2025. Aggregating single nucleotide polymorphisms improves filtering for false-positive associations postimputation. G3 (Bethesda) 15:jkaf043. https://doi.org/10.1093/g3journal/jkaf043.

Sumreddee, P., E. H. Hay, S. Toghiani, A. Roberts, S. E. Aggrey, and R. Rekaya. 2021. Grid search approach to discriminate between old and recent inbreeding using phenotypic, pedigree and genomic information. BMC Genomics 22:538. https://doi.org/10.1186/s12864-021-07872-z.

Sumreddee, P., S. Toghiani, E. H. Hay, A. Roberts, S. E. Aggrey, and R. Rekaya. 2020. Runs of homozygosity and analysis of inbreeding depression. J. Anim. Sci. 98:skaa361. https://doi.org/10.1093/jas/skaa361.

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.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.

Wiggans, G. R., and P. M. VanRaden. 1989. USDA-DHIA Animal Model Genetic Evaluations. National Cooperative Dairy Program Handbook. Fact Sheet H-2. Extension Service, USDA, Washington, DC.

Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. 94:3202–3211. https://doi.org/10.3168/jds.2010-3866.

Wootton, E., C. Robert, J. Taillon, S. D. Côté, and A. B. A. Shafer. 2023. Genomic health is dependent on long-term population demographic history. Mol. Ecol. 32:1943–1954. https://doi.org/10.1111/mec.16863.

Wright, S. 1922. Coefficients of Inbreeding and Relationship. Am. Nat. 56:330–338. https://doi.org/10.1086/279872.

Yang, J., A. Bakshi, Z. Zhu, G. Hemani, A. A. E. Vinkhuyzen, S. H. Lee, M. R. Robinson, J. R. B. Perry, I. M. Nolte, J. V. van Vliet-Ostaptchouk, H. Snieder, T. Esko, L. Milani, R. Mägi, A. Metspalu, A. Hamsten, P. K. E. Magnusson, N. L. Pedersen, E. Ingelsson, N. Soranzo, M. C. Keller, N. R. Wray, M. E. Goddard, and P. M. Visscher. 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. 47:1114–1120. https://doi.org/10.1038/ng.3390.

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.

Yengo, L., J. Yang, M. C. Keller, M. E. Goddard, N. R. Wray, and P. M. Visscher. 2021. Genomic partitioning of inbreeding depression in humans. Am. J. Hum. Genet. 108:1488–1501. https://doi.org/10.1016/j.ajhg.2021.06.005.

Yengo, L., Z. Zhu, N. R. Wray, B. S. Weir, J. Yang, M. R. Robinson, and P. M. Visscher. 2017. Detection and quantification of inbreeding depression for complex traits from SNP data. Proc. Natl. Acad. Sci. USA 114:8602–8607. https://doi.org/10.1073/pnas.1621096114.

Yuan, C., J. L. Gualdrón Duarte, H. Takeda, M. Georges, and T. Druet. 2024. Evaluation of heritability partitioning approaches in livestock populations. BMC Genomics 25:690. https://doi.org/10.1186/s12864-024-10600-y.

Zheng, H.-F., J.-J. Rong, M. Liu, F. Han, X.-W. Zhang, J. B. Richards, and L. Wang. 2015. Performance of genotype imputation for low frequency and rare variants from the 1000 Genomes. PLoS One 10:e0116487. https://doi.org/10.1371/journal.pone.0116487.

ORCIDS

Chang Xu, https://orcid.org/0009-0005-5515-5741
Christian Maltecca, https://orcid.org/0000-0002-9996-4680
Francesco Tiezzi, https://orcid.org/0000-0002-4358-9236
John B. Cole, https://orcid.org/0000-0003-1242-4401
Sajjad Toghiani, https://orcid.org/0000-0002-5090-728X
Li Ma, https://orcid.org/0000-0003-1038-1081
Jicai Jiang https://orcid.org/0000-0001-6890-7539