Genomic evaluation of crossbred dairy cattle in the United States – an update

G.R. Wiggans[1], P.M. VanRaden[2], D.J. Null[2], E.L. Nicolazzi[1], G.B. Jansen[1], and J.H. Megonigal, Jr.[2]

[1] *Council on Dairy Cattle Breeding, Bowie, MD 20716, USA*

[2] *U.S. Department of Agriculture, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350, USA*

Abstract

Genomic evaluations of crossbred dairy cattle were first released in the United States in April 2019 by the Council on Dairy Cattle Breeding (Bowie, MD). Each crossbred animal has an evaluation calculated for each of 5 breeds: Ayrshire, Brown Swiss, Guernsey, Holstein, and Jersey. The evaluations are expressed on an all-breed base and combined by weighting the breed proportions — called breed base representation (BBR) — for each animal. The BBRs are estimated with the same process used for other traits. The breed reference populations are updated annually and are composed of AI progeny-tested bulls (≥10 daughters) with at least 4 generations of same breed ancestry. A blending procedure is applied to animals that have a highest BBR of <90%. The SNP effects are estimated from animals with a BBR of ≥94%, which are considered to be purebreds. Evaluations are reported on individual breed bases. In most cases, the breed with the highest BBR is used. However, to facilitate comparisons within herd, the breed associated with an animal's identification record is used if the BBR is ≥40%. For weekly evaluations, an initial approximate BBR is calculated based on only the SNPs present on the genotyping chip so that imputation is not required. Currently, 5 groups of chips are defined, and SNP effects for breed determination are estimated for each group. To minimize the portion of animals that are incorrectly designated for blended evaluations based on the approximate BBR, animals with a BBR of ≥85% are processed with purebreds until a BBR based on an imputed genotype is calculated; then animals with a new BBR of <90% are moved to the blended group. A breed error is declared if the highest BBR is ≥85% for a breed other than that from the identification record. The blending process is not applied to type traits, which are not comparable across breeds, or to feed saved and calving traits, which are not calculated for all breeds. Recessive haplotypes are not reported for crossbreds because the largely within-breed nature of those conditions leads to less accuracy in detecting crossbred carriers. Crossbred evaluations have lower reliability than those for purebreds; however, by processing crossbreds, parentage discovery and BBRs are provided. In July 2021, 154,670 animals had blended evaluations. Of those animals, 48% were born in 2019 or later, an indication of recent increases in crossbreeding and genotyping of crossbreds.

Key words: genomic evaluation, crossbreed, breed base representation, reference population, genotype

Introduction

Genomic evaluations of crossbred dairy cattle were first released in the United States in April 2019 by the Council on Dairy Cattle Breeding (Bowie, MD) as described by Wiggans et al (2019). Since that implementation, crossbreeding has continued to increase in the United States; >260,000 (almost 7%) of the 3.7 million U.S. milk-recorded cows were crossbreds in 2020 (Norman et al., 2021). Before April 2019, crossbreds were not included in genomic
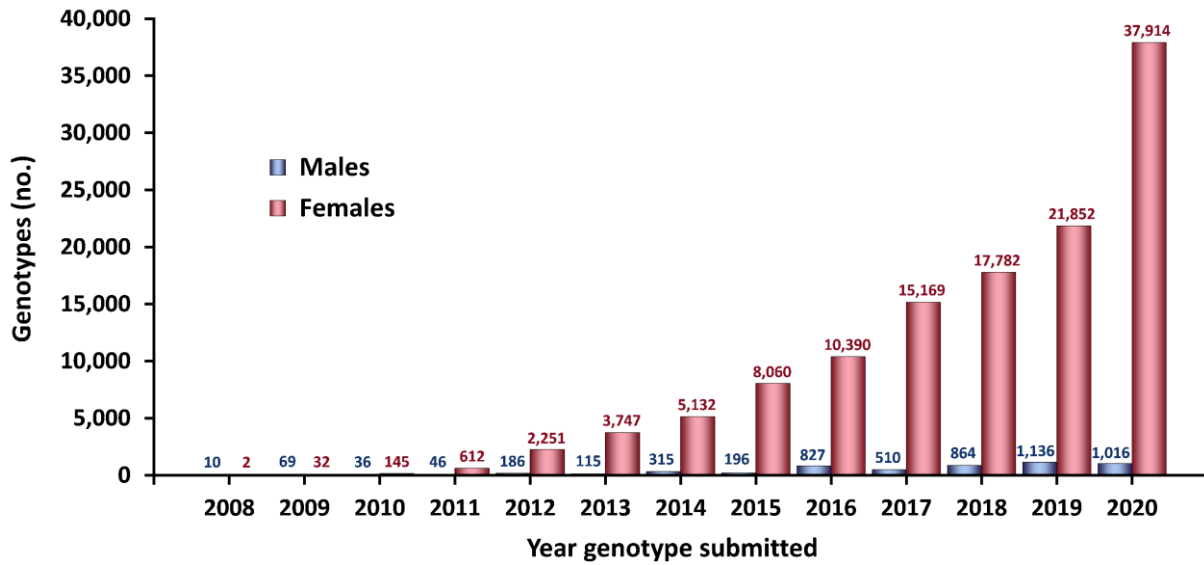
Figure 1. Frequency of crossbred genotypes by year submitted for genomic evaluation.

evaluation because estimates of marker effects, allele frequencies, and linkage differ by breed, which requires that imputation be done within breed. If parental genotypes are missing, population allele frequencies that vary by breed are used for imputation.

The success of genomic evaluation for purebreds generated interest in extending genomic evaluation to crossbreds. Figure 1 shows the increasing number of crossbred genotypes submitted each year for use in genomic evaluation. The implementation of crossbred evaluations in 2019 led to a substantial increase in the number of crossbred genotypes sub-mitted, primarily for females.

Materials and Methods

The method of evaluation is described by VanRaden et al. (2020). Each animal is evaluated in each of the 5 evaluated dairy breeds (Ayrshire, Brown Swiss, Guernsey, Holstein, and Jersey) using estimates for effects of individual-breed single nucleotide polymorphisms (SNPs) from the purebred evaluation. Those estimates then are weighted by breed proportions called breed base

representations (BBRs) to create a blended evaluation.

BBR

The SNP-effect estimates for BBR are calculated each April. For imputation, genotypes from Holstein bulls and cows with progeny and traditional evaluations and all genotypes from other breeds are included to create the haplotype library; only a random third are included for Holstein cows with only 1 progeny. For SNP-effect estimation the combined reference population includes purebred bulls from each of the 5 evaluated breeds (~36,000 bulls in April 2021). The estimation procedure and set of 79,060 SNPs used for traits that are genomically evaluated are also used for BBR. The breed reference populations are updated annually and are composed of AI progeny-tested bulls (≥10 daughters) with at least 4 generations of same-breed ancestry. Their phenotypic values are set to 0 or 100 to estimate SNP effects. Results are forced to the range of 1 to 100, and individual breed values of <2 are redistributed to the remaining breeds.

A rapid, approximate procedure that does not require imputation is employed for BBR estimation as genotypes are loaded. This allows

for immediate detection of incorrect breed assignments and an initial separation of cross-breds from purebreds. Actively used chips are assigned to a small number of subsets with SNPs common to the 79,060 SNPs used for the full BBR and to each chip included in the subset. The subsets need to be revised occasionally as new chips are adopted and older ones are retired. Separate SNP effects are estimated with the same phenotypes and imputed genotypes used for the full BBR (VanRaden et al., 2020), but the set of SNPs is restricted to those included in each subset. All data required for the application of the approximate BBR to incoming batches of genotypes, including SNP positions for each chip and SNP effects, are prepared and stored in binary format for rapid loading.

The method was tested on 194,0064 genotypes received over 3 months from June through August 2020. Four SNP subsets with 5,310 to 26,167 SNPs were formed. The correlation of approximate BBRs with subsequent full BBR (after imputation) ranged from 0.982 to 0.999, with lower correlations for a fraction of Ayrshires and Guernseys and very high correlations for Holsteins and Jerseys. Using a threshold of 89.5 for maximum BBR, no conflicting breed assignments were found between approximate and full BBRs; only 26 (0.013%) incidents of purebred designation by approximate BBR and crossbred designation by full BBR were observed along with 2,199 (1.13%) incidents of crossbred designation by approximate BBR and purebred designation by full BBR. The underestimation of maximum BBR was most prevalent for genotypes with low call rates. In subsequent testing, a lower BBR threshold of 85 was imposed to minimize incorrect initial assignment of purebreds to crossbred pro-cessing. In weekly and monthly evaluations, genotypes for animals that are designated as crossbred are imputed using the same multi-breed haplotype library as for estimation of BBRs. Non-crossbred genotyped parents are included to improve accuracy.

A genotype is processed as a crossbred if maximum initial BBR is <85%. However, if a BBR based in imputation remains <90, the genotype is processed with the crossbreds. This procedure increases the likelihood that a genotype that qualifies for processing as a purebred is so processed.

If the BBR for a different breed is ≥85, the genotype is excluded from further analysis and reported as a breed error. Genotypes that were initially assigned to purebred processing and then found to have a maximum BBR of <89.5 are reassigned as crossbred. Those BBRs initially assigned to crossbred processing and then found to have a maximum BBR of >89.5 receive an evaluation based only on the breed of the maximum BBR. All BBRs are recalculated each month, and the first monthly BBR replaces the BBR from the weekly evaluation. The stored BBRs are updated if any of the 5 breed BBRs for an animal differs from the stored BBR by ≥4. Only monthly BBRs are released because they are calculated from genotypes imputed in the appropriate breed group and, therefore, expected to be more stable.

Calculation of crossbred evaluations

Combining evaluations across breeds is possible when genetic effects are on the same multi-breed base. To make this possible, the Council on Dairy Cattle Breeding modified the genomic evaluation procedure in April 2018 to use traditional evaluations on a multi-breed base before adjusting to individual breed bases. Evaluations from a single breed are reported for type traits because evaluations are not available on an across-breed base and for feed saved, calving ease, and health traits because evaluations are not available for all breeds. Evaluation breed is the identification breed if BBR is ≥40 (beginning in December 2019) or the breed with the highest BBR. This policy allows the owner to have some control over the breed base for an animal's reported evaluation, particularly for first-generation crossbreds.

Purebred evaluations are the source of SNP-effect estimates used in crossbred evaluations. The reference population for calculation of those effects includes only animals with an evaluation-breed BBR of ≥94. This limitation ensures that SNP-effect estimates are not affected by animals of other breeds; however, it does reduce the size of the reference population.

Reliability and future inbreeding

Expected inbreeding of future progeny is calculated with relationships from purebred evaluations. The purebred reference population is used to determine both the average genomic relationship of the animal and expected future inbreeding from pedigree. The evaluation breed determines which purebred population is used.

Results & Discussion

For July 2021 evaluations, 154,670 animals received blended evaluations. For those animals, 48% were born in 2019 or later, and 86% were Holstein × Jersey crosses (sum of Holstein and Jersey BBRs ≥ 90). In an evaluation including data available in March 2019, 31,521 of 68,691 animals (46%) had not previously received evaluations through the purebred-only official evaluation system.

The exclusion of crossbred animals from purebred evaluations has little effect on pure-breds. The largest effect is for crossbreds with a BBR of <94 that previously received a purebred evaluation because their own performance did not affect the SNP effect estimates.

Recessive conditions

Most recessive characteristics are breed specific. Furthermore, imputation of crossbreds considers haplotypes from all breeds. Although the overall imputation accuracy is sufficient to ensure valid genomic evaluations, pedigree tracing of small haplotype regions is not stable or reliable enough to support their release. False positives also occur more frequently. For those reasons, recessive conditions are not yet reported for crossbreds.

Conclusions

Implementation of blending based on BBR extends genomic evaluation to more animals and provides more complete information for herds that genotype all females. Even for animals that were evaluated before implementation of genomic evaluation for crossbreds, blended evaluations more accurately reflect an animal's multi-breed origin. The calculation of BBRs as genotypes are loaded enables more accurate assignment of genotypes to purebred or blended processing as well as earlier detec-tion of incorrect breed identification. Because of the direct influence of BBRs on evaluations, they are checked monthly and updated if a large change has occurred because of genotype re-assignment or genotyping with a higher density chip. Crossbred evaluations have lower reliability than those for purebreds but provide the benefit of parentage discovery and BBRs. The breed for reporting an evaluation is based on an animal's BBR. An unfortunate consequence is that animals in a herd with extensive cross-breeding have evaluations that are not comparable because they are based on different breeds.

Acknowledgments

The Council on Dairy Cattle Breeding and USDA are equal opportunity providers and employers.

References

Norman, H.D., Walton, L.M., Dürr. J.W. 2021. Reasons that cows in Dairy Herd Improvement programs exit the milking herd (2020). https://queries.uscdcb.com/publish/dhi/dhi21/cullall.html.

VanRaden, P.M., Tooker, M.E., Chud, T.C.S., Norman, H.D., Megonigal, Jr., J.H., Haagen, I.W., Wiggans, G.R. 2020. Genomic predictions for crossbred dairy cattle. *J. Dairy Sci. 103,* 1620–1631.

Wiggans, G.R., VanRaden, P.M., Nicolazzi, E.L., Tooker, M.E., Megonigal, Jr., J.H., Walton, L.M. 2019. Extending genomic evaluation to crossbred dairy cattle – US implementation. *Interbull Bull. 55,* 46–49.