Extending genomic evaluation to crossbred dairy cattle – US implementation

G.R. Wiggans[1], P.M. VanRaden[2], E.L. Nicolazzi[1], M.E. Tooker[2,3], J.H. Megonigal, Jr.[1], and L.M. Walton[1]

[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

[3]Retired

Abstract

In April 2019, the Council on Dairy Cattle Breeding (CDCB) extended its genomic evaluation system to crossbred dairy cattle by weighting estimates for effects of individual-breed single nucleotide polymorphisms (SNPs) by breed proportions. Previously, animals detected as crossbreds were excluded because SNP effect estimates differ by breed and imputation relies on breed-specific haplotype frequencies if parental genotypes are missing. Breed proportions for all animals were estimated using a combined reference population of purebred bulls from each of the five evaluated breeds. Their phenotypic values were set to zero or 100 to estimate SNP effects for breed base representation (BBR). For genomic evaluation purposes, genotypes for animals that appeared to be crossbred were imputed using the same multi-breed haplotype library as for estimation of BBR. Non-crossbred genotyped parents were included to improve accuracy. Because combination of SNP effect estimates across breeds requires that genetic effects be on the same multi-breed base, CDCB modified the genomic evaluation procedure in April 2018 to use traditional evaluations on multi-breed base before adjusting to individual breed bases. Animals with a breed percentage of at least 90 were not included in the blending, and a minimum breed percentage of 94 was imposed for an animal to contribute to SNP effect estimation. This limitation ensured that SNP effect estimates were not affected by animals of other breeds but reduced the size of the reference population. For animals with a breed percentage below 90, weighting evaluations by BBR was not possible for type traits because they are not on a common base or for calving and health traits because they are not evaluated for all breeds. In those cases, the single-breed evaluation for the breed with the highest percentage is reported. If BBR is close to 50%, some allowance is made to use the breed reported in an animal's identification as the evaluation breed even if it is not the highest BBR. For a March 2019 test evaluation, 68,691 animals received evaluations that were weighted by BBR, and 31,521 of those animals had not previously received an official evaluation. Implementation of the weighting process extends genomic evaluation to more animals and provides more complete information for herds that genotype all females. Even for animals that received evaluations in the past, the evaluations weighted by BBR more accurately reflect an animal's multi-breed origin.

Key words: genomic prediction, reference genome, ARS-UCD1.2, imputation, genotype

Introduction

Crossbreeding has been increasing in the United States, and >200,000 (>5%) of the 3.9 million U.S. milk-recorded cows were crossbreds in 2018 (Norman et al., 2019). This change has been driven by increased demand for milk components and an attempt to improve fertility and robustness. Crossbred animals have been excluded from genomic evaluation because estimates of marker effects differ by breed, and allele frequencies and linkage also differ, 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 widespread adoption of genomic evaluation to enable early determination of genetic merit has generated a demand to extend genomic evaluation to crossbreds. The Council on Dairy Cattle Breeding (CDCB) had received >32,000 genotypes of crossbred animals that had been excluded from evaluation based on a set of over 600 markers used to identify crossbreds. Figure 1 shows the increasing number of crossbred genotypes by year. This means that >$1 million was spent in genotyping with no genomic evaluation provided. The common practice of

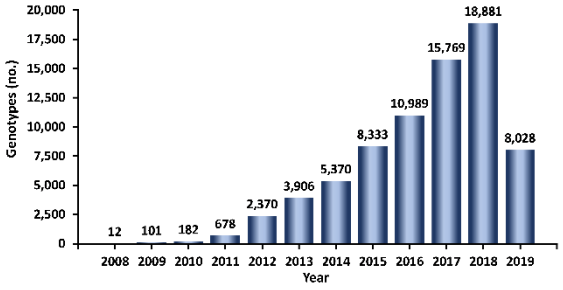

Figure 1. Frequency of crossbred genotypes received by year from 2008 through May 2019.

whole herd genotyping contributed to CDCB's receipt of genotypes for crossbred animals.

In April 2019, the Council on Dairy Cattle Breeding (CDCB) extended its genomic evaluation system to crossbred dairy cattle using the method described by Tooker et al (2017). Each animal is evaluated in each of the five 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. These estimates then are weighted by breed proportions called breed base representations (BBRs).

BBR

The SNP-effect estimates are calculated each April using a combined reference population of purebred bulls in each of the five evaluated breeds (~36,000 bulls in April 2019). Their phenotypic values are set to zero 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. Genotypes for animals that appear to be crossbred are imputed using the same multi-breed haplotype library as for estimation of BBRs. Non-crossbred genotyped parents are included to improve accuracy. The BBRs are used to determine if animals should be included in the crossbred analysis. As a first approximation, genotypes are designated as crossbred if >15% of the breed-specific SNP alleles are unlikely for the identification breed. However, if <10% of alleles are unlikely for a different breed, the genotypes are excluded from further analysis and reported as breed errors. BBR is first calculated during weekly genomic evaluation. Genotypes that were initially assigned to purebred processing and then found to have an evaluation breed BBR of <89.5 are reassigned as crossbred; those initially assigned to crossbred processing and then found to have an evaluation breed BBR of >89.5 receive an evaluation based only on their identification breed. BBR is recalculated at each monthly evaluation, and the first monthly BBR replaces the BBR from the weekly evaluation. The stored BBRs are updated if any of the five breed values for an animal differs from the stored value by ≥4. Only monthly BBRs are released because they are expected to be more stable as they were calculated from genotypes imputed in the appropriate breed group.

Calculation of crossbred evaluations

A requirement for weighting evaluations across breeds is that genetic effects must be on the same multi-breed base. CDCB 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 calving ease and health traits because evaluations are not available for all breeds. Evaluation breed is the identification breed if BBR is ≥45 (the minimum will be reduced to 40 in October 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 the SNP-effect estimates used in crossbred evaluations. The reference population for the calculation of those effects is further restricted to include only animals with an evaluation-breed BBR of ≥94. This limitation was imposed to ensure that SNP-effect estimates were not affected by animals of other breeds; however, it does reduce the size of the reference population.

For July 2019 evaluations, 74,759 animals received evaluations weighted by BBR. Based on a March 2019 test with 68,691 animals in the weighted evaluation, 31,521 had not been evaluated in the previously purebred-only official evaluation system.

The exclusion of crossbred animals from purebred evaluations had little effect on the

purebred evaluations. The largest effect was for crossbreds with a BBR of <94 that had previously received a purebred evaluation because their own performance did not affect the SNP effect estimates.

Recessive conditions

Most recessive characteristics are breed specific. For this reason, recessive conditions were not reported for crossbred evaluations initially. Development work is under way to provide this information. The list of recessive characteristics for each animal has been extended to include characteristics across breed. Bulls of other breeds progeny tested with at least 100 or 1,000 daughters depending on gene test and breed are also assumed to be noncarriers and used in the reference population. This policy prevents accidentally filling missing alleles in haplotypes in one breed because of a carrier in a different breed. Because the genotypes of parents are included to enhance imputation, they also are present to predict recessive conditions. Reporting may be augmented by an indicator to identify that carrier status is also confirmed by a detected carrier ancestor.

Reliability and future inbreeding

Calculation of reliability is based on the relationship between the animal and the predictor population. For crossbred evaluation, purebred reference populations must be accessed. Similarly, a subset of the predictor population is used to obtain expected inbreeding of future progeny (EFI), which along with heterosis is used to correct the evaluations. Initially, reliability and EFI from weighted evaluation were based on the multi-breed reference population used to determine BBR estimates because weighted evaluation does not have its own predictor population. However, this strategy had two main drawbacks: (1) animals closer to the purebred population than to other crossbreds did not benefit from the greater relationship to the predictor population because the method did not rely on BBR, and (2) reliability and EFI were underestimated for most crossbreds.

A new method to account for the different degrees of relationship between crossbreds and the purebred population was developed because of the large drop of 20 percentage points in reliability that was observed when BRR was <90 (Table 1). In the new method, the purebred (instead of multi-breed) reference population is used to determine both the average relationship of the animal and EFI. The evaluation breed determines which purebred population is used. As a result, both reliability and EFI estimates are expected to be more accurate for crossbreds. Table 1 shows differences in reliability between the current official method and the new method that will be implemented by the end of 2019.

Conclusions

Implementation of weighting based on BBR extends genomic evaluation to more animals and provides more complete information for herds that genotype all females. Even for animals that received evaluations in the past, weighted evaluations more accurately reflect an animal's multi-breed origin. Because of the direct influence of BBRs on evaluations, they

Table 1. Official and new method reliabilities and standard deviations (SDs) for purebred and crossbred Holsteins in July 2019 by BBR.

		Official method		New method	
BBR range	Animals (no.)	Reliability (%)	SD	Reliability (%)	SD
98 – 100	2,529,217	75.4	2.9	75.4	2.8
95 – <98	45,962	73.5	4.4	73.5	4.3
93 – <95	14,303	73.9	3.5	74.0	3.4
90 – <93	10,815	74.1	3.1	74.1	2.9
85 – <90	9,890	45.6	7.3	71.2	3.7
80 – <85	6,900	44.3	7.6	69.3	5.0
75 – <80	5,496	42.9	7.9	66.8	8.1
70 – <75	4,012	41.3	7.9	63.4	11.3
60 – <70	2,655	40.1	8.6	60.6	11.1
<60	2,457	39.8	8.1	58.0	10.0

are checked monthly and updated if a large change has occurred because of genotype reassignment or genotyping with a higher density chip.

References

Norman, H.D., Walton, L.M., Dürr. J.W. 2019. Reasons that cows in Dairy Herd Improvement programs exit the milking herd (2018). https://queries.uscdcb.com/publish/dhi/current/cullall.html?

Tooker, M.E., VanRaden, P.M., Fok, G.C. 2017. Genomic predictions for crossbreds from all-breed data. *J. Dairy Sci. 100* (Suppl. 2), 409–410.