Genomic Evaluations and Breed Composition for Crossbred U.S. Dairy Cattle

P.M. VanRaden and T.A. Cooper
Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350, USA

Abstract

Genomic evaluations are desired for crossbred as well as purebred populations when selection is applied to commercial and not only breeding herds. Genomic breed composition was estimated from 60 671 markers using the known breeds of daughter-proven Holstein, Jersey, Brown Swiss and Ayrshire bulls as the four traits (breed fractions) to be predicted. Genotypes of 6 296 crossbred animals were imputed from lower density chips together with either their 3 119 ancestors or all 834 367 genotyped animals. Estimates of breed composition were adjusted so that no values were negative or exceeded 100 and the four breed percentages summed to 100. The crossbreds included 733 Jersey x Holstein crossbreds with >40% of both breeds (F1 crosses), 55 Brown Swiss x Holstein F1, 2 300 Holstein backcrosses with >67% and <90% Holstein, 2 026 Jersey backcrosses, 27 Brown Swiss backcrosses and 502 other crossbreds of various mixtures. Crossbred evaluations were averages of direct genomic values computed using marker effects for each pure breed, weighted by the animal's genomic breed composition. The marker effects were estimated separately for each breed on the all-breed scale instead of the within-breed scales currently used.

Key words: genomic evaluation, multiple breeds, crossbreeding, imputation

Introduction

Most genomic evaluations are separate by breed, and crossbreds usually are not included with the exception of the multi-breed evaluation in New Zealand (Winkelman *et al*., 2015). Some evaluations include mixtures of closely related breeds (Zhou *et al*., 2014) or subpopulations within breeds that were separate for many generations (Thomasen *et al*., 2013). Many research studies have combined data from different pure breeds to improve marker estimates, but gains were mostly small and not implemented routinely yet (Hozé *et al*., 2014; Karoui *et al*., 2012; Kemper *et al*., 2015; Makgahlela *et al*., 2013; Olson *et al*., 2012).

Genomic breed composition (GBC) can be estimated most accurately using all markers (Hulsegge *et al*., 2013; VanRaden *et al*., 2011). Until now, breed check markers were used to exclude many crossbreds from US evaluations because none of the individual breed prediction equations were accurate for crossbreds. The selected markers were monomorphic in one breed and had fewer than 30% of animals homozygous for that allele in another breed (Wiggans *et al*., 2010).

Goals of this study were to 1) compute GBC for all animals in the national database, 2) examine categories of animals genotyped, 3) compare imputation strategies for crossbreds, and 4) evaluate crossbreds using GBC to weight marker effects computed for each pure breed. Evaluations for crossbreds would reduce the breeder's need to guess before genotyping if an animal will pass breed check edits and be evaluated.

Methods

Genotypes of 6 296 crossbred animals were examined for potential inclusion in routine genomic evaluations. Previously these animals were excluded because genomic predictions are computed within breed, and all-breed methods are more complex with little reliability gain for purebreds. Animals had been excluded if they had 1) a pedigree sire or dam of another breed, or 2) >40% of breed check markers not from the breed of evaluation

for medium or high density genotypes or >20% not from breed of evaluation for low density genotypes, or 3) a genotype identified as being from a completely different breed. The pedigree file for the crossbreds included 79 235 animals, but only 3 119 of the ancestors had genotypes; those were included to improve accuracy of imputation. A second analysis imputed all 834 367 purebred and crossbred animals together from March 2015 genotypes to compare to the partial data that included only the 6 296 crossbred animals and their 3 119 genotyped ancestors.

Genomic breed composition (GBC) of each animal was estimated using 60 671 markers after imputation with findhap. Only 33 of the crossbred animals had ≥50K genotypes, the other 99.5% had a variety of lower density chips containing 3K to 13K usable markers. Marker effects to predict breed percentage as a trait were estimated using genotypes of all progeny-tested bulls for each of the four breeds Holstein (HO), Jersey (JE), Brown Swiss (BS), and Ayrshire (AY), or by using just four observations, each containing one pure breed's allele frequencies. The AY analysis included all Red Dairy Cattle as if they were one breed.

Estimates of breed composition from multiple regressions can exceed 100% for a given breed or be negative for one or more other breeds. For purebreds, expected and average GBC equal 100% for the declared breed and 0% for other breeds, but individual animals fluctuate around those values due to genomic variation, similar to genomic inbreeding becoming negative for animals less related than the average for base animals. Generally the most popular animals within a breed may exceed 100% GBC because they have more of the alleles used to differentiate the breeds.

Adjusted breed composition (ABC) can force genomic estimates for each animal to sum to 100%, with no estimates <0% or above 100%. This may be needed when applying estimates to marker effects or for ease of interpretation, and supposes that all ancestors are from breeds with equations available. The mathematical steps are:

1) obtain the sum of GBC across breeds,
2) adjust the GBC mean by subtracting from each GBC value the sum of GBC divided by the number of breeds (Nbrd),
3) obtain the range of the adjusted GBC from the maximum adjusted breed GBC and minimum adjusted breed GBC,
4) obtain a SD adjustment if any adjusted GBC are >100 or <0, computed as the maximum of (largest adjusted GBC - 100 / Nbrd) / [100 * (1 – 1 / Nbrd)], or (100 / Nbrd - smallest adjusted GBC) / (100 / Nbrd), and
5) obtain ABC = 100 / Nbrd + (adjusted GBC - 100 * Nbrd) / SD.

Genomic predicted transmitting abilities (GPTAs) were computed by applying the marker effects from each of the four breeds and then weighting those GPTAs by genomic breed composition. This step requires computing all inputs on the same all-breed scale, and then converting GPTAs to the published within-breed scales. Strandén and Mäntysaari (2012) proposed a similar random regression approach but used pedigree breed composition (PBC) rather than GBC to weight the marker effects. A simpler but probably less accurate approach is to compute GPTAs for all breeds using just one set of marker effects.

Results and Discussion

Breed composition is more accurate using genotypes than pedigrees because pedigrees often are partially missing or incorrect. For example, 1 024 animals had genomic breed composition >90% pure but had one parent recorded as being from a different breed. For another 133, their own breed code differed from their genomic breed, indicating possible sample switches. The crossbreds included 733 JE x HO crossbreds with >40% of both breeds, 55 BS x HO F1s, 2,300 HO backcrosses with >67% and <90% HO, 2,026 JE backcrosses, 27 BS backcrosses, and 502 other crossbreds of various mixtures (Table 1). Many HO animals from Mexico and Chile were counted as HO backcrosses with about 10% contribution from each of JE, BS, and AY, possibly indicating local ancestors from other populations (Spanish or Zebu cattle).

Averages of GBC for the 834 367 animals in the full data were 85.7% HO, 11.5% JE, 2.3% BS, and 0.5% AY, but after adjusting each animal's GBC to sum to 100% and range between 0 and 100%, averages of ABC for each breed were slightly closer to the 25% average across breeds (Table 2). Average PBC was nearly the same as GBC. Lowest GBC were -3% to -8% for the 4 breeds, whereas highest GBC were 104% to 108%. Correlations of GBC with PBC were high, above 0.99, and correlations of GBC with ABC were even higher. Correlations of GBC with PBC for only the crossbreds were much lower, ranging from 0.72 for AY to 0.87 for JE.

Table 1. Categories of animals excluded by breed check edits.

Ani-mals	Category	Definition
733	JE x HO F1	>40% of both
55	BS x HO F1	>40% of both
2 300	HO backcrosses	>67%, <90% HO
2 026	JE backcrosses	>67%, <90% JE
27	BS backcrosses	>67%, <90% BS
502	Other crosses	Not in above
1 024	Purebreds	>90% of ID breed
133	Wrong breed	<20% of ID breed

Table 2. Means, minimums, and maximums of genomic and adjusted breed composition (%) across all animals.

Statistic	HO	JE	BS	AY
Average GBC	85.7	11.5	2.3	0.5
Average ABC	84.3	11.7	2.7	0.9
Average PBC	85.7	11.5	2.2	0.5
Minimum GBC	-8	-3	-4	-3
Minimum ABC	0	0	0	0
Maximum GBC	106	108	104	110
Maximum ABC	100	100	100	100
Corr(GBC,ABC)	.999	.999	.999	.997
Corr(GBC,PBC)	.996	.996	.998	.990

Estimates of GBC from part and full data were very similar, implying that genotypes were imputed consistently even when fewer purebred animals were included. Correlations between part and full GBC were 0.999 for each breed fraction except 0.997 for AY breed fraction. Thus, GBC could be computed fairly quickly after imputing from part data or eventually with imputation of all animals together. Estimates of GBC using only the four breed allele frequencies gave similar overall statistics for the crossbreds, but many individual purebred animals were not accurately estimated. Another possibility would be to directly estimate breed composition without imputation using only the 6 909 markers in common to most chips, but a previous study (VanRaden *et al*., 2011) found that using 3 000 markers gave lower accuracy than 43 000 markers.

Crossbred GPTAs obtained as averages of the four purebred marker effects weighted by ABC were correlated by only about 0.91 to GPTAs obtained from a common set of marker effects for all breeds. For the purebred animals, the weighted GPTAs and the single-breed official GPTAs were correlated by 0.98-0.99 for all breeds except AY, which were correlated by 0.93. Correlations were much lower using common effects for all breeds. Thus, weighting by ABC retains information within each pure breed while allowing crossbreds to share that information.

More breeders now do whole herd genotyping and may expect evaluations for all genotyped animals in the future. They already spent about $300 000 genotyping these crossbreds, and demand is growing. With the current breed edits, owners must guess if an animal will qualify for evaluation before paying for genotyping, whereas evaluations for all animals could simplify management decisions. Numbers of genotyped crossbreds may still be too small to estimate crossbred performance as a trait separate than purebred performance, but theoretical studies have developed methods that could be applied in populations where crossbreeding is more routine (Christensen *et al*., 2014; Esfandyari *et al*., 2015; Zeng *et al*., 2013).

Conclusions

Breed composition was estimated from genotypes after imputing lower density chips to 60 671 markers for all animals in the national database. Methods were developed to

adjust the initial GBC to ABC values that were limited to the range of 0-100% and summed to 100% across breeds. Correlations were high between GBC and ABC. Most of the genotyped crossbreds were Jersey x Holstein backcrosses. Genomic evaluations for crossbreds can be computed by weighting the marker effects for separate breeds by ABC instead of PBC as in some previous reports, but the marker effects must be computed on the all-breed base rather than within-breed bases. An advantage of ABC over PBC is that pedigrees are often incomplete or inaccurate for crossbred animals. Thousands of crossbred animals are being genotyped because many commercial producers now apply genomic selection to their whole herds.

Acknowledgements

The Council on Dairy Cattle Breeding (CDCB) provided the data for this research under USDA Nonfunded Cooperative Agreement 58-1245-3-228N. This research was part of USDA-ARS project 1265-31000-096-00, "Improving Genetic Predictions in Dairy Animals Using Phenotypic and Genomic Information." The authors thank Gary Fok for assistance with data processing and staff of the American Jersey Cattle Association, Zoetis and CDCB for helpful discussions.

References

Christensen, O.F., Madsen, P., Nielsen, B. & Su, G. 2014. Genomic evaluation of both purebred and crossbred performances. *Genetics Selection Evolution 46*, 23.

Esfandyari, H., Sørensen, A.C. & Bijma, P. 2015. Maximizing crossbred performance through purebred genomic selection. *Genetics. Selection Evolution 47*, 16.

Hozé, C., Fritz, S., Phocas, F., Boichard, D., Ducrocq, V. & Croiseau, P. 2014. Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. *Journal of Dairy Science 97*, 3918–3929.

Hulsegge, B., Calus, M.P.L., Windig J.J., Hoving-Bolink, A.H., Maurice-van Eijndhoven, M.H.T. & Hiemstra, S.J. 2013. Selection of SNP from 50K and 777K arrays to predict breed of origin in cattle. *Journal of Animal Science 91,* 5128-5134.

Karoui. S., Carabaño, M.J., Díaz, C. & Legarra, A. 2012. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. *Genetics Selection Evolution 44*, 39.

Kemper, K.E., Reich, C.M., Bowman, P.J., vander Jagt, C.J., Chamberlain, A.J., Mason, B.A., Hayes, B.J. & Goddard, M.E. 2015. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genetics Selection Evolution 47*, 29.

Makgahlela, M.L., Strandén, I., Nielsen, U.S., Sillanpää, M.J. & Mäntysaari, E.A. 2013. The estimation of genomic relationships using breedwise allele frequencies among animals in multibreed populations. *Journal of Dairy Science 96,* 5364-5375.

Olson, K.M., VanRaden, P.M. & Tooker, M.E. 2012. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science 95,* 5378–5383.

Strandén, I. & Mäntysaari, E.A. 2012. Use of random regression model as an alternative for multibreed relationship matrix. *Journal of Animal Breeding and Genetics 130,* 4-9.

Thomasen, J.R., Sørensen, A.C., Su, G., Madsen, P., Lund, M.S. & Guldbrandtsen, B. 2013. The admixed population structure in Danish Jersey dairy cattle challenges accurate genomic predictions. *Journal of Animal Science 91*, 3105-3112.

VanRaden, P.M., Olson, K.M., Wiggans, G.R., Cole, J.B. & Tooker, M.E. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science 94*, 5673–5680.

Wiggans, G.R., VanRaden, P.M., Bacheller, L.R., Tooker, M.E., Hutchison, J.L., Cooper, T.A. & Sonstegard, T.S. 2010. Selection and management of DNA markers for use in genomic evaluation. *Journal of Dairy Science 93,* 2287–2292.

Winkelman, A.M., Johnson, D.L. & Harris, B.L. 2015. Application of genomic evaluation to dairy cattle in New Zealand. *Journal of Dairy Science 98*, 659-675.

Zeng, J., Toosi, A., Fernando, R.L., Dekkers, J.C.M. & Garrick, D.J. 2013. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genetics Selection Evolution 45*, 11.

Zhou, L., Heringstad, B., Su, G., Guldbrandtsen, B., Meuwissen, T.H.E., Svendsen, M., Grove, H., Nielsen, U.S. & Lund, M.S. 2014. Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. *Journal of Dairy Science 97*, 4485-4496.