Predicting Breed Composition Using Breed Frequencies Of 50,000 Markers From The U.S. Meat Animal Research Center 2,000 Bull Project

L.A. Kuehn[*], J.W. Keele[*], G.L. Bennett[*], T.G. McDaneld[*], T.P.L. Smith[*], W.M. Snelling[*], T.S. Sonstegard[†] and R.M. Thallman[*]

Introduction

Breed composition would be useful for predicting heterosis, adaptability to production environments, and sorting animals into management groups. In addition, estimates of breed composition of crossbred commercial cattle would be useful for mapping loci underlying economically important traits and host resistance to disease. If it were possible to accurately estimate breed composition, unselected control allele frequencies matching a sample of cases (sick animals) could be statistically derived based on allele frequencies from reference samples for each breed.

Breed identification has been performed in cattle (e.g., Watanabe et al. (2008)) and other species (e.g., dogs in Parker et al. (2004)) using microsatellite alleles in specific genomic regions. However, prediction of breed composition in advanced generations of outcrossed populations is more difficult with limited markers because specific breed alleles are not necessarily passed on to advanced generations. The high density of the BovineSNP50 array may allow for a more robust test of breed composition.

Our objective was to evaluate the accuracy of estimating breed composition of crossbreds based on their BovineSNP50 genotypes and allele frequencies from unrelated purebred reference samples.

Material and methods

Bulls used to estimate breed frequencies. Allele frequencies were estimated for each of 16 breeds based on a sample of approximately 2,235 AI bulls expected to be influential in their breeds. Bulls were chosen by the U.S. beef cattle breed associations and semen (DNA source) was provided to the U.S. Meat Animal Research Center (USMARC) in 2008 and 2009. Representation of each breed was approximately proportional to the national herd size. Total numbers of bulls sampled per breed are listed in table 1. Each bull was genotyped using the

---

[*] USDA, ARS, U.S. Meat Animal Research Center, Clay Center, NE, 68933-0166 USA

[†] USDA, ARS, Bovine Functional Genomics Laboratory, Beltsville, MD 20705 USA

Illumina[‡] BovineSNP50 array (Matukumalli et al. (2009)). This sample of bulls and their genotypic data has been informally termed the USMARC 2,000 Bull Project.

In general, the aim of the 2,000 Bull Project was to provide a conduit to transfer research results to the beef cattle industry. In principle, estimates of SNP associations can be combined with genotypes from the 2,000 Bull Project to provide predicted genetic merit for traits not routinely collected by industry such as feed intake and resistance to disease.

Table 1: Numbers of bulls sampled per breed in the USMARC 2,000 Bull Project

British derived breeds		Continental European derived breeds		United States derived breeds	
Angus	403	Braunvieh	27	Beefmaster	65
Hereford	491[a]	Chianina[b]	47	Brahman	53
Red Angus	175	Charolais	125	Brangus	53
Shorthorn	86	Gelbvieh	146	Santa Gertrudis	54
		Limousin	141		
		Maine- Anjou	59		
		Salers	41		

[a] 180 Hereford bulls from L1 pedigree based at USDA-ARS Ft. Keogh Livestock and Range Research Laboratory.
[b] Most are Chianina by Angus composites with variable levels of base breed percentages.

Genotyping was successful for 52,156 markers using the Illumina BovineSNP50 array. Frequencies of each allele were summarized for each marker on a within breed basis. Resulting breed frequencies were used to test whether breed composition of crossbred animals could be predicted.

Crossbred test population. Steers and heifers from Cycle VII of the USMARC Germplasm Evaluation Project had previously been genotyped using the Illumina BovineSNP50 (Snelling et al. (2010)). Cycle VII was initiated by sampling 150 bulls and mating them to Angus, Hereford, or MARCIII composite (¼ Angus, ¼ Hereford, ¼ Pinzgauer, ¼ Red Poll) females to produce F_1 progeny. These F_1 progeny were subsequently mated in multiple-sire mating groups to produce 2-, 3-, and 4-breed cross progeny termed F_1^2 (i.e., crosses of F_1); breed proportions were therefore multiples of 25%. BovineSNP50 genotypes were obtained on the original purebred bulls, the F_1 sires (73), and 2,014 F_1^2. Paternity was assigned using BovineSNP50 results; consequently, expected pedigree breed composition of F_1^2 progeny was established (assuming dams were identified correctly).

Statistical analysis. To predict breed composition, genotypes of F_1^2 animals were converted to copies of Illumina allele B. Each individual animal's genotypes (copies of allele B divided by 2;

[‡] Reference herein to any specific commercial products by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government, and shall not be used for advertising or product endorsement purposes.

0, 0.5, or 1) were then predicted (**y**) using the following model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

where **X** is a 52,156 by 16 matrix of frequencies of allele B for each breed, **b** is a vector of regression coefficients representing the percentage contribution of each breed to the animal in **y**, and **e** was a random residual vector. Bulls that produced grandprogeny in the F_1^2 test population were not included in the breed frequency calculation for this demonstration.

Predicted breed contributions in **b** were compared to breed contributions estimated from pedigree records by correlating the respective vectors within each animal. Because the Pinzgauer and Red Poll portions of the MARCIII composite were not part of the 2,000 bull project sample, the portions of MARCIII belonging to Hereford and Angus were added to the pedigree Hereford and Angus contributions (25% x %MARCIII for each breed). Red Poll and Pinzgauer portions were considered their own composite breed (with no opportunity to be predicted). Also, the predicted frequencies for composite breeds (Beefmaster, Brangus, and Santa Gertrudis) were split among their respective base breeds (described further in results). Correlations were examined as the basis of whether breed composition could be accurately predicted from the 2,000 bull project breed frequencies.

Results and discussion

The mean and median estimates of within animal correlations before splitting composite breed predictions into their founder breeders were 0.903 and 0.918, respectively. After splitting the composite breeds, the mean and median were increased marginally to 0.913 and 0.926, respectively. The distribution of correlation coefficients after splitting the composite breeds is shown in figure 1.

Agreement between breed composition predicted by genotypes by genotypes and estimated from pedigree was high for most animals. Upon closer inspection of the data, 13 animals had a correlation between predicted and pedigree breed composition of less than 0.60. Of these, 11 had discrepant maternal breeds predicted, suggesting possible misassignment of the animal’s dam. Of the 627 animals with correlation coefficients between 0.60 and 0.90, 441 were ¼ MARCIII composite. Correlation coefficients of animals with MARCIII composites could be improved if the prediction data set contained a source of Pinzgauer and Red Poll or MARCIII germplasm. If the Pinzgauer x Red Poll portion of breed composition was zeroed from the pedigree breed composition of MARCIII derived F_1^2 animals (making the sum of pedigree breed compositions less than 1), the correlation coefficient mean and median rose to 0.936 and 0.952, respectively.

Several of the remaining animals had Red Angus or Angus breed contributions. Of all the breeds, these two were the most difficult to predict separately likely due to a shorter period since breed divergence. Similarly, many of the Angus derived F_1^2 animals were predicted to have Chianina breed composition to a small extent (usually no more than 5%), likely because Chianina bulls sampled in the 2,000 bull project were generally >50% Angus.

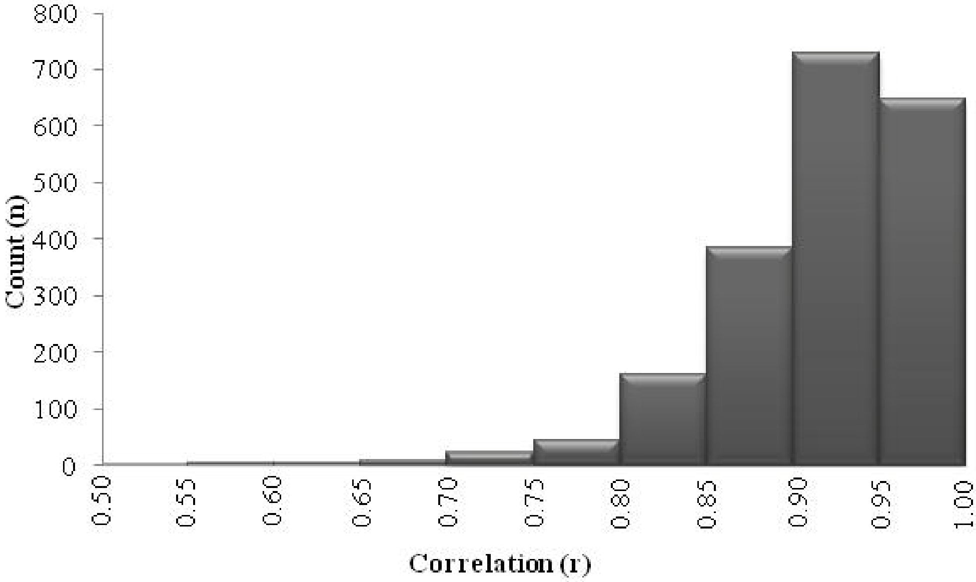

Figure 1: Distribution of correlation coefficients between pedigree and predicted breed composition

Conclusion

These results indicate that breed frequencies predicted from a high-density SNP panel can be used to predict breed composition of crossbred animals. While the breed frequency estimates were not all-encompassing relative to beef cattle breeds (and notably doesn't include prominent dairy breeds), these results do indicate that even when breeds are missing from the prediction breeds (i.e., Red Poll and Pinzgauer in the example data set), breed contribution can still be accurately predicted for the remaining breeds. Breed differences can be exploited in genetic selection and management; a tractable method of predicting breed composition in animals with unknown pedigrees can improve marker discovery and be utilized in marker assisted management programs, although use of the BovineSNP50 array for this purpose would currently be cost prohibitive.

References

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D. *et al.* (2009). *PLoS ONE*, 4:e5350.

Parker, H.G., Kim, L.V., Sutter, N.B. *et al.* (2004). *Science*, 304:1093-1095.

Snelling, W.M., Allan, M.F., Keele, J.W. *et al.* (2010). *J. Anim. Sci.*, (in press).

Watanabe, T., Hirano, T., Takano, Y. *et al.* (2008). *Anim. Genet.*, 39:374-382.