

Across-breed genomic prediction for body weight in Siberian cattle populations

Burak KARACAÖREN* 

Department of Animal Science, Akdeniz University, Antalya, Turkey

Received: 29.11.2019 • Accepted/Published Online: 23.03.2020 • Final Version: 02.06.2020

Abstract: Body weight (BW) is an important heritable phenotype and related to other functional and production traits in cattle. The past decade has seen an increase in emphasis on genome wide association studies (GWAS) for detecting single nucleotide polymorphisms (SNPs) that are associated with quantitative phenotypes. Prediction of phenotypes using across-breed GWAS information [genomic prediction (GP)] is an also important research area but received less attention from the community. Understanding the link between major genes and common ancestors within and between breeds will contribute to a deeper understanding of GP across breeds. The aims of the present study were two-fold: 1) to examine genetic structure and to detect associated SNPs for BW using various single and multiple locus genomic models and 2) genomic prediction of BW using Siberian cattle populations based on across-breed genomic information. The most obvious finding to emerge from this study was the increase in the across-GP accuracy when gene segregation in both related populations was found. These findings have significant implications for the understanding of the way in which common ancestors and/or the presence of quantitative trait loci might affect the accuracy of the GP results.

Keywords: Body weight, genomic selection, genome wide association analyses

1. Introduction

Body weight (BW) is an important heritable phenotype and related to other functional and production traits in cattle [1]. Previous studies have reported positive genetic correlations between milk yield and BW [2]. [1] predicted longitudinal genetic correlations between BW and other functional traits, including dry matter intake and milking speed over days. Heritability estimates of BW ranged from 0.32 to 0.61 [2,3] while estimates of BW change over the range from 0.10 to 0.34 [4,5]. Furthermore, BW plays critical role in fertility- and survival-related phenotypes [2].

The past decade has seen an increase in emphasis on genome-wide association studies (GWAS) for detecting single nucleotide polymorphisms (SNPs) associated with quantitative phenotypes. There is a growing body of literature that recognizes the importance of assumptions for underlying genetic structure of the phenotypes for the GWAS. The common model of choice, the single regression model (SRM) [6], assumes the existence of major gene(s) in association with phenotype. While an SRM is useful for detecting major genes, it does have drawbacks regarding multiple hypothesis testing, disregarded correlations among SNPs, and low power to detect genes with small effects [7]. [8] assumed a small correlation in gene effects for human height and was able to more thoroughly explain

the much higher genomic variance when compared with the SRM [9]. Since then, many other studies using this model or other multi-locus Bayesian models were done to investigate genetic structure of phenotypes [10–12].

Prediction of phenotypes using across-breed GWAS information [genomic prediction, (GP)] is an also important research area but has received less attention from the community [13]. Prediction accuracy will depend on genetic structure of the trait [14,15]; hence, the assumptions for the size of gene effects, number of genes, experiment/testing population sizes and linkage disequilibrium define which statistical model should be used in practice. Studies of [16] and [17] showed the importance of marker preselection and/or marker densities for degree of GP accuracy across breeds. It has been demonstrated by [18,19] that ancestral relationships within and between experiment (training) and test datasets result in different prediction accuracies in both within and between GPs. [20] studied the presence of effect of large quantitative trait locus (QTL) in the reference set for the across-breed GP. It can therefore be assumed that common ancestors and the presence of large quantitative trait locus (QTL) in the experiment and testing populations might lead to higher genomic prediction accuracy in across-breed GPs.

* Correspondence: burakkaracaoren@akdeniz.edu.tr

By using SRM and multiple hypothesis testing procedures in Siberian cattle populations, [21] identified several major QTLs associated with BW. Their findings further support the idea of common ancestors in Siberian cattle populations. Both populations were found to be related and were in Siberia for several decades [21]. Understanding the link between major genes and common ancestors within and between the breeds will contribute to a deeper understanding of GP across breeds. The main aim of the present study was two-fold: (1) to examine genetic structure and to detect associated SNPs for BW using various single and multiple locus genomic models and (2) genomic prediction of BW using Siberian cattle populations based on across-breed genomic information [21].

2. Material and methods

2.1. Materials

BW for 174 genotyped animals (150 Hereford with 92 dams and 58 bulls and 24 Kazakh with 4 dams and 20 bulls) from the Siberian regions were obtained from [21]. Single nucleotide polymorphisms (SNPs) were genotyped using a GGP HD150K array as described by [21]. Unannotated SNPs and sex chromosomes were removed. SNPs were removed from the data if the call rate was < 90%, minor allele frequency < 5%, and Hardy Weinberg equilibrium ($P < 10^{-6}$). Finally, 107,550 SNPs were collected for statistical analyses of the genotypes. Details of the published dataset can be found in [21].

2.2. Methods

[22] showed the genomic relationship between Kazakh and Hereford breeds. However, there are certain drawbacks associated with the use of admixed populations in GWAS. Due to the effect of common ancestors in the two populations, genetic stratification needs to be taken into account in order to avoid false positive results. We used a linear mixed model to take into account the effects of the admixture as was implemented in GenABEL [23] using GWAS with a mixed model and regression (GRAMMAR-Gamma) [23,24] approach in R software [25].

The linear mixed model can be expressed by the equation:

$$y = Xb + Za + e \quad (1)$$

in which y contains the number of observations, b is the breed and square root of age effects, a is the additive genetic effect, matrices X and Z are incidence matrices, and e is a vector containing residuals.

$$\text{Var} \begin{pmatrix} a \\ e \end{pmatrix} \sim N \left[\mathbf{0}; \begin{pmatrix} \mathbf{A}\sigma_a^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_e^2 \end{pmatrix} \right]$$

For the random effects, it is assumed that A is the coefficient of coancestry obtained from genotype of animals, I is an identity matrix, and σ_a^2 and σ_e^2 are the

additive genetic and residual variances, respectively. The use of genomic principal components for correcting a population stratification has a relatively long tradition in GWAS. [26] suggested using principal components for detection and correction of population admixture in a linear mixed model (1). We used the approach of [26] for GWAS as was implemented in GenABEL [23] based on the highest genomic loadings. However, there are certain drawbacks associated with the use of GWAS p values without correcting for the number of hypothesis tests. The main disadvantage of a huge number of hypothesis is an inflated number of false positive genomic signals [9]. One advantage of the false discovery rate (FDR) approach is avoidance of the problem of false positive genomic signals by increasing significance levels to 0.05/number of SNPs. However, interval-based methods can be more useful for identifying and taking into account linkage disequilibrium for GWAS p values over correlated SNPs [27]. In his article, [27] described a truncated product method (TPM) for combining test statistics over chromosomal locations of SNPs using different window sizes.

[28] used sparse and larger variances to model SNPs effects as “Bayesian sparse linear mixed models” (BSLMM). [28] used a mixture of two normal distributions and additional random effects to yield a more flexible model compared with other Bayesian models.

We used BSLMM for prediction of SNP effects:

$$y_i = \text{breed} + \text{age} + \sum_{j=1}^n (z_{ij}a_j\delta_j) + e_i \quad (2)$$

in which y_i is the phenotypes of the i th animal, z_{ij} is an indicator variable (small or major effects from the two normal distributions) for the i th animal, j th SNP locus and k th allele, a_j is marker of locus effects, δ_j indicates whether SNP has an effect (or not), and e_i is the residual for animal i . To determine whether the various assumptions regarding genetic structure of the body weight gave different results, the number of mixtures were increased. BayesR [29] assumed a mixture of four normal distributions for predictions of SNP effects (assumed to be 0.00001, 0.0001, 0.001, and 0.01 genetic variances) in model (2). For each phenotype, the Markov Chain Monte Carlo (MCMC) algorithm was run for 1.000.000 samples, and the first 2000 samples were discarded as burn in period. We collected each tenth sample from each realization of the MCMC as the thinning period.

One of the most well-known models for assessing polygenic effects in GP is the use of a genomic relationship matrix in (1) in which a refers to animals termed as genomic best linear unbiased prediction (GBLUP) [30]. We used GBLUP, BayesR, and BSLMM for prediction of Kazakh phenotypes using Hereford genotypes based on their breeding values (BV) or small gen effects (ALPHA) [28]. The whole genomic dataset was partitioned by Hereford

and Kazakh sets (reference and validation, respectively). BW measurements of the Kazakh breed in the validation set were assumed to be missing. Phenotypes of the Kazakh breed were predicted with the information from the Hereford breed in the reference set. A random sample of reference set (2/3 of the Hereford animals, $n = 105$) was used to create predictive equations. This procedure was repeated 10 times. The correlation coefficient between the predicted and realized phenotypes of the Kazakh animals was calculated over 10 replications.

3. Results

The mean autosomal heterozygosity for SNPs was estimated as 0.3877 (0.1109). Mean identity by state for the animals was estimated as 0.3890 (0.0197). Genomic heritability was found to be 0.2115.

The first set of investigations aimed to interrogate SNP existence in association with BW by single SNP regression models. A major problem with the GWAS is the presence of genotypic clusters (or population stratification). The results obtained from the genomic principal component analysis of genotypes detected breed specific clusters (results are not shown). We used genomic relationship matrix (GRAMMAR-Gamma), genomic principal components (EGSCORE), and a Bayesian linear mixed model (BSLMM) for correcting population stratification in regression models. Tables 1–3 show an overview of the association results based on the genomic models. After taking the multiple hypothesis correction due to huge number of hypothesis in GWAS into account, FDR-corrected P values were also calculated (results not shown). No SNPs were significant after FDR correction.

The second aim of this study was to investigate the usefulness of GP with additional information from ancestrally-related Siberian cattle populations and presence of major genes for BW. Recent developments in

the field of GP have led to a renewed interest in genetic structure of the phenotypes. A much-debated hypothesis is whether BW could be explained by genes with small and/or major genes in GP. To this end, we used different GP models for taking into account various genetic structures for BW (Table 4). Table 4 compares Pearson's correlation coefficients for within and between GP using different models. On average, within-breed GPs were shown to have low correlations coefficients (Table 4). A closer inspection of Table 4 shows that GBLUP_ALPHA predicted the highest correlation of 0.3373 for the across-breed GP.

4. Discussion

The first set of analyses examined the impact of major genes on BW using single SNP regression models. The results of the GRAMMAR, EGSCORE, and BSLMM analyses are presented in Tables 1–3. A comparison of the Tables 1–3 reveals that statistically significant results were obtained from chromosomes 1, 2, 5, 12, and 20. Strong genomic signals from chromosome 5 were found using all of the GWAS models. Interestingly, a significant genomic signal on chromosome 1 was observed based on the results of the Bayesian model (Table 3). Prior studies have noted the importance of multiple hypothesis testing procedures in GWAS for reducing false positive findings [7,9]. A strong relationship between number of hypothesis (SNPs) and false positives has been reported in the literature [31]. All of the SNPs in Tables 1–3 turned out to be insignificant after applying the FDR control ($P > 0.05$). One of the most significant current discussions in multiple hypothesis testing correction is the linkage disequilibrium effects of adjacent SNPs on the test statistics [32]. We used TPM to correct for dependency among adjacent SNPs over genomic locations based on windows sizes of 2, 4, and 6. Strong evidence of genomic signaling from chromosome 5 at the vicinity of base pair 106,987,567 was found by

Table 1. Summary of GRAMMAR-Gamma model.

SNP	Chromosome	Position	Chi-square	P-value
Hapmap38028-BTA-122265	12	2058297	17.5334	2.82E-05
BovineHD2000011846	20	41207894	15.9122	6.63E-05
BovineHD0500030494	5	106294449	15.3161	9.09E-05
ARS-BFGL-NGS-106674	5	106296860	15.3161	9.09E-05
BovineHD0500030342	5	105647645	15.0857	0.000103
BovineHD0500030763	5	106987567	14.8007	0.000119
BovineHD0500000856	5	3365443	14.7395	0.000123
BovineHD4100004048	5	106888999	14.7118	0.000125
ARS-BFGL-NGS-107504	2	27505377	14.6287	0.000131
BovineHD0500030501	5	106329896	14.5855	0.000134

Table 2. Summary of EGSCORE model.

SNP	Chromosome	Position	Chi-square	P-value
Hapmap38028-BTA-122265	12	2058297	18.8547	1.41E-05
BovineHD2000011846	20	41207894	18.2927	1.89E-05
BovineHD0500030711	5	106905471	17.8554	2.38E-05
BovineHD0500030727	5	106938388	17.8554	2.38E-05
BovineHD0500030728	5	106939259	17.8554	2.38E-05
BovineHD0500030730	5	106941930	17.8554	2.38E-05
BovineHD0500030494	5	106294449	17.7094	2.57E-05
ARS-BFGL-NGS-106674	5	106296860	17.7094	2.57E-05
BovineHD0500030693	5	106857445	17.6240	2.69E-05
BovineHD0500030763	5	106987567	17.2323	3.31E-05

Table 3. Summary of BSLMM model.

SNP	Chromosome	Position	Regression coefficient	P-value
Hapmap38028-BTA-122265	12	2058297	36.0748	5.78E-06
BovineHD2000011846	20	41207894	-74.2981	4.2E-05
ARS-BFGL-NGS-107504	2	27505377	40.6875	5.15E-05
BovineHD0500030494	5	106294449	52.8970	5.66E-05
ARS-BFGL-NGS-106674	5	106296860	52.8970	5.66E-05
BovineHD0500030342	5	105647645	45.4067	7.26E-05
BovineHD0500000856	5	3365443	-34.050	7.55E-05
BovineHD0500030763	5	106987567	40.6463	7.93E-05
Hapmap54019-rs29023016	1	156538750	-41.541	8.12E-05
BovineHD0500030501	5	106329896	51.3744	9.29E-05

TPM at windows size of 2 [$-\log(P)=6.48e^{-10}$]. This finding is consistent with those of other studies that reported a weight gain gene *CCND2* at the same genomic location [2,21,33]. Estimation of the proportion of genetic variance explained by detected SNPs found to be around 13% based on the Bayesian model.

To assess within-Hereford breed prediction accuracy, a 10-fold cross validation was used. Table 4 presents the correlation among the four different GP methods. Overall, cross validation found differences in correlations within GP due to small sampling size (range from -0.00204 to 0.2230 for BSLMM_1DD, for instance). In Table 4, we can see that BSLMM_2alfa resulted in the highest correlation values of 0.0614 for within-breed GP. On average, Bayesian models were shown to have a higher correlation when compared with the other models for within-breed GP (Table 4). These results reflect those of [34], who also found that 0.06 correlations of Hereford within GP using

yearling weight. However, our results differ from that of [34]'s estimate of within-GP Hereford correlation of 0.42 obtained for weaning weight. A possible explanation for this discrepancy might be due to differences among genetic structure of the phenotypes between current and studies by [34]. In addition possible sources of sampling error (due to small sampling size) could also have affected the results of the current study.

Across-GP results in Table 4 are revealing in several ways. It can be seen in Table 4 that across-breed GPs resulted in much higher GPs when compared with the within-breed GP. On average GBLUP models were shown to have higher correlations compared with the other models for the across-breed GP. It has previously been observed that across-breed GP with GBLUP accuracies was around 0 [17]. Interestingly, Table 4 shows that the GBLUP_alpha resulted in the highest accuracy (0.3373). These results corroborate the ideas of [20, 35] in which

Table 4. Pearson correlations of genomic predictions obtained by different models.

Method	Within breed genomic prediction	Across breed genomic prediction
BayesR	0.0574	0.2589
BSLMM_BV	0.0482	0.3138
GBLUP_BV	0.0297	0.3359
BSLMM_ALPHA	0.0614	0.3251
GBLUP_ALPHA	0.0298	0.3373

the presence and use of QTL information in across-breed GP was demonstrated. In addition, comparisons of the findings with others [18,19] confirm the effects of common ancestors on the across-breed GP.

References

- Karacaören B, Jaffrézic F, Kadarmideen HN. Genetic parameters for functional traits in dairy cattle from daily random regression models. *Journal of Dairy Science* 2016; 89 (2): 791-798. doi: 10.3168/jds.S0022-0302(06)72141-5
- Yin T, König S. Genome-wide associations and detection of potential candidate genes for direct genetic and maternal genetic effects influencing dairy cattle body weight at different ages. *Genetics Selection Evolution* 2019; 51: doi: 10.1186/s12711-018-0444-4
- Koenen EPC, Veerkamp RF. Genetic covariance functions for live weight, condition score, and dry-matter intake measured at different lactation stages of Holstein Friesian heifers. *Livestock Production Science* 1998; 57: 67-77. doi: 10.1016/S0301-6226(98)00159-6
- Veerkamp RF. Selection for economic efficiency of dairy cattle using information on live weight and feed intake: a review. *Journal of Dairy Science* 1998; 81 (4): 1109-1119. doi: 10.3168/jds.S0022-0302(98)75673-5
- Veerkamp RF, Oldenbroek JK, Van Der Gaast HJ, Van Der Werf JHJ. Genetic correlation between days until start of luteal activity and milk yield, energy balance, and live weights. *Journal of Dairy Science*, 2000; 83: 577-583 doi: 10.3168/jds.S0022-0302(00)74917-4
- Yi H, Breheny P, Imam N, Liu Y, Hoeschele I. Penalized multimarker vs. single-marker regression methods for genome-wide association studies of quantitative traits. *Genetics* 2015; 199 (1): 205-222. doi: 10.1534/genetics.114.167817
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA et al. Finding the missing heritability of complex diseases. *Nature* 2009; 461 (7265): 747-753. doi: 10.1038/nature08494
- Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* 2011; 43 (6): 519-25. doi: 10.1038/ng.823
- Tam V, Patel N, Turcotte M, Bossé Y, Paré G et al. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics* 2019; 20 (8): 467-484. doi: 10.1038/s41576-019-0127-1
- Campos G, Gianola D, Allison DB. Predicting genetic predisposition in human: the promise of whole-genome markers. *Nature Reviews Genetics* 2010; 11: 880-886. doi: 10.1038/nrg2898
- Lello L, Avery SG, Tellier L, Vazquez AI, de los Campos G, Hsu, SD. Accurate genomic prediction of human height. *Genetics* 2018; 210 (2): 477-497. doi: 10.1534/genetics.118.301267
- Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ, UCLEB Consortium. Reevaluation of SNP heritability in complex human traits. *Nature Genetics* 2017; 49 (7): 986-992. doi: 10.1038/ng.3865
- De Roos AP, Hayes BJ, Goddard ME. Reliability of genomic predictions across multiple populations. *Genetics* 2009; 183 (4): 1545-1553. doi: 10.1534/genetics.109.104935
- Ober U, Ayroles JF, Stone EA, Richards S, Zhu D et al. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genetics* 2012; 8 (5): e1002685. doi: 10.1371/journal.pgen.1002685
- Daetwyler HD, Calus MP, Pong-Wong R, de los Campos G, Hickey JM. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 2013; 193 (2): 347-365. doi: 10.1534/genetics.112.147983
- Iheshiulor OO, Woolliams JA, Yu X, Wellmann R, Meuwissen TH. Within-and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *Genetics Selection Evolution* 2016; 48 (1): 15. doi: 10.1186/s12711-016-0193-1

The main goal of the current study was to determine the effects of the QTL and presence of common ancestors for the within- and across-GP BW in cattle. The most obvious finding to emerge from this study is the increase in across-GP accuracy when QTL segregation in both correlated populations occurs. These findings have significant implications for the understanding of the way in which common ancestors and/or presence of QTL might affect the accuracy of the GP results.

Acknowledgement

This work was supported by Akdeniz University, Scientific Research Projects Unit (BAP) under project number FBA-2019-5016.

Conflict of Interest

The authors are declared that there is no conflict of interest.

17. Raymond B, Bouwman AC, Wientjes YC, Schrooten C, Houwing-Duistermaat J et al. Genomic prediction for numerically small breeds, using models with pre-selected and differentially weighted markers. *Genetics Selection Evolution* 2018; 50 (1): 49. doi: org/10.1186/s12711-018-0419-5
18. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density Ö Msingle nucleotide polymorphism panels. *Journal of Dairy Science* 2012; 95 (7): 4114-4129. doi.:10.3168/jds.2011-5019
19. Calus MP, Huang H, Vereijken A, Visscher J, Ten Napel J, Windig JJ. Genomic prediction based on data from three layer lines: a comparison between linear methods. *Genetics Selection Evolution* 2014; 46 (1): 57. doi: 10.1186/s12711-014-0057-5
20. Hoze C, Fritz S, Phocas F, Boichard D, Ducrocq V, Croiseau P. Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. *Journal of Dairy Science* 2014; 97 (6): 3918-3929. doi: 10.3168/jds.2013-7761
21. Igoshin AV, Yudin NS, Belonogova NM, Larkin DM. Genome-wide association study for body weight in cattle populations from Siberia. *Animal Genetics* 2019; 50(3): 250-253. doi: 10.1111/age.12786.
22. Yurchenko A, Yudin N, Aitnazarov R, Plyusnina A, Brukhin V, et al. Genome-wide genotyping uncovers genetic profiles and history of the Russian cattle breeds. *Heredity* 2018; 120 (2): 125-137. doi: 10.1038/s41437-017-0024-3
23. Aulchenko YS, De Koning DJ, Haley C. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 2007; 177: 577-585. doi.: 0.1534/genetics.107.075614
24. Svishcheva GR, Axenovich TI, Belonogova NM, van Duijn CM, Aulchenko YS. Rapid variance components-based method for whole-genome association analysis. *Nature Genetics* 2012; 44:1166-1170. doi: 10.1038/ng.2410
25. R Development Core Team. R: A language and environmental for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013.
26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 2006; 38: 904-909. doi: 10.1038 / ngl847
27. Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS. Truncated product method for combining P-values. *Genetic Epidemiology* 2002; 22, 170-185. doi: 10.1002/gepi.0042
28. Zhou X, Carbonetto P, Stephens M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics* 2013; 9 (2): e1003264. doi: 10.1371/journal.pgen.1003264
29. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM. Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genetics* 2015; 11 (4): e1004969. doi: 10.1371/ journal.pgen.1004969
30. Habier D, Fernando RL, Garrick DJ, Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* 2013; 194: 597-607. doi: 10.1534/genetics.113.152207
31. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nature Reviews Genetics* 2014; 15 (5): 335-346. doi: 10.1038/nrg3706
32. Hu X, Zhang W, Zhang S, Ma S, Li Q. Group-combined P-values with applications to genetic association studies. *Bioinformatics* 2016; 32(18): 2737-2743. doi: 10.1093/bioinformatics/btw314
33. Seabury C M, Oldeschulte DL, Saatchi M, Beever JE, Decker JE et al. Genome-wide association study for feed efficiency and growth traits in US beef cattle. *BMC Genomics* 2017; 18 (1): 386. doi: 10.1186/s12864-017-3754-y
34. Kachman SD, Spangler ML, Bennett GL, Hanford KJ, Kuehn LA et al. Comparison of molecular breeding values based on within-and across-breed training in beef cattle. *Genetics Selection Evolution* 2013; 45 (1): 30. doi: 10.1186/1297-9686-45-30
35. Lee J, Kim JM, Garrick DJ. Increasing the accuracy of genomic prediction in pure-bred Limousin beef cattle by including cross-bred Limousin data and accounting for an F94L variant in MSTN. *Animal Genetics* 2019; 50 (6): 621-633. doi: 10.1111/ age.12846 [Epub ahead of print].