final report

Project code: B.NBP.0762

Prepared by: Matthew Kelly, Andrew Swan,
Marina Fortes, Michel Farah
University of Queensland

Date published: November 2013

ISBN: 9781740362740

PUBLISHED BY
Meat & Livestock Australia Limited
Locked Bag 991
NORTH SYDNEY NSW 2059

Improved methods for genotypic data analysis

- Accuracy of genomic selection for age at puberty in a multi breed population of tropically adapted beef cattle

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Abstract

Genomic selection is becoming a standard tool in livestock breeding programs, particularly for traits that are hard to measure. Accuracy of genomic selection can be improved by increasing quantity and quality of data and potentially by improving analytical methods. Adding genotypes and phenotypes from additional breeds or crosses often improves the accuracy of genomic predictions, but will require specific methodology. A method was developed to incorporate breed composition estimated from genotypes into genomic selection models. This method was applied to age-at-puberty data (as estimated from age at first observation of a *corpus luteum*) from a mix of Brahman and Tropical Composite beef cattle. In this data set the new model incorporating breed composition did not increase the accuracy of genomic selection. However the breeding values exhibited slightly less bias (as assessed by deviation of regression of phenotype and genomic breeding values from the expected value of 1). Adding additional Brahman animals to the Tropical Composite analysis increased the accuracy of genomic predictions and did not affect the accuracy of the Brahman predictions.

Executive summary

Genomic selection is becoming a standard tool in livestock breeding programs, particularly for traits that are hard to measure. Accuracy of genomic selection can be improved by increasing the amount and quality of phenotypic data used and potentially by improving analytical methods. Adding genotypes and phenotypes from additional breeds or crosses can, in theory, be used to improve the accuracy of genomic predictions but will require specific methodology. This is particularly pertinent for smaller breed groups such as the tropical composite breeds used in Northern Australia where limited data are captured compared to Brahmans. The primary aim of this study was to investigate the possibility of increasing the accuracy of genomic selection for Tropical Composites by using data from Brahmans. To better combine information from the two breed groups, the breed composition (content of *Bos indicus* genes) was used to model population structure, including in the computation of the genomic relationship matrix (GRM) used to estimate genomic breeding values (GEBV).

Age at the first *corpus luteum* (AGECL, days) was the measurement used to define age at puberty in heifers in a previous study - Brahman (BB, n = 980) and Tropical Composite (TC, n = 1074) heifers with AGECL measurements were genotyped and imputed such that all animals had genotype estimates for the Illumina High-Density Bovine BeadChip (BovineHD) (containing approximately ~770 thousand SNP).

The Brahman percentage (BB%) content of individual animals was estimated in the current study using a reference population that included 1000 Brahman animals and 2000 *Bos taurus* animals. The resulting phenotypic and genotypic data were used to estimate genomic breeding values using models with and without adjustments for differences in phenotypic variance between breeds, adding BB% as a fixed effect, and adjusting the GRM for differences in BB%. Cross validation was used to estimate realised accuracies with and without these adjustments in a factorial approach.

Adding Brahman animals to the Tropical Composite analysis increased the accuracy of genomic selection. Each group added to the analysis increased the accuracy, although the magnitude diminished with each additional group of Brahmans added. Adjusting the genomic relationship using Brahman percentage predicted from genotypes was tested as a potential way to improve genomic selection accuracy and reduce bias. There was no increase in accuracy associated with the GRM adjusted for Brahman content. However the regression between adjusted phenotype and GEBV was slightly less variable when BB% was used to adjust the GRM. Thus adjusting GRM for breed composition may lead to more robust genomic predictions. Adding genomic predictions of Brahman content as a fixed effect in the genetic analysis in most cases did change the accuracy of genomic selection. The exception was the accuracy of genomic selection when predicting Brahman from Tropical Composite data alone.

In conclusion, for breeds with limited data availability it may be possible to combine data with other related breeds to increase the accuracy of genomic

selection. For the data set used in this study the addition of Brahman percentage as a fixed effect in genetic evaluation and adjusting the Genomic relationship matrix for breed composition had little impact on the accuracy of genomic predictions. The slope of the relationship slope between genomic predictions and phenotype was quite different in the Brahmans and Tropical Composites. Thus care must be taken when combining data from different breeds. Accounting for breed in the genetic evaluation made the predictions more stable than not accounting for this effect, particularly in the Tropical Composites.

Table of contents

List of Tables

List of Figures

1. Background

Improved genomic selection for fertility and other economically important traits for beef production will be reliant on the availability of genotyped reference populations with accurate phenotypes, and the development of better analytical methods. There is a need to test alternative methods of genomic prediction and estimation of individual gene effects, given the multi-breed scenario that is typical of the beef industry in Northern Australia. Most of the methods used to date have been based on those implemented in the dairy industry, and have therefore been developed and tested within a single breed (Holstein). The Australian beef industry in contrast consists of a mix of breeds, especially in tropical regions where adaption traits are important and animals with varying degrees of *Bos indicus* genetics are widely used (Bolormaa et al., 2013b; Burrow, 2012). Tropical Composite is a term used to define a breed that is a stable cross of Zebu (*Bos indicus*) and Taurine (*Bos taurus*) breeds, which is prominent in in Northern Australia (Burns et al., 2013; Prayaga et al., 2009). Recent studies have analysed Tropical Composite cattle and have considered them to be a single population. Alternative prediction methods have been proposed for use in multibreed dairy cattle populations (Erbe et al., 2012; Harris and Johnson, 2010; Olson et al., 2012). These methods were shown to increase accuracy of genomic selection for Jerseys where additional Holstein data were added to the analysis. Both studies also suggest that the methods could be further modified to account for crossbred animals. Accordingly methods have been proposed by Harris and Johnson (2010) that will accommodate both multiple breeds and crosses.

Because of the complexity of multibreed populations, there is increased potential for biases in genomic breeding values if models do not account for breed of origin (Misztal et al., 2013). Better understanding of the factors that degrade predictive power in multibreed populations is necessary in order to increase the accuracy of estimated genomic breeding values. Therefore, this project had two objectives, firstly to understand the breed composition of the Tropical Composite beef cattle population of Northern Australia, and secondly to develop genomic prediction methods to model the diverse nature of the population.

Genome assisted breeding values have been utilised heavily in the Dairy industry and have been available for Australian Angus, a temperate beef breed, for a number of a number of years. However the key barrier to adoption of this technology with potentially high value to northern breeds of cattle is the accuracy of prediction and cost. This project addressed key issues affecting the accuracy of this technology in the north, namely the ability to develop accurate genomic breeding values in cross-bred cattle (Tropical composites). Its objectives were to:

1. Improve the value of genomic selection due to more genetic variation explained, and/or reduction in biases caused by complex breed structure of tropical composites;

2. Provide better understanding of the factors that degrade predictive power in a multi-breed population particularly within the tropical composites; and

2. Methods

Phenotype and genotype data

The trait used for this study was age at the first *corpus luteum* (AGECL, days) recorded on 2054 genotyped females that consisted of Brahman (BB, n=980) and Tropical Composite breeds (TC, n=1074). AGECL is used as an indicator of the age at puberty in beef cattle. These cattle represent a subset of the population established by the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC). This population and its phenotypes have been described in detail previously (Barwick et al., 2009; Burns et al., 2013; Hawken et al., 2012b; Johnston et al., 2009). A key feature of the population structure relevant to our study is that the Tropical Composite animals used were formed by crossing *Bos indicus* (Brahman) and *Bos taurus* breeds. The relative contribution from genes of each group (*Bos indicus* and *Bos taurus*) was established for the Tropical Composite animals in our study, and used as a central component of the analyses.

All individuals have high density SNP genotypes available, either directly genotyped or imputed from lower density genotypes. Animals were genotyped using the Illumina Bovine SNP50 bead chip (Matukumalli et al., 2009) version 1 (containing approximately 50,000 SNP). Imputation was performed using a reference set of 917 animals genotyped with the high density BovineHD. The imputation was performed using BEAGLE and the methods, number of animals used and accuracy is described in detail in Bolormaa et al.(2013a). All SNP chips were processed according to the manufacturer’s protocols. Repeated samples were included in the genotyping for quality assurance, and Bead Studio software (Illumina, Inc.) was used to determine genotype calls. Quality control analysis methods and results have been reported previously (Hawken et al., 2012a).

Genomic analysis methods

Genomic breeding values were estimated using GBLUP, based on the following general mixed model:

$$y = X\beta + Zu + e$$

were *y* is the vector of AGECL phenotypes; *X* is an incidence matrix for fixed effects; β is a vector of fixed effects; *Z* is an incidence matrix for genomic breeding values; *u* is a vector of random genomic breeding values for each animal ($Var(u) = G\sigma_u^2$where G is a genomic relationship matrix as described below and σ_u^2 is the variance of genomic breeding values), and *e* is a vector of residual random effects ($Var(e) = I\sigma_e^2$ where I is an identity matrix and $\sigma_e^2$is the residual variance).

The model was fitted with one of two genomic relationship matrices, genomic relationships using allele frequencies calculated as a single breed group GRM_{SB} and GRM_{XB} with allele frequency adjusted for breed, for the 2054 recorded and genotyped females. The GRMs were calculated following an adaptation of the methods described by (Harris and Johnson, 2010; VanRaden et al., 2011; Yang et al., 2010):

$$Gr = \frac{(WW')}{n}$$

where $W = M - 2P$, in which M is the $n \times m$ matrix of genotypes for $n = 2054$ animals and m SNP, with values of 0 for the homozygous genotype of the first allele, 1 for the heterozygous genotype, and 2 for the homozygous genotype of the second allele. P is the $m \times n$ matrix containing the frequencies of the second allele of each SNP (p_i) expressed as the frequency multiplied by 2.

For GRM_{SB}, allele frequencies for each SNP in P were calculated from the group of 2054 analysis females, irrespective of breed. Therefore, rows of P are the same for all animals.

For GRM_{XB}, P was calculated as QC, where Q is a $n \times 2$ matrix describing the fraction of genes of Brahman and *Bos taurus* origin (columns) for each of the 2054 analysis animals (rows). Each row of Q sums to 1. C is a $2 \times m$ matrix containing the allele frequencies of each SNP (columns) for BB and *Bos taurus* populations (rows). Both Q and C were derived from analyses using the software package Admixture (Alexander and Lange, 2011; Alexander et al., 2009)**,** as described below. Apart from the multibreed formulation of QC a key difference between GRM_{XB} and GRM_{SB} is that allele frequencies in GRM_{XB} were estimated in the Admixture analysis from animals of known breed not including the analysis animals, whereas allele frequencies in GRM_{SB} were estimated directly from the analysis animals. Harris and Johnson (2010) describe a similar method for deriving a multibreed GRM, although in their study the breed fractions (Q) were derived from pedigree rather than genomic information.

Such genomic relationships matrices are positive semi-definite, and often singular (Forni et al., 2011). So, to enable inversion, genomic relationship matrices were weighted following (VanRaden, 2008):

$$G = wGr + (1 - w)A_{22},$$

where, G is the final genomic relationship matrix to be used in the analysis; Gr is the initial genomic relationship matrix as described above and based only on genotypic information, *w* is a weighting factor equal to 0.95 (Aguilar et al., 2010); and $A_{22}$is the subset of the pedigree based numerator relationship matrix for the genotyped females in the analysis.

Estimation of Brahman content

The Brahman and *Bos taurus* contents (Q) for each animal were estimated using a supervised Admixture analysis as described previously in (Alexander and Lange, 2011; Alexander et al., 2009). The dataset used to estimate *Brahman* content (BB%) consisted of training animals from six *Bos taurus* breeds (Angus, Murray Grey, Charolais, Hereford, and Shorthorn) with 2000, 200, 400, 500 and 500 cattle respectively. The *Bos indicus* training set included 2000 Brahman cattle. Both groups are part of the same Beef CRC experimental population, but excluded the 2054 analysis females used in this study. To obtain the estimates of breed content required for Q the analysis females were added to the Admixture analysis with their breed masked. The analysis was performed considering the 6 *Bos taurus* breeds as a single breed, and compared with the Brahman animals. Thus the number of breeds (the 'k' parameter) in Admixture was set to 2, and all other parameters set to their default values (Alexander and Lange, 2011; Alexander et al., 2009).

Estimation of genomic breeding values

Variance components for σ_u^2 and σ_e^2 used in GBLUP analyses were estimated by restricted maximum likelihood (REML) using the Wombat software package (Meyer, 2007). The variance estimates used in GBLUP were calculated based on all animals with phenotype and genotype data using an animal model fitted with the inverse of the pedigree based numerator relationship matrix. Fixed effects fitted included cohort (year of birth and farm), origin (O), month of birth (BM), sire breed (Sg), dam breed (Dg) and the interactions between BM*O, cohort*O, Sg*Dg, BM*Sg. The estimate of Brahman percentage(BB%) obtained form the admixture analysis was also included in the model as a covariate. Variance estimates from these models are presented in Table 4 and were used in the estimation of breeding values for the GBLUP cross validation analysis. The GBLUP analyses were also fitted in Wombat using the same fixed effects and the two GRM previously described (GRM_{SB} and GRM_{XB}).

Scenarios tested

Cross validation was used to evaluate the impact of data and model factors on accuracy and bias of genomic evaluation. To study the impact of data on Tropical Composite predictions, increasing numbers of records on Brahman females were added to the analyses. The model factors studied were: fitting GRM_{SB} compared to GRM_{XB}, fitting Brahman content (BB%) as a covariate, and pre-adjustment of data by breed to the same phenotypic variance.

A series of cross validation analysis were performed to estimate the effect of each of the four factors on accuracy and bias of genomic predictions. Cross validation groups were formed within each breed group (Brahmans and Tropical Composites) by randomly selecting sire families into one of four groups, stratified by number of sibs with genotypes to ensure reasonably similar sized groups.

The cross validation strategies are described in Table 1. Standard cross validation where one of the four groups was omitted from the analysis to use as a validation group was performed within Brahman and Tropical Composites (denoted 3BB and 3TC respectively). A series of cross validation analysis was then run where additional groups were added to the Tropical Composite cross validation. In each case all possible combinations of BB groups were run in cross validation. At the end of the analysis for each of the cross validation runs the correlation between adjusted phenotype and GEBV was estimated for animals that were not included in training the model for each combination. The mean correlation and regression was then estimated from the group estimates.

Table 1 Example of cross validation strategy used for each scenario examined. All possible combinations of groups were run within BB when < 3 groups were included in training

Training Strategy Name	TC groups				Number of TC CV groups	BB groups				Number of BB CV groups	Number of analyses
	1	2	3	4		1	2	3	4		
0TC+3BB	V	V	V	V	0	T	T	T	V	3	4
3TC	T	T	T	V	3	V	V	V	V	0	4
3TC+1BB	T	T	T	V	3	T	V	V	V	1	4*4=16
3TC+2BB	T	T	T	V	3	T	T	V	V	2	4*6=24
3TC+3BB	T	T	T	V	3	T	T	T	V	3	4*4=16
3TC+4BB	T	T	T	V	3	T	T	T	T	4	4

3. Results

Estimation of Brahman content (BB%)

Figure 1 represents the absolute value of the difference in allele frequency between Brahman (BB) and *Bos taurus* (BT). The smaller difference between the frequencies show similarity between the frequencies in both population. This Figure shows that a high proportion of SNP have similar frequencies in both Brahmans and *Bos taurus* .

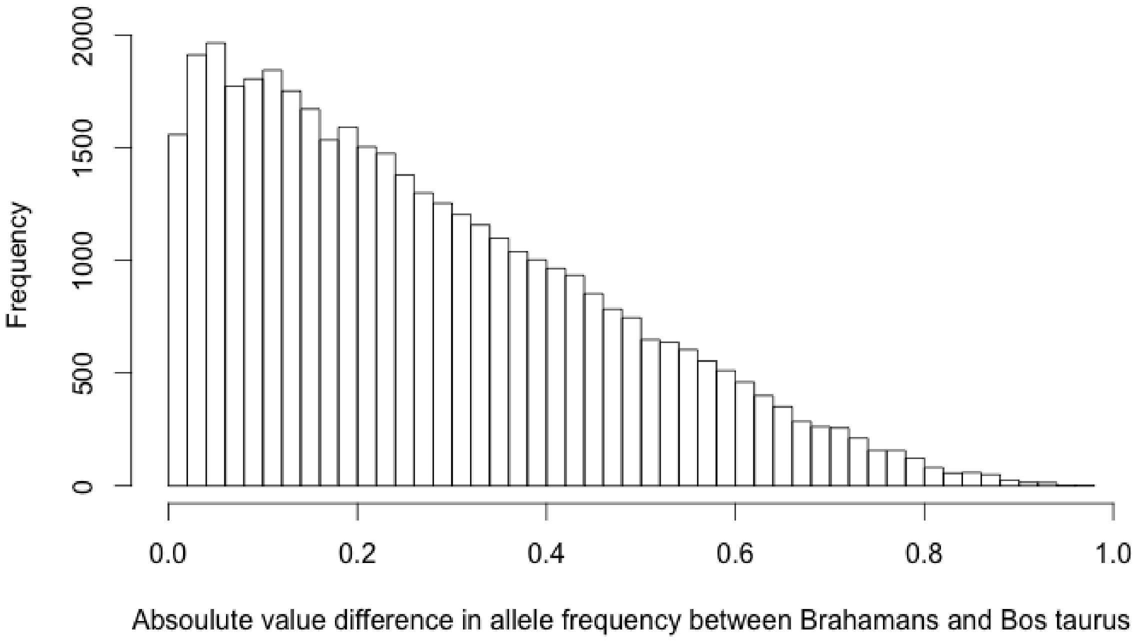

Figure 1 Histogram of the absolute value of the difference in allele frequency between Brahman (BB%) and Bos Taurus (BT) for individual SNP (calculated across 6 BT breeds)

The proportion of BB% and BT% in all animals was estimated using the Admixture software package on a reference population of 2000 Brahman and 3650 *Bos taurus* cattle. For the animals included in training the estimated breed proportions were fixed at 1 for their respective breeds (Table 2). The estimated BB% of Brahman and *Bos taurus* animals not included in training was slightly lower with averages of 0.974 and 0.002 respectively. The average BB% of Tropical Composite animals was 0.41, but the estimated proportions for individual animals covered a wide range (Figure 2).

Table 2 Average and standard deviation of estimated Brahman content results from Admixture, training animals were used in development of predictions and testing animals were excluded from training analysis**.**

Training			
Population	Mean	SD	N
BB	1.000	0.000	2000
BT	0.000	0.000	3650
Testing			
Population	Mean	SD	N
BB	0.974	0.048	3045
BT	0.002	0.011	1435
TC	0.412	0.086	1788

BB is a Brahman population; BT is a *Bos taurus* population; TC is a Tropical Composite population; SD is a Standard Deviation; and N is the total of animals used in Admixture

* NRM – Pedigree based relationship matrix; GRM_{SB} elements adjusted by average allele frequency of the dataset - ; GRM_{XB} elements of the GRM adjusted by individual animals breed proportion thus including breed allele frequencies- .

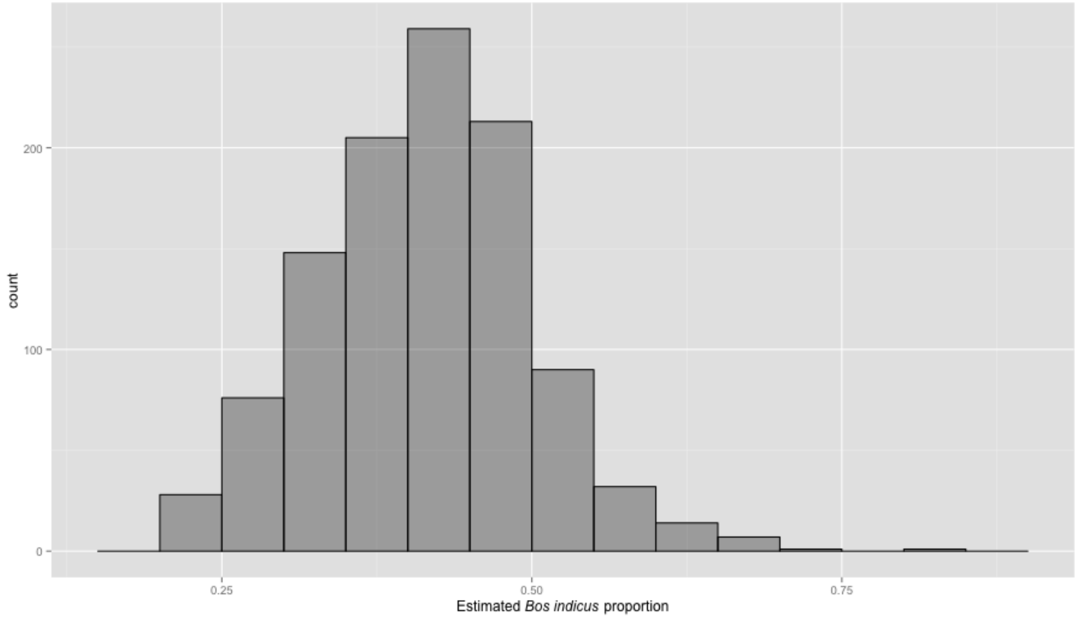

Figure 2 Histogram demonstrating the diversity of Brahman proportion estimates within Tropical Composite beef cattle.

Comparison of different GRM methods

Statistics of relationship coefficients are represented in Table 3. For the diagonal elements both genomic matrices (GRM_{SB}, GRM_{XB}) were similar and were smaller than the pedigree relationship matrix (NRM). The variances of these elements were very small (close to zero) for all matrices. The off-diagonals were impacted by the different GRM methods. The average, minimum and maximum off-diagonal was smaller when allele frequencies were adjusted for breed composition (GRM_{xb}) in both the Tropical Composites and the Brahmans. The off diagonals linking BB and TC animals were increased slightly by adjusting for breed composition.

Table 3 Statistics of relationship coefficients for Brahman (BB), Tropical Composite (TC), between Brahman and Tropical Composite (BBTC) and all the population (FULL) using pedigree and genomic information

Diagonal

	FULL			BB			TC			BBTC		
	NRM	GRM_{SB}	GRM_{XB}	NRM	GRM_{SB}	GRM_{XB}	NRM	GRM_{SB}	GRM_{XB}	NRM	GRM_{SB}	GRM_{XB}
Average	1.002	0.766	0.760	1.002	0.796	0.789	1.001	0.738	0.734	-	-	-
Min.	1.000	0.689	0.692	1.000	0.742	0.741	1.000	0.689	0.692	-	-	-
Max.	1.266	0.899	0.888	1.266	0.899	0.888	1.158	0.864	0.861	-	-	-
Var.	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	-	-	-

Off-Diagonal

	FULL			BB			TC			BBTC		
	NRM	GRM_{SB}	GRM_{XB}	NRM	GRM_{SB}	GRM_{XB}	NRM	GRM_{SB}	GRM_{XB}	NRM	GRM_{SB}	GRM_{XB}
Average	0.004	0.338	0.339	0.008	0.473	0.465	0.008	0.319	0.316	0.000	0.286	0.294
Min.	0.004	0.338	0.339	0.008	0.473	0.465	0.008	0.319	0.316	0.000	0.286	0.294
Max.	0.511	0.654	0.643	0.511	0.654	0.643	0.454	0.553	0.550	0.000	0.433	0.430
Var.	0.001	0.006	0.005	0.001	0.001	0.001	0.002	0.001	0.001	0.000	0.001	0.001

NRM – Pedigree based relationship matrix; GRM_{SB} elements adjusted by average allele frequency of the singlebreed dataset; GRM_{XB} elements of the GRM adjusted by individual animals breed proportion thus including breed allele frequencies.

Table 4 presents variance component estimates from each breed group and for the combined dataset using each of the relationship matrices. The variance components from the full model were used in the estimation of genomic breeding values (GEBV).

Table 4 Averages of heritability and genetic parameters to 4 validation family groups for Brahman (BB), Tropical Composite (TC) and both (FULL) breeds using different relationship matrices

	NRM		
	h^2	σ_e^2	σ_a^2
FULL	0.546	5543.300	6671.800
BB	0.661	4636.425	9058.200
TC	0.464	5388.600	5706.450

Accuracy and precision of genomic selection

Table 5 presents the correlations between phenotype and GEBV predicted using a range of models and including different numbers of cross validation groups. The accuracy of predicting Tropical Composites from Brahmans alone was similar to that when predicting Tropical Composites from Tropical Composites alone. Adding Brahmans increased the accuracy (from 0.14 to 0.22). There was no difference in the correlations observed between the two GRMs (<0.003), adding BB% (<0.03), or rescaling the phenotypes (<0.03).

Table 5 Realised correlations between genomic breeding values (GEBV) and adjusted phenotypes considering increasing numbers of Brahman animals in training (Row Q No/ Yes indicates BB% includes as covariate in analysis; Rescale Yes/No indicates phenotypes Brahman and Tropical Composite animals rescaled to the same phenotypic variance; BD SB and XB indicate Single breed allele frequency and adjusted for breed specific allele frequency respectively)

Q	No				Yes			
Rescale	No		Yes		No		Yes	
GRM	SB	XB	SB	XB	SB	XB	SB	XB
Tropical Composites								
3BB	0.142	0.144	0.142	0.144	0.131	0.137	0.131	0.137
TC	0.151	0.151	0.151	0.151	0.178	0.177	0.178	0.177
TC+1BB	0.174	0.174	0.173	0.173	0.191	0.191	0.191	0.191
TC+2BB	0.196	0.195	0.194	0.193	0.205	0.206	0.205	0.206
TC+3BB	0.213	0.212	0.211	0.210	0.217	0.219	0.217	0.218
TC+4BB	0.227	0.226	0.225	0.223	0.226	0.230	0.226	0.229
Brahman								
3BB	0.335	0.334	0.335	0.334	0.336	0.335	0.336	0.335
TC	0.086	0.091	0.086	0.091	0.135	0.133	0.135	0.133
TC+1BB	0.242	0.243	0.237	0.238	0.266	0.265	0.263	0.262
TC+2BB	0.316	0.316	0.312	0.312	0.330	0.329	0.328	0.327
TC+3BB	0.334	0.333	0.332	0.332	0.344	0.343	0.344	0.342

*TC is cross validation with 3 groups included in training; Number preceding BB represents the number of BB cross validation groups included in training

The accuracy of predicting Brahman animals from Tropical Composites was low. Adding as little as one BB group into the analysis increased the accuracy substantially (from 0.086 to 0.242). Additional groups increased the accuracy to around 0.33. The accuracy of genomic predictions when 3 groups of both breeds were included in the analysis was similar to the results from the Brahman only analysis, thus adding Tropical Composite data to Brahman analysis did not reduce accuracy of prediction within Brahmans. There was no difference in the accuracy between the two relationships matrices. In contrast to the Tropical Composite results, adding BB% had a small impact in some scenarios, but when three groups of Brahmans were included in the analysis there was no difference (scenarios 3BB and 3TC + 3BB). However, if less than three BB groups were included in the analysis the inclusion of BB% increased the correlation. The correlation was increased by 0.04-0.05 for the TC only analysis and by a smaller amount for the other training scenarios (0.01-0.03). Rescaling the phenotypes had no impact on the correlation.

Table 6 presents the slope of the regression coefficients between GEBV and adjusted phenotypes. In general the regression coefficients were closer to 1 for the Tropical Composite animals and well above 1 for the BB animals. Within the Tropical Composite animals adding Brahmans increased the regression coefficient when BB% was not included in the model. When BB% was included in the model the regression coefficient was either stable when phenotypes were rescaled, or decreasing when not rescaled. Lastly, the regression coefficient was slightly more stable when considering GRM_{XB} compared to GRM_{SB}.

Table 6 Regression coefficient between genomic breeding values (GEBV) and adjusted phenotypes considering increasing numbers of Brahman animals in training (Row Q No/ Yes indicates BB% includes as covariate in analysis; Rescale Yes/No indicates phenotypes Brahman and Tropical Composite animals rescaled to the same phenotypic variance; BD SB and XB indicate Single breed allele frequency and adjusted for breed specific allele frequency respectively)

Q	No				Yes			
Rescale	No		Yes		No		Yes	
BD	SB	XB	SB	XB	SB	XB	SB	XB
Tropical Composites								
TC	0.783	0.789	0.746	0.752	1.018	1.015	0.971	0.968
TC+1BB	0.885	0.881	0.844	0.841	0.992	1.007	0.963	0.974
TC+2BB	0.944	0.936	0.908	0.900	0.972	0.995	0.955	0.975
TC+3BB	0.971	0.964	0.943	0.936	0.948	0.978	0.941	0.968
3BB	0.970	1.006	1.027	1.065	0.830	0.895	0.879	0.948
TC+4BB	0.976	0.972	0.957	0.952	0.919	0.956	0.921	0.954
Brahman								
TC	1.036	1.101	0.987	1.049	1.704	1.680	1.624	1.601
TC+1BB	1.895	1.906	1.879	1.892	2.054	2.051	2.063	2.059
TC+2BB	1.921	1.920	1.965	1.966	1.945	1.950	2.004	2.008
TC+3BB	1.690	1.687	1.757	1.754	1.693	1.696	1.767	1.769
3BB	1.749	1.750	1.852	1.853	1.729	1.734	1.831	1.835

*TC is cross validation with 3 groups included in training; Number preceding BB represents the number of BB cross validation groups included in training

Within the BB animals, the regression coefficients were lowest (and closest to 1) when no Brahmans were included in training. Adding Brahman animals increased the regression coefficients. Adding BB% as a covariate reduced the range in regression coefficients across all other scenarios, particularly when no BB animals were included in the analysis. There was little difference in the regression coefficients between the two GRMs.

4. Discussion

Genetic evaluation in mixed or admixed breed populations is complicated by estimation of the effect of the ancestral breeds on each trait. The breed proportion in traditional analysis is calculated by tracing parental breed through the pedigree. Using this approach each animal is given the average proportion of its parents, however through recombination the actual proportion inherited may vary from this due to Mendelian segregation. It has been proposed that breed component be estimated from genomic information for use in genetic evaluation (Porto Neto et al., 2013; Thomasen et al., 2013). Accuracies of breed composition estimates from high density genotypes are high (Frkonja et al., 2012; Kuehn et al., 2011) thus it would be expected that using these values in place of pedigree based estimates of breed proportions may increase accuracy. Accordingly Thomasen et al. (2013) added breed proportion as a covariate in analysis of genomic data using random regression. In this case the accuracy of genomic selection was not improved, however in this study the divergence between the breeds was rather small as the two breeds (Danish and US Jersey populations) had only been separated for 100 years (Thomasen et al., 2013). This is in contrast to Brahman and the *Bos taurus* component of Tropical Composites which are estimated to have diverged hundreds of thousands of years ago. Accordingly Porto Neto et al. (2013) suggested that Zebu content could be added to genetic evaluation programs that include Tropical Composites.

Genomic predictions across breeds mostly have low accuracy, particularly for breeds not represented within the training population (Erbe et al., 2012; Garrick, 2011). However where a minor breed is represented in both the training and validation populations often the accuracy is similar to or slightly better than training on the smaller population. For example Erbe et al. (2012) and Pryce et al. (2012) found that adding Holstein animals to a Jersey reference increased accuracy with either no reduction or a small reduction in Holstein accuracies depending on the trait. Similarly Zhang et al. (2014) found that adding Brahman animals to TC increases accuracy for Tropical Composites, and this also was observed in our analysis: adding additional groups of Brahmans to the training population lead to consistent increases in realised accuracy.

This study confirmed that adding BB can lead to increases in accuracy of TC genomic evaluations. Adding breed specific GRMs did not improve the accuracy of genomic evaluation however it did improve the regression coefficient for TC animals. This impact will be particularly important where animals do not have links to animals in the current genetic evaluation. Such animals need to be placed into appropriate genetic groups.

The effect of incorrect genetic grouping can have substantial impact on breeding value estimates (Misztal et al., 2013).

As noted it was observed that the Brahman regression coefficient was inflated, while the Tropical Composite was around 1. An additional analysis was performed where the variances were adjusted so the Brahman regression coefficient was closer to 1, however under these parameters the Tropical Composite regressions were well below 1 (data not shown). Thus it is does not seem possible to obtain correct regressions for both traits under a univariate model. The alternative would be to consider each AGECL as a separate but correlated trait for Brahmans and Tropical composites.

Porto Neto et al. (2013) estimated the Zebu content of animals within the same population using a different set of reference animals: in their study 81 Angus and 29 Nelore were used as reference animals. The Brahman animals used in our study would contain a proportion of *Bos taurus* genes as a consequence of the grading up process, where a small number of imported Brahman sires were crossed to Australian *Bos taurus* animals to produce the current industry Brahman herds. This is reflected in the Zebu average content of 95% in the analysis of (Porto Neto et al., 2013). The contrasts with the estimate using Brahman animals as reference population (BB%=98) the estimate of BB% in the Tropical Composites was also slightly lower (43%) than the estimate of Porto Neto et al. (2013).

5. Implications

There was a clear benefit in adding Brahman animals to Tropical Composite genomic evaluations. However using breed composition in the development of the GRM did not have an impact on the accuracy of prediction. The model did influence the precision of genomic evaluations and therefore highlights the importance of correctly accounting for breed in genetic evaluation. It is suggested that future work would examine the effect of BB% on multibreed GEBVs in more detail and examine the effect of additional data sets.

Recommendations:

- With care it is possible to combine genotypic datasets including different breeds.
    - This is most pertinent for smaller breeds, which are in part, are derived from a major breed.
- When combining genotypic information from different breeds care must be taken to ensure that relationship between predictions with breeding values remain at one.

6. Acknowledgements

The authors acknowledge that this research uses resources built by the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC).

7. References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93:743-752.

Alexander, D. H., and K. Lange. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC bioinformatics 12:246.

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome research 19:1655-1664.

Barwick, S. A., D. J. Johnston, H. M. Burrow, R. G. Holroyd, G. Fordyce, M. L. Wolcott, W. D. Sim, and M. T. Sullivan. 2009. Genetics of heifer performance in 'wet' and 'dry' seasons and their relationships with steer performance in two tropical beef genotypes. Anim. Prod. Sci. 49:367.

Bolormaa, S., J. E. Price, K. Kemper, K. Savin, B. J. Hayes, W. Barendse, Y. Zhang, B. Reich, B. A. Mason, R. J. Bunch, B. E. Harrison, A. Reverter, R. M. Herd, B. Tier, H.-U. Graser, and M. E. Goddard. 2013a. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in Bos taurus, Bos indicus and composite beef cattle. J. Anim. Sci 91:3088-3104.

Bolormaa, S., J. E. Pryce, K. E. Kemper, B. J. Hayes, Y. Zhang, B. Tier, W. Barendse, A. Reverter, and M. E. Goddard. 2013b. Detection of quantitative trait loci in Bos indicus and Bos taurus cattle using genome-wide association studies. Genetics, selection, evolution : GSE 45:43.

Burns, B. M., N. J. Corbet, D. H. Corbet, J. M. Crisp, B. K. Venus, D. J. Johnston, Y. Li, M. R. McGowan, and R. G. Holroyd. 2013. Male traits and herd reproductive capability in tropical beef cattle. 1. Experimental design and animal measures. Anim. Prod. Sci. 53:87.

Burrow, H. M. 2012. Importance of adaptation and genotype x environment interactions in tropical beef breeding systems. Animal 6:729-740.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114-4129.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genetics, selection, evolution : GSE 43:1.

Frkonja, A., B. Gredler, U. Schnyder, I. Curik, and J. Solkner. 2012. Prediction of breed composition in an admixed cattle population. Anim Genet 43:696-703.

Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genetics, selection, evolution : GSE 43:17.

Harris, B. L., and D. L. Johnson. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J Dairy Sci 93:1243-1252.

Hawken, R. J., Y. D. Zhang, M. R. Fortes, E. Collis, W. C. Barris, N. J. Corbet, P. J. Williams, G. Fordyce, R. G. Holroyd, J. R. Walkley, W. Barendse, D. J. Johnston, K. C. Prayaga, B. Tier, A. Reverter, and S. A. Lehnert. 2012a. Genome-wide association studies of female reproduction in tropically adapted beef cattle. J Anim Sci 90:1398-1410.

Hawken, R. J., Y. D. Zhang, M. R. S. Fortes, E. Collis, W. C. Barris, N. J. Corbet, P. J. Williams, G. Fordyce, R. G. Holroyd, J. R. W. Walkley, W. Barendse, D. J. Johnston, K. C. Prayaga, B. Tier, A. Reverter, and S. A. Lehnert. 2012b. Genome-wide association studies of female reproduction in tropically adapted beef cattle. J. Anim. Sci 90:1398-1410.

Johnston, D. J., S. A. Barwick, N. J. Corbet, G. Fordyce, R. G. Holroyd, P. J. Williams, and H. M. Burrow. 2009. Genetics of heifer puberty in two tropical beef genotypes in northern Australia and associations with heifer- and steer-production traits. Anim. Prod. Sci. 49.

Kuehn, L. A., J. W. Keele, G. L. Bennett, T. G. McDaneld, T. P. Smith, W. M. Snelling, T. S. Sonstegard, and R. M. Thallman. 2011. Predicting breed composition using breed frequencies of 50,000 markers from the US Meat Animal Research Center 2,000 Bull Project. J Anim Sci 89:1742-1750.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PloS one 4:e5350.

Meyer, K. 2007. WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). Journal of Zhejiang University. Science. B 8:815-821.

Misztal, I., Z. G. Vitezica, A. Legarra, I. Aguilar, and A. A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. Journal of Animal Breeding and Genetics:n/a-n/a.

Olson, K. M., P. M. VanRaden, and M. E. Tooker. 2012. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. J Dairy Sci 95:5378-5383.

Porto Neto, L. R., S. A. Lehnert, M. R. S. Fortes, M. Kelly, and A. Reverter. 2013. Population Stratification and Breed Composition of Australian Tropically Adapted Cattle. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20:4.

Prayaga, K. C., N. J. Corbet, D. J. Johnston, M. L. Wolcott, G. Fordyce, and H. M. Burrow. 2009. Genetics of adaptive traits in heifers and their relationship to growth, pubertal and carcass traits in two tropical beef cattle genotypes. Anim. Prod. Sci. 49:413-425.

Pryce, J. E., B. J. Hayes, and M. E. Goddard. 2012. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. J Dairy Sci 95:377-388.

Thomasen, J. R., A. C. Sorensen, G. Su, P. Madsen, M. S. Lund, and B. Guldbrandtsen. 2013. The admixed population structure in Danish Jersey dairy cattle challenges accurate genomic predictions. J. Anim. Sci 91:3105-3112.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J Dairy Sci 91:4414-4423.

VanRaden, P. M., K. M. Olson, G. R. Wiggans, J. B. Cole, and M. E. Tooker. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J Dairy Sci 94:5673-5682.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature genetics 42:565-569.

Zhang, Y. D., D. J. Johnston, S. Bolormaa, R. J. Hawken, and B. Tier. 2014. Genomic selection for female reproduction in Australian tropically adapted beef cattle. Anim. Prod. Sci. 54:16.