animal

Possibilities to improve the genetic evaluation of a rare breed using limited genomic information and multivariate BLUP

G. E. Pollott[†], A. Charlesworth and D. C. Wathes

Royal Veterinary College, Royal College Street, London, NW1 0TU, UK

The use of molecular genetic information in the evaluation of livestock has become more common. This study looks at the efficacy of using such information to improve the genetic evaluation of a rare breed of dual-purpose cattle. Data were available in the form of pedigree information on the Gloucester cattle breed in the United Kingdom and recorded milk and beef performance on a small number of animals. In addition, molecular genetic information in the form of multi-marker, multiple regression results converted to a 1 to 10 score (Igenity scores) and 123 single nucleotide polymorphism (SNP) genotypes for 199 non-recorded animals were available. Appropriate mixed-animal models were explored for the recorded traits and these were used to calculate estimated breeding values (EBV), and their accuracies, for 6527 animals in the breed's pedigree file. Various ways to improve the accuracy of these EBV were explored. This involved using multivariate BLUP analyses, genomic estimated breeding values (GEBV) and combining Igenity scores with recorded traits in a series of bivariate genetic analyses. Using the milk recording traits as an example, the accuracy of a number of traits could be improved using multivariate analyses by up to 14%, depending on the combination of traits used. The level of increase in accuracy largely corresponded to the absolute difference between the genetic and residual correlations between two traits, but this was not always symmetrical. The use of GEBV did not increase the accuracy of milk trait EBV owing to the low proportion of variance explained by the 101 SNPs used. Using Igenity scores in bivariate analyses with the recorded data was more successful in increasing EBV accuracy. The largest increases were found in genotyped animals with no recorded performance (e.g. a 58% increase in fat weight in milk); however, the size of the increase depended on the level of the genetic correlation between the recorded trait and the Igenity score for that trait. Lower levels of improvements in accuracy were seen in animals that were recoded but not genotyped, and ancestors which were neither genotyped nor recorded. This study demonstrated that it was possible to improve the accuracy of EBV estimation by including Igenity score information in genetic analyses but it also concluded that increasing the level of performance recording in the breed would be beneficial.

Implications

Rare breeds of livestock are an important source of biodiversity in the agricultural sector. Their survival depends on finding a suitable economic role. Genetic improvement is one way to make such breeds more relevant and the use of molecular genetic information offers the possibility to speed up their improvement. This paper demonstrates that it is possible to achieve faster rates of improvement in some traits by combining recorded information on productivity with genetic scores derived from analysing large numbers of animals for key traits. This approach needs to be combined with recording a larger number of animals in the breed to make good progress.

† E-mail: gpollott@rvc.ac.uk

Introduction

Rare breeds, by their very nature, suffer from a number of issues which may be a drawback when trying to implement genetic improvement programmes. These include small numbers of animals in the breed, limited numbers of sires, possible reduced genetic variation, potentially high levels of inbreeding and, in many cases, ownership by part-time farmers in small herds in a breed of relatively low economic value. The 'genomic revolution' is changing the way mainstream breeds are selected, and it is of interest to investigate how such methods might help owners of rare breeds to improve their animals.

The situation of the Gloucester cattle breed (Gloucester Cattle Society (GCS), 2013; Oklahoma State University, 2013; RBST, 2013) in the United Kingdom was such that the

685

pedigree information was quite extensive, performance data for dairy, beef and fertility traits was limited, and genomic information from a small number of animals with 123 single nucleotide polymorphisms (SNP) genotypes was available. The basic question addressed in this paper was how to maximise the benefit of such information to Gloucester breeders. In addition, it was interesting to investigate the contribution of various sources of information to the genetic evaluation of the animals, to see what might be done to improve the breed and to consider what else could be done to improve its genetic evaluation. This study investigated the use of 123 SNPs from the panel used by Merial to derive multi-marker, multiple regression results converted to a 1 to 10 score (Igenity scores) for a range of traits on cattle in the United States (Merial, 2008 and 2009). These scores, along with other more traditional genetic evaluation methods, were evaluated as an aid to determining the genetic potential of a rare breed of cattle.

Material and methods

Several sources of information were available for this study. The basic data comprised the identities of the animals registered with the breed society and that of their ancestors (pedigree data). Some breeders also recorded information on different aspects of productivity (performance data). In addition, several breeders had data on Igenity scores produced by Merial (now part of the Neogen Corporation) and based on genotype information for 123 SNP (genotype data).

Breed society pedigree data
Gloucester cattle are an ancient UK breed, numerous in the Severn Vale and throughout Gloucestershire, from as early as the 13th century (GCS, 2013). They are considered dual-purpose, being valued for their meat and milk (producing cheese), and were also used as strong draught oxen in former times. Achieving the peak of their popularity in about 1750, Gloucester cattle were found from Devon (south-west England) to Essex (eastern England) and to the Welsh coast. Later they were depleted by disease in the 18th Century, the development of other breeds, arable farming taking grazing land and the continuing sales of established herds in the early 20th Century, resulting in only one substantial herd remaining by 1972. Originally formed in 1919, the GCS was revived in 1973 initially to provide for the survival of the breed. The Society has been very successful and breed numbers have now grown to over 700 registered females. The pedigree file used in these analyses consisted of 6527 animals, with an average of 247 being registered annually from 1990 to the present day. Pedigrees were checked for consistency using Pedigree Viewer (Version 5.5; Kinghorn, 2012).

Gloucester performance data
Initially, farms were selected for data collection based on the first 166 DNA samples sent to Merial for genotyping. Subsequently, farms were also selected on the basis that they kept animals that were related to the genotyped animals.

This resulted in records from 21 farms being collected on about 200 animals. The recorded traits are summarised in Table 1 for both dairy and beef traits. Not all traits were recorded on all animals.

The dairy data used in this study were collected from the milk records kept on the farms and follow the trait descriptors commonly used by organisations such as National Milk Records Ltd (Harrogate, North Yorkshire, UK). Information on herd, calving date and lactation number was also collected for each lactation. Both somatic cell count and calving interval (CI) data were $\log_{10}$ transformed to make them more normally distributed for analysis.

Beef characteristics were collected for up to 214 animals from slaughterhouse records, together with appropriate breed society and farm data. Age at slaughter was calculated as the difference between the date of birth and date of slaughter and expressed in months. Carcase weight (CWT) was the cold CWT from the abattoir records. Both the conformation and fat score data were generated from the EUROP scheme classifications (English Beef and Lamb Executive, 2012). The conformation grades E, U, R, O and P, also assigned a − or + designation, were scored in ascending order from the poorest to the best conformation on a 1 to 10 scale; hence, P − = 1 up to E + = 10. In the same way, the fat scores were on a 1 to 7 scale with Fat Classes 4 and 5 being split into L (low) and H (high). Clearly these new scores, like the original classes, were split into arbitrary values; however, it was assumed in these analyses that a unit increase in both the fat and conformation scores was equal across the range used. Additional data on herd, date of birth and sex were collected and used as explanatory variables in the analyses.

Genotype data in the form of Igenity scores and SNP information
Starting in 2009, the GCS obtained DNA from 199 animals: 37 male and 162 female. These samples were used to obtain profiles for, each individual, of their predicted beef and dairy characteristics based on an Igenity score (IG). This was a system developed by Merial which produced scores for each animal for 10 beef traits (e.g. tenderness, average daily live weight gain) and 12 dairy traits (e.g. milk yield, somatic cell count) based on 123 SNP genotypes. This scoring system was produced using the evidence of associations between the DNA profile (genotype) and the actual measured trait (phenotype) from other breeds of cattle. Several beef breeds were used to compile the beef index, whereas the dairy index was nearly all derived from Holsteins. These were designed by Merial on the basis of research carried out in the United States and Canada. A comparison between a number of such scores for beef traits has been reported by Van Eenennaam *et al.* (2007). A study outlining one possible method for developing SNP-panel scores has been described by Tang *et al.* (2011). The mean and standard deviation of the Igenity scores used in this paper are shown in Supplementary Table S1 for 19 traits covering beef, dairy and functional characteristics. Each trait was scored on a 1 to 10 scale, with 1 being the lowest and 10 the highest genetic merit for each trait separately.

Genetic parameter and breeding value estimation
All mixed model testing and breeding value estimation (BLUP analyses) were carried out using ASReml (Gilmour *et al.*, 2009) which utilises an average information-restricted maximum likelihood algorithm (REML). Each recorded trait was tested using a step-down approach from fitting all available fixed and random effects, and covariates in a mixed model:

$$T_{ijklmno} = \mu + H_i + Y_j + S_k + L_l + \text{DIM} + a_m + pe_n + e_{ijklmno}$$
(1)

The dairy traits were analysed using Equation (1) where $T_{ijklmno}$ was one of milk yield, fat weight, protein weight, fat %, protein % or $\log_{10}$ somatic cell count. The fixed effects were the ith herd (H_i), the jth year (Y_j), the kth season of calving (S_k) and the lth lactation number (L_l). Interactions between herd and year and herd by year by season were tested for each trait. The length of the lactation (days in milk, DIM) was fitted as a covariate. Random effects of the mth animal in the pedigree file, estimating the additive genetic variance, and the permanent environmental effect of the nth cow were fitted with variance σ_a^2 and σ_{pe}^2 respectively. The residual term was assumed to be normally distributed with a mean of 0 and variance of σ_e^2. $\log_{10}$ CI was analysed with a similar model but included a term for the month of calving and omitted the covariate DIM.

$$T_{ijklmn} = \mu + H_i + Y_j + M_k + S_l + \text{Age} + \text{CWT} + a_m + e_{ijklmn}$$
(2)

The beef traits were analysed using equation (2) where T_{ijklmn} was one of growth rate, cold CWT, fat score or conformation score. The fixed effects were the ith herd (H_i), the jth year (Y_j), the kth month of birth (M_k) and the lth sex (S_l). Interactions between herd and year and herd by year by season were tested for each trait. The age at slaughter (Age) was fitted as a covariate within sex using a second-order polynomial, and CWT was fitted as a covariate to all traits except CWT. These models were used following a model selection procedure, which tested different orders of polynomial fitted both overall and within sex. An upward selection procedure using a log–likelihood test to indicate a significant difference between a model and the previous reduced model was used. Residuals were inspected for non-systematic patterns indicating the use of a poor model. The random effect of the mth animal in the pedigree file, estimating the additive genetic variance, was fitted with variance σ_a^2. The residual term was assumed to be normally distributed with a mean of 0 and variance of σ_e^2.

Effects were discarded in the step-down process on the basis of their probability being >0.05 for an *F*-ratio, in order of the size of the probability with the largest discarded first. Estimated breeding values (EBV) and their standard errors were recovered from the final run for any trait comprising the significant effects and covariates decided by the step-down approach. The phenotypic variance was calculated as the sum of all the available random variance terms for a trait. The heritability of each trait was calculated as the ratio of the additive to phenotypic variance. The repeatability of a trait was calculated as the sum of the additive and permanent environmental variance divided by the phenotypic variance. The standard errors of both heritability and repeatability were calculated as described by Gilmour *et al.* (2009).

The Igenity scores were analysed with a random effects model (Equation (3)).

$$T_{mn} = \mu + a_m + e_{mn}$$
(3)

T_{mn} was one of the 19 Igenity scores and a_m was the random term fitting the pedigree-based relationship matrix to estimate the additive genetic variance (σ_a^2). The heritabilities and their standard errors were calculated as described above.

The EBV correlations and the correlations between EBV and Igenity scores were calculated as Pearson product moment correlation coefficients. Their standard errors were calculated as $(1 - r^2)/(n - 1)$, where r was the correlation and n the number of pairs of values on which the correlation was based. The genetic and residual correlations were calculated using a bivariate mixed model in ASReml comprising the final model for each trait as used in the BLUP analyses. Their standard errors were calculated as described by Gilmour *et al.* (2009).

The genetic correlations of the five milk trait Igenity scores with the equivalent recorded trait were calculated using ASReml and the appropriate final model for each of the two traits. All aspects of these analyses were carried out using ASReml (Gilmour *et al.*, 2009) as described above.

For the purpose of comparing accuracy values under different methods or using multivariate BLUP, the additive, permanent environmental and residual variances were fixed at their univariate values. Only the covariance parameters were allowed to vary in the multi-trait analyses. The standard error of each EBV was calculated as the square root of the prediction error variance, derived from the reciprocal of the diagonal element of the coefficient matrix after absorption of all other effects (Mrode, 1996). All accuracy values were derived from the standard errors (SE) of the EBV or genomic estimated breeding values (GEBV) calculated by the respective programs as $\sqrt{(1 - (\text{SE}^2/(1 + f) \; \sigma_a^2))}$, where f was the inbreeding coefficient of the animal, and σ_a^2 the additive genetic variance for the trait (Mrode, 1996; Gilmour *et al.*, 2009). Values were adjusted for repeated records as appropriate. Where mean accuracy values were compared, for specific groups of animals, a two-tailed paired-comparison *t*-test was used to signify the probability of a difference between the two sets of results.

Genomic estimated breeding values
Genomic estimated breeding values and their standard errors were calculated using the single-step genomic BLUP (GBLUP) methodology of Misztal *et al.* (2009) which utilises an expectation-maximisation algorithm. This approach combines pedigree, genomic and performance details from animals with differing amounts of information; animals may have any combination of the types of data, pedigree, genomic and performance. Genotype data for 123 SNPs from the 199

animals analysed by Merial to derive the Igenity scores were coded. The 22 monomorphic SNPs were removed from the data set. The remaining 101 SNPs were used in conjunction with the full pedigree and performance data for all 11 recorded traits to estimate the GEBV and their standard errors. Because this was a small set of SNP, compared to those more commonly used in GBLUP calculations, preliminary analyses were carried out to determine the proportion of the genetic variance accounted for by the set of SNP (α in the terminology of Misztal et al., 2009), for each trait. Successive analyses were run for each trait until the maximum log–likelihood value was found by varying α between 0 and 1 (Christensen and Lund, 2010). All subsequent analyses for a given trait were run at the appropriate α value. All traits were analysed using the final models obtained from the BLUP analyses described above.

Results and discussion

Gloucester performance
The Gloucester is a rare, dual-purpose breed of cattle hence data were available for meat, milk and fertility traits. The level of inbreeding found in these analyses was 0.12 ± 0.04 for the cows with dairy records and 0.14 ± 0.04 for the animals with Igenity scores. The data from the 176 lactations recorded on 81 Gloucester cows are summarised in Table 1. These animals had a mean lactation yield of 2707 kg from lactations averaging 242 days in length with a mean fat and protein % of 3.67 and 3.4, respectively. They had an average lactation number of 2.21 and were 53.9 months of age at calving. The data were derived from 9 herds, 26 years, 11 months and 37 herd/year/season groups. The carcase data are also summarised in Table 1. There were 214 animals slaughtered at an average age of 42 months and a CWT of 287 kg. The animals grew at an average of 8.84 kg/month (291 g/days) and averaged between O− and O+ for conformation and 4L for fat on the EUROP classification scale. These data included both younger animals raised for beef production and older breeding animals culled from the herds. The carcase data were derived from 8 herds, 14 years, 12 months and 3 sexes (entire male, castrates and female). The mean of 501 CI from 141 cows kept in 19 herds was 416 days ($\pm$147.1). Igenity scores were provided by Merial for 199 animals and these are summarised for 19 traits in Supplementary Table S1.

Clearly, the Gloucester does not have as high a level of performance as the major commercial beef or dairy breeds but acts as a dual-purpose breed kept on small farms producing milk suitable for cheese making. In comparison, UK Holstein cows averaged 9091 kg/305 days lactation with mean fat and protein % of 3.93 and 3.16, respectively, and CI of 423 days (Holstein UK, 2013). Key beef breeds in the United Kingdom grow from 750 g/days liveweight and produce carcases of 275 kg at 450 + days of age (Kempster and Southgate, 1984). The Gloucester will not achieve these levels of performance but genetic improvement has been shown to lead to steady, accumulative increases in performance in several farmed species once recording

Table 1 The mean and standard deviation of the milk, beef and fertility traits recorded on Gloucester cattle used in this study

	Mean	s.d.
Milk traits[1]		
Lactation number	2.21	1.56
Age (months)	53.9	23.9
Lactation length (days)	242	77
Total milk yield (kg)	2707	1354
Fat weight (kg)	99.9	51.7
Protein weight (kg)	91.8	44.6
Fat %	3.67	0.54
Protein %	3.40	0.26
Somatic cell count ('000)	127	270
Beef traits[2]		
Age at slaughter (months)	42.0	30.1
Carcase growth rate (kg/month)	8.84	3.51
Cold carcase weight (kg)	287	68
Fat score	3.96	1.35
Conformation score	3.50	1.12
Fertility trait[3]		
Calving interval (days)	416	147

[1]Milk production records from 176 lactations by 81 Gloucester cows in 9 herds and 38 herd/year/season groups. Data from 156 somatic cell count records.
[2]Beef data from 214 animals in 8 herds and 42 herd/year groups.
[3]Fertility data from 501 calving intervals by 141 cows in 19 herds.

and genetic evaluation schemes have been implemented. One important set of results in Table 1 and Supplementary Table S1 is the number of animals recorded in this data set. Genetic evaluations become more accurate and useful as more animals are recorded. This should be a key objective of the breed, to record as much information on as many animals as possible.

Mixed model analysis of recorded data
The 11 recorded traits were analysed by mixed-animal models to derive genetic parameters. A summary of the significant ($P < 0.05$) fixed effects and covariates is shown in Supplementary Table S2. The genetic analyses of each trait were carried out using the fixed effects and covariates shown as significant effects in Supplementary Table S2 in the mixed model. These analyses highlight the key factors that need to be recorded to use an appropriate model in genetic analyses. All the final models used in these analyses had residuals evenly distributed about the mean over the range of values analysed. One additional point is that, in these analyses, the carcase traits were derived from all animals from the breed sent for slaughter. If it is important for the breed to consider young beef animals as a different product from older culled animals, then different models for these two types of beef animal may be necessary.

The variance components and genetic parameters, with their standard errors, derived from 11 univariate analyses are summarised in Table 2. The heritability values shown in Table 2 indicate that there was a good level of genetic variation in most traits except somatic cell count, CI and conformation score, which showed very little genetic variation.

Table 2 *The variance estimates and genetic parameters of the 11 recorded Gloucester traits derived from fitting a mixed animal model to each trait*

Trait	Animal	Permanent environmental	Residual	Heritability	s.e.	Repeatability	s.e.
Milk yield (kg)	200 113	28 073	190 892	0.48	0.20	0.54	0.10
Fat weight (kg)	182.4	17.1	470.4	0.27	0.20	0.30	0.11
Protein weight (kg)	129.1	54.1	239.2	0.31	0.21	0.43	0.11
Fat %	0.1552	0.0032	0.1407	0.51	0.29	0.52	0.12
Protein %	0.0297	0.0000	0.0251	0.54	0.09	0.54	0.09
Log. SCC ('000)	0.0029	0.0203	0.2261	0.01	0.23	0.09	0.14
Log. CI (days)	0.0000	0.0018	0.0097	0.00	0.10	0.16	0.06
Growth rate (kg/month)	1.12		1.89	0.37	0.22		
Carcase weight (kg)	682		545	0.55	0.23		
Fat score	0.644		0.524	0.55	0.33		
Conformation score	0.0206		0.9290	0.02	0.23		

SCC = somatic cell count; CI = calving interval.

Table 3 *The correlations between the estimated breeding values of the 11 recorded Gloucester traits from 6423 animals in the pedigree file (lower triangle) and genetic correlations and their s.e. (below the correlation) between five milk traits derived from the animal-model bivariate analyses (upper triangle)*

	Milk yield	Fat weight	Protein weight	Fat %	Protein %	Log. SCC	Log. CI	Growth rate	Carcase weight	Fat score
Milk yield		**0.96**	**0.98**	**− 0.64**	**− 0.66**					
		0.066	0.134	0.201	0.180					
Fat weight	**0.96**[1]		**0.99**	− 0.22	− 0.24					
			0.135	0.313	0.314					
Protein weight	**0.94**	**0.90**		**− 0.71**	**− 0.37**					
				0.270	0.285					
Fat %	− 0.25	− 0.05	**− 0.34**		0.65					
					0.152					
Protein %	− 0.29	− 0.11	**− 0.32**	**0.64**						
Log. SCC	**0.31**	**0.34**	**0.39**	− 0.11	0.14					
Log. CI	− 0.08	− 0.13	− 0.08	− 0.24	− 0.21	**− 0.45**				
Growth rate	− 0.23	− 0.22	− 0.19	− 0.02	− 0.06	**− 0.34**	0.04			
Carcase weight	0.06	0.04	0.10	− 0.20	− 0.18	− 0.26	0.25	**0.68**		
Fat score	0.18	0.14	0.24	− 0.19	− 0.04	0.15	0.15	0.04	**0.31**	
Conformation score	**− 0.37**	**− 0.37**	**− 0.40**	− 0.04	− 0.08	− 0.24	0.02	**0.49**	**0.40**	− 0.20

SCC = somatic cell count; CI = calving interval.
The standard errors of the correlations in the lower triangle range from 0.00237 (correlation ± 0.9) to 0.0123 (correlation ± 0.1).
[1]Correlations >0.3 or <− 0.3 shown in bold.

Although the precision of these estimates was somewhat low, the heritability of the milk and component weight traits was higher than found in Holstein studies. For example, Banos *et al.* (2012) quote heritabilities of milk traits as 0.22 (milk yield), 0.20 (fat weight), 0.16 (protein weight), 0.68 (fat percentage) and 0.55 (protein percentage). Similarly, the carcase traits were found to have similar or slightly higher heritabilities than the average number of estimates found from a range of commercial breeds, with the exception of conformation score (Rios-Utrera, 2004).

Two types of genetic correlation were estimated in this study. Correlations between each pair of derived BLUP EBV (EBV correlations) for the 11 recorded traits using 6423 animals in the pedigree file are shown in the bottom segment of Table 3. Actual genetic correlations for the five milk traits, derived from bivariate mixed-animal model analyses are shown in the top segment of Table 3. The level of concordance between the two sets of results for the milk traits gives confidence in the remainder of the correlations as shown in Table 3. The high correlations between the milk weight traits were found in other studies and there was a moderate correlation between milk yield and somatic cell count (see e.g. Kadarmideen *et al.*, 2000), a common feature of intensive dairy breeds. Interestingly, there was no relationship between milk yield and CI in this breed. Not surprisingly, growth rate was positively correlated with both CWT and conformation, and CWT was correlated with both fat and conformation score. Dual-purpose breeds allow the relationships between milk and meat traits to be evaluated. In this case, there was very little relationship between the two, except for the fact that carcase conformation was moderately negatively correlated with milk production.

Table 4 *The correlations between the Igenity scores and recorded trait estimated breeding values (EBV) for the 199 Gloucester animals with both sets of data*[1]

	EBV from records										
Igenity Scored traits	Milk yield	Fat weight	Prot weight	Fat %	Protein %	Log. SCC	Log. CI	Growth rate	Carcase weight	Fat score	Conformation score
Tenderness	0.08	0.09	0.04	**0.21**	0.15	0.17	− 0.15	**− 0.24**	**− 0.28**	− 0.03	− 0.19
Marbling score	0.13	0.12	0.10	0.02	0.04	− 0.13	0.04	0.02	0.15	0.16	− 0.06
Average daily gain	0.04	0.05	0.08	0.05	0.13	**0.22**	− 0.16	**− 0.20#**	**− 0.24**	− 0.06	− 0.04
Red meat yield	− 0.19	**− 0.20**	**− 0.28**	0.05	− 0.06	− 0.09	0.03	− 0.01	− 0.07	− 0.11	0.04
Carcase fat	0.11	0.10	0.11	0.03	0.08	− 0.09	− 0.07	− 0.06	− 0.04	0.11#	0.01
Ribeye area	0.10	0.07	0.02	0.04	− 0.06	0.14	− 0.16	− 0.11	− 0.12	− 0.16	− 0.05
Heifer pregnancy rate	0.08	0.11	0.04	0.11	− 0.15	0.04	0.03	0.03	0.01	0.06	0.01
Longevity	0.14	**0.20**	0.17	**0.22**	0.17	**0.27**	**− 0.22**	**− 0.20**	− 0.18	0.16	**− 0.24**
Maternal calving ease	0.14	0.14	0.17	− 0.03	− 0.08	0.07	0.03	0.12	0.05	0.04	0.07
Docility	− 0.08	− 0.09	− 0.11	− 0.09	**− 0.27**	0.03	− 0.05	0.10	**0.23**	0.00	0.15
Productive life	0.15	0.19	0.13	0.30	**0.23**	0.11	**− 0.21**	− 0.12	**− 0.26**	0.04	− 0.08
Fertility	0.10	0.14	0.08	**0.27**	0.18	0.11	− 0.18#	**− 0.25**	**− 0.38**	− 0.18	− 0.12
SCC	0.08	0.03	0.10	− 0.10	0.06	0.05#	0.15	− 0.15	− 0.02	0.10	**− 0.25**
Dairy form	0.16	0.16	0.15	0.02	− 0.02	− 0.03	0.01	− 0.02	0.14	**0.27**	− 0.07
Milk yield	**0.23#**	**0.20**	**0.22**	− 0.08	− 0.18	0.10	− 0.07	− 0.16	0.07	− 0.05	− 0.07
Fat weight	**0.34**	**0.36#**	**0.26**	**0.31**	**0.26**	0.16	**− 0.24**	**− 0.20**	− 0.18	**0.27**	− 0.19
Fat %	0.12	0.16	0.07	**0.33#**	**0.36**	0.08	− 0.16	− 0.07	**− 0.20**	**0.30**	− 0.13
Protein weight	**0.21**	0.18	**0.28#**	− 0.13	− 0.08	0.18	− 0.05	− 0.08	0.10	0.06	− 0.16
Protein %	0.11	0.11	0.18	0.07	**0.27#**	0.11	0.08	− 0.13	**− 0.21**	0.15	**− 0.35**

SCC = somatic cell count; CI = calving interval.
The standard errors of the correlations in the lower triangle range from 0.00237 (correlation ± 0.9) to 0.0123 (correlation ± 0.1).
[1]Figures followed by # are correlations of the same traits measured as both Igenity scores and EBV. Figures in bold are > 0.2 or < − 0.2.

Table 5 *Genetic parameters of the Igenity scores for five milk traits and their genetic correlation with the equivalent recorded trait obtained from bivariate animal model genetic analyses*

Trait	Animal variance	Residual variance	Heritability	s.e.	Genetic correlation	s.e.
Milk yield	1.591	0.091	0.95	0.066	0.66	0.295
Fat weight	2.238	0.560	0.80	0.061	0.99	0.205
Protein weight	3.144	0.208	0.94	0.066	0.84	0.334
Fat %	1.883	0.217	0.90	0.084	0.61	0.274
Protein %	0.984	0.121	0.89	0.072	0.12	0.378

Igenity score analyses
As Igenity scores were derived from the SNP genotypes carried by the sampled animals, they are effectively a direct genetic measure of each animal, according to Merial's protocols. The scores for the 19 traits from 199 animals were used to derive bivariate correlations between each pair of traits; these are shown in Supplementary Table S1 and may be considered as a type of genetic correlation. A further set of correlations were derived by analysing the 19 Igenity scores with the 11 recorded-trait EBV for all 199 animals, which were genotyped. The results of these 209 correlations are shown in Table 4; these too could be considered to be a type of genetic correlation. Milk production Igenity scores were also analysed by animal model to estimate their heritabilities. They were also analysed in a series of bivariate analyses with the equivalent recorded trait. The results of both sets of analyses are shown in Table 5.

The correlations between the Igenity scores and EBV for the 199 genotyped animals ranged from − 0.38 (CWT EBV with fertility) to 0.36 (milk fat weight EBV and IG score; milk fat % EBV and protein %). Thus, the range in correlations was lower than that found between EBV (Table 3) and also between Igenity scores (Supplementary Table S3). As a number of traits were found in both the data sets, it is noteworthy that the correlation between the same traits measured in different ways was low. However, the EBV were derived largely from the relatives of the genotyped animals, and therefore would have had a lower accuracy than if they had been measured directly on the animals, and the SNP used to calculate the Igenity scores only accounted for ∼3% of the trait variation. The low or negative relationships between the carcase traits in both the data sets were also worth noting.

Table 6 *The percentage change in accuracy from univariate to bivariate BLUP for 5 milk traits using the 81 recorded cows*

Correlated traits		Genetic correlation	Residual correlation	Abs. difference	Percentage change in accuracy				
					Milk yield[1]	Fat %	Protein %	Fat weight	Protein weight
Milk yield	Fat weight	0.96	0.81	0.15	0			0.07	
Milk yield	Protein weight	0.98	0.92	0.06	0				0.09
Milk yield	Fat %	− 0.64	0.40	1.03	0.07	0.04			
Milk yield	Protein %	− 0.66	− 0.07	0.59	0.04		0.01		
Fat weight	Protein weight	0.99	0.84	0.15				0.02	0.01
Fat weight	Fat %	− 0.22	0.58	0.80		0		0.03	
Fat weight	Protein %	− 0.24	0.11	0.35			0.03	0	
Protein weight	Protein %	− 0.37	0.18	0.55			0.01		0.03
Protein weight	Fat %	− 0.71	0.55	1.26		0			0.14
Fat %	Protein %	0.65	0.45	0.20		0	0.01		

[1]All percentage change in accuracy values > 0 shown in this table were significantly different when tested using a two-tailed paired-comparison *t*-test (*P* < 0.05).

The results of analysing the Igenity scores (Table 5) resulted in very high heritability values (0.80 for fat weight to 0.95 for milk yield). Although much higher than those derived from phenotypic performance, these high heritability values are understandable as related animals will have similar SNP genotypes at the 123 SNPs used to derive the Igenity scores, and therefore relatives will have very similar genotypes. MacNeil *et al.* (2010) found similar high heritability values for Igenity scores of carcase traits, as did Tang *et al.* (2011) for a range of beef and milk traits. The genetic correlations between Igenity score and the same dairy trait recorded on the Gloucester cattle ranged from 0.12 (protein %) to 0.99 (fat weight). Tang *et al.* (2011) also reported the genetic correlation between milk production score and deregressed EBV of 0.14. In Table 5 this was 0.66.

*Accuracy of EBV using multi-trait BLUP*
One issue with rare breeds is the level of accuracy to which EBV are estimated owing to the small numbers of animals in the data set, in particular the small half-sib groups from sires. The use of bivariate and trivariate BLUP was investigated with the milk traits to see the effect on accuracy of including other traits in a multi-trait analysis. These results are summarised in Supplementary Table S4 for two groups of animals, recorded cows and the ancestors of recorded cows back three generations. Table 6 summarises the difference between the univariate and bivariate accuracies for recorded cows, along with the genetic and residual correlations between the five milk traits and their absolute difference.

The main objective of this paper was to see how genetic evaluations of a rare breed might be improved and the chosen method to evaluate this was the accuracy of the EBV estimates. Accuracy of an EBV is the correlation between the true breeding value (an unknown quantity) and the EBV (Henderson, 1975). It is defined as $\sqrt{(1 - C^{22}\alpha)}$ where C^{22} is the diagonal element of the inverse of the coefficient matrix used in the EBV mixed-model equations for any given animal and α is σ^2_e/σ^2_a, the ratio of the residual variance to

additive genetic variance (Mrode, 1996; this is a different use of α from that quoted earlier by Misztal *et al.*, 2009). If the variance ratio is fixed for all analyses of a given trait, then the change in accuracy is due to the change in the inverse of the coefficient matrix. Mrode (1996) suggests that multivariate BLUP analyses can lead to improvements in accuracy for any given trait provided that the absolute difference between the genetic and residual correlations of two traits is large. Table 6 indicates that both fat and protein % might contribute to the improvement in the accuracy of milk yield (absolute differences of 1.03 and 0.59, respectively), fat % might contribute to higher accuracy of fat weight (0.80 difference), fat % and protein % might improve the accuracy of protein weight (1.16 and 0.55 differences) and milk yield, and fat weight and protein weight (1.03, 0.80 and 1.16 differences) might improve the accuracy of fat %.

Taking milk yield as an example, Supplementary Table S4 shows that the mean accuracy of the recorded cows was 0.573 from a univariate BLUP analysis, but this changed to 0.636 when milk yield was analysed in a bivariate BLUP run with fat %. This is an increase of 0.07 (Table 6). The only other trait expected to increase the accuracy of milk yield was protein %, which had a mean accuracy for recorded cows of 0.609. Combining both fat and protein % with milk yield in a trivariate BLUP run further increased the accuracy of milk yield to 0.645. Thus, it was possible to increase the accuracy of milk yield EBV by including other traits in the analysis and this led to an increase in accuracy of 0.0720 or 12.5% of the univariate accuracy value. The mean accuracy of milk yield EBV for three generations of ancestors was 0.300 (Supplementary Table S4) and increased in a similar manner to that of the recorded cows when additional traits were added to the model. The accuracy of ancestors was lower than recorded cows because of the large effect of the animal's own record on accuracy (Falconer and Mackay, 1996) and older animals mostly did not have their own record.

Analyses of all five milk traits in the same manner are summarised in Table 6 for all bivariate combinations of
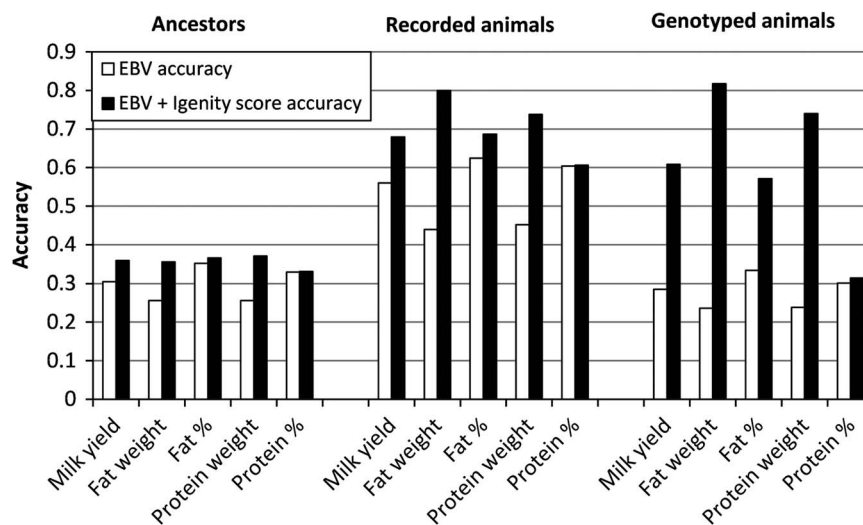
Figure 1 The mean accuracy of estimated breeding values (EBV) for recorded cows, ancestors and genotyped animals of milk traits estimated by BLUP and bivariate BLUP including the Igenity score for that trait. All comparisons of mean EBV accuracies with and without Igenity scores, within a trait and animal group, were significantly different when tested with a paired-comparison t-test ($P < 0.05$).

the traits. Mrode (1996) suggests that the improvement in accuracy between a univariate and bivariate analysis of a given trait depends on the absolute difference between the genetic and residual correlations between the two traits and their heritability, with the trait with a lower heritability benefitting more than the other trait. In this data set, fat and protein weight had lower heritabilities than milk yield, fat and protein % (Table 2). The results in Table 6 support Mrode's (1996) assertion to some extent but not completely. Two bivariate analyses had an absolute difference of more than 1: milk yield with fat %, and protein weight with fat %. Despite milk yield and fat % having similar heritabilities, milk yield accuracy improved by 7%, but fat yield only increased by 4% when analysed together. Protein weight accuracy improved by 14%, but fat % accuracy was unchanged when analysing these two traits together. Curiously, protein % accuracy increased by 9% when analysed with milk yield, despite the absolute difference between genetic and residual correlation being only 0.06.

Accuracy of EBV using GBLUP
A similar approach was used to investigate the use of single-step GBLUP to improve accuracy. These results are summarised in Supplementary Table S5 for recorded cows and for genotyped animals. There was very little difference between the mean accuracies of BLUP and GBLUP EBV for any trait or either group of animals shown. At first sight, this is a disappointing result, but the 101 SNP used to genotype the 199 animals only accounted for between 3% and 3.5% of the variation of any of the 11 recorded traits (see Supplementary Table S6). Thus, the SNP could not add very much information compared with the other sources.

Accuracy of EBV using bivariate analyses with Igenity score
A third method for improving accuracy was investigated that used Igenity scores and the recorded data in a series of

bivariate analyses. These results are shown in Supplementary Table S7 and summarised in Figure 1 for recorded cows, ancestors of recorded cows and genotyped animals. In this data set, the recorded cows were rarely genotyped and the genotyped animals rarely recorded; thus, for discussion purposes, they could be considered as two different groups of animals. The improvement in accuracy for the recorded cows when Igenity scores were included in a bivariate analysis was in relation to the correlation between the two traits, as shown in Table 5. Thus, the order of improvement was fat weight > protein weight > milk yield > fat % > protein %. The improvement in accuracy for recorded cows ranged from 12.6% to 0%. For the parents of recorded cows this ranged from 9.9% to 0.2% and for genotyped animals from 58% to 1.4%. Clearly, the use of Igenity scores improved the accuracy of genotyped animals much more than for recorded cows or the ancestors of recorded cows. However, for no trait did the level of accuracy of genotyped animals become greater than that of recorded cows. Incidentally, a small number of animals were both recorded and genotyped. Their accuracies increased when using the Igenity scores were also in line with the correlation shown in Table 5.

Genetic evaluation of rare breeds and EBV accuracy
There has been a growing recognition over the last 25 years that rare breeds of livestock are an important component of global biodiversity and play a critical role in sustainable development in all types of economies worldwide (Scherf, 2000). By their very nature rare breeds have certain critical characteristics that make the improvement of animals within these breeds more challenging than may be the case in 'mainstream' breeds. These include small numbers of animals, lower levels of investment, less precise estimates of genetic merit, small contemporary groups and, often, the small scale of the farming operation involved. Nevertheless, the owners of rare-breed animals have a legitimate need to

improve their animals for a number of reasons: economic, cultural, survival, moral and historical. This paper attempts to investigate how recent developments in molecular genetic methodology may be able to improve one aspect of rare-breed management, genetic evaluation of the animals, by using an example of a rare breed kept, recorded and managed in the United Kingdom.

The basic approach in this study has been to use data recorded on the farms maintaining animals from the Gloucester cattle breed in a series of analyses both with and without the use of molecular genetic information and seeing what differences the use of modern technology might make. To aid the comparison of results from different BLUP analyses, the variance used in multi-trait BLUP runs have been fixed at those found for each trait in the univariate analyses.

The first aspect of the accuracy calculations carried out here worth noting is that the conclusions drawn depend on the characteristics of the population and data used. Combinations of traits used in these calculations affected the changes in accuracy, in particular the genetic parameters of the population. For example, the effect of a second trait used in a multi-trait BLUP run depended on the genetic and residual correlations between the traits. Fat % had a large effect on the accuracy of milk yield, when used together, but protein % did not. The heritability of a trait in a population is another specific characteristic affecting the outcomes. Three traits had low to zero heritability and were therefore difficult or impossible to use in some of these analyses. Clearly, a trait with no genetic variation cannot be improved by selection, no matter how accurate the EBV.

The accuracy calculations carried out in this study confirmed many of the widely known relationships in genetic evaluations. First, an animal's own record is a valuable contributor to the accuracy of its EBV; any comparison between the mean accuracy of the recorded animals and that of their ancestors will confirm this. In addition, the absolute difference between the residual and genetic correlations between two traits largely indicates when improvements in accuracy will occur by using a bivariate BLUP run. Interestingly, this relationship does not work both ways. For example, milk yield and fat % had an absolute difference between the genetic and residual correlations of 1.03, but milk yield accuracy increased by 7% and fat % by only 4% when used together, compared with their univariate recorded-animal accuracies. Even more surprising was the combination of milk yield and protein weight; their correlation difference was only 0.06, yet protein weight accuracy improved by 9%, whereas milk yield was unchanged.

The use of the molecular genetic information provided some rather unexpected results. First, the use of GBLUP with this limited set of SNP did not improve the accuracy of any trait. However, the effect of the Igenity scores was marked with improvements in accuracy reaching 12.6% for fat weight in recorded cows and 58.2% for genotyped animals. For the 12 cows that were both recorded and genotyped, their mean accuracy for fat weight went from 0.44 to 0.80 (univariate BLUP *v.* bivariate BLUP with Igenity score; data not shown). The response in other traits was dependent on the genetic correlation between the Igenity scores and the recorded traits; protein % showed no or very little change in accuracy, depending on the group of animals considered, because their genetic correlation was only 0.12. This contrasts with the value of 0.99 for the genetic correlation between the two measurements of fat weight genetic merit.

The reason for this contrast in the outcome of using Igenity scores and GBLUP to improve accuracy is worth considering. The SNP set used here only accounted for 3.5%, or less, of the genetic variation found in these traits. Its small effect in GBLUP is therefore not surprising. The effect of the SNP genotypes could have been artificially bolstered by setting the 'Misztal α' to a higher value, say 40%. This would have given a different balance to the information coming from the SNP set and the pedigree. The use of the Igenity scores in bivariate analyses may have effectively done this, as the scores, on a 1 to 10 basis, were the only measure of the trait used and may well have amplified the importance of the SNP genotypes used in each Igenity score's calculation. Another aspect may be that the small set of animals genotyped may not have estimated the effect of the individual SNP genotypes very well in GBLUP, whereas the effect of the same genotypes when used in the Igenity score may have been calculated by Merial from a much larger set of data. A larger Gloucester data set and more genotyped SNPs (say using the Illumina 50K Beadchip) may help improve this situation.

Conclusions and recommendations for a rare breed

These analyses have used the milk traits as a model for investigating a range of methods for increasing the accuracy of EBV calculated on a subset of animals in a rare breed. The first recommendation is that more animals should be recorded. This should improve the genetic evaluation of the breed in two ways; first, the genetic links created by an excellent pedigree structure and recording within the breed would be better utilised. Second, the best way to increase the accuracy of an animal's EBV is to record it for that trait.

The 11 traits used in this study are all valuable to the breed and should be recorded on as many animals as possible. Future analyses might consider growth and carcase characteristics of the young animal separately from that of the older animals. However, as this is a rare breed, the production of beef from young animals may be a luxury that cannot yet be implemented.

The decision about whether to carry on genotyping animals may revolve around the cost of genotyping. The analyses carried out here have shown that genotyping unrecorded animals has had a marked affect on the accuracy of some traits more than others. This approach could be extended to the carcase traits recorded here, but not analysed, once more data have been collected. The outcome would depend on the genetic correlation between the recorded trait and the equivalent Igenity score.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/doi:10.1017/S1751731114000330.

References

Banos G, Coffey MP, Veerkamp RF, Berry DP and Wall E 2012. Merging and characterising phenotypic data on conventional and rare traits from dairy cattle experimental resources in three countries. Animal 6, 1040–1048.

Christensen OF and Lund MS 2010. Genomic prediction when some animals are not genotyped. Genetics, Selection and Evolution 42, 2.

English Beef and Lamb Executive 2012. Marketing prime beef cattle for better returns. English Beef and Lamb Executive, Stoneleigh, UK.

Falconer DS and Mackay TFC 1996. Introduction to quantitative genetics, 4th edition. Longman, Harlow, UK.

Gilmour AR, Gogel BJ, Cullis BR and Thompson R 2009. ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK. www.vsni.co.uk

Gloucester Cattle Society 2013. History. Retrieved August 14, 2013, from http://www.gloucestercattle.org.uk/

Henderson CR 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423–447.

Holstein UK 2013. Breed performance statistics, Holstein UK, 1999–2009. Retrieved August 14, 2013, from http://www.ukcows.com/thecdi/docs/BrdStats2012_Summary HOL.pdf

Kadarmideen HN, Thompson R and Simm G 2000. Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. Animal Science 71, 411–419.

Kempster AJ and Southgate JR 1984. Beef breed comparisons in the UK. Livestock Production Science 11, 491–501.

Kinghorn B 2012. Pedigree Viewer. Retrieved August 14, 2013, from http://www-personal.une.edu.au/~bkinghor/pedigree.htm

MacNeil MD, Northcutt SL, Schnabel RD, Garrick DJ, Woodward BW and Taylor JF 2010. Genetic correlations between carcass traits and molecular breeding values in Angus cattle. Proceedings 9th World Congress on Genetics Applied to Livestock Production, Leipzig, pp. 2–148.

Merial Ltd 2008. Igenity results key for dairy. Merial Ltd, Duluth, GA.

Merial Ltd 2009. Igenity results key for beef. Merial Ltd, Merial Canada Inc, Montreal.

Misztal I, Legarra A and Aguilar I 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. Journal of Dairy Science 92, 4648–4655.

Mrode RA 1996. Linear models for prediction of animal breeding values. CAB International, Chippenham, UK.

Oklahoma State University 2013. Retrieved August 14, 2013, from http://www.ansi.okstate.edu/breeds/cattle/

RBST 2013. RBST Fact Sheet – Gloucester. Retrieved August 14, 2013, from https://www.rbst.org.uk/sitemanager/uploads/ck_files/files/Gloucester%20-%20Fact%20Sheet.pdf

Ríos-Utrera A 2004. Genetic evaluation of carcass traits: looking at the effects of slaughter endpoints. Proceedings of the Beef Improvement Federation, 36th Annual Research Symposium and Annual Meeting, Sioux Falls, SD. Iowa State University, Ames, pp 79–103.

Scherf BD 2000. WorldWatch list for animal diversity, 3rd edition. FAO, Rome, Italy.

Tang G, Li X, Plastow G, Moore SS and Wang Z 2011. Developing marker-assisted models for evaluating growth traits in Canadian beef cattle genetic improvement. Livestock Science 138, 62–68.

Van Eenennaam AL, Li J, Thallman RM, Quaas RL, Dikeman ME, Gill CA, Franke DE and Thomas MG 2007. Validation of commercial DNA tests for quantitative beef quality traits. Journal of Animal Science 85, 891–900.