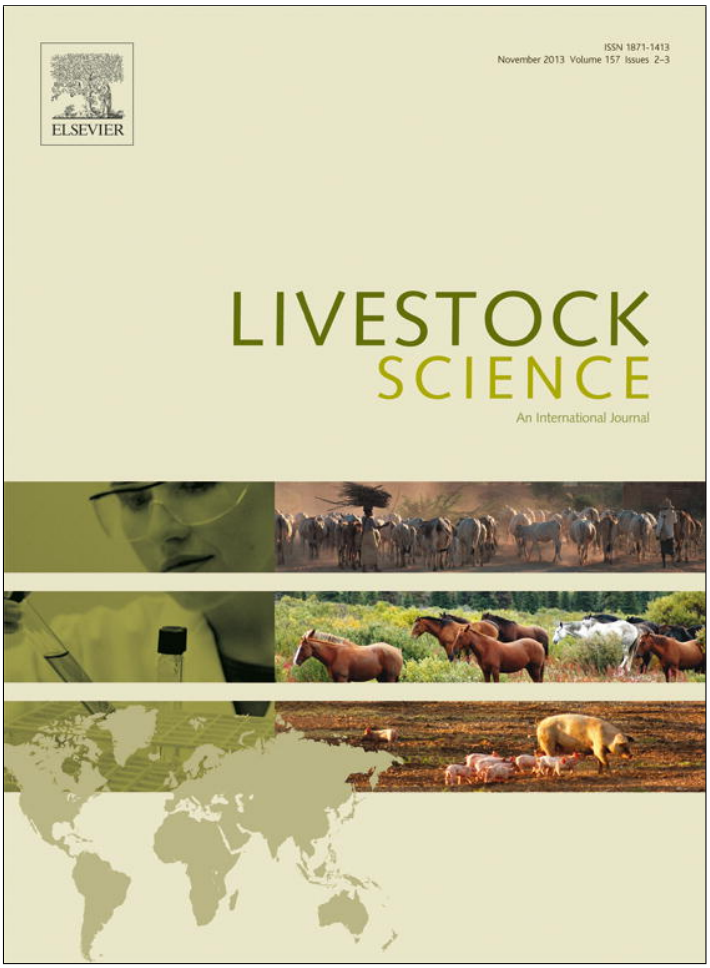
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/authorsrights

Author's personal copy

Livestock Science 157 (2013) 421–426

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Short communication

Accuracy of genomic prediction for milk production traits with different approaches in a small population of Slovenian Brown bulls

M. Špehar [a,b,*], K. Potočnik [b], G. Gorjanc [b]

[a] Croatian Agricultural Agency, Ilica 101, Zagreb 10 000, Croatia
[b] University of Ljubljana, Biotechnical Faculty, Animal Science Department, Groblje 3, Domžale 1 230, Slovenia

ARTICLE INFO

Article history:
Received 26 October 2012
Received in revised form
2 August 2013
Accepted 7 August 2013

Keywords:
Genomic selection
Breeding value
Accuracy
Bivariate model

ABSTRACT

Use of genome-wide single-nucleotide polymorphism (SNP) data enables accurate estimates of breeding value (EBV) for young animals when a sufficiently large number of animals in a population are genotyped. The objective of this study was to evaluate the accuracy of integrating genomic data into the national evaluation system for milk traits in a small population of Slovenian Brown bulls using univariate national evaluation based on phenotype and pedigree data (U), international direct genomic value (DGV), and bivariate national evaluation incorporating DGV as a correlated trait (B). Comparison of approaches was assessed separately for training and validation subset of bulls using theoretical and empirical accuracy. Genetic correlation between the phenotype based EBV and DGV was between 0.79 and 0.86 confirming the utility of DGV for prediction. Use of DGV did not improve already high accuracy (0.98) for proven bulls due to the substantial number of daughters per bull. In young bulls, inclusion of DGV in B analysis has increased theoretical accuracy of prediction from 0.58 to 0.89 and further from 0.92 to 0.96 when these bulls were progeny tested. Empirical correlations on the validation subset confirmed the observed increases in theoretical accuracy although values were considerably lower due to the low and variable number of daughters per bull in the validation subset. When combining both progeny and DGV data, correlation between the U and B evaluation was 0.92 in validation subset confirming the usefulness of integrating both data sources. Integration of all the available information is not only beneficial for the use of all the data, but also to simplify publication since all information can be combined in a single breeding value.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The advantage of using genome-wide single-nucleotide polymorphism (SNP) data in dairy cattle is to increase the accuracy of estimated breeding value (EBV) for young non-phenotyped animals allowing to shorten generation interval and to increase genetic gain per year (Schaeffer, 2006). The highest increase in accuracy can be expected in populations with a large number of genotyped and phenotyped animals (Goddard, 2009). When populations are small prediction equations for EBV using SNP data can be developed on the experimental populations of reasonable size and later blended with the national evaluation system as proposed in beef cattle populations (Johnston et al., 2012; MacNeil et al., 2010). An alternative is to develop prediction equations on a larger combined international data as in Brown breed, where an international reference

* Corresponding author at: Croatian Agricultural Agency, Ilica 101, Zagreb 10 000, Croatia. Tel.: +385 1 3903 176; fax: +385 1 3903 192.
E-mail address: mspehar@hpa.hr (M. Špehar).

1871-1413/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.livsci.2013.08.008

Author's personal copy

population had been setup with almost 8000 bulls from Austria, France, Germany, Italy, Slovenia, Switzerland, and the United States of America via the InterGenomics consortium (Jorjani et al., 2012).

Within the scope of the InterGenomics consortium specific prediction equations for each country are developed based on all available SNP data and multiple across country evaluation (MACE) EBV on a country's specific scale (Jorjani et al., 2012). Each bull has therefore several evaluations at the national level (parent average for young bulls and progeny test EBV for proven bulls) and at the international level (SNP genotype based EBV called direct genomic value (DGV) and a blended genomically enhanced breeding value of DGV and parent average EBV for young bulls and progeny test MACE EBV for proven bulls). All these evaluations complicate the publication of results and inhibit the use of all the available data.

The objective of this study was to evaluate the accuracy of progeny and genomic evaluations in a small population of Slovenian Brown bulls utilizing all the available data. Specifically, the accuracy of evaluations was assessed using: univariate national evaluation based on phenotype and pedigree data, international DGV, and bivariate national evaluation incorporating international DGV as a correlated trait.

2. Materials and methods

2.1. Phenotype and genotype data

Phenotypic data from the Slovenian Brown breed population consisted of 1,342,134 test-day records for milk, fat, and protein yield from 56,670 cows that were progeny of 736 bulls. Among these, 184 bulls born between 1990 and 2007 were genotyped with the Illumina 50 K BovineSNP chip. In addition, international direct genomic value (DGV) was available from the InterGenomics consortium for the analyzed traits (Jorjani et al., 2012). This information was considered for 399 bulls that had daughters in the national data, namely for 184 Slovenian and 215 foreign bulls (Table 1).

2.2. Evaluation methods

Based on the available data the following approaches were used for the evaluation of EBVs. The first approach was univariate repeatability test-day model (U) based on the national phenotypic and pedigree data:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}_a\boldsymbol{a} + \boldsymbol{Z}_c\boldsymbol{c} + \boldsymbol{Z}_p\boldsymbol{p} + e,$$

Table 1
Description of data sets used in evaluation.

Data set	No. of bulls	Birth year	Origin of bulls with direct genomic value	
			Slovenian	Foreign
Complete	736	1990–2007	184	215
Training	701	1990–2003	149	215
Validation	35	2004–2007	35	–

where $\boldsymbol{y}$ is a vector of phenotypic observations, $\boldsymbol{b}$ is a vector of parameters for fixed effects, $\boldsymbol{a} \sim N(0, \boldsymbol{A}\sigma_a^2)$ is a vector of parameters for breeding values with pedigree relationship matrix $\boldsymbol{A}$, $\boldsymbol{c} \sim N(0, \boldsymbol{I}\sigma_c^2)$ is a vector of parameters for herd effect, $\boldsymbol{p} \sim N(0, \boldsymbol{I}\sigma_p^2)$ is a vector of parameters for permanent environment effect, and $\boldsymbol{e} \sim N(0, \boldsymbol{I}\sigma_e^2)$ is a vector of residuals, while $\boldsymbol{X}$, $\boldsymbol{Z}_a$, $\boldsymbol{Z}_c$, and $\boldsymbol{Z}_p$ are incidence matrices linking $\boldsymbol{y}$ and $\boldsymbol{b}$, $\boldsymbol{a}$, c, and $\boldsymbol{p}$.

The second approach was based on the bivariate model (B) combining the U approach and DGV as a correlated trait into a national estimate (Kachman, 2008; Mäntysaari and Strandén, 2010):

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{dgv} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_y & 0 \\ 0 & \boldsymbol{X}_{dgv} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \mu \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{a_y} & 0 \\ 0 & \boldsymbol{Z}_{a_{dgv}} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_y \\ \boldsymbol{a}_{dgv} \end{bmatrix} + \boldsymbol{Z}_{c_y}\boldsymbol{c}_y + \boldsymbol{Z}_{p_y}\boldsymbol{p}_y + \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_{dgv} \end{bmatrix},$$

where DGV is incorporated via covariance $\sigma_{a_y,a_{dgv}}$:

$$\begin{bmatrix} \boldsymbol{a}_y \\ \boldsymbol{a}_{dgv} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{a_y}^2 & \sigma_{a_y,a_{dgv}} \\ \sigma_{a_y,a_{dgv}} & \sigma_{a_{dgv}}^2 \end{bmatrix} \otimes \boldsymbol{A} \right).$$

Estimation of required (co)variance components was performed with the residual maximum likelihood method. Conditional on these estimates, breeding values were evaluated. All computations were performed with VCE (Kovač et al., 2002). In addition to the U and B approach, we have used the international DGV as a direct predictor of a national proof.

2.3. Validation of evaluations

For the purpose of validation, the complete data was divided into training and validation subset by removing daughter phenotypes in the validation subset (Table 1). Training subset was comprised of 701 bulls born before the year 2004. Validation subset was comprised of 35 genotyped bulls born in the years between 2004 and 2007. Evaluations using the U and B approaches and different data sets were named: U_t – the U approach using the training subset, U_c – the U approach using the complete data set; B_t – the B approach using the training subset, and B_c – the B approach using the complete data set. EBVs evaluated with a certain approach (e.g., U_c) were always obtained for all bulls and were denoted accordingly (e.g., $EBV_{U,c}$). Validation of predictions was assessed via theoretical and empirical accuracies separately for proven and validation bulls and separately for different evaluations ($EBV_{U,t}$, $EBV_{U,c}$, $EBV_{B,t}$, $EBV_{B,c}$). Theoretical accuracy was evaluated from the variance of prediction errors and additive genetic variance in base generation, while empirical correlations were computed as a correlation between EBVs from different evaluations.

3. Results and discussion

Heritability estimates were 0.28 for milk, 0.21 for fat, and 0.25 for protein yield and were the same for both the U and the B approach (Appendix Table A1). In the B approach, heritabilities for DGV were equal to 1.00 (after

Author's personal copy

rounding) since DGV is in principle a fully heritable trait with a complete penetrance due to the same SNP equation being used for all animals. Genetic correlation between the phenotype based EBV and DGV was 0.86 for milk, 0.80 for fat, and 0.79 for protein yield.

These results suggest that DGV is a useful early predictor. Applications of such a bivariate approach are not present in the literature for dairy traits. For beef populations, lower genetic correlations between phenotypic based EBV and DGV are found. The average of reported values was 0.48 in Angus (Saatchi et al., 2011), 0.50 in Simmental (Saatchi et al., 2012), and 0.55 in Limousine breed (Saatchi et al., 2012). These correlations are lower than in our case which can be attributed to several factors. Training set used for the development of DGV equation was larger in the InterGenomics consortium. In addition, dairy populations tend to have a smaller effective population size with stronger linkage disequilibrium leading to tighter linkage between SNP markers and potential QTLs (Goddard, 2009). Finally, part of the data used for the development of DGV equation was also used in this study, which leads to double counting of data in the presented analysis. However, the number of such bulls is small compared to the total number of bulls in the InterGenomics consortium leading to the negligible amount of double counting. Mäntysaari and Strandén (2010) proposed a method for the correction of double counting for the B approach and found that required corrections tend to be small. Therefore, the issue of double counting was neglected in this analysis. The use of different sources of information can be performed via the modification of a prior distribution for the additive genetic effect. In this way, correction for double counting can be applied directly opening opportunity to improve current analyses (Vandenplas and Gengler, 2012).

The average theoretical accuracy of proven bulls was 0.98 for all milk traits using the U_c approach and the inclusion of DGV via the B_c approach did not lead to the significant increase in the accuracy (Table 2). This can be attributed to the sizeable number of daughters per bull (231 on average). However, for a small number of bulls with a low number of daughters, a significant increase in accuracies was observed. The average theoretical accuracy of parent average information for bulls in the validation set (U_t) was 0.58 for milk and protein yield, and 0.52 for fat yield (Table 2). The inclusion of DGV in the B_t analysis led to an increase in the accuracy up to 0.89. Further increases in the theoretical accuracy were achieved using both, the U_c and B_c analyses, when these validation bulls were progeny tested.

In line with the theoretical accuracies in proven bulls there were also empirical correlations between EBVs from different approaches (Fig. 1). For proven bulls high correlations were observed between the national phenotype based evaluation ($EBV_{U,c}$) and genomic prediction using DGV (0.98) for milk yield. With the B approach correlations were similar. Results for fat and protein yield were similar to those of milk yield (Appendix Fig. A1 and A2). Since proven bulls had a sufficient number of daughters, blending of DGV with phenotype evaluation did not change their EBV significantly.

Empirical correlations in a set of validation bulls were lower. Correlation between parent average ($EBV_{U,t}$) and progeny evaluation ($EBV_{U,c}$) was 0.49 in milk yield (Fig. 1). Using DGV as a sole predictor had better correlation (0.56) with progeny evaluation than pedigree prediction via parent average. Inclusion of DGV in the B_t approach led to a small increase in correlation between the progeny evaluation and prediction via the genomically enhanced B_t approach ($cor(EBV_{U,c}, EBV_{B,t})=0.61$). Such a low correlation can be attributed to the low number of daughters per bull in the validation set, decreasing the correlation between these two evaluations due to the initial low accuracy of the U_c evaluation (Table 2). Finally, when combining both progeny and DGV data correlation between the $EBV_{U,c}$ and $EBV_{B,c}$ was 0.92. Similar trends in empirical correlations in the validation set of bulls were observed also for fat and protein yield, with higher values for fat yield and lower values for protein yield (Appendix Fig. A1 and A2).

Empirical correlations show the benefit of combining all the available information via the B approach in the analyzed population characterized by a small population size and a small number of daughters per tested bull. The availability of genomic information from a larger consortium provided the opportunity to implement genomic prediction in this population even though the size of own reference population is not sufficiently large. This paves the way of integrating the external genomic information into other small populations of dairy cattle with

Table 2
Average theoretical accuracies (minimum and maximum) for milk traits using different approach and data for a group of proven and validation bulls.

Bulls	No. of daughters	Data set[a]	Accuracy		
			Milk yield	Fat yield	Protein yield
Proven ($n=149$)	231 (40–1067)	U_c	0.98 (0.52–0.99)	0.98 (0.43–0.99)	0.98 (0.51–0.99)
		B_c	0.99 (0.87–0.99)	0.98 (0.84–0.99)	0.99 (0.83–0.99)
Validation ($n=35$)	57 (1–213)	U_t	0.58 (0.30–0.68)	0.52 (0.38–0.64)	0.58 (0.29–0.68)
		B_t	0.89 (0.86–0.91)	0.86 (0.82–0.88)	0.85 (0.81–0.88)
		U_c	0.92 (0.62–0.99)	0.90 (0.56–0.99)	0.91 (0.61–0.99)
		B_c	0.96 (0.89–0.99)	0.95 (0.86–0.99)	0.95 (0.85–0.99)

[a] U_c – phenotypic and pedigree data used in the national genetic evaluation on complete data set; B_c – bivariate analysis based on U_c and DGV for all genotyped bulls in the national pedigree on complete data set; U_t – phenotypic and pedigree data used in the national genetic evaluation on training subset; B_t – bivariate analysis based on U_t and DGV for all genotyped bulls in the national pedigree on training subset.

Author's personal copy

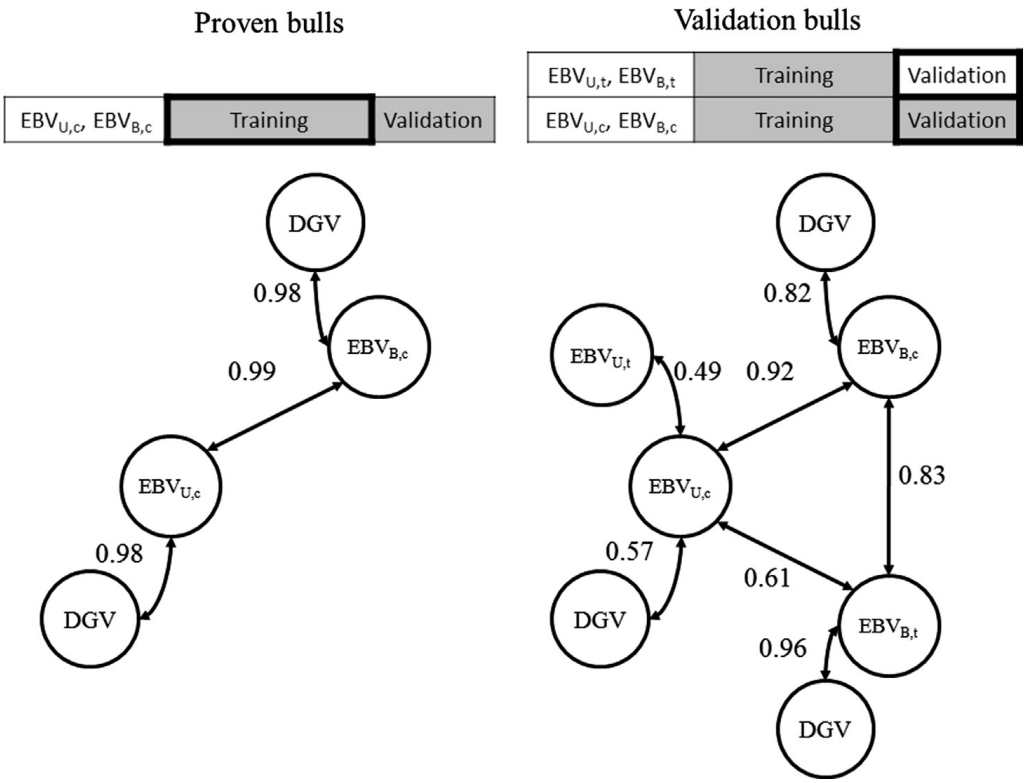


Fig. 1. Empirical correlations between EBVs from different approaches and dataset for proven and validation bulls for milk yield.

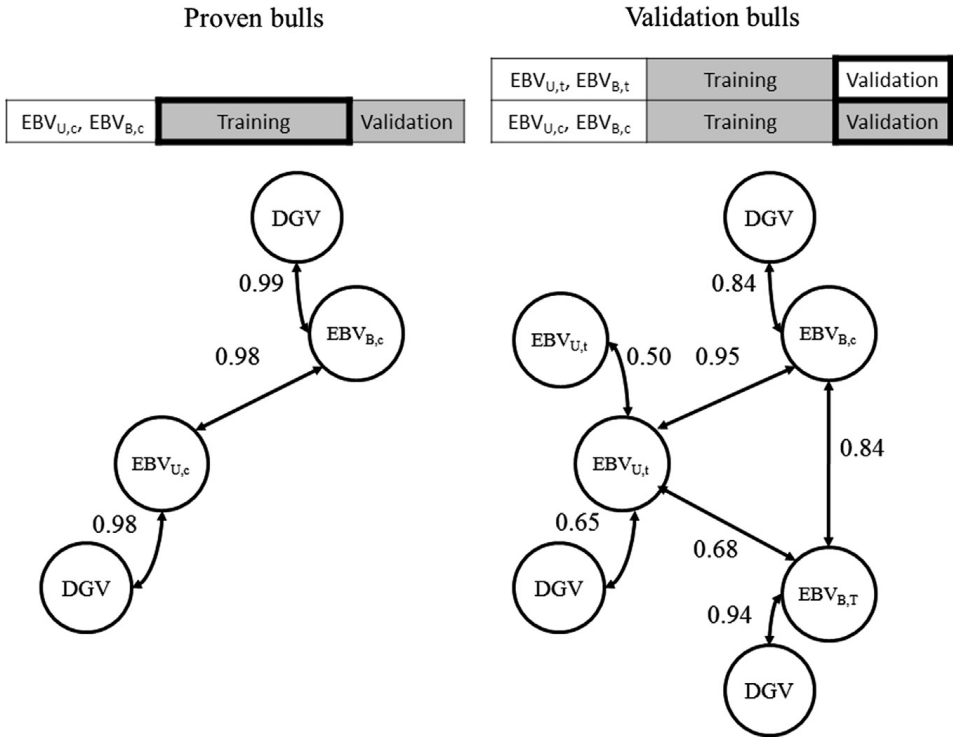


Fig. A1. Empirical correlations between EBVs from different approaches and dataset for proven and validation bulls for fat yield.

the national evaluation system. Additional benefit of the B approach is the unified system integrating all the available data in a single EBV instead of reporting several EBV from different data sources for the same animal.

4. Conclusion

Different approaches of integrating genomic information into a national evaluation system for small population

Author's personal copy

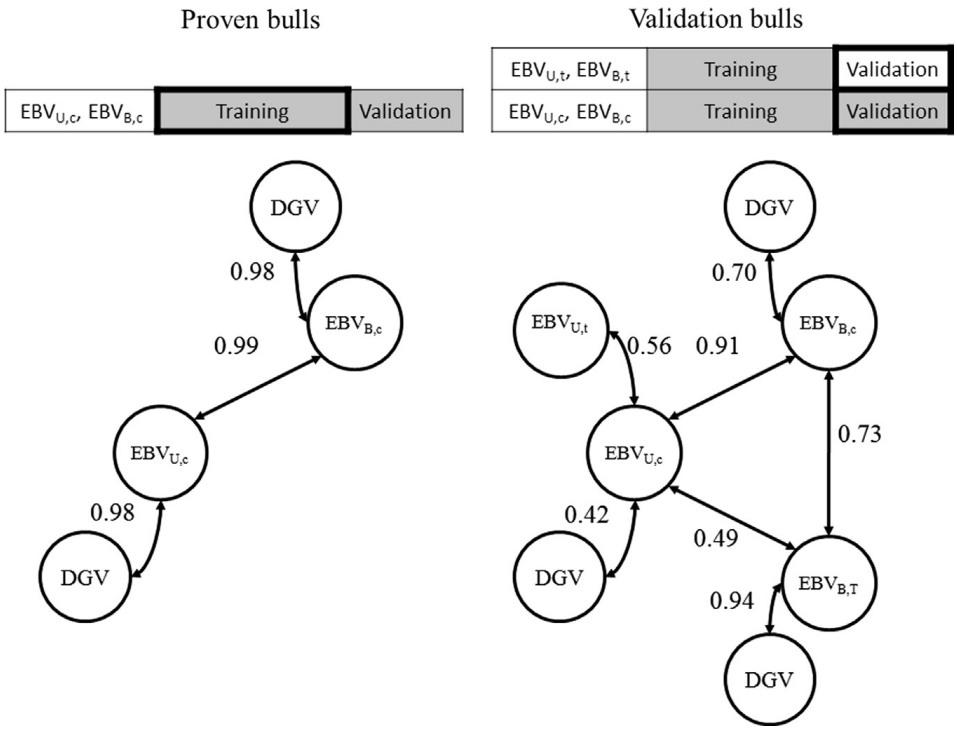


Fig. A2. Empirical correlations between EBVs from different approaches and dataset for proven and validation bulls for protein yield.

Table A1
Estimates of (co)variance component ratios and correlations (± standard errors) for breeding values based on the national phenotype information (h^2_y), breeding values based on the direct genomic value(h^2_{dgv}), herd (c), permanent environment (p) effects, and residual (e) using two approaches[a] (U_c, B_c) and complete data set.

Item	Trait					
	Milk yield		Fat yield		Protein yield	
	U_c	B_c	U_c	B_c	U_c	B_c
h^2_y	0.28 ± 0.003	0.28 ± 0.003	0.21 ± 0.001	0.21 ± 0.002	0.25 ± 0.003	0.25 ± 0.003
h^2_{dgv}	–	1.00 ± 0.000	–	1.00 ± 0.000	–	1.00 ± 0.000
c^2_y	0.22 ± 0.005	0.23 ± 0.005	0.21 ± 0.001	0.21 ± 0.005	0.28 ± 0.005	0.28 ± 0.006
p^2_y	0.18 ± 0.002	0.18 ± 0.002	0.15 ± 0.001	0.15 ± 0.001	0.15 ± 0.002	0.15 ± 0.001
e^2_y	0.32 ± 0.002	0.32 ± 0.002	0.43 ± 0.001	0.43 ± ± 0.003	0.32 ± 0.003	0.32 ± 0.003
$Cor(a_y, a_{dgv})$	–	0.86 ± 0.030	–	0.80 ± 0.030	–	0.79 ± 0.003
$Cor(e_y, e_{dgv})$	–	0.00 ± 0.000	–	0.00 ± 0.000	–	0.00 ± 0.000

[a] U_c – phenotypic and pedigree data used in the national genetic evaluation; B_c – bivariate analysis based on national genetic evaluation (U_c) and DGV for all genotyped bulls in the national pedigree.

of Brown breed were evaluated in comparison to the conventional evaluation based on phenotype and pedigree data. Results indicate that integration of DGV from a large consortium into the national evaluation as a correlated trait enabled combination of all the available data. In addition, this approach provides a way to automatically blend all the results in a single value avoiding the need to publish several estimates of breeding values per animal.

Conflict of interest statement

There are no known conflicts of interest.

Appendix A

See Figs. A1 and A2.
See Table A1.

References

Goddard, M.E., 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245–257.
Johnston, D.J., Tier, B., Graser, H.U., 2012. A beef cattle breeding in Australia with genomics: opportunities and needs. Anim. Prod. Sci. 52, 100–106.
Jorjani, H., Jakobsen, J., Hjerpe, E., Palucci, V., Dürr, J., 2012. Status of genomic evaluation in the Brown Swiss populations. Interbull Bull. 46, 46–54.

Author's personal copy

Kachman, S.D., 2008. Incorporation of marker scores into national genetic evaluations. In:Proceedings of the 9th Genetic Prediction Workshop Beef Improvement Federation, Kansas City, Missouri, pp. 92–98.

Kovač, M., Groeneveld, E., Garcia Cortes, L., 2002. VCE-5, a Package for the Estimation of Dispersion Parameters. In: Proceedings of the 7th WCGALP, Montpellier, France, pp.741–742.

MacNeil, M.D., Northcutt, S.L., Schnabel, R.D., Garrick, D.J., Woodward, B. W., Taylor, J.F., 2010. Genetic correlations between carcass traits and molecular breeding values in Angus cattle. In: Proceedings of the 9th WCGALP, Leipzig, Germany, pp. 482–485.

Mäntysaari, E.A., Strandén, I., 2010. Use of bivariate EBV-DGV model to combine genomic and conventional breeding values evaluations. In: Proceedings of the 9th WCGALP, Leipzig, Germany, pp. 353–356.

Saatchi, M., Schnabel, R.D., Rolf, M.M., Taylor, J.F., Garrick, D.J., 2012. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. Genet. Sel. Evol. 44, 38.

Saatchi, M., McClure, M.C., McKay, S.D., Rolf, M.M., Kim, J.W., Decker, J.E., Taxis, T.M., Chapple, R.H., Ramey, H.R., Northcutt, S.L., Bauck, S., Woodward, B., Dekkers, J.C.M., Fernando, R.L., Schnabel, R.D., Garrick, D.J., Taylor, J.F., 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet. Sel. Evol. 43, 40.

Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123, 218–223.

Vandenplas, J., Gengler, N., 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. J. Dairy Sci. 95, 1513–1526.