Joint genomic evaluation of cows and bulls with BayesD for prediction of genotypic values and dominance deviations

Robin Wellmann[1], Johann Ertl[2], Reiner Emmerling[2], Christian Edel[2], Kay-Uwe Götz[2] and Jörn Bennewitz[1].
[1]Institute of Animal Husbandry and Breeding, University Hohenheim, Germany
[2]Bavarian State Research Center for Agriculture, Institute of Animal Breeding, Grub, Germany

ABSTRACT: The inclusion of dominance effects into models for genomic selection enables the utilization of non-additive genetic variance in mate selection programs. Usually, only genotyped individuals with own phenotypes are used for prediction. This study investigated whether the accuracies of genotypic values, dominance deviations, and breeding values can be increased by a joint evaluation of genotyped individuals with phenotyped offspring and with own phenotypic records. Daughter yield deviations for fat yield from 6,858 Fleckvieh bulls and yield deviations from 1,986 cows were used to estimate marker effects of 629,028 loci with BayesD. The correlation between estimated genotypic values (EGV) and yield deviations (YD) increased by a joint evaluation of cows and bulls. BayesD2 with a moderately heavy tailed distribution of marker effects provided the highest correlation between EGV and YD (0.398), which was 0.027 larger than for G-BLUP with dominance.
Keywords: BayesD; Genomic selection; Prediction of genotypic values

Introduction

With the availability of high density genotypes from thousands of individuals, the interest of breeders in genome wide evaluations for optimization of mate allocation increases. This requires the inclusion of non-additive marker effects into models for genomic selection. Several models for estimating additive and dominance effects have been proposed. Modelling these effects as independent leads for many loci to overdominant alleles, which is not considered to be realistic for most traits. Overdominant alleles are not required to explain a considerable amount of dominance variance (Wellmann and Bennewitz (2011)). Indeed, including dominance effects in genomic selection models may result in equal or lower prediction accuracies if the additive and dominance effects are modelled independently (Zhao et al. (2013)). Wellmann and Bennewitz (2012) showed for simulated data that including dominance effects into models for genomic selection can increase the accuracy of genotypic values considerably if the model makes correct assumptions about the joint distribution of additive effects and dominance effects. This requires, however, many phenotyped and genotyped individuals, a high density marker panel, and prior knowledge about the true joint distribution of additive and dominance effects. By using the scaling argument of Meuwissen (2009) they predicted that 30.000 individuals with phenotypes are needed to get an accuracy of the dominance deviation of approx. 0.7. The number of genotyped females with phenotypic records in Fleckvieh cattle, however, is still much smaller.

The aim of this study was to investigate, whether the accuracy of genotypic values and dominance deviations can be increased by a joint evaluation of bulls and cows. Additionally, the sensitivity of the BayesD models with respect to the choice of the hyper-parameters was studied.

Materials and Methods

Yield deviations (YD) for fat yield (FY) from 1986 genotyped Bavarian Fleckvieh cows, and daughter yield deviations (DYD) from 6858 genotyped Fleckvieh bulls were analysed with BayesD. The cows and 1485 bulls were genotyped with the Illumina BovineHD Genotyping BeadChip. The remaining bulls were imputed with fImpute (Sargolzaei et al. (2011)). After quality control, 629,028 loci remained in the data set. The effective numbers of own performances (EOP) were used as weights for the YD and the DYD.

The BayesD model described in Wellmann and Bennewitz (2012) was extended to allow for the joint evaluation of bulls and cows. For the cows we used the model

$$y_1 = \mu_1 1 + Z_{A1}a + Z_{D1}d + E_1,$$

where y_1 is the vector with YD of the analysed trait, the fixed effect μ_1 is the intercept for the YD, a is the vector with additive marker effects, d is the vector with the dominance effects of the markers, Z_{A1} is the gene content matrix for the cow genotypes with entries 0, 1, and 2, Z_{D1} is the indicator matrix for heterozygosity, and the vector E_1 with residuals was normally distributed with $V(E_1) = \sigma_1^2 F_1$ (given σ_1^2), where F_1 is a diagonal matrix with the reciprocals of the EOP in the diagonal.
For the bulls we used the model

$$y_2 = \mu_2 1 + Z_{A2}a + Z_{D2}d + E_2,$$

where y_2 is the vector with DYD, μ_2 is the intercept for the DYD, Z_{A2} is the gene content matrix for the bull genotypes with entries 0, 1, and 2, and E_2 is a normally distributed vector with covariance matrix $V(E_2) = \sigma_2^2 F_2$, where F_2 is a diagonal matrix with the reciprocals of the EOP in the diagonal. The matrix Z_{D2} has a different definition for DYD than for YD. Take $Z_{A2(j)}$ and $Z_{D2(j)}$ to be the *j*th column of matrix Z_{A2} and

Z_{D2}. Since the contribution of a QTL j to the breeding value of an individual is $Z_{A2(j)}(a_j + (q_j - p_j)d_j)$, where p_j is the frequency of the 1-allele, and q_j is the frequency of the 0-allele, we defined $Z_{D2(j)} = Z_{A2(j)}(q_j - p_j)$.

The joint prior distribution of the additive effect a_j and the dominance effect d_j of marker j is described in detail in Wellmann and Bennewitz (2012). It is a mixture of two distributions with different variances. The probability of a marker effect to come from the distribution with large variance is p_{LD}. The absolute additive effect $|a_j|$ comes from a mixture of two folded t-distributions with $\nu > 2$ degrees of freedoms. The smaller ν is, the more heavy tailed is the distribution of the marker effects. For $\nu \to \infty$, the t-distribution approaches the normal distribution, so for $p_{LD} = 1$, the models BayesD0 without dominance effects and the model BayesD1 with dominance effects approaches the G-BLUP models. Therefore, results from BayesD0 and BayesD1 using $p_{LD} = 1$ and $\nu = 50$ are called the G-BLUP results in this paper.

The parameter α determines the prior probability that additive and dominance effect have the same sign. For $\alpha = 0$ they are independent, whereas for $\alpha \approx 1$,

$$sign(a_j) = -sign(q_j - p_j)sign(d_j)$$

holds with high probability, which reflects the assumption that it is unlikely that a QTL has a frequency for which its contribution to the additive variance is large. For BayesD2 the dominance coefficient $\delta_j = d_j / |a_j|$ is normally distributed with mean $E(\delta_j)$ and variance $Var(\delta_j)$.

The hyper-parameters of the model were computed from the variance components V_A and V_D estimated by Ertl et al. (2013) from the same data. These authors estimated larger dominance variance than commonly reported in the literature. In order to be conservative, we assumed that the dominance variance was half as large as published in Ertl et al. (2013). For computation of the hyper-parameters, an estimate of the inbreeding depression was also needed. It is defined as the expected decrease of the trait value if the inbreeding coefficient increases from 0 to 1. In accordance with Miglior et al. (1995) we assumed an inbreeding depression of 2 phenotypic standard deviations.

The cow data set was split into a validation set consisting of 399 cows and an estimation set consisting of the remaining 1587 cows. Marker effects were estimated from 5,000 iterations of the MCMC algorithm.

Results and Discussion

Table 1 shows the correlations of estimated genotypic values (EGV), breeding values (EBV), and dominance deviations (EDV) with the yield deviations of the cows in the validation set. The correlation between EGV and YD increased by a joint evaluation of cows and bulls. This increase, however, is mainly due to more accurately estimated breeding values. For FY, BayesD2 models assuming a moderately heavy tailed distribution for the marker effects and $\alpha \approx 1$ provided the highest correlation between EDV and YD. It is possible that the two outlier results with Cor(EGV,FY)<0.32 would improve with a larger number of iterations.

It was expected that for BayesD models with a heavy tailed distribution of marker effects, the inclusion of bulls increases the correlation between YD and EDV because BayesD2 assumes that only loci with a non-zero additive effect can have dominance effects. Thus, a more precise estimation of substitution effects using bull records reduces the number of loci for which dominance effects need to be estimated. However, increased accuracies were not clearly observed in the results. We are currently working on a faster algorithm with a better mixing of the Markov chain. It is possible that this would improve the results for such data sets with large numbers of markers. We will apply the improved algorithm to other traits.

Estimated marker effects for FY from the joint evaluation of cows and bulls are shown in Figure 1, whereby the effects were averaged over all results from BayesD2 with $\alpha = 0.9$. The markers with largest substitution effect were on chromosomes 14, 20, and 19. No

Table 1. Correlations of estimated genotypic values (EGV), breeding values (EBV), and dominance deviations (EDV) with yield deviations for the trait fat yield (FY).

	Parameters					Cor(EGV, FY)		Cor(EBV, FY)		Cor(EDV, FY)	
	v	p_{LD}	α	$E(\delta_j)$	$Var(\delta_j)$	C	C&B	C	C&B	C	C&B
BayesD2	2.5	0.05	0.9	0.04	0.43	0.363	0.387	0.360	0.379	0.101	0.097
BayesD2	2.5	0.10	0.9	0.03	0.43	0.317	0.387	0.301	0.379	0.118	0.128
BayesD2	3	0.10	0.9	0.03	0.43	0.314	0.398	0.300	0.390	0.116	0.112
BayesD2	3	0.10	0	0.03	0.52	0.357	0.368	0.353	0.371	0.092	0.034
G-BLUP D	50	1.00	0	-	-	0.352	0.371	0.346	0.368	0.102	0.095
G-BLUP	50	1.00	0	0	0	0.356	0.370	0.356	0.370	-	-

Correlations of estimated genotypic values (EGV), breeding values (EBV), and dominance deviations (EDV) with yield records for the trait fat yield (FY) for different models and parameter settings. For column 'C' the marker effects were estimated from YD of cows, whereas for column 'C&B' they were estimated from cows and bulls jointly.

overdominant alleles with a large dominance effect were observed.

Conclusion

A joint evaluation of genotyped individuals with phenotyped offspring and with own phenotypic records using BayesD models is a promising strategy to increase the accuracy of estimated genotypic values. With genotyping more cows, large scale data sets will become available that allow for more accurate prediction of dominance deviations.

Acknowledgement

Data were provided by AgroClustEr "Synbreed—Synergistic plant and animal breeding" (grant no. 0315628 H), funded by the German Federal Ministry of Education and Research (Bonn, Germany). The Förderverein Biotechnologieforschung e.V. (Bonn, Germany) and the organizations contributing to the German-Austrian pool of Fleckvieh genotypes are gratefully acknowledged for permitting the use of 50K genotypes. The development of BayesD was supported by the German Research Foundation (DFG).

Literature Cited

Ertl, J., Legarra, A., Vitezica, Z. G., et al. (2013). Interbull Bulletin No 47, Nantes, France

Meuwissen, T. H. E. (2009). Genet. Sel. Evol. **41**, 35

Miglior, F., Burnside, E. B., and Kennedy, B. W. (1995). J Dairy Sci **78**: 1174-1180.

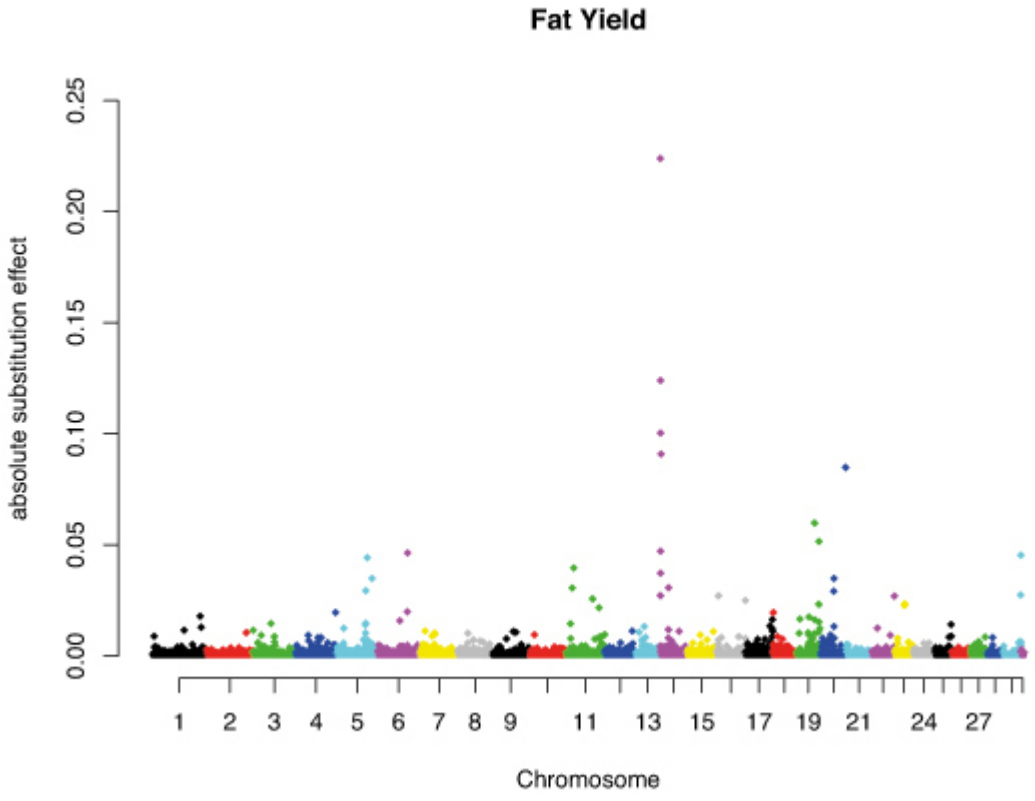
Sargolzaei, M., Chesnais, J.P., and Schenkel, F.S. (2011). J. Dairy Sci., 94: 421

Wellmann, R., and Bennewitz, J. (2011). Genet. Res. Camb. **93**: 139-154

Wellmann, R., and Bennewitz, J. (2012). Genet. Res. Camb. **94**: 21-37

Zhao, Y., Zeng, J., Fernando, R., et al. (2013). Crop Science. **53**: 802-810

Figure 1. Marker effects for fat yield



The figure shows allele substitution effects in phenotypic standard deviations, estimated for the trait FY from bulls and cows. These effects were averaged over the results from the three BayesD2 methods that provided the highest accuracy (i.e. $\alpha = 0.9$ was used).