Genomic Analysis of Dominance Effects in Milk Production and Conformation Traits of Fleckvieh Cattle

J. Ertl[1], A. Legarra[2], Z. G. Vitezica[3], L. Varona[4], C. Edel[1], R. Emmerling[1], K.-U. Götz[1]

[1] *Bavarian State Research Centre of Agriculture, 85586 Poing-Grub, Germany*

[2] *UR 631 SAGA, Institut National de la Recherche Agronomique, 31326 Castanet-Tolosan, France*

[3] *UMR 1289 TANDEM, INRA/INPT-ENSAT/ENVT, Université de Toulouse, 31326 Castanet-Tolosan, France*

[4] *Departamento de Anatomía, Embriología y Genetíca, Universidad de Zaragoza, 50013 Zaragoza, Spain*

Abstract

Dominance is the phenomenon of intra-locus interaction of alleles. Dominance variance in dairy cattle estimated from pedigree data shows a large variation and amounts to up to 50% of the genetic variance in conformation traits and 43% in milk production traits. With SNP genotypes of cows, dominance variance can be estimated both on the marker level and on the animal level using genomic dominance relationship matrices. Variance components of nine milk production and conformation traits were estimated in additive and dominance models by REML estimation and Gibbs sampling. Estimated dominance variance amounted to between 3.3% and 50.5% of the total genetic variance. REML and Gibbs sampling estimates showed good concordance. Although standard errors of dominance variance were rather large, dominance variance in milk, fat, protein yield, somatic cell score and milkability was significant.

Key words: dominance, variance component

Introduction

Dominance arises when the allele effect at a locus is not just additive, but alleles are interacting so that the value of heterozygote genotypes deviates from the mean of the homozygote values. Estimates of dominance variance in dairy cattle range from 7.3% to 49.8% of the total genetic variance in conformation traits (Misztal *et al.*, 1997; Tempelman and Burnside, 1990a) and from 3.4% to 42.9% in milk production traits (Miglior *et al.*, 1995; Tempelman and Burnside, 1990b; Van Tassell *et al.*, 2000).

At an individual level, dominance is not used in animal breeding (Misztal *et al.*, 1998), in spite of including a relevant part of genetic variation. The reason is the computational demand of large scale genetic evaluations for dominance, the relatively low accuracy, and the complexity of planning and computing the outcome of planned matings (Varona and Misztal, 1999).

With the availability of SNP genotypes of animals with own phenotypic records, dominance at the marker allele can be readily determined. Further, genomic dominance covariance matrices can be calculated similarly to genomic additive relationship matrices, which are widely used in genomic selection, such that dominance effects can be estimated in a GBLUP model (Su *et al.*, 2012; Vitezica *et al.*, submitted).

In this work, we estimated variance components including dominance variance in a dataset of genotyped Bavarian Fleckvieh cows.

Material and Methods

Cows were genotyped with Illumina BovineHD Genotyping BeadChip at 777,962 loci. SNPs with call rate <0.9, a minor allele frequency <0.005 and highly significant deviation ($p<10^{-5}$) from the Hardy Weinberg equilibrium and SNPs that were not annotated (UMD3) on the autosomes or the pseudo-autosomal region of the X-chromosome were excluded from the analysis. 629,028 loci remained in the dataset after editing. High-density genotypes and yield deviations (YD) in nine traits (milk yield, fat yield, protein yield, somatic cell score, milkability, stature, udder score, udder depth and feet and legs score) from 1996 Bavarian Fleckvieh cows were

available to estimate variance components including dominance variance in a GBLUP framework. The effective number of own performances (EOP; Edel *et al.,* 2008) was provided as weight for the YD.

Additive-genetic (σ_A^2) and residual (σ_E^2) variance components were estimated with models MA and MG.

MA: $\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{e}$; $V(\mathbf{u}) = \mathbf{A}\sigma_A^2$; $V(\mathbf{e}) = \mathbf{F}\sigma_E^2$

MG: $\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{e}$; $V(\mathbf{u}) = \mathbf{G}\sigma_A^2$; $V(\mathbf{e}) = \mathbf{F}\sigma_E^2$

y is a vector of YD, μ is the overall mean, **Z** is a design matrix relating YD to breeding values, **u** is a vector of breeding values of cows and **e** is a vector of residuals. Covariance matrices of additive effects are $V(\mathbf{u}) = \mathbf{A}\sigma_A^2$ in model MA and $V(\mathbf{u}) = \mathbf{G}\sigma_A^2$ in model MG, where **A** is the numerator relationship matrix and **G** is the genomic relationship matrix. The genomic relationship matrix $\mathbf{G}^*$ was calculated following the approach of VanRaden (2008) using PREGSF90 (Aguilar *et al.,* 2011). $\mathbf{G}^*$ was scaled so that the means of diagonals and off-diagonals are the same as in **A** (Vitezica *et al.,* 2011; Christensen, 2012) and finally combined with **A** to $\mathbf{G} = 0.95\,\mathbf{G}^* + 0.05\,\mathbf{A}$ in order to improve numerical stability. The variance matrix of residual effects is $V(\mathbf{e}) = \mathbf{F}\sigma_E^2$ in both models, where **F** is a diagonal matrix with reciprocals of EOP as weights. Extending MG for a dominance effect leads to model MGD.

MGD: $\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{Zd} + \mathbf{e}$; $V(\mathbf{u}) = \mathbf{G}\sigma_A^2$; $V(\mathbf{d}) = \mathbf{D}\sigma_D^2$; $V(\mathbf{e}) = \mathbf{F}\sigma_E^2$

d is a vector of individual dominance deviations. The covariance matrix of dominance effects is $V(\mathbf{d}) = \mathbf{D}\sigma_D^2$ where **D** is the genomic dominance relationship matrix and σ_D^2 is the dominance variance. $\mathbf{D}^*$ was calculated as:

$$\mathbf{D}^* = \frac{\mathbf{WW}'}{4\sum_{k=1}^{m} p_k^2(1-p_k)^2}$$

where **W** has a dimension of the number of individuals (*n*) by the number of loci (*m*) and takes the values $-2(1-p_k)^2$ and $-2p_k^2$ for homozygous and $2p_k(1-p_k)$ for heterozygous genotypes. p_k is the allele frequency at locus *k*. $\mathbf{D}^*$ was finally combined with the identity matrix **I** as $\mathbf{D} = 0.95\,\mathbf{D}^* + 0.05\,\mathbf{I}$ to improve numerical stability.

Estimation of variance components was performed with REMLF90 (Misztal *et al.,* 2002). The superiority of model MGD over model MG was tested by means of a likelihood ratio test. The test statistics was calculated as χ^2 = -2ln(likelihood for MG) + 2ln(likelihood for MGD). The likelihood ratio follows a mixture of the χ^2-distributions with 0 and 1 degrees of freedom (Visscher, 2006). Variance components of model MGD were additionally estimated by Gibbs Sampling using GIBBS1F90 software in order to get standard errors of the estimates. Additive and dominance variance components on the marker level (σ_a^2 and σ_d^2) were estimated with GS3 software (Legarra *et al.,* 2010) in a Markov Chain Monte Carlo algorithm, using a model on the marker level:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Ta} + \mathbf{X\delta} + \mathbf{e}$$

where **a** and **δ** are additive and dominant effects of the SNPs, and **T** and **X** are incidence matrices coded as {-1, 0, 1} and {0, 1, 0} for the three possible genotypes. The assumed covariance structure is $V(\mathbf{a}) = \mathbf{I}\sigma_a^2$ and $V(\boldsymbol{\delta}) = \mathbf{I}\sigma_d^2$. From the estimates, additive and dominance variance components on the animal level were calculated as

$\sigma_A^2 = \sum_{k=1}^{m}[2p_k(1-p_k)]\,\sigma_a^2 + \sum_{k=1}^{m}\{2p_k(1-$
$pk)1-pk-pk^2\sigma d2$
and $\sigma_D^2 = \sum_{k=1}^{m}\left[4p_k^2(1-p_k)^2\right]\sigma_d^2$.

Results and Discussion

Means of off-diagonals of **G** (before scaling) and **D** are 0.000 (This implies that the complete population is in Hardy-Weinberg equilibrium). The standard deviation of off-diagonals of **G** is 0.036. This is 5 times as large as the standard deviation of off-diagonals of **D** which is 0.007. The proportion of off-diagonals that are smaller than -0.05 or larger than 0.05 is 6.27% for **G** but only 0.02% for **D**. Therefore, matrix **D** is less informative than **G**. Estimated variance components for model MGD are shown in Table 1. Dominance variance (expressed as proportion of total genetic variance) ranged from 3.3% in stature

to 50.5% in somatic cell score. Dominance variance was larger in milk production traits than in conformation traits. The estimates of additive variance with the dominance model were not very different (between -5.5% in protein yield and 1.3% in stature) from the estimates with the genomic additive model. The only exception was milkability where the estimated additive variance was 43.5% smaller in the dominance model.

Estimates of variance components in model MGD with Gibbs Sampling or at a SNP level with GS3 were very similar to REML estimates and the results are not shown here. Estimates of the ratio between dominance and phenotypic variance have standard errors around 0.10, which is fairly good in such a small dataset.

Table 1. Estimated variance components (REMLF90) in model MGD.

Trait	$\sigma_A^{2\,a}$	$\sigma_D^{2\,a}$	$\sigma_E^{2\,a}$	$\frac{\sigma_D^2}{\sigma_A^2+\sigma_D^2}$ [a]
Milk yield	208900 ± 28797	92640 ± 45132	164700 ± 33330	0.308 ± 0.107
Fat yield	267 ± 37	104 ± 53	198 ± 367	0.281 ± 0.104
Protein yield	166 ± 26	115 ± 44	154 ± 29	0.409 ± 0.104
Somatic cell score	0.256 ± 0.067	0.261 ± 0.120	0.555 ± 0.094	0.505 ± 0.152
Milkability	0.0122 ± 0.0019	0.0076 ± 0.0017	0.0029 ± 0.0010	0.390 ± 0.081
Stature	5.80 ± 0.87	0.20 ± 0.49	6.51 ± 0.80	0.033 ± 0.065
Udder score	1.99 ± 0.52	0.27 ± 0.78	9.29 ± 0.92	0.118 ± 0.160
Udder depth	0.380 ± 0.061	0.119 ± 0.095	0.517 ± 0.102	0.238 ± 0.120
Feet and legs	1.19 ± 0.45	0.21 ± 0.67	9.89 ± 0.82	0.153 ± 0.192

[a] Estimate ± standard error (estimated with Gibbs sampling in GIBBS1F90)

The results agree with Misztal *et al.* (1998) who reported more dominance variation for production than for type traits. The estimates of additive and dominance variances, expressed as proportions of the total variance, of milk production traits in this study are larger than in Miglior *et al.* (1995), Van Tassell *et al.* (1999) and Van Tassell *et al.* (2000). One reason for this might be that a part of phenotypic variation in our dataset was removed in the process of calculation of YD. Tempelman and Burnside (1990b) estimated dominance variance in fat yield to 24% of the phenotypic variance, which is a bit larger than in our results. The ratio of dominance to additive variance in our dataset was between 0.39 and 0.69 in milk production traits which is considerably larger than in Van Tassell *et al.* (2000). The difference might be explained by two reasons. First, Fleckvieh is genetically more diverse than Holstein as indicates the considerably larger effective population size of the Fleckvieh breed (Pausch *et al.,* 2013). Second, all estimates of dominance variance available in the literature were obtained using covariance matrices based on pedigree data. The use of genomic information is expected to improve the estimation of dominance relationships which can explain the larger estimates of dominance variance obtained in this study. Estimates of dominance variance in conformation traits are quite small, except for udder depth. This is in analogy with Misztal *et al.* (1997) who reported dominance variance in 10 conformation traits to amount to less than 5% of phenotypic variance.

In all traits, model MG, that exploited genomic information, fitted the data better than model MA, that included pedigree information only. The superiority of model MGD, containing the dominance effect, over MG was significant in the traits milk yield, fat yield, protein yield, somatic cell score and milkability as tested with the likelihood ratio test.

Conclusions

Genomic estimates of variation due to dominance in dairy cattle agree with pedigree-based estimates and the computational complexity and modeling is straightforward.

The impact of modeling dominance on genomic predictions and the use of dominance effects in planned matings should be investigated.

Acknowledgements

This research was funded by the German Federal Ministry of Education and Research (Bonn, Germany) within the AgroClustEr "Synbreed – Synergistic plant and animal breeding" (Grant no: 0315628 H).

References

Aguilar, I., Misztal, I., Legarra, A. & Tsuruta, S. 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet. 128*, 422-428.

Christensen, O.F. 2012. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genetics Selection Evolution 44:37.*

Edel, C., Emmerling, R. & Götz, K.-U. 2008. Optimized Aggregation of Phenotypes for MA-BLUP Evaluation in German Fleckvieh. *Interbull Bulletin 40,* 178-183.

Legarra, A., Ricard, A. & Filangi, O. 2010. GS3 Genomic Selection – Gibbs Sampling – Gauss Seidel (and BayesCπ). http://genoweb.toulouse.inra.fr/~alegarra/manualgs3_last.pdf (5.7.2013)

Miglior, F., Burnside, E.B. & Kennedy, B.W. 1995. Production Traits of Holstein Cattle: Estimation of Nonadditive Genetic Variance Components and Inbreeding Depression. *J. Dairy Sci. 78,* 1174-1180.

Misztal, I., Lawlor, T. J. & Gengler, N. 1997. Relationships among estimates of inbreeding depression, dominance and additive variance for linear traits in Holsteins. *Genetics Selection Evolution 29*, 319-326.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T. & Lee, D.H. 2002. BLUPF90 and related programs (BGF90). *Proceedings of the 7th World Congress Applied to Livestock Production.*

Misztal, I., Varona, L., Culbertson, M., Bertrand, J.K., Mabry, J., Lawlor, T.J., Van Tassell, C.P. & Gengler, N. 1998. Studies on the value of incorporating the effect of dominance in genetic evaluations of dairy cattle, beef cattle and swine. *Biotechnol. Agron. Soc. Environ. 2*, 227-233.

Pausch, H., Aigner, B., Emmerling, R., Edel, C., Götz, K.-U. & Fries, R. 2013. Imputation of high-density genotypes in the Fleckvieh cattle population. *Genetics Selection Evolution 45:3.*

Su, G., Christensen, O.F., Ostersen, T., Henryon, M. & Lund, M.S. 2012. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLoS ONE 7(9),* e45293.

Tempelman, R.J. & Burnside, E.B. 1990a. Additive and Nonadditive Genetic Variation for Conformation Traits in Canadian Holsteins. *J. Dairy Sci. 73,* 2214-2220.

Tempelman, R.J. & Burnside, E.B. 1990b. Additive and Nonadditive Genetic Variation for Production Traits in Canadian Holsteins. *J. Dairy Sci. 73,* 2206-2213.

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

Van Tassell, C.P., Misztal, I. & Varona, L. 2000. Method R Estimates of Additive Genetic, Dominance Genetic, and Permanent Environmental Fraction of Variance for Yield and Health Traits of Holsteins. *J. Dairy Sci. 83,* 1873-1877.

Van Tassell, C.P., Wiggans, G.R. & Norman, H.D. 1999. Method R Estimates of Heritability for Milk, Fat, and Protein Yields of United States Dairy Cattle. *J. Dairy Sci. 82,* 2231-2237.

Varona, L. & Misztal, I. 1999. Prediction of Parental Dominance Combinations for Planned Matings, Methodology, and Simulation Results. *J. Dairy Sci. 82,* 2186-2191.

Visscher, P.M. & Thompson, R. 1992. Comparisons between genetic variances estimated from different types of relatives in dairy cattle. *ANIMAL PRODUCTION-GLASGOW- 55,* 315-320.

Vitezica, Z.G., Aguilar, I., Misztal, I. & Legarra, A. 2011. Bias in genomic predictions for populations under selection. *Genet. Res., Camb. 93,* 357-366.