TECHNISCHE UNIVERSITÄT MÜNCHEN

Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Lehrstuhl für Tierzucht

Investigations on genomic evaluations using high-density genotypes in Fleckvieh with consideration of dominance effects

Johann Ertl

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Agrarwissenschaften

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Wilhelm Windisch

Prüfer der Dissertation:
1. Prof. Dr. Hans-Rudolf Fries
2. Hon.-Prof. Dr. Kay-Uwe Götz
3. Prof. Dr. Henner Simianer

Die Dissertation wurde am 13.11.2017 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 13.03.2018 angenommen.

Publications arising from this thesis

Peer reviewed

Ertl J, Edel C, Emmerling, R, Pausch H, Fries R, Götz K-U (2014): On the limited increase in validation reliability using high-density genotypes in genomic best linear unbiased prediction: Observations from Fleckvieh cattle. *Journal of Dairy Science*, 97:487-496.

Ertl J, Legarra A, Vitezica Z G, Varona L, Edel C, Emmerling R, Götz K-U (2014): Genomic analysis of dominance effects in milk production and conformation traits of Fleckvieh cattle. *Genetics Selection Evolution*, 46:40.

Ertl J, Edel C, Pimentel E C G, Emmerling R, Götz K-U (2018): Considering dominance in reduced single-step genomic evaluations. *Journal of Animal Breeding and Genetics*, 135:151-158.

Conference Proceedings

Ertl J, Edel C, Neuner S, Emmerling R, Götz K-U (2012): Comparative analysis of linkage disequilibrium in Fleckvieh and Brown Swiss cattle. Book of abstracts of the 63rd Annual Meeting of the European Federation of Animal Science, Bratislava.

Ertl J, Edel C, Neuner S, Emmerling R, Götz, K-U (2012): Nutzen von HD-Genotypen für die genomische Zuchtwertschätzung beim Fleckvieh – erste Ergebnisse. Tagungsband der DGfZ/GfT-Gemeinschaftstagung 2012, Halle (Saale).

Ertl J, Legarra A, Vitezica Z G, Varona L, Edel C, Emmerling R, Götz K-U (2013): Genomic analysis of dominance effects in milk production and conformation traits of Fleckvieh cattle. Interbull meeting, Nantes.

Ertl J, Edel C, Emmerling R, Pausch H, Fries R, Götz K-U (2013): Validation accuracy of genomic breeding values with HD genotypes in Fleckvieh cattle. Book of abstracts of the 64th Annual Meeting of the European Federation of Animal Science, Nantes.

Ertl J, Legarra A, Vitezica Z G, Varona L, Edel C, Emmerling R, Götz K-U (2013): Genomische Untersuchung von Dominanzeffekten in Milchleistungs- und Exterieurmerkmalen bei der Rasse Fleckvieh. Tagungsband der DGfZ/GfT-Gemeinschaftstagung 2013, Göttingen.

Wellmann R, **Ertl J**, Emmerling R, Edel C, Götz K-U, Bennewitz J (2014): Joint genomic evaluation of cows and bulls with BayesD for prediction of genotypic values. Proceedings, 10th World Congress of Genetics Applied to Livestock Production, Vancouver.

Ertl J, Edel C, Pimentel E, Emmerling R, Götz K-U (2014): Berücksichtigung von Dominanz im One-Step-Verfahren der Zuchtwertschätzung. Tagungsband der DGfZ/GfT-Gemeinschaftstagung 2014, Dummerstorf.

Table of contents

1st Chapter

General introduction

Estimation of breeding values and selection in dairy cattle

During the history of animal breeding, availability of selection criteria has evolved with progress in performance testing, computing facilities and the development of the statistical toolbox. Animal breeders have always been eager to use new and more precise sources of information in order to make breeding decisions as accurate as possible. In the beginning of animal breeding, qualitative or quantitative phenotypes served as selection criteria. Response to selection depended essentially on the heritability of a trait in this context. Consequently, for centuries, genetic gain has remained quite small. After important contributions had been made by Fisher (1918), Wright (1921a-e) and Haldane (1932) in the development of quantitative genetics theory, the selection index (Hazel, 1943) was another step forward to analyze the genetically determined part of phenotypic observations and enabled the selection of animals based on their breeding values estimated from phenotypic information of the animal itself and its relatives. Main disadvantages of the selection index were that environmental effects had to be known (which is often not the case in practice) and that the phenotypic covariance matrix had to be constructed for each different constellation of information sources, which led to suboptimal approximations only for the sake of simplicity. The development of best linear unbiased prediction (BLUP; e.g. Henderson, 1973, 1975) using the mixed model equations was a quantum leap in genetic evaluation. Simultaneous estimation of fixed and random effects with the BLUP individual animal model assured unbiased estimates for many kinds of real-life data structures. Henderson (1976) developed an algorithm to build the inverse relationship matrix directly, which enabled the application of the individual animal model to real world populations. BLUP has worked successfully for the past decades and has assured remarkable genetic gains in a variety of traits and in different livestock species. Several adaptations of the BLUP animal model have been developed to account for special data structures and in order to reduce computational demands in specific situations. For example, breeding values of sires

with progeny records were often estimated by means of a sire model and multiple-trait models have enabled simultaneous evaluation of several traits using information from genetically correlated traits mutually (e.g. Henderson and Quaas, 1976). Fixed and random regression models have been developed to evaluate longitudinal data, e.g. lactation curves based on test day records (e.g. Schaeffer, 2004).

Evolution of genomic evaluation

Marker-assisted selection

In the 1980s, first DNA-based markers began to become available. Minisatellites (variable number of tandem repeats), microsatellites (simple sequence repeats) and restriction fragment length polymorphisms were the typical markers at this time. Reflections about how to use this new type of genetic information in order to improve genetic progress led to the concept of marker-assisted selection (MAS; e.g. Fernando and Grossman, 1989). Quantitative traits have a continuous distribution and the observed genetic variance of quantitative traits is caused by a multitude of gene loci which are called quantitative trait loci (QTL) (Geldermann, 1975). Linkage as well as linkage disequilibrium between markers and (unknown) QTL is exploited in MAS when effects of markers are estimated as surrogates for QTL effects. While at the beginning of MAS, the number of QTL causing genetic variation of a specific trait was assumed to be small to medium (e.g. Hayes and Goddard, 2001), many newer results indicate that genetic architecture of quantitative traits is closer to the infinitesimal model: each out of many loci contributes a tiny effect to the genetic variation of the trait (e.g. Reed et al., 2008). Although sophisticated statistical approaches have been developed to estimate marker effects and to map QTL positions (e.g. Churchill and Doerge, 1994; Sillanpää and Corander, 2002; Meuwissen and Goddard, 2004), effects have often been overestimated and could rarely be exploited consistently in real breeding schemes. Until now France has remained the only

country that implemented MAS for several years in its breeding scheme (Guillaume et al., 2008).

Genomic selection

During the years of improvement and refinement of MAS, the insight rose, that many more markers than assumed in the MAS concept were required to accurately capture genetic variation of a specific trait. Meuwissen et al. (2001) elaborated the conceptual base of genomic selection by estimating breeding values based on the effects of genotypes at many thousands of markers covering the whole genome. The pioneering work by Meuwissen et al. (2001) has initiated the development of procedures for genomic evaluation of breeding values in the following decade. In the genomic selection concept, each QTL is expected to be in high linkage disequilibrium with at least one of many markers. Marker effects are estimated simultaneously in genomic selection without imposing a significance threshold in order to avoid the Beavis effect (Beavis, 1998), the overestimation of the largest marker effects.

The complete sequencing of the bovine genome in 2009 (e.g. Liu *et al.*, 2009; Zimin *et al.*, 2009) laid the technological base for the development of single nucleotide polymorphism (SNP) chips to efficiently genotype individuals for thousands of SNP distributed all over the genome. SNP are the markers of choice for the implication of genomic selection. They are located at a single nucleotide, are usually biallelic and numerously distributed all over the genome. Re-sequencing of 129 Holstein, 47 Angus, 43 Fleckvieh and 15 Jersey animals has identified 26.7 million SNP in the cattle genome (Daetwyler et al., 2014). Commercial high-throughput and high-density genotyping chips (Gunderson et al., 2005; Steemers et al., 2006; Steemers and Gunderson, 2007) are offered e.g. by the companies Illumina Inc. and Affymetrix Inc. Illumina's Bovine50 BeadChip comprising around 54,000 SNP (Illumina Inc., 2015a) was

the first and has remained up to now the most common SNP chip for genotyping of cattle. More recently, Illumina developed denser (~777,000 SNP; Illumina Inc., 2015b) and less dense (Illumina Inc., 2015c: ~8,000 SNP; Wiggans et al., 2012: ~3,000 SNP) SNP chips for cattle. Affymetrix offers a genotyping array comprising ~640,000 SNP (Affymetrix Inc., 2015).

Most valuable contributions to the development of genomic selection have been the suggestions of VanRaden (2008) to calculate the genomic or marker-based relationship matrix and the work of both Legarra et al. (2009) and Christensen and Lund (2010) who independently developed the combination of genomic and pedigree relationships in the single-step method of genomic evaluation. Already Meuwissen et al. (2001) have suggested a ridge regression model – a linear model where SNP effects follow a normal distribution and *a priori* have equal variance – for genomic evaluation. The genomic BLUP (GBLUP) model has proven to be equivalent to the ridge regression model by Habier et al. (2007). In the GBLUP model, the numerator relationship matrix is replaced by the relationship matrix calculated from marker genotypes (e.g. VanRaden, 2008). Realized relationships are calculated from genotypes of markers covering the whole genome that trace Mendelian sampling during meiosis – in contrast to numerator relationships based on pedigree information. Genomic breeding values for selection candidates are much more reliable than parent averages (VanRaden et al., 2009) because they estimate Mendelian sampling terms based on realized relationships with all proven bulls.

Pre-correction and aggregation of phenotypes to different types of pseudo-phenotypes has the favorable effect that the proportion of genetic variance compared to total variance increases. This can increase the power of genomic analyses (e.g. genomic prediction, genome-wide association study) or, alternatively, smaller samples are sufficient to obtain significant results. Genetic evaluation is the most consequent method to separate the genetic part from the total variation of a phenotype. Estimated breeding values of proven bulls rely on records of large

numbers of daughters and are highly reliable. Estimated breeding values are frequently used as phenotypes e.g. in genome-wide association studies (Pausch et al., 2011). In the estimation process, breeding values are regressed towards the population mean depending on their information content. De-regression of breeding values restores the original variance of records (e.g. Thomsen et al., 2001; Garrick et al., 2009). Garrick et al. (2009) recommended removing parent average effects before de-regressing to avoid double-counting of information. De-regressed breeding values, also known as de-regressed proofs (DRP; Garrick et al., 2009), are frequently used in genomic analyses because they can be obtained relatively easily from estimated breeding values and the respective reliabilities. Daughter yield deviations (DYD; VanRaden and Wiggans, 1991) are the aggregated phenotypes that are the closest to real observations. Phenotypes of daughters are corrected for fixed effects, non-genetic random effects as well as the breeding value of their dam and averaged to obtain the DYD of a bull.

A GBLUP model is a linear model as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},$$

where **y** is a vector of pseudo-phenotypes, typically DRP or DYD. **X** is an incidence matrix and **b** is a vector of fixed effects. Usually, a single fixed effect is modelled as intercept because other non-genetic effects have been removed during pre-correction of phenotypes. **Z** is a design matrix that relates **y** to genomic breeding values. Genomic breeding values **u** have the covariance matrix $\mathbf{G}\sigma_A^2$, where **G** is the genomic relationship matrix calculated from marker genotypes and σ_A^2 is the additive genetic variance. Genomic breeding values for calibration and validation animals are predicted by means of BLUP:

$$\hat{\mathbf{u}} = \mathbf{G}\sigma_A^2\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}),$$

where $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{V}^{-1}\mathbf{y}$, and **V**, the covariance matrix of **y**, is calculated as

$$\mathbf{V} = \mathbf{Z}(\mathbf{G}\sigma_A^2)\mathbf{Z}' + \mathbf{F}\sigma_E^2,$$

where σ_E^2 is the error variance and **F** is a diagonal matrix with reciprocals of the equivalent number of own performances (EOP). EOP are calculated as follows:

$$EOP = \frac{\sigma_e^2}{\sigma_a^2}\frac{R_y^2}{1-R_y^2},$$

where R_y^2 is the reliability of pseudo-phenotypes **y**.

VanRaden (2008) suggested calculating the genomic relationship matrix in the following way:

$$\mathbf{G} = \frac{\mathbf{WW'}}{2\sum_{k=1}^{m} p_k(1-p_k)},$$

where **W** is a genotype matrix having a dimension of the number of individuals (*n*) by the number of loci (*m*) and is calculated as **W** = **M** - **P**. The elements of **M** are -1 and 1 for opposite homozygous genotypes and 0 for heterozygous genotypes. Column *k* of the matrix **P** is $2(p_k - 0.5)$ and p_k is the allele frequency at locus *k*. Preferably, the allele frequencies of the unselected base population are used as p_k to assure consistency with pedigree-based evaluation. Base allele frequencies can be estimated with the method of Gengler et al. (2007). Meuwissen et al. (2011) suggested scaling of the genomic relationship matrix towards the numerator relationship matrix to account for e.g. genetic variance not totally explained by marker genotypes. Additional combination with a small proportion of the numerator relationship matrix can be useful to improve numerical stability of the genomic relationship matrix (e.g. VanRaden, 2008).

The United States were the first country to implement GBLUP based on 50K genotypes of insemination bulls and candidate bulls in national genetic evaluation of Holstein, Jersey and Brown Swiss in 2009 (Wiggans et al., 2011). Many other important cattle breeding countries followed soon and started to estimate and publish genomic breeding values, predominantly using GBLUP. In Germany and Austria, official genomic breeding values for Fleckvieh and Brown Swiss were published in 2011 for the first time. In the beginning of genomic evaluation,

only 50K genotypes were available because the 50K chip was the first commercial SNP chip. When Illumina released a low-density chip with 2,900 SNP (3K) and a high-density chip with 777,962 SNP (HD) in 2010, it became necessary to include also these other marker sets in genomic evaluation. Different software packages are available for the imputation from sparse to dense marker sets. The most frequently used imputation programs in animal breeding are: BEAGLE (Browning and Browning, 2007), findhap (VanRaden et al., 2011, 2013), FImpute (Sargolzaei et al., 2011) and AlphaImpute (Hickey et al., 2011). Although imputation programs use information about linkage and linkage disequilibrium between markers intensively, a small percentage of imputation errors occurs with every program. Chen et al. (2011) reported that between 1.6% and 3.3% of alleles were not correctly imputed from 3K to 50 K. Error rate in imputation from 3K to 50K ranged from 2.1% to 5.5% in a study of Dassonneville et al. (2011). With higher marker density, it is more probable that a marker is in high linkage disequilibrium with a QTL. Several studies examined the hypothesis that accuracy of genomic evaluation should increase with higher marker density. However, only minor gains could be realized when using (imputed) HD genotypes instead of 50K genotypes in different Holstein populations (Harris et al., 2011; Erbe et al., 2012; Su et al., 2012; VanRaden et al., 2013). In these studies, the question was not addressed if the reported small gains in reliability were significant.

Breeding program for Fleckvieh in Bavaria

The current breeding program for Fleckvieh animals in Bavaria (in effect since 2012) is schematically depicted in Figure 1. The breeding program is based on 605,000 Fleckvieh cows in the herd-books of Bavarian Fleckvieh breeding associations. Each year, based on their breeding values, around 25,000 cows are suggested as potential dams of artificial insemination (AI) bulls. From these 25,000 potential bull dams, 12,000 are selected to breed candidate bulls.

Some candidates are excluded because of undesired properties and out of around 5,000 male candidates, 270 bulls are selected for AI ("young genomic AI bulls"). These young AI bulls are bred to about 30% of the cow population and after performance-testing of daughters, 75 bulls are selected for broad AI use. Each year, 40 young genomic bulls and 20 performance-tested bulls are selected as bull sires and are bred to 40% and 60% of 12,000 selected bull dams, respectively.

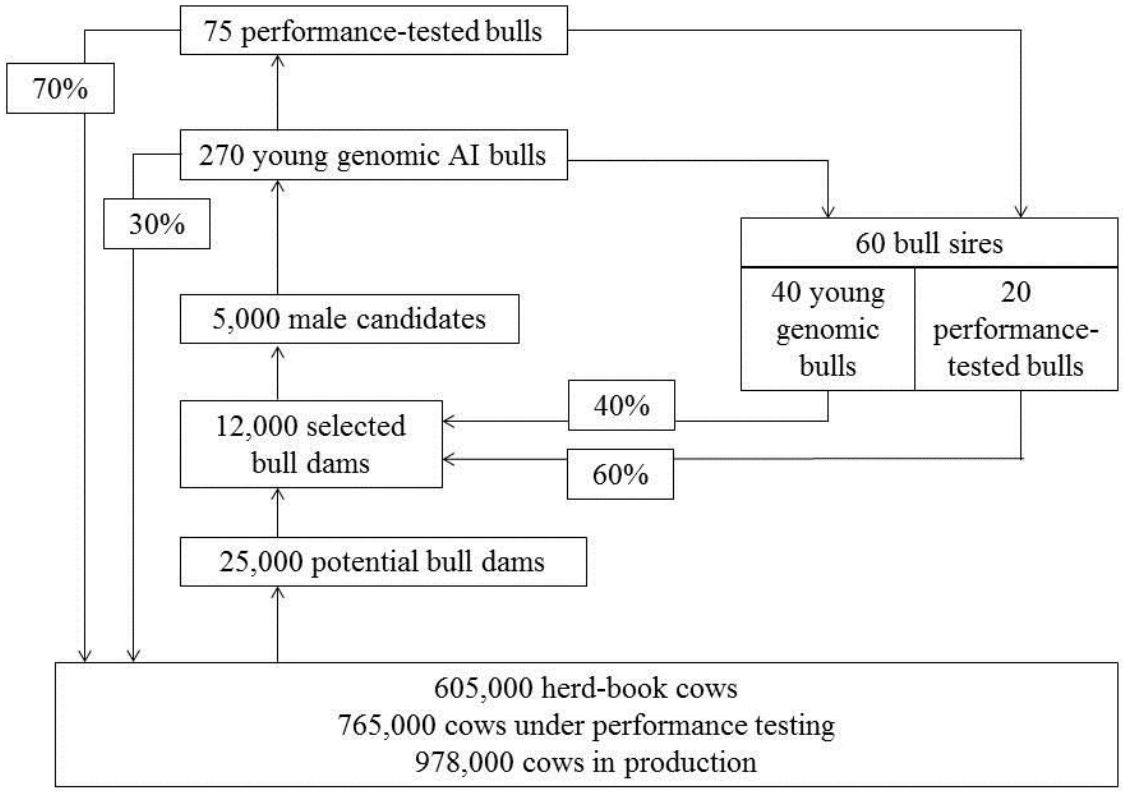

Figure 1: Current breeding program for Fleckvieh in Bavaria [adapted from Bayerisches Staatsministerium für Ernährung, Landwirtschaft und Forsten (2015)]

Dominance – new applications in the genomic era?

Dominance results from the interaction of alleles at a locus and is - together with additivity and epistasis - an important component of genetic variance. Estimates of dominance variance in

dairy cattle that are based on pedigree data range from 7.3% to 49.8% of the total genetic variance for conformation traits (Tempelman and Burnside, 1990b; Misztal et al., 1997) and from 3.4% to 42.9% for milk production traits (Tempelman and Burnside, 1990a; Miglior et al., 1995; Van Tassell et al., 2000). Classically, dominance variation has been ignored in pure-breeding programs for several reasons: First, the computational demand of large-scale genetic evaluations for dominance is challenging. Second, the accuracy of estimates for dominance effects is relatively low. Third, using dominance variance in planned matings is complex and computationally demanding. With the availability of SNP genotypes, there are new possibilities for the evaluation of dominance effects analyzing heterozygosity and differences in heterozygosity at marker loci. Analyses can be performed at the marker level, as suggested e.g. by Toro and Varona (2010) and Wellmann and Bennewitz (2012). An alternative is to calculate marker-based dominance relationships in a GBLUP framework (Su et al., 2012; Vitezica et al., 2013).

The above mentioned GBLUP model can be extended for dominance as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Zv} + \mathbf{e},$$

where **y** is a vector of phenotypes, which are usually pre-corrected for environmental effects. For genomic analyses of dominance effects, direct records from the genotyped animals are required and it is necessary that the dominance effect is not removed from the pseudo-phenotype during pre-correction. Yield deviations (YD; VanRaden and Wiggans, 1991) contain dominance deviations and can therefore be used as phenotypes in dominance models. DYD or DRP, however, cannot be used because they do not contain a dominance component. **Z** is a design matrix that relates **y** to both genomic breeding values and genomic dominance deviations. Genomic dominance deviations **v** have the covariance matrix $\mathbf{G_D}\sigma_D^2$, where $\mathbf{G_D}$ is

the genomic dominance relationship matrix calculated from marker genotypes and σ_D^2 is the dominance variance, and are predicted by means of BLUP:

$$\hat{\mathbf{v}} = \mathbf{G_D}\sigma_D^2\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}),$$

where $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{V}^{-1}\mathbf{y}$, and $\mathbf{V}$, the covariance matrix of $\mathbf{y}$, is calculated as:

$$\mathbf{V} = \mathbf{Z}(\mathbf{G}\sigma_A^2)\mathbf{Z}' + \mathbf{Z}(\mathbf{G_D}\sigma_D^2)\mathbf{Z}' + \mathbf{W}\sigma_E^2,$$

where σ_E^2 is the error variance and $\mathbf{W}$ is a diagonal matrix with reciprocals of the equivalent number of own performances (EOP). EOP are calculated as follows:

$$EOP = \frac{\sigma_E^2}{\sigma_A^2+\sigma_D^2}\frac{R_y^2}{1-R_y^2},$$

where R_y^2 is the reliability of pseudo-phenotypes $\mathbf{y}$.

Vitezica et al. (2013) suggested calculating $\mathbf{G_D}$ in the following way:

$$\mathbf{G_D} = \frac{\mathbf{W}_d\mathbf{W}_d{}'}{4\sum_{k=1}^{m} p_k^2 q_k^2},$$

where $\mathbf{W_d}$ has dimensions of the number of individuals (n) by the number of loci (m), with elements that are equal to $-2q_k^2$ for genotype A_1A_1, $2p_kq_k$ for genotype A_1A_2, and $-2p_k^2$ for genotype A_2A_2.

Variance components can be estimated both with the restricted maximum likelihood (REML) method and with Gibbs sampling. The BLUPF90 package (Misztal et al., 2002) contains the relevant FORTRAN programs REMLF90 and GIBBS1F90. With REML estimation, the likelihood ratio test evaluates if extending the model with an additional random effect (e.g. dominance) fits the data significantly better. Standard errors of variance component estimates can be calculated from Gibbs sampling chains.

Although usually the interest in genomic evaluation lies on genetic values for animals, the setup of a marker model (ridge regression BLUP; RR-BLUP) is equivalent to the GBLUP model as shown by Habier et al. (2007), Goddard (2009) and Hayes et al. (2009). A RR-BLUP model of the type

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Ta} + \mathbf{Xd} + \mathbf{e}$$

can be solved e.g. with GS3 software (Legarra et al., 2014). **a** and **d** are vectors of additive and dominance effects of the SNP, and **T** and **X** are incidence matrices coded as {-1, 0, 1} and {0, 1, 0} for the three possible genotypes. The assumed variance-covariance structure is V(**a**) = $\mathbf{I}\sigma_a^2$ and V(**d**) = $\mathbf{I}\sigma_d^2$. The quantities σ_a^2 and σ_d^2 are additive and dominance variance components at the marker level. Assuming that markers are in linkage equilibrium and marker effects are not correlated, animal level variance components can be calculated from marker level variance components and allele frequencies (Gianola et al., 2009; Vitezica et al., 2013):

$$\sigma_A^2 = \sum_{k=1}^{m} (2p_k q_k)\, \sigma_a^2 + \sum_{k=1}^{m} [2p_k q_k (q_k - p_k)^2]\, \sigma_d^2$$

$$\sigma_D^2 = \sum_{k=1}^{m} (4p_k^2 q_k^2)\, \sigma_d^2$$

The other way round, marker level variance components can be calculated from variance components of the animal model:

$$\sigma_d^2 = \frac{\sigma_D^2}{\sum(4p_k^2 q_k^2)}$$

$$\sigma_a^2 = \frac{\sigma_A^2 - \sum[2p_k q_k (q_k - p_k)^2]\sigma_d^2}{\sum(2p_k q_k)}$$

Instead of deducing from animal model parameters, marker level variance components can be estimated e.g. by means of a Markov Chain Monte Carlo algorithm as e.g. implemented in GS3.

If dominance variance amounts to a relevant part of genetic variation, it should be interesting to use this extra genetic variance for management purposes. For example, in the case that dominance deviations and total genetic values can be predicted from genotypic data with acceptable accuracy, this information could be used to select calves for dairy production. The remaining calves with less promising total genetic values in the relevant traits would then be used for beef or veal production. This early step of selection could anticipate the culling of cows that are already in milk but do not fulfill the farmer's demands.

The next step straightforward could be to select production cows not only as calves but already when the mating decision is taken for the dam. Matings could be planned to optimize expected production performance of resulting offspring if the prediction of expected total genetic value of a mating is possible.

Toro and Varona (2010) suggested to predict the total genetic value g_{ij} of progeny from a mating between bull *i* and cow *j* as follows:

$$\hat{g}_{ij} = \sum_k [Pr_{ijk}(AA)\hat{a}_k + Pr_{ijk}(Aa)\hat{d}_k - Pr_{ijk}(aa)\hat{a}_k],$$

where $Pr_{ijk}()$ is the probability of the corresponding genotype at locus *k*. Analogously, the breeding value u_{ij} of progeny from a mating between bull *i* and cow *j* can be predicted as:

$$\hat{u}_{ij} = \sum_k [Pr_{ijk}(AA)(2-2p_k)\hat{\alpha}_k + Pr_{ijk}(Aa)(1-2p_k)\hat{\alpha}_k + Pr_{ijk}(aa)(-2p_k)\hat{\alpha}_k],$$

where $\hat{\alpha}_k = \hat{a}_k + \hat{d}_k(q_k - p_k)$.

Matings can be selected on $\hat{u}$ to maximize additive genetic gain or on $\hat{g}$ to maximize total genetic superiority. The latter maximizes the productive performance of the offspring, which might be a farmer's interest. However, $\hat{g}$ can be maximized only for the next generation because a gain in the dominance part of $\hat{g}$ cannot be accumulated in subsequent generations.

Selection on $\hat{u}$ leads to maximum additive gain, which can be accumulated in subsequent generations, and thus optimizes cumulative multi-generational genetic gain. A desirable objective might be to maximize $\hat{g}$ of matings and at the same time to keep the expected $\hat{u}$ of the offspring as high as possible. This could be realized by pre-selection of bulls on their breeding value and subsequent optimization of matings based on expected total genetic values.

Joint evaluation of bulls' and cows' genotypes

As soon as cows are genotyped in addition to bulls, phenotypes of genotyped cows have to be integrated in the genomic evaluation scheme. Basically, genomic evaluation can be based directly on phenotypes that are typically recorded on cows. This procedure is known as one-step or single-step evaluation and was independently developed by Legarra et al. (2009) and Christensen and Lund (2010). The model is classical BLUP augmented by a genomic component:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_A}\begin{bmatrix}\mathbf{u_1}\\ \mathbf{u_2}\end{bmatrix} + \mathbf{e},$$

where **y** is a vector of observations, **b** is a vector of fixed effects, **X** is a design matrix, $\mathbf{u} = \begin{bmatrix}\mathbf{u_1}\\ \mathbf{u_2}\end{bmatrix}$ is a vector of breeding values (subscripts 1 and 2 indicate non-genotyped and genotyped animals, respectively) and **e** is a vector of residual errors [$\mathbf{e} \sim N(0; \mathbf{I}\sigma_E^2)$]. $\mathbf{Z_A}$ is a design matrix that connects observations with the corresponding animals. The variance-covariance structure of breeding values is $\mathbf{u} \sim N(0; \mathbf{H}\sigma_A^2)$, where **H** is the single-step (combined pedigree and genomic) relationship matrix. The combined relationship matrix **H** integrates pedigree and genomic information and is calculated in the following way from pedigree and genomic relationships (Aguilar et al., 2010):

$$\mathbf{H} = \begin{bmatrix} \mathbf{H_{11}} & \mathbf{H_{12}} \\ \mathbf{H_{21}} & \mathbf{H_{22}} \end{bmatrix} = \begin{bmatrix} \mathbf{A_{11} - A_{12}A_{22}^{-1}A_{21} + A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21}} & \mathbf{A_{12}A_{22}^{-1}G} \\ \mathbf{GA_{22}^{-1}A_{21}} & \mathbf{G} \end{bmatrix}$$

A is the pedigree-based relationship matrix between all animals in the pedigree and **G** is a genomic relationship matrix between genotyped animals, which can be calculated e.g. by the method of VanRaden (2008).

Aguilar et al. (2010) showed the following way to calculate in the inverse of **H** directly:

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G^{-1} - A_{22}^{-1}} \end{bmatrix}$$

In the multi-step procedure, evaluation of genotyped animals is based on pseudo-phenotypes (DYD/YD, DRP, EBV) that have been estimated in a previous conventional genetic evaluation (e.g. VanRaden 2008; cf. chapter 1). Genomic breeding values are recombined with conventional estimates by means of a selection index. A challenge in multi-step evaluation is the correct scaling of genomic breeding values as they are estimated only from the subset of genotyped animals but not from the entire data set. Different attempts have been made to scale the genomic relationship matrix in order to have genomic and conventional breeding values on the same scale (e.g. VanRaden 2008; Meuwissen et al., 2011; Vitezica et al., 2011). Despite these scaling techniques it has been reported that the dispersion of predicted breeding values was closer to the expected values with the single-step procedure using all information simultaneously (Gao et al., 2012; Su et al., 2012). In turn convergence problems have been reported with large single-step evaluations (Harris et al., 2013; Liu et al., 2014). In order to overcome the difficulties of full single-step analysis and at the same time to profit maximally from single-step mechanisms, different groups of authors (Gao et al., 2012; Su et al., 2012; Harris et al., 2013) have applied adapted approaches analyzing pseudo-phenotypes of genotyped and non-genotyped animals jointly. We will call this type of model ‘reduced single-step’ in this thesis. When bulls’ information is to be included in the evaluation of breeding

values and dominance deviations, a single-step model has to be applied because dominance effects have to be associated with individual genotypes of cows. In addition to convergence problems, the inversion of the dominance relationship matrix might not be feasible with large single-step data sets and the direct calculation of the inverse dominance relationship matrix via sire-dam subclasses (Hoeschele and VanRaden, 1991) would even increase the model size. A reduced single-step model including DYD of bulls and YD of a subset of (genotyped) cows might be a useful alternative for this type of evaluation. In chapter 4 of this thesis, estimated breeding values and dominance deviations have been compared between reduced and full single-step models.

How to measure reliability of breeding values?

Model-based reliability

Both the amount of information available for an animal (from own records and/or from performance of relatives) and the variance components determine the prediction error variance (PEV) of a breeding value. The prediction error variance is the variance of the difference between true and predicted breeding value and thus related to the reliability of the predicted breeding value. Prediction error variance of the breeding value of animal *i* can be calculated directly from the inverted coefficient matrix of the mixed model equation (Henderson, 1975):

$$\mathbf{PEV_i} = var(\mathbf{u_i} - \hat{\mathbf{u}}_\mathbf{i}) = \mathbf{C_{ii}^{22}}\sigma_E^2 = (1 - \mathbf{r_i}^2)\sigma_A^2,$$

where $\mathbf{C_{ii}^{22}}$ is the *i*-th element of the block referring to breeding values of the generalized inverse of the coefficient matrix and $\mathbf{r_i}^2$ is the model-based reliability of the *i*-th breeding value. From the above equation,

$$\mathbf{r_i}^2 = 1 - \mathbf{C_{ii}^{22}}\frac{\sigma_E^2}{\sigma_A^2}.$$

In practice, inversion of the coefficient matrix is not feasible because of its size. Iterative procedures have been developed to approximate the diagonal elements $\mathbf{C_{ii}^{22}}$ (e.g. Meyer, 1989).

Cross-validation reliability

Cross-validation is an empirical method to assess prediction reliability. For cross-validation, the data set is repeatedly divided in a training set and a validation set. Genomic breeding values are predicted by only using phenotypic information from the training set. The predicted genomic breeding values of validation animals are compared with their aggregated phenotypes or breeding values including records from the validation set. The squared correlation between true and predicted breeding values of validation animals is a measure of prediction reliability (e.g. Legarra et al., 2008). As the true breeding value can only be known in simulation studies but not in real data analyses, cross-validation reliability is computed as the squared correlation between the pseudo-phenotype (e.g. DYD) and the predicted breeding value divided by the reliability of the pseudo-phenotype. When pseudo-phenotypes of validation animals have different reliabilities, a weighted regression should be calculated and the coefficient of determination of this regression – divided by the mean reliability of pseudo-phenotypes – can be used as cross-validation reliability. Predicted genomic breeding values, dominance deviations and total genetic values in chapter 3 of this thesis were validated by cross-validation.

Forward prediction reliability

Forward prediction is a special type of validation without replication. The youngest animals are defined as validation set and genomic breeding values are predicted for validation animals only using phenotypic information from the training set. For validation, also phenotypes of

validation animals are included. Forward prediction is a realistic imitation of predicting genomic breeding values for young bulls after genotyping and comparing these GBV with EBV when daughter information is available. The Interbull GEBV test is calculated by means of forward prediction (Mäntysaari et al., 2010). Empirical reliability from forward prediction is substantially smaller than the model-based reliability from the inverse of the BLUP coefficient matrix when the validation animals are selected (Bijma, 2012; Edel et al., 2012).

Genomic breeding values predicted in chapter 2 as well as genomic breeding values, dominance deviations and total genetic values of validation cows in chapter 4 of this thesis were validated in a forward prediction analysis.

The objectives of this thesis were to investigate the potential benefits from genome-wide SNP genotypes at higher marker density and from female genotypes in addition to bulls' genotypes in genomic evaluation of the Fleckvieh breed.

References

Affymetrix Inc. (2015): http://media.affymetrix.com/support/technical/datasheets/axiom_gw_bos1_arrayplate_datasheet.pdf (accessed 1 September 2015).

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

Bayerisches Staatsministerium für Ernährung, Landwirtschaft und Forsten (2015): Zuchtprogramm Fleckvieh in Bayern 2012. http://www.stmelf.bayern.de/mam/cms01/landwirtschaft/dateien/zuchtprogramm_fleckvieh.pdf (accessed 28 August 2015).

Beavis, W. D. (1998): QTL analyses: power, precision, and accuracy, pp. 145–162 in Molecular Dissection of Complex Traits, edited by A. H. Paterson. CRC Press, New York.

Bijma, P. (2012): Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. *J. Anim. Breed. Genet.* **129:**345-358.

Browning, S. R., and Browning, B. L. (2007): Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am. J. Hum. Genet.* **81:**1084-1097.

Chen, J., Liu, Z., Reinhardt, F., and Reents, R. (2011): Reliability of genomic prediction using imputed genotypes for German Holsteins: Illumina 3K to 54K bovine chip. *Interbull Bulletin* **44:**51-54.

Christensen, O. F., and Lund, M. S. (2010): Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **42:**2.

Churchill, G. A., and Doerge, R. W. (1994): Empirical Threshold Values for Quantitative Trait Mapping. *Genetics* **138:**963-971.

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R. F., Liao, X., Djari, A., Rodriguez, S. C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M.-N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P. J., Coote, D., Chamberlain, A. J., Anderson, C., VanTassell, C. P., Hulsegge, I., Goddard, M. E., Guldbrandtsen, B., Lund, M. S., Veerkamp, R. F., Boichard, D. A., Fries, R., and Hayes, B. J. (2014): Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**:858-865.

Dassonneville, R., Brøndum, R. F., Druet, T., Fritz, S., Guillaume, F., Guldbrandtsen, B., Lund, M. S., Ducrocq, V., and Su, G. (2011): Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *J. Dairy Sci.* **94:**3679-3686.

Edel, C., Neuner, S., Emmerling, R., and Götz, K.-U. (2012): A Note on using ‚Forward Prediction' to Assess Precision and Bias of Genomic Predictions. *Interbull Bull.* **46:**16-19.

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

Falconer, D. S., and Mackay, T. F. C. (1996): Introduction to Quantitative Genetics. Fourth edition. Pearson Education Limited, Harlow.

Fernando, R. L., and Grossman, M. (1989): Marker-assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* **21:**467-477.

Fisher, R. A. (1918): The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh* **52:**399-433.

Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009): Deregressing estimated breeding values and weighting information for genomic analyses. *Genet. Sel. Evol.* **41:**55.

Geldermann, H. (1975): Investigations on inheritance of quantitative characters in animals by gene markers I. Methods. *Theor. Appl. Genet.* **46:**319-330.

Gengler, N., Mayeres, P., and Szydlowski, M. (2007): A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal* **1:**21-28.

Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. (2009): Additive genetic variability and the Bayesian alphabet. *Genetics* **183:**347-363.

Goddard, M. E. (2009): Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136:**245-257.

Guillaume, F., Fritz, S., Boichard, D., and Druet, T. (2008): *Short Communication*: Correlations of Marker-Assisted Breeding Values with Progeny-Test Breeding Values for Eight Hundred Ninety-Nine French Holstein Bulls. *J. Dairy Sci.* **91:**2520-2522.

Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G., and Chee, M. S. (2005): A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genetics* **37:**549-554.

Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007): The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics* **177:**2389-2397.

Haldane, J. B. S. (1932): The Causes of Evolution. Longmans, Green, London.

Harris, B. L., Creagh, F. E., Winkelman, A. M., and Johnson, D. L. (2011): Experiences with the Illumina high density Bovine BeadChip. *Interbull Bulletin* **44:**3-7.

Harris, B. L., Winkelman, A. M., and Johnson, D. L. (2013): Impact of including a large number of female genotypes on genomic selection. *Interbull Bulletin* **47:**23-27.

Hayes, B. J., and Goddard, M. E. (2001): The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.* **33:**209-229.

Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009): Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res. Camb.* **91:**47-60.

Hazel, L. N. (1943): The genetic basis for constructing selection indexes. *Genetics* **28:**476-490.

Henderson, C. R. (1973): Sire evaluation and genetic trends. Pages 10-412 in Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. J. L. Lush. American Society of Animal Science and American Dairy Science Association, Champaign, IL.

Henderson, C. R. (1975): Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31:**423-447.

Henderson, C. R. (1976): A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics* **32:**69-83.

Henderson, C. R., and Quaas, R. L. (1976): Multiple Trait Evaluation Using Relatives' Records. *J. Anim. Sci.* **43:**1188-1197.

Hickey, J.M., Cleveland, M., Gorjanc, G., Tier, B., van der Werf, J.H.J., and Kinghorn, B. (2011): An Imputation Strategy which Results in an Alternative Parameterization of the Single Step Genomic Evaluation. *Interbull Bulletin* **44:**38-41.

Illumina Inc. (2015a): http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_bovine_snp5O.pdf (accessed 1 September 2015).

Illumina Inc. (2015b): http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_bovineHD.pdf (accessed 1 September 2015).

Illumina Inc. (2015c): http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_bovineLD.pdf (accessed 1 September 2015).

Legarra, A., Aguilar, I., and Misztal, I. (2009): A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* **92:**4656-4663.

Legarra, A., Ricard, A., and Filangi, O. (2014): **GS3 Genomic Selection – Gibbs Sampling – Gauss Seidel (and BayesCπ).** [http://genoweb.toulouse.inra.fr/~alegarra/manualgs3_last.pdf]

Liu, Z., Goddard, M. E., Reinhardt, F., and Reents, R. (2014): A single-step genomic model with direct estimation of marker effects. *J. Dairy Sci.* **97:**5833-5850.

Legarra, A., Robert-Granié, C., Manfredi, E., and Elsen, J.-M. (2008): Performance of Genomic Selection in Mice. *Genetics* **180:**611-618.

Liu, Y., Qin, X., Song, X.-Z. H., Jiang, H., Shen, Y., Durbin, K. J., Lien, S., Kent, M. P., Sodeland, M., Ren, Y., Zhang, L., Sodergren, E., Havlak, P., Worley, K. C., Weinstock, G. M., and Gibbs, R. A. (2009): *Bos taurus* genome assembly. *BMC Genomics* **10:**180.

Mäntysaari, E. A., Liu, Z., VanRaden, P. M. (2010): Interbull validation test for genomic evaluations. *Interbull Bulletin* **41:**17-22.

Meyer, K. (1989): Approximate accuracy of genetic evaluation under an animal model. *Livestock Production Science* **21:**87-100.

Meuwissen, T. H. E., and Goddard, M. E. (2004): Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol.* **36:**261-279.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001): Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157:**1819-1829.

Meuwissen, T. H. E., Luan, T., and Woolliams, J. A. (2011): The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. *J. Anim. Breed. Genet.* **128:**429-439.

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., and Gengler, N. (1997): Relationships among estimates of inbreeding depression, dominance and additive variance for linear traits in Holsteins. *Genet. Sel. Evol.* **29:**319-326.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. H. (2002): BLUPF90 and related programs (BGF90). In *Proceedings of the 7th World Congress Applied to Livestock Production: 19-23 August 2002; Montpellier*; 28-07.

Reed, D. R., Lawler, M. P., and Tordoff, M. G. (2008): Reduced body weight is a common effect of gene knockout in mice. *BMC Genet.* **9:**4.

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2011): FImpute – An efficient imputation algorithm for dairy cattle populations. *J. Dairy Sci.* **94**(E-Suppl. 1):421.

Schaeffer, L. R. (2004): Application of random regression models in animal breeding. *Livest. Prod. Sci.* **86:** 35-45.

Sillanpää, M. J., and Corander, J. (2002): Model choice in gene mapping: what and why. *Trends Genet.* **18:**301-307.

Steemers, F. J., Chang, W., Lee. G., Barker, D. L., Shen, R., and Gunderson, K. L. (2006): Whole-genome genotyping with the single-base extension assay. *Nature Methods* **3:**31-33.

Steemers, F. J., and Gunderson, K. L. (2007): Whole genome genotyping technologies on the BeadArray™ platform. *Biotechnol. J.* **2:**41–49.

Su, G., Brøndum, R. F., Ma, P., Guldbrandtsen, B., Aamand, G. P., and Lund, M. S. (2012): Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *J. Dairy Sci.* **95:**4657-4665.

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:**e45293.

Tempelman, R. J., and Burnside, E. B. (1990a): Additive and nonadditive genetic variation for production traits in Canadian Holsteins. *J. Dairy Sci.* **73:**2206-2213.

Tempelman, R. J., and Burnside, E. B. (1990b): Additive and nonadditive genetic variation for conformation traits in Canadian Holsteins. *J. Dairy Sci.* **73:**2214-2220.

Toro, M. A., and Varona, L. (2010): A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol.*, **42:**33.

VanRaden, P.M., Null, D.J., Sargolzaei, M., Wiggans, G.R., Tooker, M.E., Cole, J.B., Sonstegard, T.S., Connor, E.E., Winters, M., van Kaam, J.B.C.H.M., Valenti, A., Van Doormaal, B.J., Faust, M.A., and Doak, G.A. (2013): Genomic imputation and evaluation using high-density Holstein genotypes. *J. Dairy Sci.* **96:**668–678.

VanRaden, P. M., O'Connell, J. R., Wiggans, G. R., and Weigel, K. A. (2011): Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* **43:**10.

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., and Schenkel, F. S. (2009): Invited review: Reliability of genomic predictions

for North American Holstein bulls. *J. Dairy Sci.* **92:**16–24

VanRaden, P. M., and Wiggans, G. R. (1991):. Derivation, calculation, and use of national animal model information. *J. Dairy Sci.* **74:**2737-2746.

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.

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

Vitezica, Z. G., Varona, L., Legarra, A. (2013): On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* **195:**1223-1230.

Wellmann, R., and Bennewitz, J. (2012): Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genet. Res. (Camb.)* **94:**21-37.

Wiggans, G. R., Cooper, T. A., VanRaden, P. M., Olson, K. M., and Tooker, M. E. (2012): Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. *J. Dairy Sci.* **95:**1552-1558.

Wiggans, G. R., VanRaden, P. M., and Cooper, T. A. (2011): The genomic evaluation system in the United States: Past, present, future. *J. Dairy Sci.* **94:**3202–3211.

Wright, S. (1921a): Systems of mating. I. The biometric relations between parent and offspring. *Genetics* **6:**111-123.

Wright, S. (1921b): Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* **6:**124-143.

Wright, S. (1921c): Systems of mating. III. Assortative mating based on somatic resemblance. *Genetics* **6:**144-161.

Wright, S. (1921d): Systems of mating. IV. The effects of selection. *Genetics* **6:**162-166.

Wright, S. (1921e): Systems of mating. V. General considerations. *Genetics* **6:**167-178.

Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C. P., Sonstegard, T. S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J. A., and Salzberg, S. L. (2009): A whole-genome assembly of the domestic cow, *Bos taurus. Genome Biol.* **10:**R42.

2nd Chapter

On the limited increase in validation reliability using high-density genotypes in genomic best linear unbiased prediction: Observations from Fleckvieh cattle

Johann Ertl*, Christian Edel*, Reiner Emmerling*, Hubert Pausch[§], Ruedi Fries[§], Kay-Uwe Götz*

* Institute of Animal Breeding, Bavarian State Research Centre for Agriculture, 85586 Poing, Germany

[§] Chair of Animal Breeding, Technische Universität München, 85354 Freising, Germany

published in *Journal of Dairy Science* 97:487-496

Abstract

This study investigated reliability of genomic predictions using medium-density (40,089; 50K) or high-density marker sets (HD; 388,951). We developed an approximate method to test differences in validation reliability for significance. Model based reliability and the effect of HD genotypes on inflation of predictions were analyzed additionally. Genomic breeding values were predicted for at least 1,321 validation bulls based on phenotypes and genotypes of at least 5,324 calibration bulls by means of a linear model in milk, fat and protein yield; somatic cell score; milkability; muscling; udder, feet and legs score as well as stature. In total, 1,485 bulls were actually HD genotyped and HD genotypes of the other animals were imputed from 50K genotypes using FImpute software. Validation reliability was measured as the coefficient of determination of the weighted regression of daughter yield deviations on predicted breeding values divided by the reliability of daughter yield deviations and inflation was evaluated by the slope of this regression. Model based reliability was calculated from the model. Distributions for validation reliability of 50K markers were derived by repeated sampling of 50,000-marker samples from HD to test differences in validation reliability statistically. Additionally, the benefit of HD genotypes in validation reliability was tested by repeated sampling of validation groups and calculation of the difference in validation reliability between HD and 50K genotypes for the sampled groups of bulls. The mean benefit in validation reliability of HD genotypes was 0.015 compared with real 50K genotypes and 0.028 compared with 50K samples from HD affected by imputation error and was significant for all traits. The model based reliability was, on average, 0.036 lower and the regression coefficient was 0.036 closer to the expected value with HD genotypes. The observed gain in validation reliability with HD genotypes was similar to expectations based on the number of markers and the effective number of segregating chromosome segments. Sampling error in the marker-based relationship coefficients causing overestimation of the model based reliability was smaller with HD

genotypes. Inflation of the genomic predictions was reduced with HD genotypes, accordingly. Similar effects on model based reliability and inflation, but not on the validation reliability, were obtained by shrinkage estimation of the realized relationship matrix from 50K genotypes.

Introduction

The use of genomic information for the prediction of breeding values is now widespread in dairy cattle breeding all over the world. Genomic breeding values for selection candidates that are predicted with dense marker genotypes are much more reliable than parent averages (VanRaden et al., 2009). Breeding animals are genotyped with high-throughput technologies for thousands of SNP covering the whole genome. Genotyped markers are used as surrogates for unknown QTL in the estimation of the realized relationship matrix. Realized relationships between animals deviate from the expected numerator relationships because of Mendelian sampling during meiosis. In a genomic BLUP (**GBLUP**) model (VanRaden, 2008), the numerator relationship matrix is replaced by the marker-based relationship matrix. This improves the reliability of breeding values of selection candidates that are predicted based on realized relationships with proven bulls that have phenotypic information. Although currently used assays enable efficient genotyping of more than 50,000 SNP (**50K**), a high-density (**HD**) assay was developed for genotyping more than 777,000 SNP, providing genomic information at a higher resolution (Illumina Inc., San Diego, CA). In several investigations, reliabilities of predicted breeding values were compared when breeding values were predicted either from 50K or from HD genotypes using a GBLUP model, but only minor, if any, gains in validation reliability were observed (Erbe et al., 2012; Su et al., 2012). In this study, we examined the gain in validation reliability from HD genotypes in the Fleckvieh breed, which is known to be genetically more diverse with lower levels of linkage disequilibrium between pairs of markers than Holstein (Pryce et al., 2011). We inferred distributions for the validation reliability of

genomic breeding values predicted with varying 50K marker sets to test a potential advantage of HD genotypes for statistical significance. Furthermore, we analyzed the effect of higher marker density on model-based reliability and inflation of the predictions. The observed difference in validation reliability with HD genotypes was compared with theoretical expectations.

Materials and Methods

Data

Genotypes for 21,092 Fleckvieh bulls and candidates were available from the German-Austrian joint genomic evaluation program. The genotypes were generated with Illumina Bovine SNP50 (v1 and v2) Genotyping BeadChips (Illumina Inc.). A total of 1,492 bulls and 2,038 cows were additionally genotyped with an Illumina BovineHD Genotyping BeadChip. The following editing criteria were applied to 50K markers: SNP that had a mean call rate below 0.95 or deviated from Hardy-Weinberg equilibrium with $P < 10^{-5}$ were excluded from the data set. Genotypes with a minor allele frequency below 0.02 were also discarded. The same editing criteria were applied to HD markers with an adjusted minor allele frequency threshold of 0.005 due to the smaller number of HD genotyped animals. In total 41,274 50K and 624,892 HD markers passed the quality criteria and were used for imputation. Consistency between genomic and pedigree information was checked by a 2-step approach. In step 1, the genotypes of parent-offspring pairs were compared directly and conflicting genotype-configurations were eliminated. In step 2, relationships to genotyped grandsires and (or) within half- and full-sib groups were checked based on marker-based identity-by-descent (IBD)-coefficients (Wang, 2007). Conflicting genotypes were either partly or completely set to missing or the sire or the dam was removed from the animal's pedigree following a defined protocol. Imputation of 50K

to HD genotypes was performed with the program FImpute (Sargolzaei et al., 2011). After imputation, genotypes of 629,028 SNPs were available. Although all available genotypes were used for imputation to obtain maximum imputation accuracy by exploiting comprehensive pedigree information, the analysis was based on 50K and imputed HD genotypes of 10,240 bulls registered at AI stations and, thus, potentially possessing phenotypic information. Among these 10,240 bulls, 1,492 were actually HD genotyped and HD genotypes of the remaining 8,748 bulls were imputed from 50K genotypes; 50K and imputed HD genotypes were additionally checked for redundancy. A SNP was considered as redundant when it was in very high linkage disequilibrium (r^2 >0.99) with an adjacent locus. From redundant SNP, the first with regard to UMD3 assembly (http://www.cbcb.umd.edu/research/bos_taurus_assembly.shtml) position was kept in the data set. For computational reasons, the redundancy check was performed only within chromosomes but not across the whole genome. After exclusion of redundant SNP, 388,951 loci remained in the imputed HD dataset and 40,089 remained in the 50K data set.

Phenotypes for the investigation were daughter yield deviations (**DYD**; VanRaden and Wiggans, 1991) for milk, fat and protein yield, SCS, stature, muscling, udder and feet and legs and deregressed proofs (**DRP**; Garrick et al., 2009) for milkability from the December 2012 German-Austrian routine evaluation. A fraction of 2,265 bulls did not have phenotypes yet.

Genomic Prediction

Forward predictions were computed using GBLUP (VanRaden, 2008). Bulls born before April 2005 were assigned to the calibration group and the younger bulls served as validation animals. Calibration and validation groups were defined separately for each trait according to the amount of phenotypic information. Only data from validation bulls with phenotypic information equivalent to at least 20 effective daughter contributions (Wilmink and Dommerholt, 1985) in the respective trait were used to validate the genomic predictions. For

the different traits, calibration and validation groups consisted of at least 5,324 and 1,321 bulls, respectively. Birth years of analyzed bulls are shown in Figure 1. Direct genomic values (**DGV**) were predicted for validation bulls based on genotypes and phenotypes of the calibration group. The genomic relationship matrix $\mathbf{G}^*$ was calculated following the approach of VanRaden (2008):

$$\mathbf{G}^* = \frac{\mathbf{ZZ}'}{2\sum_{k=1}^{m} p_k(1-p_k)},$$

where **Z** is a genotype matrix having a dimension of the number of individuals (n) by the number of loci (m) and is calculated as **Z** = **M** - **P**. The elements of **M** are -1 and 1 for opposite homozygous genotypes and 0 for heterozygous genotypes. Column k of the matrix **P** is $2(p_k - 0.5)$ and p_k is the base allele frequency of locus k estimated according to the approach of Gengler et al. (2007); $\mathbf{G}^*$ was additionally scaled toward the numerator relationship matrix **A** (Meuwissen et al., 2011) and finally combined with **A** as $\mathbf{G} = 0.99\mathbf{G}^* + 0.01\mathbf{A}$ to improve numerical stability.

The following model was applied to predict genomic breeding values **g**:

$$\mathbf{y} = \mu + \mathbf{Dg} + \mathbf{e},$$

where **y** is a vector of DRP (milkability) or DYD (other traits) of calibration animals, μ is the intercept and **D** is a design matrix that relates **y** to breeding values; **g** is a vector of DGV for all animals with genotypes in **D**, including calibration and validation animals. Genomic breeding values were predicted for calibration and validation animals by best linear unbiased prediction (Henderson, 1973):

$$\hat{\mathbf{g}} = \mathbf{G}\sigma_a^2\mathbf{D}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{1_n}\hat{\mu}),$$

where $\mathbf{1}_n$' is a vector of 1, $\hat{\mu} = (\mathbf{1_n}'\mathbf{V}^{-1}\mathbf{1_n})^{-1}\mathbf{1_n}'\mathbf{V}^{-1}\mathbf{y}$, σ_a^2 is the additive genetic variance and **V**, the covariance matrix of DYD or DRP, was calculated as

$$\mathbf{V} = \mathbf{D}(\mathbf{G}\sigma_a^2)\mathbf{D}' + \mathbf{W}\sigma_e^2,$$

where σ_e^2 is the error variance and **W** is a diagonal matrix with reciprocals of the equivalent number of own performances (**EOP**); EOP are calculated as follows: $EOP = \frac{\sigma_e^2}{\sigma_a^2}\frac{R_{DYD}^2}{1-R_{DYD}^2}$, where R_{DYD}^2 is the reliability of the DYD. The parameters σ_a^2 and σ_e^2 were adopted from the German-Austrian routine evaluation.

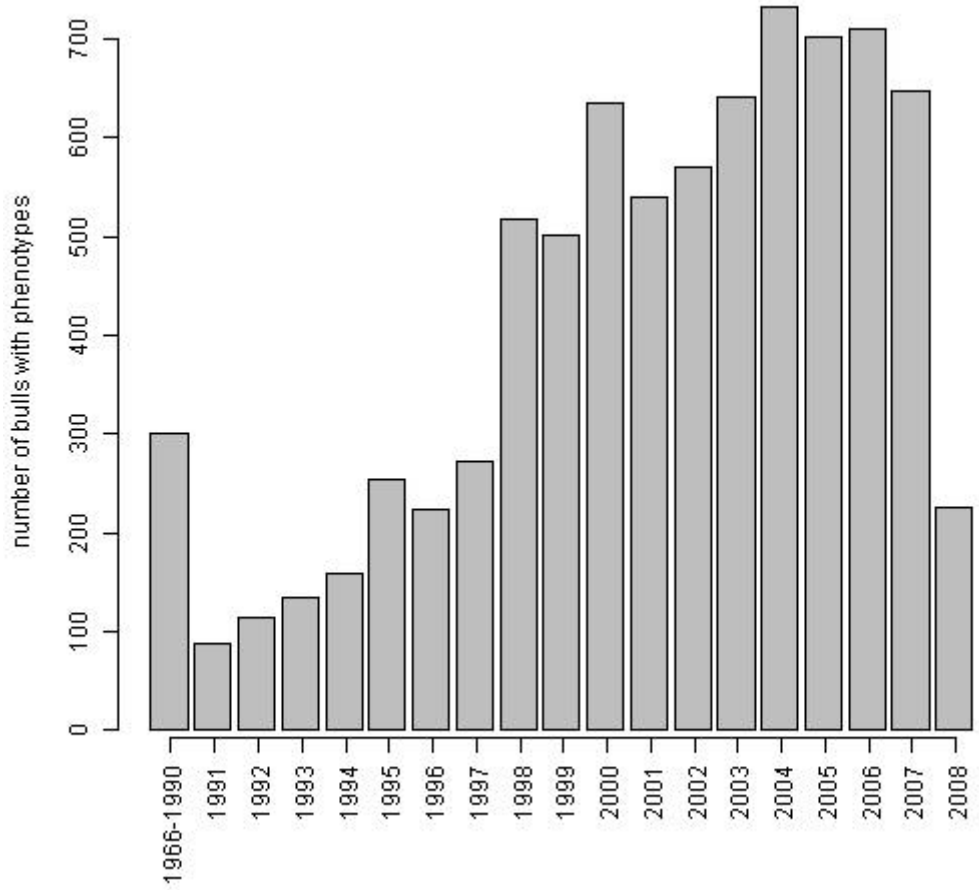

Figure 1. Birth years of bulls with known phenotypes.

The reliability of the prediction as observed in the validation group was measured by means of the coefficient of determination of the weighted regression of DYD on DGV divided by the average reliability of DYD [R^2 (validation)] of validation bulls. We will refer to this reliability as "validation reliability" hereinafter. The model-based reliability [R^2 (model based)] was computed by direct inversion of the genomic equation system. The weighted regression coefficient of DYD on DGV [**b(DYD, DGV)**] was calculated to estimate the inflation of the predictions. Inflation means that b(DYD, DGV) is not equal to its expectation (Mäntysaari et al., 2010). For milkability, DYD were not available and DRP were used instead of DYD to calculate R^2 (validation) and b(DYD, DGV).

To create an empirical distribution of validation reliabilities of 50K predictions, 500 random subsets of 40,089 markers (the number of 50K markers in this analysis; designated as 50K) were sampled from the HD markers and used for calibration and prediction of DGV. This kind of comparison is regarded as fair because 50K samples are affected by imputation error to the same extent as HD genotypes in contrast to real 50K markers that were actually genotyped and thus not affected by imputation error. To obtain samples with a structure similar to the original 50K Illumina Bead Chip, the chromosomes were subdivided in bins of 1-Mb length. The number of markers sampled per 1-Mb bin was determined by the number of 50K markers observed in this 1-Mb bin. Predictions based on sampled 50K markers were performed to obtain an empirical distribution of validation reliability for 50K marker sets. For each trait, we calculated the probability that a value from the empirical distribution of 50K validation reliabilities exceeds the observed reliability from GBLUP using HD genotypes.

For computational reasons, we performed the sampling of 50K markers in all traits with the results from only 1 validation group. However, because the difference in validation reliability between HD and 50K genotypes might also be influenced by the structure of the validation group, we additionally sampled 50 groups of 500 animals with replacement from all potential

validation animals. The difference in R^2 (validation) between HD and real 50K in each trait was calculated for each validation sample. This difference is the result of the relevant comparison valid in the current situation that all selection candidates are genotyped with 50K and only a fraction of bulls are HD genotyped. By means of a one-sided *t*-test, we tested the null hypothesis that validation reliability with HD genotypes is not larger than with 50K genotypes.

Both Goddard et al. (2011) and Endelman and Jannink (2012) proposed to reduce the sampling error of genomic relationship coefficients, which is basically a function of the number of markers, by shrinking the coefficients toward a target. Goddard et al. (2011) recommended shrinking against the pedigree-based numerator relationship matrix. The shrinkage target of Endelman and Jannink (2012) is a diagonal matrix with elements 1 + *f*, where *f* is the mean inbreeding coefficient of the current population. We evaluated the impact of shrinkage estimation of the realized relationship matrix in GBLUP with 50K genotypes, following the approach of Endelman and Jannink (2012), on the validation and model based reliability as well as on the inflation of DGV and compared the results to the HD results. For this approach, the realized relationship matrix was first calculated as described above using estimates of current allele frequencies. In a second step, the matrix was shrunk toward 1 + *f*. According to Endelman and Jannink (2012), the genomic relationship matrix (calculated from marker genotypes) $\mathbf{S} = \frac{\mathbf{ZZ}'}{m} - \frac{1}{m}\sum_{k=1}^{m}\mathbf{Z}_{\bullet k} \cdot \frac{1}{m}\sum_{k=1}^{m}\mathbf{Z}'_{\bullet k}$ has to be shrunk unless the number of markers is much larger than the number of animals to reduce the mean squared error of the genomic relationship coefficients. Shrinkage of **S** leads to the realized relationship matrix

$$\mathbf{G}^{*} = \frac{\delta\frac{1}{n}\sum_{i=1}^{n}\mathbf{S}_{ii}\mathbf{I} + (1-\delta)\mathbf{S} + \frac{1}{m}\sum_{k=1}^{m}\mathbf{Z}_{\bullet k} \cdot \frac{1}{m}\sum_{k=1}^{m}\mathbf{Z}'_{\bullet k}}{2\frac{1}{m}\sum_{k=1}^{m}p_k(1-p_k)},$$

where δ is the optimal shrinkage intensity ranging from 0 to 1 with the heuristic $\delta \sim \frac{n}{m \cdot CV^2}$.

n is the number of genotyped animals, **I** is an identity matrix of dimension *n*, *m* the number of markers and *CV* is the coefficient of variation of the eigenvalues of **S**. The optimal shrinkage intensity was estimated dependent on the number of markers.

Results

The validation reliability of genomic breeding values when mainly imputed high-density genotypes were used ranged from 0.313 in udder to 0.559 in SCS and was larger than the reliability from 50K genotypes for all traits (Table 1). The differences in validation reliability between DGV based on mainly imputed HD and actually genotyped 50K markers, however, were not large and ranged from 0.008 to 0.023. The mean increase in validation reliability of HD over all traits was 0.015. In contrast to the validation reliability, the model-based reliability with HD genotypes was lower than with 50K genotypes. The model-based reliability with HD genotypes ranged from 0.547 in feet and legs to 0.662 in milk yield and was 0.626, on average. The mean model-based reliability for 50K genotypes was 0.662.

The slopes of the regression of DYD on DGV for HD genotypes ranged from 0.689 in protein yield to 1.024 in milkability and were in all traits closer to the expected value when HD genotypes were used instead of 50K genotypes. The mean of b(DYD, DGV) over all traits was 0.839 for HD and 0.803 for 50K genotypes. The mean difference in b(DYD, DGV) indicates less inflation of the predictions from HD genotypes. The correlation between 50K and HD predictions was between 0.959 and 0.974 for the different traits. This leads to moderate differences in the ranking of the bulls.

For all analyzed traits, the validation reliability from HD genotypes was clearly larger than the mean of the distribution of R^2 (validation) of sampled 50K marker sets. As an example, the

results from the sampling of 50K markers are shown in Figure 2 for the traits milk yield, SCS, udder, and stature. The dotted line indicates the 95% quantile of the distribution of validation reliability from sampled 50K genotypes. The observed validation reliability from HD genotypes exceeds this 95% quantile for all traits. The probabilities that a sample of 50K SNP from HD genotypes results in at least the same validation reliability as HD genotypes are given in Table 2 and they are always smaller than 1% for the different traits.

Table 1. Validation [R^2 (validation); adjusted for reliability of daughter yield deviations (DYD)] and model-based reliability [R^2 (model based)] and regression coefficients [b(DYD, DGV), where DGV = direct genomic value] of genomic predictions with 50,000-SNP (50K) and high-density (HD) genotypes

	R^2 (validation)		R^2 (model based)		b(DYD, DGV)	
Trait	50K	HD	50K	HD	50K	HD
Milk yield	0.398	0.414	0.701	0.662	0.700	0.739
Fat yield	0.419	0.427	0.696	0.657	0.777	0.810
Protein yield	0.378	0.392	0.691	0.653	0.658	0.689
SCS	0.548	0.559	0.679	0.643	0.818	0.837
Milkability	0.458	0.481	0.685	0.648	0.962	1.024
Muscling	0.480	0.501	0.627	0.594	0.821	0.846
Udder	0.297	0.313	0.633	0.599	0.845	0.886
Feet and legs	0.315	0.330	0.575	0.547	0.824	0.858
Stature	0.366	0.379	0.669	0.632	0.824	0.863
Average	0.407	0.422	0.662	0.626	0.803	0.839

The validation reliability for sampled 50K subsets averaged 0.394 (Table 2) and was 0.013 smaller than that observed from the real 50K chip. The difference in validation reliability between HD and the mean of sampled 50K genotypes ranged from 0.021 to 0.036 with an

average of 0.028. The sampling of validation animals resulted in statistically significant ($P < 0.001$) differences in validation reliability between HD and 50K genotypes for every trait analyzed.

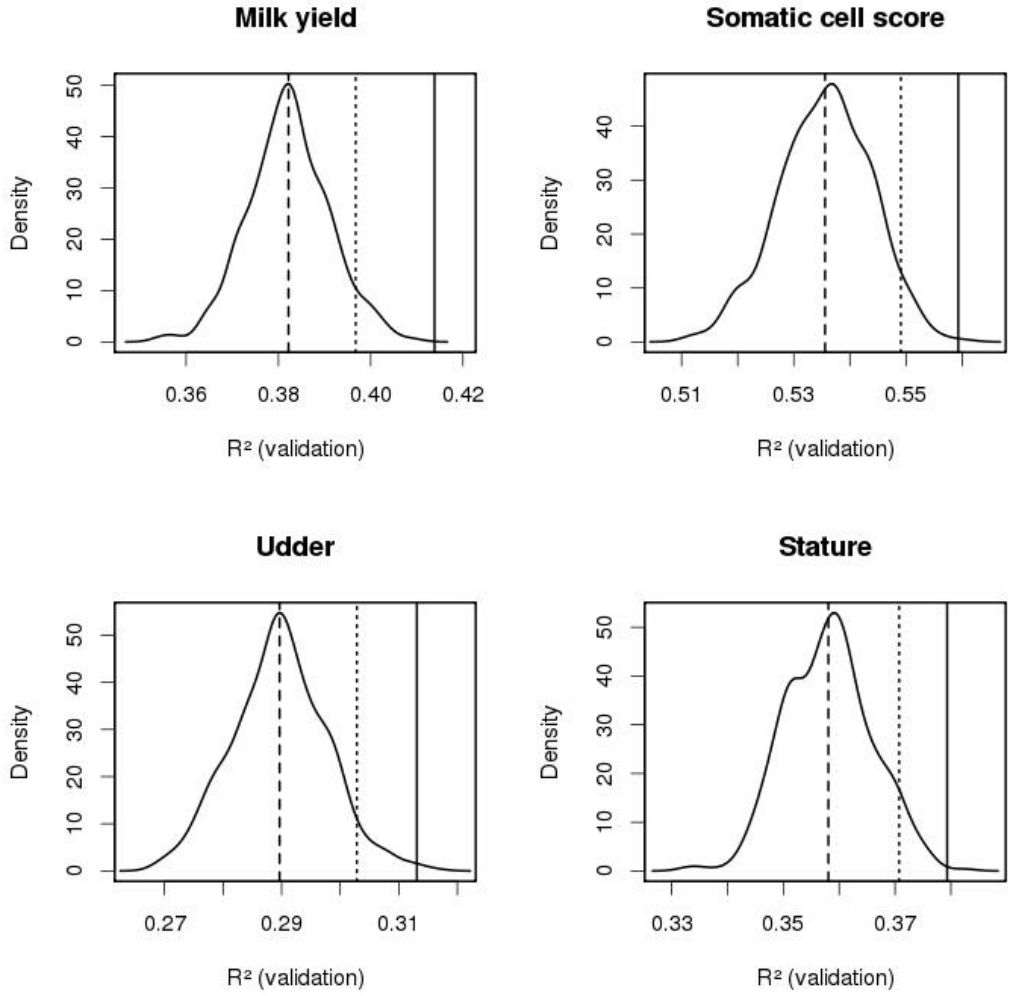

Figure 2. Validation reliability of genomic predictions [R² (validation); adjusted for reliability of daughter yield deviations (DYD)] resulting from high-density (HD) markers (solid line) and from sampled 50,000-SNP (50K) subsets from HD (solid curve; dashed line = mean; dotted line = 95% quantile) in milk yield, SCS, udder, and stature.

Applying shrinkage estimation to 50K genotypes, had almost no effect on the validation reliability of the genomic prediction. The validation reliability resulting from 50K genotypes with shrinkage amounted to 0.409, on average, and was just 0.002 larger than without

shrinkage (Table 3). Model-based reliability from 50K genotypes, however, decreased by 0.026 with shrinkage estimation and approached the model-based reliability from HD genotypes. In the same manner, the regression coefficient b(DYD, DGV) from 50K genotypes with shrinkage approached the result from HD genotypes. On average, it amounted to 0.832 and was 0.029 larger than without shrinkage. The remaining difference to HD genotypes was only 0.007, on average.

Table 2. Validation reliability [adjusted for reliability of daughter yield deviations (DYD)] of genomic predictions resulting from high-density (HD) markers and from sampled 50,000-SNP (50K) subsets from HD, and *P*-values for the statistical test of the hypothesis that any value from the distribution of 50K markers exceeds the validation reliability from HD markers

Trait	50K subset		HD	Difference (HD - 50K subset)	*P*-value
	Mean	95% quantile			
Milk yield	0.382	0.397	0.414	0.032	1.9×10^{-4}
Fat yield	0.393	0.408	0.427	0.034	9.4×10^{-5}
Protein yield	0.359	0.374	0.392	0.033	1.5×10^{-4}
SCS	0.536	0.549	0.559	0.023	2.0×10^{-3}
Milkability	0.445	0.458	0.481	0.036	5.6×10^{-6}
Muscling	0.474	0.487	0.501	0.027	2.4×10^{-4}
Udder	0.290	0.303	0.313	0.023	1.8×10^{-3}
Feet and legs	0.308	0.322	0.330	0.022	5.7×10^{-3}
Stature	0.358	0.371	0.379	0.021	2.9×10^{-3}
Average	0.394	-	0.422	0.028	-

Table 3. Validation [R^2 (validation); adjusted for reliability of daughter yield deviations (DYD)] and model-based reliability [R^2 (model based)] and regression coefficients [b(DYD, DGV), where DGV = direct genomic value] of genomic predictions from 50,000-SNP (50K) genotypes with shrinkage estimation (50K shr.) compared with high-density (HD) genotypes

	R^2 (validation)		R^2 (model based)		b(DYD, DGV)	
Trait	50K shr.	HD	50K shr.	HD	50K shr.	HD
Milk yield	0.403	0.414	0.672	0.662	0.730	0.739
Fat yield	0.422	0.427	0.667	0.657	0.810	0.810
Protein yield	0.382	0.392	0.662	0.653	0.684	0.689
SCS	0.547	0.559	0.651	0.643	0.845	0.837
Milkability	0.466	0.481	0.657	0.648	1.004	1.024
Muscling	0.480	0.501	0.605	0.594	0.837	0.846
Udder	0.298	0.313	0.610	0.599	0.872	0.886
Feet and legs	0.314	0.330	0.556	0.547	0.848	0.858
Stature	0.370	0.379	0.645	0.632	0.856	0.863
Average	0.409	0.422	0.636	0.626	0.832	0.839

Discussion

The aim of the present study was to investigate whether a change from 50K to HD genotypes in the German-Austrian routine evaluation would present an immediate benefit that justifies the additional costs. Therefore, we used the standard GBLUP model of the routine evaluation and the same variance components. The use of GBLUP in the genomic analysis is justified, as no clear evidence exists that the infinitesimal model does not apply to the majority of quantitative traits in dairy cattle populations (VanRaden et al., 2009, 2013; Erbe et al., 2012; Su et al., 2012) and especially in the Fleckvieh breed (Gredler et al., 2010; Pausch et al., 2011; Pryce et al., 2011). Moreover, GBLUP has the practical advantage that genetic base population and

variance components do not need to be changed when it is assured that the genomic relationship matrix has the same scale as the numerator relationship matrix which should be the case in our study. In this context, the objective of this study was to analyze to what extent the higher marker density improves the estimate of a large genomic relationship matrix and reliability and inflation of genomic predictions in a Fleckvieh population given the GBLUP model.

The use of mainly imputed HD genotypes as compared with 50K genotypes resulted in a small gain in validation reliability of 0.015, on average. Other groups, which compared the reliabilities of genomic breeding values of 50K and HD markers predicted with GBLUP, found similar or even smaller differences in reliability between HD and 50K genotypes (Erbe et al., 2012; Su et al., 2012). Gains in validation reliability with HD genotypes were not much larger when nonlinear models were applied (Harris et al., 2011; Erbe et al., 2012; Su et al., 2012; VanRaden et al., 2013). The investigations show that only small, if any, gains in the validation reliability can be obtained with HD genotypes using a GBLUP model. The advantage of 0.015 in validation reliability that we achieved with mainly imputed HD genotypes in our study might not be the true gain which is attainable with HD in the GBLUP model. High-density genotypes were affected by imputation error and this should be taken into account when comparing with results from 50K genotypes. Likewise, the HD genotypes in the studies cited above were imputed from 50K and thus affected by imputation error. The effect of genotype error on the validation reliability is reported in the Appendix. These results show that the decrease of 0.013 in validation reliability with 50K samples compared with real 50K genotypes was caused by imputation error in 50K samples because introduction of a genotype error equivalent to imputation error in Fleckvieh led to a very similar decay of 0.015 compared with original 50K not affected by imputation error.

By means of sampling of 50K marker sets and comparison of the resulting distribution of validation reliability with HD genotypes and sampling of validation animals, we were able to show that the advantage of HD over 50K is statistically significant in all 9 analyzed traits. To assess the quality of the 50K samples for the estimation of genomic breeding values, we compared the mean of all samples to that of the original 50K set. The difference in validation reliability between sampled and original 50K marker sets is likely caused by imputation error affecting imputed HD and sampled 50K genotypes, but not real 50K genotypes. The results in Table 1 represent the current situation that all bulls are genotyped with the 50K chip and only a fraction of bulls are genotyped with the HD chip. Imputed HD genotypes of the remaining bulls are affected by imputation error in contrast to their 50K genotypes. Table 2 summarizes the fair comparison when both marker sets are equally affected by imputation error. These differences apply equally to a setting where all bulls are genotyped with both marker sets or where imputation without error is possible. By means of sampling of validation bulls, we were able to show that the difference between HD and real 50K genotypes is significant even under the unfavorable conditions that 50K markers are genotyped and HD markers are imputed with some level of error.

The observed small but significant increase in validation reliability with HD genotypes in the GBLUP model should be compared with theoretical expectations. The effective number of chromosome segments segregating in the population (Goddard, 2009; Hayes et al., 2009) is a common measure for haplotype diversity and determined by the amount of linkage disequilibrium in the respective population (Goddard et al, 2011). Goddard et al. (2011) derived a theoretical expectation of the reliability of predicted breeding values dependent on the number of calibration animals, mean heritability of the phenotypes, number of markers, and the effective number of chromosome segments segregating in the population. We calculated the expected relative gain from HD markers based on this formula to obtain the expected

validation reliability from HD markers. This expectation was consistent with the observed validation reliability from HD markers (Table 4). Contrary to the small gain in validation reliability with HD genotypes, the model-based reliability decreased with the larger number of markers. In a simulation study, Goddard et al. (2011) demonstrated that an increase in the variance of sampling errors of realized relationship coefficients [V(E)] causes model-based reliability to be overestimated. As V(E) is reciprocally related to the number of markers (Yang et al., 2010), overestimation of the model-based reliability is reduced with HD genotypes. Using HD genotypes has thus the effect that is intended by the use of shrinkage of estimated genomic relationships (Goddard et al., 2011; Endelman and Jannink, 2012): it reduces the variance of the relationship coefficients and induces accordance of model-based and true reliability.

Table 4. Observed validation reliability [adjusted for reliability of daughter yield deviations (DYD)] with 50,000-SNP (50K) and high-density (HD) genotypes and expectation of the validation reliability for HD genotypes

	50K	HD	
Trait	Observed	Expected	Observed
Milk yield	0.398	0.407	0.414
Fat yield	0.419	0.428	0.427
Protein yield	0.378	0.386	0.392
SCS	0.548	0.560	0.559
Milkability	0.458	0.468	0.481
Muscling	0.480	0.490	0.501
Udder	0.297	0.303	0.313
Feet and legs	0.315	0.322	0.330
Stature	0.366	0.374	0.379
Average	0.407	0.415	0.422

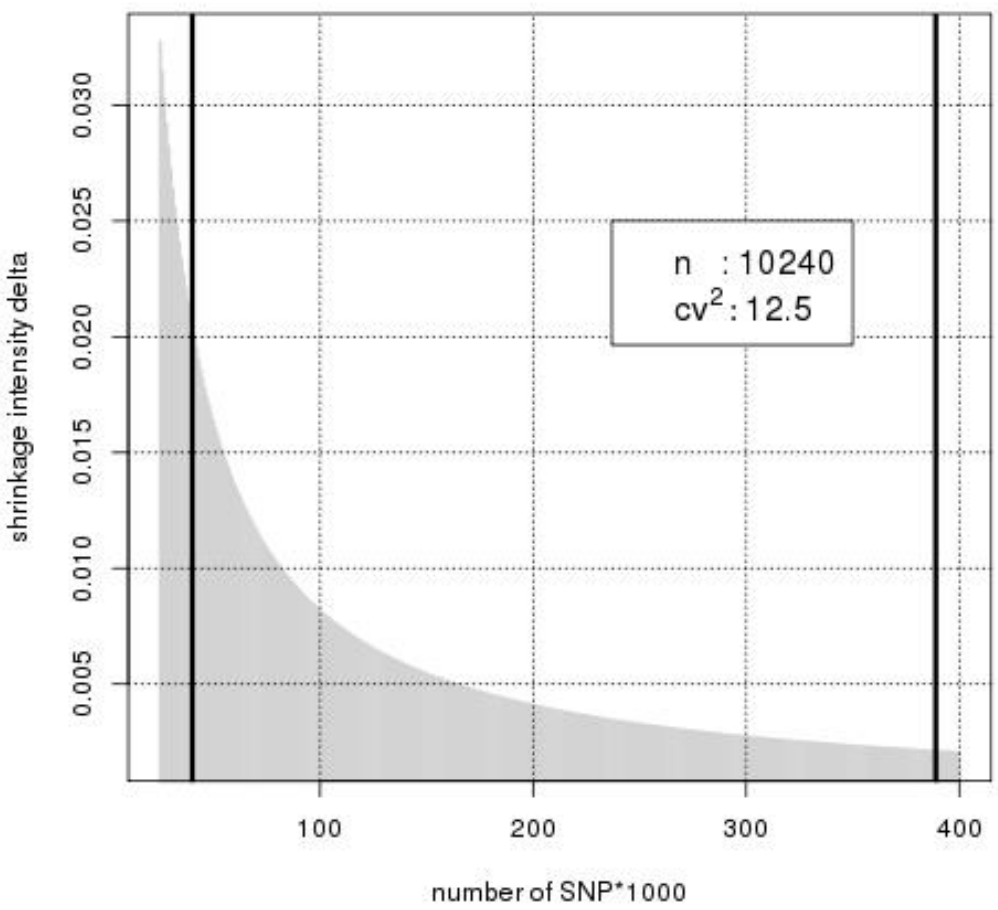

Figure 3. Dependency of the optimal shrinkage intensity on the number of markers in our Fleckvieh data set.

Endelman and Jannink (2012) showed that the sampling error of the relationship coefficients depends on the ratio of *n* (number of animals) and CV^2 (squared coefficient of variation of the eigenvalues of the realized relationship matrix). Notably, with fixed *m* and *CV*, the sampling error will become larger if the number of genotyped animals increases. Even though the shrinkage target in the method of Endelman and Jannink (2012) is different from that of Goddard et al. (2011), the aim is likewise to reduce the estimation error of realized relationship coefficients. The optimal shrinkage was estimated from our data set for different numbers of markers (depicted in Figure 3). Because the optimal shrinkage intensity decreases hyperbolically with the number of markers, it is negligible for our HD data set. From this point

of view, the relationships estimated from HD genotypes should reflect the realized relationships at QTL quite well, provided that the distribution of QTL follows approximately the infinitesimal model and that QTL and markers do not differ systematically in their properties.

Shrinkage estimation resulted in a decrease in model-based reliability, whereas validation reliability remained unchanged. Endelman and Jannink (2012) did not detect any increase in prediction accuracy in validation animals with shrinkage estimation either. In contrast to higher marker density, shrinkage estimation does not provide additional information for the prediction of breeding values but ensures that the variance of the realized relationship coefficients is not overestimated. Nevertheless, model-based reliabilities in this investigation are still considerably larger than validation reliabilities. Possible reasons for this discrepancy are markers not capturing all of the additive genetic variance, preselection of validation bulls to genotyping and error in validation reliability due to the limited size of the validation group (VanRaden et al., 2009). Preselection of validation bulls to genotyping and errors in the validation reliability due to limited sample size are independent of the number of markers and affect concordance of model-based and validation reliability irrespective of the marker set. In the German-Austrian genomic routine evaluation, trait-optimized proportions of polygenic variance ranging from 10 to 25% are included in the analysis to account for genetic variance that is not captured by markers. Therefore, the reported results do not reflect necessarily the situation in the routine evaluation.

Direct genomic values that were predicted with HD genotypes showed less inflation than predictions from 50K genotypes. This finding is confirmed by Su et al. (2012) who predicted genomic values with 50K and HD genotypes and reported that the regression coefficient of deregressed proofs on predicted genomic values was larger when HD genotypes were used. As already pointed out, the variance of estimated relationship coefficients is higher with a limited

number of markers. The consequence is that the variance of predicted breeding values is inflated. This results in b(DYD, DGV) farther from the expectation. With an increasing number of markers, this variance decreases and, consequently, the inflation of the predictions is reduced. However, the inflation of DGV that is still observable with HD probably indicates that even HD does not capture all of the additive genetic variance assumed by the model (Goddard et al., 2011).

Conclusions

Prediction of genomic breeding values in the Fleckvieh breed with HD genotypes instead of 50K genotypes in a GBLUP model leads to small gains in validation reliability and reduces the inflation of predicted breeding values. Model-based reliabilities that are overestimated with 50K genotypes decrease with HD genotypes because the sampling error of the realized relationship matrix is reduced. Similar effects on inflation and model-based reliabilities are obtained by shrinkage estimation of the realized relationship matrix from 50K genotypes. However, whereas HD genotypes increase validation reliability, shrinkage estimation only reduces inflation and potentially overestimated model-based reliability.

Acknowledgments

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). We thank M. Sargolzaei for providing his imputation software FImpute. 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 manuscript has benefited from critical comments of 2 anonymous reviewers.

References

Chen, J., Z. Liu, F. Reinhardt, and R. Reents. 2011. Reliability of genomic prediction using imputed genotypes for German Holsteins: Illumina 3K to 54K bovine chip. The 2011 Interbull Open Meeting, Stavanger, Norway. Interbull, Uppsala, Sweden.

Dassonneville, R., R. F. Brøndum, T. Druet, S. Fritz, F. Guillaume, B. Guldbrandtsen, M. S. Lund, V. Ducrocq, and G. Su. 2011. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. J. Dairy Sci. 94:3679-3686. http://dx.doi.org/10.3168/jds.2011-4299.

Endelman, J. B., and J.-L. Jannink. 2012. Shrinkage estimation of the realized relationship matrix. G3 (Bethesda) 2:1405-1413. http://dx.doi.org/10.1534/g3.112.004259.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114-4129. http://dx.doi.org/10.3168/jds.2011-5019.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic analyses. Genet. Sel. Evol. 41:55. http://dx.doi.org/10.1186/1297-9686-41-55.

Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. Animal 1:21-28. http://dx.doi.org/10.1017/S1751731107392628.

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica 136:245-257. http://dx.doi.org/10.1007/s10709-008-9308-0.

Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J. Anim. Breed. Genet. 128:409-421. http://dx.doi.org/10.1111/j.1439-0388.2011.00964.x.

Gredler, B., H. Schwarzenbacher, C. Egger-Danner, C. Fuerst, R. Emmerling, and J. Sölkner. 2010. Accuracy of genomic selection in dual purpose Fleckvieh cattle using three types of methods and phenotypes. Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Gesellschaft für Tierzuchtwissenschaften e.V., Gießen, Germany.

Harris, B. L., F. E. Creagh, A. M. Winkelman, and D. L. Johnson. 2011. Experiences with the Illumina high density Bovine BeadChip. Interbull Bull. 44:3-7.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92:433-443. http://dx.doi.org/10.3168/jds.2008-1646.

Henderson, C. R. 1973. Sire evaluation and genetic trends. Pages 10-412 in Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. J. L. Lush. American Society of Animal Science and American Dairy Science Association, Champaign, IL.

Mäntysaari, E., Z. Liu, and P. VanRaden. 2010. Interbull validation test for genomic evaluations. Interbull Bull. 41:17-22.

Meuwissen, T. H. E., T. Luan, and J. A. Woolliams. 2011. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. J. Anim. Breed. Genet. 128:429-439. http://dx.doi.org/10.1111/j.1439-0388.2011.00966.x.

Pausch, H., K. Flisikowski, S. Jung, R. Emmerling, C. Edel, K.-U. Götz, and R. Fries. 2012. Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle. Genetics 187:289-297. http://dx.doi.org/10.1534/genetics.110.124057.

Pryce, J. E., B. Gredler, S. Bolormaa, P. J. Bowman, C. Egger-Danner, C. Fuerst, R. Emmerling, J. Sölkner, M. E. Goddard, and B. J. Hayes. 2011. Short communication: Genomic selection using a multi-breed, across-country reference population. J. Dairy Sci. 94:2625-2630. http://dx.doi.org/10.3168/jds.2010-3719.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2011. FImpute – An efficient imputation algorithm for dairy cattle populations. J. Dairy Sci. 94(E-Suppl. 1):421. (Abstr.)

Su, G., R. F. Brøndum, P. Ma, B. Guldbrandtsen, G. P. Aamand, and M. S. Lund. 2012. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. J. Dairy Sci. 95:4657-4665. http://dx.doi.org/10.3168/jds.2012-5379.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423. http://dx.doi.org/10.3168/jds.2007-0980.

VanRaden, P. M., D. J. Null, M. Sargolzaei, G. R. Wiggans, M. E. Tooker, J. B. Cole, T. S. Sonstegard, E. E. Connor, M. Winters, J. B. C. H. M. van Kaam, A. Valentini, B. J. Van Doormaal, M. A. Faust, and G. A. Doak. 2013. Genomic imputation and evaluation using high-density Holstein genotypes. J. Dairy Sci. 96:668-678. http://dx.doi.org/10.3168/jds.2012-5702.

VanRaden, P. M., C. P. Van Tassell, C. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16-24. http://dx.doi.org/10.3168/jds.2008-1514.

VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737-2746. http://dx.doi.org/10.3168/jds.S0022-0302(91)78453-1.

Wang, J. 2007. Triadic IBD coefficients and applications to estimating pairwise relatedness. Genet. Res. 89:135-153. http://dx.doi.org/10.1017/S0016672307008798.

Wilmink, J. B. M., and J. Dommerholt. 1985. Approximate reliability of best linear unbiased prediction in models with and without relationships. J. Dairy Sci. 68:946-952.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565-569. http://dx.doi.org/10.1038/ng.608.

Appendix

We found that shrinkage of the identity-by-state (IBS) matrix results in lower model based reliability while validation reliability remains constant. To assess the effect of wrongly imputed genotypes on the validation reliability, we introduced different degrees of error into the 50K genotypes. For each animal, a proportion of genotypes was drawn at random and substituted by false genotypes. When the original genotype was heterozygous, it was replaced by one of the homozygotes with equal probability. Homozygous genotypes were replaced by either the heterozygote or the alternative homozygote with equal probability. Thus, 50K datasets were created with 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0% false genotypes and used for the prediction of genomic breeding values. The validation reliability was calculated as the coefficient of determination of the regression of DYD on DGV of validation animals divided by the reliability of DYD of validation animals. For each error percentage, 30 replicates were performed to obtain a reliable estimate of R^2 (validation).

The effect of erroneous genotypes on the validation reliability is presented in Figure 4 for different error rates. The validation reliability decreased almost linearly with increasing error percentage. The extent of the decrease differed between the traits. Fat and protein yield were most and milkability was least affected by genotype errors. Validation reliabilities are reported for 1.5% false genotypes in Table 5 and compared with mean validation reliability from 50K samples and validation reliability from the 50K chip. Validation reliability from 50K genotypes with 1.5% errors was, on average, 0.015 lower than validation reliability from original 50K

genotypes. The decay in validation reliability with 1.5% error was very similar to the decrease with sampled 50K subsets from HD genotypes. This indicates that the imputation error was approximately equivalent to a homogeneous genotype error of 1.5%. Analysis of imputation from 50K to HD with FImpute in Fleckvieh resulted in 1.6% genotype error. Similar reductions of validation reliability with imputation error were reported by Chen et al. (2011) and Dassonneville et al. (2011).

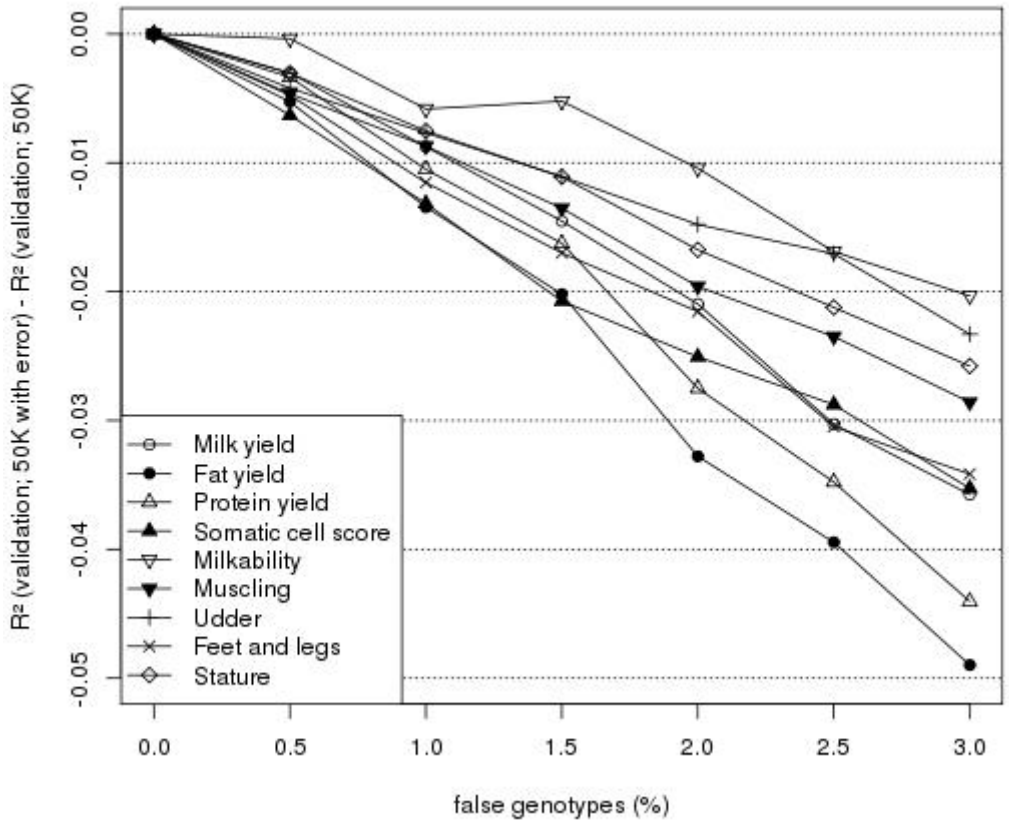

Figure 4. Decay in validation reliability [adjusted for reliability of daughter yield deviations (DYD)] for different percentages of false genotypes. 50K = 50,000 SNP.

Table 5. Validation reliability [adjusted for reliability of daughter yield deviations (DYD)] from original 50,000-SNP (50K) genotypes, 50K genotypes with 1.5% homogeneous error, and sampled 50K genotypes from high density (HD)

Trait	50K	50K with 1.5% error	50K subsets
Milk yield	0.398	0.384	0.382
Fat yield	0.419	0.399	0.393
Protein yield	0.378	0.362	0.359
SCS	0.548	0.527	0.536
Milkability	0.458	0.452	0.445
Muscling	0.480	0.466	0.474
Udder	0.297	0.286	0.290
Feet and legs	0.315	0.298	0.308
Stature	0.366	0.355	0.358
Average	0.407	0.392	0.394

Summary and Authors' Contributions

This study investigated reliability of genomic predictions with medium-density (40,089; 50K) or high-density marker sets (HD; 388,951). Differences in validation reliability were tested for significance by means of an approximate method. Model based reliability and the effect of HD genotypes on inflation of predictions were analyzed additionally. Genomic breeding values were predicted for at least 1,321 validation bulls based on at least 5,324 calibration bulls. In total, 1,485 bulls were actually HD genotyped and HD genotypes of the other animals were imputed from using FImpute software. Distributions for validation reliability of 50K markers were derived by repeated sampling of 50,000-marker samples from HD to test differences in validation reliability. The benefit of HD genotypes in validation reliability was tested by repeated sampling of validation groups and calculating the difference between marker densities for the samples. The mean benefit in validation reliability of HD genotypes was 0.015 compared with real 50K genotypes and 0.028 compared with 50K samples from HD affected by imputation error and was significant for all traits. The model based reliability was, on average, 0.036 lower and the regression coefficient was 0.036 closer to the expected value with HD genotypes. Sampling error in the marker-based relationship coefficients causing overestimation of the model based reliability was smaller with HD genotypes. Inflation of the genomic predictions was reduced with HD genotypes. Similar effects on model based reliability and inflation, but not on the validation reliability, were obtained by shrinkage estimation of the realized relationship matrix from 50K genotypes.

J. Ertl performed the analysis using own-written R programs and drafted the manuscript. J. Ertl, C. Edel and K.-U. Götz designed the study. C. Edel and R. Emmerling prepared phenotypic and genotypic data. C. Edel imputed HD genotypes. C. Edel, R. Emmerling, H. Pausch, R. Fries and K.-U. Götz revised the manuscript.

3rd Chapter

Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle

Johann Ertl*, Andrés Legarra[§+], Zulma G. Vitezica[+§], Luis Varona[#], Christian Edel*, Reiner Emmerling*, Kay-Uwe Götz*

* Institute of Animal Breeding, Bavarian State Research Centre for Agriculture, 85586 Poing, Germany

[§] INRA, UMR 1388 Génétique, Physiologie et Systèmes d'Elevage, CS 52627, 31326 Castanet-Tolosan, France

[+] Université de Toulouse INPT ENSAT, UMR 1388 Génétique, Physiologie et Systèmes d'Elevage, 31326 Castanet-Tolosan, France

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

published in *Genetics Selection Evolution* 46:40

Abstract

Background

Estimates of dominance variance in dairy cattle based on pedigree data vary considerably across traits and amount to up to 50% of the total genetic variance for conformation traits and up to 43% for milk production traits. Using bovine SNP (single nucleotide polymorphism) genotypes, dominance variance can be estimated both at the marker level and at the animal level using genomic dominance effect relationship matrices. Yield deviations of high-density genotyped Fleckvieh cows were used to assess cross-validation accuracy of genomic predictions with additive and dominance models. The potential use of dominance variance in planned matings was also investigated.

Results

Variance components of nine milk production and conformation traits were estimated with additive and dominance models using yield deviations of 1996 Fleckvieh cows and ranged from 3.3% to 50.5% of the total genetic variance. REML and Gibbs sampling estimates showed good concordance. Although standard errors of estimates of dominance variance were rather large, estimates of dominance variance for milk, fat and protein yields, somatic cell score and milkability were significantly different from 0. Cross-validation accuracy of predicted breeding values was higher with genomic models than with the pedigree model. Inclusion of dominance effects did not increase the accuracy of the predicted breeding and total genetic values. Additive and dominance SNP effects for milk yield and protein yield were estimated with a BLUP (best linear unbiased prediction) model and used to calculate expectations of breeding values and total genetic values for putative offspring. Selection on total genetic value instead of breeding value would result in a larger expected total genetic

superiority in progeny, i.e. 14.8% for milk yield and 27.8% for protein yield and reduce the expected additive genetic gain only by 4.5% for milk yield and 2.6% for protein yield.

Conclusions

Estimated dominance variance was substantial for most of the analyzed traits. Due to small dominance effect relationships between cows, predictions of individual dominance deviations were very inaccurate and including dominance in the model did not improve prediction accuracy in the cross-validation study. Exploitation of dominance variance in assortative matings was promising and did not appear to severely compromise additive genetic gain.

Background

Dominance arises when the effects of alleles at a locus are not only additive, but interact so that the value of the heterozygous genotypes deviates from the mean of the values of the homozygous genotypes. With a and $-a$ being the genotypic values of homozygous genotypes A_1A_1 and A_2A_2, let d be the genotypic value of the heterozygous genotype A_1A_2 [1]. If $d = 0$, there is no dominance action at the locus and the genotypic values at the locus are purely additive. The additive effects of genotypes at a locus are expressed as breeding values, which include part of the dominance effect because animals pass alleles, not genotypes, to their offspring. Breeding values are $2q[a + d(q-p)]$ for genotype A_1A_1, $(q-p)[a + d(q-p)]$ for genotype A_1A_2 and $-2p[a + d(q-p)]$ for genotype A_2A_2, where p is the frequency of allele A_1 in the population and q the frequency of allele A_2. The dominance deviation for a given genotype at the locus is the difference between genotypic value and breeding value, and is equal to $-2q^2d$, $2pqd$ and $-2p^2d$ for genotypes A_1A_1, A_1A_2 and A_2A_2, respectively [1].

Until recently, studies on dominance deviations were sparse because without genomic information, the availability of large datasets with sufficient proportions of individuals with non-null dominance effect relationships, such as full-sibs, is essential for accurate estimation of dominance variance. Estimates of dominance variance in dairy cattle that are based on pedigree data range from 7.3% to 49.8% of the total genetic variance for conformation traits [2,3] and from 3.4% to 42.9% for milk production traits [4-6].

At the individual animal level, dominance is hardly used in animal breeding [7], although it contains a relevant part of genetic variation. The reasons are the heavy computational demand of large-scale genetic evaluations for dominance, the relatively low accuracy of resulting estimates of dominance effects, and the complexity of planning and computing the outcome of planned matings [8].

With the availability of SNP genotypes, dominance at a marker locus can be readily determined, dominance effects of markers can be estimated [9,10] and computing the expected outcome of planned matings based on SNP genotypes is straightforward [9]. Furthermore, covariance matrices of genomic dominance effects among individuals can be calculated, similar to matrices of genomic additive relationships, which are widely used in genomic selection, such that dominance effects can be estimated in a GBLUP (genomic best linear unbiased prediction) model [11,12].

In this work, we explored the possibilities of including dominance effects in genomic evaluation and furthermore in planned matings in dairy cattle. We estimated variance components, including dominance variance, in a dataset of genotyped Bavarian Fleckvieh cows, analyzed the predictions of breeding and total genetic values using cross-validation, and predicted total genetic values of specific matings.

Methods

Estimation of variance components

First-lactating cows from 145 Bavarian dairy herds (all first-lactating cows of each herd were genotyped), born in 2008 and 2009, were genotyped with the Illumina BovineHD Genotyping BeadChip that includes 777 962 SNPs. SNPs with a call rate lower than 0.9, a minor allele frequency higher than 0.005 and a highly significant deviation ($P < 10^{-5}$) from the Hardy Weinberg equilibrium, and SNPs that were not annotated (UMD3) on the autosomes or on the pseudo-autosomal region of the X-chromosome were excluded from the analysis. A total of 629 028 SNPs remained in the dataset after editing. High-density SNP genotypes and yield deviations (YD) for 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 (a) estimate variance components, including dominance variance and (b) perform cross-validation in order to evaluate the predictive ability of a model with dominance effects in comparison to a purely additive model. Both studies were done within a GBLUP framework. YD were calculated based on test-day observations adjusted for non-genetic effects, but not for permanent environmental effects, for each lactation and interpolated by the method of best prediction [13,14]. A weighted mean was calculated across lactation YD of a cow in order to obtain one multi-lactation YD per cow. The effective number of own performances (EOP) [15] was provided as a weight for the multi-lactation YD. For conformation traits, a permanent environmental effect was not modeled because repeated measurements are not available for cows.

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

MA: $\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{e}$

MG: $\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{e}$,

where **y** is a vector of multi-lactation 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 were $V(\mathbf{u}) = \mathbf{A}\sigma_A^2$ 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 based on the approach of VanRaden [16] using PREGSF90 [17]:

$$\mathbf{G}^* = \frac{\mathbf{W}_a \mathbf{W}_a{}'}{2\sum_{k=1}^{m} p_k q_k},$$

where matrix $\mathbf{W}_a$ has dimensions of the number of individuals (n) by the number of loci (m), with elements that are equal to $2\text{-}2p_k$ and $-2p_k$ for opposite homozygous and $1\text{-}2p_k$ for heterozygous genotypes, p_k is the minor allele frequency of locus k, and $q_k = 1\text{-}p_k$. Matrix $\mathbf{G}^*$ was scaled so that the means of diagonals and off-diagonals are the same as in **A** [18,19] and then 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 was $V(\mathbf{e}) = \mathbf{F}\sigma_E^2$ for both models, where **F** is a diagonal matrix with reciprocals of the EOP as weights. Extending model MG with dominance effects leads to model MGD:

MGD: $\mathbf{y} = \mu + \mathbf{Zu} + \mathbf{Zv} + \mathbf{e}$,

where **v** is a vector of dominance deviations of cows. $V(\mathbf{u})$ and $V(\mathbf{e})$ are defined as in model MG. The covariance matrix of dominance effects is $V(\mathbf{v}) = \mathbf{D}\sigma_D^2$, where **D** is the genomic dominance relationship matrix and σ_D^2 is the dominance variance. Matrix $\mathbf{D}^*$ was calculated as:

$$\mathbf{D}^* = \frac{\mathbf{W}_d \mathbf{W}_d{}'}{4\sum_{k=1}^{m} p_k^2 q_k^2},$$

where $\mathbf{W_d}$ has dimensions of the number of individuals (n) by the number of loci (m), with elements that are equal to $-2q_k^2$ for genotype A_1A_1, $2p_kq_k$ for genotype A_1A_2, and $-2p_k^2$ for genotype A_2A_2. Matrix $\mathbf{D}^*$ was then combined with the identity matrix $\mathbf{I}$ as $\mathbf{D} = 0.95\,\mathbf{D}^* + 0.05\,\mathbf{I}$ to improve numerical stability.

Estimation of variance components was performed with REMLF90 [20]. Goodness of fit of the respective models to the data was measured by the likelihood. The superiority of model MGD over model MG was tested by a likelihood ratio test, which was calculated as -2ln(likelihood for MG) + 2ln(likelihood for MGD). The likelihood ratio follows a mixture of χ^2-distributions with 0 and 1 degree of freedom [21]. In addition, variance components of model MGD were estimated by Gibbs sampling using the GIBBS1F90 software [20] in order to compare them with REML results and to calculate standard errors of the estimates. A total of 200 000 iterations of the sampler were run, with the first 20 000 iterations discarded as burn-in samples and every 50th sample included in the posterior analysis. Convergence to the final distribution was checked with the Geweke diagnostics [22] of the R package coda [23,24].

Additive and dominance variance components at the marker level (σ_a^2 and σ_d^2) were also estimated with the GS3 software [25] in a Markov chain Monte Carlo algorithm, using a model at the marker level (referred to as the MGD-SNP model hereinafter), in contrast to the previous animal level models:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Ta} + \mathbf{Xd} + \mathbf{e},$$

where **a** and **d** are vectors of 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 variance-covariance structure was $V(\mathbf{a}) = \mathbf{I}\sigma_a^2$ and $V(\mathbf{d}) = \mathbf{I}\sigma_d^2$. From the resulting estimates, additive and dominance variance components on the animal level were calculated as:

$\sigma_A^2 = \sum_{k=1}^{m}(2p_k q_k)\,\sigma_a^2$

$+ \sum_{k=1}^{m}[2p_k q_k (q_k - p_k)^2]\,\sigma_d^2$

and $\sigma_D^2 = \sum_{k=1}^{m}(4p_k^2 q_k^2)\,\sigma_d^2$ [12].

A total of 300 000 iterations of Gibbs sampling were performed for each trait. The first 20 000 iterations were discarded as burn-in samples and from the remaining 280 000 every 50th sample was considered for analysis of the posterior distribution.

Prediction of breeding values and total genetic values – cross-validation

Genotyped cows with YD for the respective traits were randomly divided in ten groups in order to perform cross-validation analysis. Typically, splitting at random implies that some validation animals have descendants in the training dataset, which means that the cross-validation is based on descendants, a case of no interest in reality and which will inflate accuracies [26]. In our dataset, genotyped cows were from a single generation. Therefore, a predicted cow could not have daughters (but, e.g., half- or full-sibs) in the training dataset – hence limiting upward bias in the estimation caused by progeny of validation animals in the training data. In this setting, the cross-validation accuracy measures the accuracy to predict contemporary cows including half- and full-sibs of training cows. Each group served once as validation group and the calibration group consisted of the other nine groups. Breeding values and total genetic values for the validation group were predicted based on models MA, MG, and MGD with their respective variance components estimated with REMLF90. The correlation between predicted breeding values and YD in the validation group [$r(YD, \hat{u})$] was calculated, as well as the regression of YD on predicted breeding values [$b(YD, \hat{u})$]. For model MGD, the correlation between predicted total genetic values and YD [$r(YD, \hat{g})$] and the regression of YD on predicted total genetic values [$b(YD, \hat{g})$] were also calculated. These measures were averaged over the ten validation groups.

Prediction of total genetic values of matings

Genotype probabilities and expectations of purely additive breeding values (u) and total genetic values (g), that include dominance deviations, were calculated for the offspring of all possible matings between 1996 cows and 50 bulls for milk yield and protein yield. The bulls were genotyped and selected for the respective trait on their conventional breeding value after progeny test (including the records of 1996 genotyped cows) from the German-Austrian genetic evaluation. SNP effects *a* and *d* were estimated in a BLUP model (BLUP-SNP; equal to model MGD-SNP but with variance components known) using GS3. Variance components σ_a^2 and σ_d^2 were fixed to values calculated from REMLF90 variance components σ_A^2 and σ_D^2 (model MGD):

$$\sigma_d^2 = \frac{\sigma_D^2}{\sum(2^2 p_k^2 q_k^2)}; \sigma_a^2 = \frac{\sigma_A^2 - \sum[2p_k q_k (q_k - p_k)^2]\sigma_d^2}{\sum(2p_k q_k)}.$$

The total genetic value g_{ij} of progeny from a mating between bull *i* and cow *j* was predicted as in Toro and Varona [9]:

$$\hat{g}_{ij} = \sum_k [Pr_{ijk}(AA)\hat{a}_k + Pr_{ijk}(Aa)\hat{d}_k - Pr_{ijk}(aa)\hat{a}_k],$$

where $Pr_{ijk}()$ is the probability of the corresponding genotype at locus *k*. Analogously, the breeding value u_{ij} of progeny from a mating between bull *i* and cow *j* was predicted as:

$$\hat{u}_{ij} = \sum_k [Pr_{ijk}(AA)(2 - 2p_k)\hat{\alpha}_k$$
$$+ Pr_{ijk}(Aa)(1 - 2p_k)\hat{\alpha}_k + Pr_{ijk}(aa)(-2p_k)\hat{\alpha}_k],$$

where $\hat{\alpha}_k = \hat{a}_k + \hat{d}_k(q_k - p_k)$.

Matings can be selected on $\hat{u}$ to maximize additive genetic gain or on $\hat{g}$ to maximize total genetic superiority. The latter maximizes the productive performance of the offspring, which

might be a farmer's interest. However, $\hat{g}$ can be maximized only for the next generation because gain in the dominance part of $\hat{g}$ cannot be accumulated in subsequent generations. In our example, additive gain is assured by pre-selection of bulls on their conventional breeding value. Selection on $\hat{u}$ leads to maximum additive gain, which can be accumulated in subsequent generations, and thus optimizes cumulative multi-generational genetic gain. A desirable objective might be to maximize $\hat{g}$ of matings and at the same time to keep the expected $\hat{u}$ of the offspring as high as possible.

In order to compare the results of these two possible selection strategies, $\hat{g}$ and $\hat{u}$ of all possible matings between the 1996 cows and 50 bulls were calculated for milk and protein yields. For each cow, the top mating was selected with respect to $\hat{g}$ or $\hat{u}$, with the restriction that a single bull was not mated to more than 200 cows. The expected additive genetic gains and total genetic superiorities with selection on $\hat{u}$ or $\hat{g}$ were calculated as the difference between the mean $\hat{u}$ or $\hat{g}$ of selected matings and the mean $\hat{u}$ or $\hat{g}$ of all possible matings.

Results

Estimation of variance components

Figure 1 shows the histograms of off-diagonal elements of the additive and dominance genomic relationship matrices. Means of off-diagonals of **G** (before scaling) and **D** were equal to 0, which implies that the population was in Hardy-Weinberg equilibrium. The standard deviation of off-diagonals of **G** was equal to 0.036, which is five times larger than the standard deviation of off-diagonals of **D**, i.e. 0.007. The proportion of off-diagonals that were smaller than -0.05 or larger than 0.05 was 6.27% for **G** but only 0.02% for **D**. Therefore, matrix **D** was less informative than **G**.

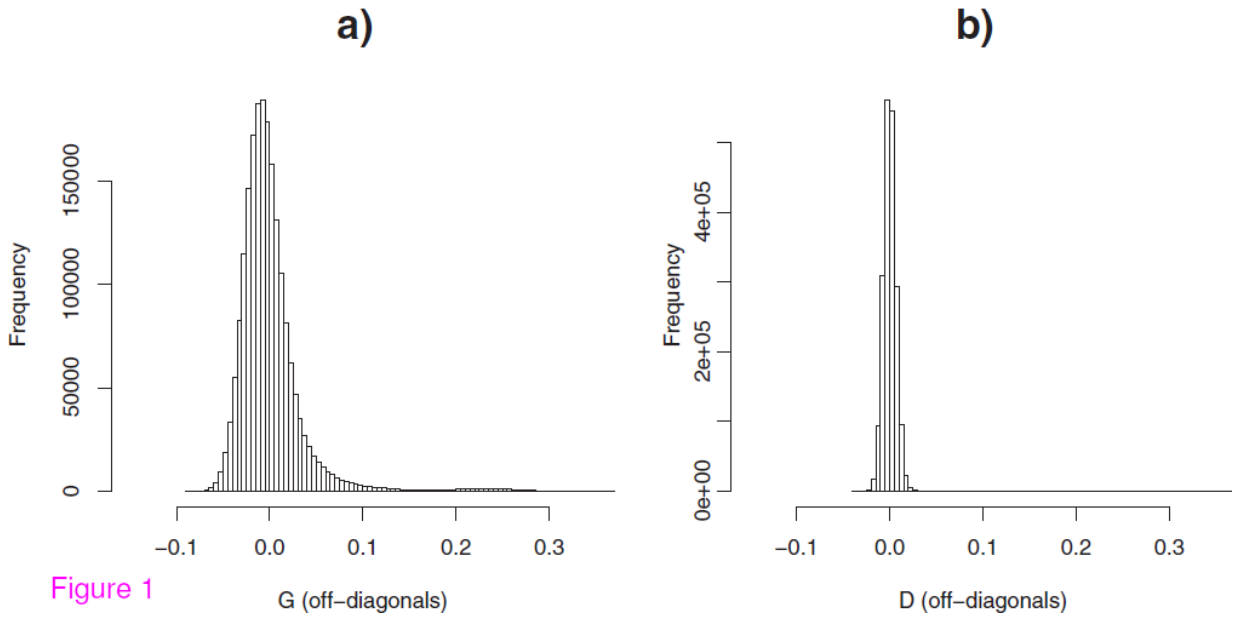

Figure 1 Histograms of off-diagonal elements of relationship matrices G (unscaled) (a) and D (b)

Table 1 Estimates of additive and dominance variance components obtained using REMLF90 for models MA, MG and MGD

Trait	MA	MG	MGD			
	σ_A^2	σ_A^2	σ_A^2	σ_D^2	σ_E^2	$\frac{\sigma_D^2}{\sigma_A^2 + \sigma_D^2}$
Milk yield	261500	214200	208900	92640	164700	0.308
Fat yield	279.4	274.4	267	104	198	0.281
Protein yield	213.4	175.4	166	115	154	0.409
Somatic cell score	0.2302	0.2640	0.256	0.261	0.555	0.505
Milkability	0.0193	0.0216	0.0122	0.0076	0.0029	0.390
Stature	3.412	5.728	5.80	0.20	6.51	0.033
Udder score	1.845	1.998	1.99	0.27	9.29	0.118
Udder depth	0.313	0.380	0.380	0.119	0.517	0.238
Feet and legs score	1.323	1.198	1.19	0.21	9.89	0.153

Estimated variance components for model MGD are in Table 1. Dominance variance (expressed as a percentage of total genetic variance) for milk production traits ranged from 28.1% for fat yield to 40.9% for protein yield. For somatic cell score and milkability, dominance variance was estimated at 39.0 and 50.5% of the genetic variance. Estimates of dominance variance for conformation traits were quite small, except for udder depth, ranging

from 3.3% for stature to 15.3% for feet and legs score. For udder depth, dominance variance was estimated at 23.8% of the genetic variance. For comparison, additive variances estimated with models MA and MG are also in Table 1. With the exception of milkability, the estimates of additive variance from model MG were consistent with additive variance estimates from the dominance model. Estimates of additive variance obtained with the pedigree model MA differed to some extent from those obtained with the genomic models. Estimates of variance components obtained using Gibbs sampling with model MGD and with an equivalent MGD-SNP model are in Table 2 and were similar to REML estimates with model MGD. Geweke statistics [22] showed convergence for model MGD but for the MGD-SNP model, the Gibbs chains did not converge even after 300 000 iterations. However, the means of the Gibbs chains for the MGD-SNP model were similar to those for the MGD model. For stature, udder score and feet and legs score, the estimated dominance variance was clearly larger with both Gibbs sampling analyses than with REML estimation because of a skewed posterior distribution of the Gibbs samples. Estimates of the ratio between dominance and total genetic variance had standard errors around 0.10, which is fairly good for such a small dataset.

For all traits, model MG, which exploited genomic information, fitted the data better than model MA, which included pedigree information only. The superiority of model MGD, which included a dominance effect, compared to model MG was significant for milk yield, fat yield, protein yield, somatic cell score and milkability, based on the likelihood ratio test. Likelihood measures and statistics of the likelihood ratio test between models MG and MGD are in Table 3. The likelihood ratio test statistics were asymptotically χ^2-distributed [27]. The χ^2-distribution function can take only non-negative values because it is defined as a sum of squared values. For two traits (stature and udder score), the likelihood ratio test statistics were negative (but very close to 0), which was due to numerical rounding or not finding the mode of the likelihood exactly.

Table 2 Estimates of additive and dominance variance components from Gibbs sampling for models MGD and MGD-SNP

Trait	MGD				MGD-SNP			
	σ_A^2	σ_D^2	σ_E^2	$\frac{\sigma_D^2}{\sigma_A^2+\sigma_D^2}$	σ_A^2	σ_D^2	σ_E^2	$\frac{\sigma_D^2}{\sigma_A^2+\sigma_D^2}$
Milk yield	211124 ±28668	98430 ±45503	161657 ±33376	0.306 ±0.108	202367 ±23929	115345 ±32156	152862 ±27651	0.358 ±0.067
Fat yield	270.2 ±36.9	112.5 ±53.3	193.5 ±36.9	0.283 ±0.105	261.9 ±31.0	119.1 ±34.4	192.4 ±29.6	0.308 ±0.064
Protein yield	168.2 ±26.3	117.8 ±43.7	152.7 ±28.9	0.401 ±0.105	166.2 ±23.0	105.6 ±30.5	160.9 ±25.0	0.383 ±0.072
Somatic cell score	0.268 ±0.067	0.261 ±0.121	0.554 ±0.096	0.471 ±0.155	0.220 ±0.070	0.130 ±0.068	0.680 ±0.088	0.352 ±0.101
Milkability	0.01228 ±0.00188	0.00735 ±0.00169	0.00315 ±0.00102	0.375 ±0.082	0.0116 ±0.00119	0.00724 ±0.00149	0.00397 ±0.00123	0.382 ±0.061
Stature	5.874 ±0.869	0.616 ±0.493	6.119 ±0.802	0.091 ±0.065	5.754 ±0.749	1.325 ±0.497	5.517 ±0.809	0.184 ±0.058
Udder score	2.016 ±0.521	1.089 ±0.787	8.527 ±0.921	0.322 ±0.161	2.010 ±0.498	1.134 ±0.452	8.466 ±0.819	0.352 ±0.070
Udder depth	0.3852 ±0.0608	0.1656 ±0.0952	0.4730 ±0.1024	0.285 ±0.120	0.387 ±0.055	0.181 ±0.059	0.457 ±0.082	0.312 ±0.067
Feet and legs score	1.212 ±0.451	0.947 ±0.676	9.209 ±0.831	0.407 ±0.192	1.320 ±0.382	0.936 ±0.370	9.118 ±0.700	0.408 ±0.070

The results are given as estimate ± standard error.

Table 3 Goodness of fit of models MA, MG, and MGD and likelihood ratio test (χ^2-value and P-value) between models MG and MGD

	-2 log likelihood			Likelihood ratio test	
	MA	MG	MGD	χ^2-value	P-value
Milk yield	31531.5	31488.1	31484.3	3.8	0.026
Fat yield	18363.1	18299.5	18295.9	3.6	0.029
Protein yield	17852.5	17824.6	17817.6	7.0	0.004
Somatic cell score	6072.9	6055.0	6050.6	4.4	0.018
Milkability	-1243.5	-1297.6	-1323.9	26.3	$1.46*10^{-7}$
Stature	9979.0	9907.9	9908.4	-0.5	1.000
Udder score	9916.5	9902.4	9902.5	-0.1	1.000
Udder depth	5287.9	5239.7	5238.5	1.2	0.137
Feet and legs score	9884.9	9880.9	9880.9	0.0	1.000

The measures of goodness of fit (-2 log likelihood) for models MA, MG and MGD are reported as well as the likelihood ratio test statistics (χ^2-value = $-2\ln\frac{likelihood\ for\ MG}{likelihood\ for\ MGD}$) between models MG and MGD and the corresponding P-values.

Prediction of breeding values and total genetic values – cross-validation

Mean accuracies of predicted breeding values [$r(YD, \hat{u})$] and slopes of the regression of YD on predicted breeding values [$b(YD, \hat{u})$] are in Table 4. For model MA, $r(YD, \hat{u})$ ranged from 0.102 for somatic cell score to 0.228 for fat yield, with an average of 0.165. Replacing pedigree with genomic relationships increased $r(YD, \hat{u})$ to between 0.108 (feet and legs) and 0.327 (milkability), with an average of 0.242. $r(YD, \hat{u})$ did not change when dominance effects were added to the model. Average standard errors of $r(YD, \hat{u})$ were equal to 0.024, 0.021 and 0.021 in models MA, MG and MGD, respectively. $r(YD, \hat{g})$ with the dominance model ranged from 0.109 for feet and legs score to 0.325 for fat yield. The difference between $r(YD, \hat{g})$ and $r(YD, \hat{u})$ in model MGD ranged from -0.004 for protein yield to 0.003 for udder score. The standard errors of $r(YD, \hat{g})$ were similar to those for $r(YD, \hat{u})$, with a mean of 0.021.

For models MA and MG, $b(YD, \hat{u})$ ranged from 0.563 (milkability) to 1.201 (feet and legs score) and from 0.744 (milkability) to 1.068 (fat yield), respectively, with means of 0.964 and 0.971. $b(YD, \hat{u})$ for model MGD ranged from 0.924 (protein yield) to 1.085 (fat yield), with a mean of 1.016. The standard errors of $b(YD, \hat{u})$ were rather large, with means of 0.151, 0.106 and 0.111 for models MA, MG and MGD, respectively. The slope of the regression of YD on predicted total genetic values ranged from 0.889 (protein yield) to 1.060 (feet and legs score), with a mean of 0.995 and was slightly smaller than $b(YD, \hat{u})$ for most traits for the same model. The fact that slopes were generally not significantly different from 1 suggests that predictions were essentially unbiased, except for milkability.

Prediction of total genetic values of matings

For milk yield, 16 bulls were chosen as mating partners when matings were selected on $\hat{g}$. The restriction of at most 200 cows per bull was reached for seven bulls. The remaining nine bulls were mated to 197, 147, 139, 86, 19, 4, 2, 1 and 1 cows. When matings were selected on $\hat{u}$,

nine bulls were mated to the maximum number of 200 cows and two other bulls to 176 and 20 cows, respectively. For protein yield, 24 bulls were chosen as mating partners when matings were selected on $\hat{g}$. The restriction of 200 cows per bull was reached for seven bulls. The remaining 17 bulls were mated to 134, 115, 114, 63, 62, 29, 26, 21, 8, 7, 4, 3, 3, 3, 2, 1 and 1 cows. When matings were selected on $\hat{u}$, eight bulls were mated to the maximum number of 200 cows and the four other bulls to 190, 157, 33 and 16 cows.

Table 4 Accuracies and regression coefficients of predicted breeding values and total genetic values for models MA, MG, and MGD

Trait	$r(YD,\hat{u})$[1]			$r_2(YD,\hat{g})$	$b(YD,\hat{u})$[3]			$b(YD,\hat{g})$[4]
	MA	MG	MGD	MGD	MA	MG	MGD	MGD
Milk yield	0.221 ± 0.029	0.277 ± 0.030	0.278 ± 0.031	0.275 ± 0.032	0.925 ± 0.109	0.955 ± 0.099	0.967 ± 0.101	0.950 ± 0.104
Fat yield	0.228 ± 0.018	0.325 ± 0.020	0.325 ± 0.019	0.325 ± 0.019	1.031 ± 0.085	1.068 ± 0.078	1.085 ± 0.079	1.072 ± 0.075
Protein yield	0.202 ± 0.031	0.236 ± 0.016	0.238 ± 0.016	0.234 ± 0.016	0.958 ± 0.148	0.889 ± 0.070	0.924 ± 0.072	0.889 ± 0.069
Somatic cell score	0.102 ± 0.018	0.169 ± 0.020	0.169 ± 0.019	0.168 ± 0.015	0.866 ± 0.165	1.007 ± 0.131	1.031 ± 0.133	0.973 ± 0.107
Milkability	0.133 ± 0.042	0.327 ± 0.025	0.324 ± 0.028	0.322 ± 0.027	0.563 ± 0.182	0.744 ± 0.057	1.053 ± 0.099	1.004 ± 0.087
Stature	0.180 ± 0.017	0.308 ± 0.014	0.308 ± 0.014	0.308 ± 0.014	1.082 ± 0.117	1.030 ± 0.059	1.023 ± 0.059	1.021 ± 0.059
Udder score	0.121 ± 0.017	0.159 ± 0.022	0.159 ± 0.022	0.158 ± 0.022	1.023 ± 0.144	1.004 ± 0.142	1.007 ± 0.142	1.002 ± 0.146
Udder depth	0.192 ± 0.017	0.269 ± 0.020	0.269 ± 0.020	0.272 ± 0.020	1.031 ± 0.119	0.988 ± 0.095	0.991 ± 0.094	0.988 ± 0.095
Feet and legs score	0.106 ± 0.026	0.108 ± 0.023	0.108 ± 0.023	0.109 ± 0.024	1.201 ± 0.290	1.055 ± 0.221	1.063 ± 0.223	1.060 ± 0.225

The results are given as mean ± standard error.
[1] $r(YD,\hat{u})$ = accuracy of predicted breeding values (cross-validation correlation between YD and predicted breeding values).
[2] $r(YD,\hat{g})$ = accuracy of predicted total genetic values (cross-validation correlation between YD and predicted total genetic values).
[3] $b(YD,\hat{u})$ = regression coefficient of YD on predicted breeding values.
[4] $b(YD,\hat{g})$ = regression coefficient of YD on predicted total genetic values.

Expected total genetic superiorities and additive genetic gains obtained with the selected matings are in Table 5, both in absolute numbers and relative to the standard deviations (SD) of

$\hat{u}$ and $\hat{g}$. When matings were selected on $\hat{g}$ for milk yield, the expected total genetic superiority was estimated to be equal to 165.2 kg, which is equivalent to 1.01 SD of $\hat{g}$. The expected total genetic superiority was reduced to 143.8 kg (0.88 SD) when matings were selected on $\hat{u}$. The expected additive genetic gain was less sensitive to the selection criterion applied since it was only slightly reduced when selection was done on $\hat{g}$ (137.7 kg; 0.85 SD) instead of on $\hat{u}$ (143.8 kg; 0.89 SD). The results were similar for protein yield. With selection on $\hat{g}$, the expected additive genetic gain was slightly smaller (0.74 vs. 0.76 SD) but the expected total genetic superiority was clearly larger (1.01 vs. 0.79 SD) compared to selection on $\hat{u}$.

Table 5 Expected total genetic superiority (ΔG) and additive genetic gain (ΔU) with selection on total genetic value ($\hat{g}$) or breeding value ($\hat{u}$)

	ΔG		**ΔU**	
	absolute (kg)	**relative to SD**	**absolute (kg)**	**relative to SD**
Milk yield				
Selection on $\hat{g}$	165.2	1.01	137.7	0.85
Selection on $\hat{u}$	143.8	0.88	143.8	0.89
Protein yield				
Selection on $\hat{g}$	4.15	1.01	3.09	0.74
Selection on $\hat{u}$	3.24	0.79	3.16	0.76

Expected total genetic superiority (ΔG) and expected additive genetic gain (ΔU) for the alternative selection criteria total genetic value ($\hat{g}$) and breeding value ($\hat{u}$) in absolute value (kg) and relative to the standard deviations (SD) of $\hat{g}$ and $\hat{u}$ of all possible matings; the maximum number of matings per bull was restricted to 200.

Discussion

This study analyzed the importance of dominance variation for several milk production and conformation traits in the Fleckvieh breed using the GBLUP methodology. Additive and dominance genomic relationship matrices were calculated similar to Su et al. [11], except that standard quantitative genetic approaches were used, with the dominance variance at locus *k*

defined as $[2p_k\ q_k\ d]^2$ [1,12]. This resulted in the reported estimates of dominance variance to be compatible with pedigree-based estimates.

Independence between **u** and **v** is the classical treatment [1] and it is convenient because it allows orthogonality of the estimates and thus an easy translation into variances and covariances of **u** and **v**. However, this independence is contradictory with the phenomena of inbreeding depression and hybrid vigor; presence of inbreeding depression indicates that dominance is directional, e.g. [28]. Wellmann and Bennewitz [10,29] reviewed biological information on milk yield and productive life in Holstein cattle to suggest *a priori* dependencies between *a* and *d* (which would result in dependencies between **u** and **v**) and Bayesian regression models that could accommodate those dependencies. The treatment of dependencies between breeding values and dominance deviations is rather complex and the computational requirements are large, thus, we did not consider this method although it should be a field of further research.

Estimates of dominance variance varied from 3.3 to 50.5% of total genetic variance for the analyzed traits. Estimated dominance variance (as a percentage of total genetic variance) was greater for milk production traits than for conformation traits. These results agree with those of Misztal et al. [7], who found larger dominance variance for production than for conformation traits. Moreover, Misztal et al. [3] reported estimates of dominance variance in US Holstein cattle for 14 conformation traits that ranged from 7.3 (rump angle) to 22.3% (strength) of the total genetic variance. This is comparable to the estimates of dominance variance for the conformation traits analyzed in this study. In the literature, reported estimates of dominance variance for milk production traits of Holstein cattle vary considerably ranging from 1.4 to 42.9% of the total genetic variance [4-7], which are within the same range but smaller than those found in our study. Two reasons may explain the relatively large estimates of dominance

variance for milk production traits obtained in our study compared to values reported in the literature: (1) Fleckvieh cattle are genetically more diverse than Holstein cattle, as reflected by the considerably larger effective population size of the Fleckvieh breed [30], which is expected to result in more heterozygosity and in QTL alleles with more intermediate frequencies; (2) all estimates of dominance variance available in the literature were obtained using relationship matrices based on pedigree data; the use of genomic information is expected to improve estimates of dominance effect relationships and reduce potential confounding with additive effects and residuals which is likely to result in different estimates.

Although moderate changes in estimates of additive variance were observed between pedigree and genomic models, estimates of additive variance were consistent for genomic additive and dominance models, except for milkability. Su et al. [11] reported a small difference in estimates of additive variance between additive and dominance models. However, the additive and dominance variances reported in Su et al. [11] result from an alternative partitioning of genetic variance and are thus not directly comparable to the classical partitioning of genetic variance [12]. In studies based on pedigree information, estimates of additive variance have been similar between additive and dominance models [5,6,31].

Both Gibbs sampling with model MGD and at the marker level with the MGD-SNP model resulted in estimated variance components that were comparable with REML estimates for most traits. The relative standard error (calculated as standard error divided by the estimate) of dominance variance was on average 2.7 times larger than the relative standard error of the estimated additive variance, which is expected based on the properties of **G** and **D**. However, in other studies the ratio between relative standard errors of dominance and additive variances was even larger, i.e. 4.1 in Misztal [32] and 4.5 in Su et al. [11]. In order to estimate dominance variance more accurately, more dominance specific information is needed. This could be

achieved, e.g., by increasing the number of full-sibs in the dataset. The present dataset contained 3% full-sibs.

Despite the large estimates of dominance variance for most analyzed traits (significantly larger than 0 for five traits), prediction accuracy of breeding values and total genetic values did not change when dominance effects were included in the model. Estimates of additive variance did not differ much between models MG and MGD, which means that additive variance is already captured quite accurately in the additive model. Thus, additive effects are relatively well predicted, whether the dominance effect is modeled or not. The accuracy of predictions of total genetic values in cross-validation was not higher with the dominance than with the additive model because the proportion of full-sibs and dominance effect relationship coefficients between the training and validation datasets were small. Thus, little information was transferred from the reference to the validation group in cross-validation for prediction of dominance effects. Su et al. [11], who analyzed non-additive effects for average daily gain with a dataset of 1911 purebred pigs, observed that the estimates of the additive variance with the additive and dominance models remained fairly constant and that gains in accuracies of predicted breeding values and predicted total genetic values reached only 0.004 and 0.011 with the dominance model. The proportion of full-sibs in the pig dataset was not reported in Su et al. [11] but is expected to be substantially larger than in our cow dataset, which might be the reason for the gain in accuracy of predicted total genetic values with inclusion of dominance in the model. Based on a simulation study, Varona et al. [33] observed that relevant changes in breeding values when switching from an additive to a dominance model were obtained only for animals that had full-sibs or full-sib progeny and little other information. A cow dataset with a larger proportion of full-sibs would contain more information in order to accurately estimate dominance effects but in practice such data is not available. Analysis of full-sib progeny from

elite animals, which generally are available, would not be representative for the whole population.

The regression coefficient of YD on predicted breeding values was generally close to 1, with a few exceptions. With the dominance model, this regression coefficient was slightly closer to 1 for most traits but differences were small, which is similar to the data reported by Su et al. [11], i.e. 0.927 and 0.983 with the additive and dominance models, respectively. In our study, the regression coefficient of YD on predicted total genetic values for model MGD was slightly smaller than the regression on predicted breeding values, which agrees with Su et al. [11], but it remained close to the expectation, which means that predictions were unbiased. In general, bias can originate from preferential treatment, unrecognized pre-selection of validation animals, or inappropriate modeling of predictions (i.e. using incorrect variance components).

The results show that selection of matings on $\hat{g}$ instead of $\hat{u}$ led to 14.8% (milk yield) and 27.8% (protein yield) greater expected total genetic superiorities and maximized expected productive performance of the offspring. Although the accuracy of estimates of total genetic values was not greater than that of estimates of breeding values, as indicated by the cross-validation results (Table 4), expected total genetic superiority was not impaired by this result because predicted genetic values are best linear unbiased predictions and therefore unbiased expectations [34]. Toro and Varona [9] reported that expected total genetic superiority with optimized mate allocation was 16% greater than with selection on the breeding value only, for a trait with additive and dominance variances amounting to 40 and 10% of the phenotypic variance. Expected additive genetic gain was reduced by only 4.5% for milk yield and by 2.6% for protein yield with selection of matings on $\hat{g}$ instead of $\hat{u}$. Thus, optimization of $\hat{g}$ of the offspring appears to be feasible without a great loss in $\hat{u}$. Our considerations of optimized matings are limited to the first generation offspring. Toro and Varona [9] found that response

from assortative mating was only realized in the first generation without any additional response in subsequent generations. Thus, optimization of matings with respect to total genetic value has to be applied in each generation, otherwise the dominance-specific advantage is lost. Toro and Varona [9] pre-selected males and females on their estimated breeding values and then optimized the total genetic value of matings between these pre-selected animals. In our example, only bulls were pre-selected on their conventional breeding value and the optimal bull was determined for each cow based on the expected total genetic value of the offspring. However, the potential of assortative mating to exploit dominance variance optimally by combining mates that are expected to produce offspring with large total genetic values is limited even for these two traits with sizeable dominance variation. This can be caused either by cancellation effects across the genome (i.e., it is extremely unlikely to combine all positive dominance effects) or by a reduced accuracy of the dominance deviation of a mating because of uncertainty about the resulting marker genotypes.

Conclusions

Estimates of genomic variance due to dominance in Fleckvieh cattle ranged from 3 to 50% of the genetic variance and were within the range of published pedigree-based estimates for dairy cattle. The computational complexity and modeling were straightforward. Predictive ability of breeding and total genetic values by cross-validation was not improved when dominance effects were included in the prediction model, probably because of the limited size of the dataset and the small proportion of full-sibs. There is potential to exploit dominance variance in planned matings in order to increase total genetic value of the offspring (i.e. future performance) without compromising additive genetic gain. Use of planned matings could also be a way to motivate farmers that are otherwise not interested in using genomic breeding values for breeding schemes.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

JE performed the analysis and drafted the manuscript. JE, AL and KUG designed the study. AL, ZGV and LV developed methods. CE and RE prepared phenotypic and genotypic data. AL, ZGV, LV, CE, RE and KUG revised the manuscript. All authors read and approved the final manuscript.

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). A Legarra and ZG Vitezica acknowledge funding from SelGen metaprogram action X-Gen. We are grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrénées for providing computing and storage resources. We thank H Anzenberger, W Heinrichs, A Krämer, S Neuner, R Schnagl, L Schweiger, S Schweiger, H Strasser and H Trager for sampling nasal swabs of cows for genotyping and Prof. R Fries and his team at the Chair of Animal Breeding of Technische Universität München for genotyping the DNA samples. We want to thank two anonymous reviewers, the associate editor, H Hayes and J Dekkers for their valuable comments.

References

1. Falconer DS, Mackay TFC: *Introduction to Quantitative Genetics.* Fourth edition. Harlow: Pearson Education Limited; 1996.

2. Tempelman RJ, Burnside EB: **Additive and nonadditive genetic variation for conformation traits in Canadian Holsteins.** *J Dairy Sci* 1990, **73**:2214-2220.

3. Misztal I, Lawlor TJ, Gengler N: **Relationships among estimates of inbreeding depression, dominance and additive variance for linear traits in Holsteins.** *Genet Sel Evol* 1997, **29**:319-326.

4. Tempelman RJ, Burnside EB: **Additive and nonadditive genetic variation for production traits in Canadian Holsteins.** *J Dairy Sci* 1990, **73**:2206-2213.

5. Miglior F, Burnside EB, Kennedy BW: **Production traits of Holstein cattle: Estimation of nonadditive genetic variance components and inbreeding depression.** *J Dairy Sci* 1995, **78**:1174-1180.

6. Van Tassell CP, Misztal I, Varona L: **Method R estimates of additive genetic, dominance genetic, and permanent environmental fraction of variance for yield and health traits of Holsteins.** *J Dairy Sci* 2000, **83**:1873-1877.

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

8. Varona L, Misztal I: **Prediction of parental dominance combinations for planned matings, methodology, and simulation results.** *J Dairy Sci* 1999, **82**:2186-2191.

9. Toro MA, Varona L: **A note on mate allocation for dominance handling in genomic selection.** *Genet Sel Evol* 2010, **42**:33.

10. Wellmann R, Bennewitz J: **Bayesian models with dominance effects for genomic evaluation of quantitative traits.** *Genet Res (Camb)* 2012, **94**:21-37.

11. Su G, Christensen OF, Ostersen T, Henryon M, Lund MS: **Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers.** *PLoS ONE* 2012, **7**:e45293.

12. Vitezica ZG, Varona L, Legarra A: **On the additive and dominant variance and covariance of individuals within the genomic selection scope.** *Genetics* 2013, **195**:1223-1230.

13. VanRaden PM: **Lactation yields and accuracies computed from test day yields and (co)variances by best prediction.** *J Dairy Sci* 1997, **80**:3015-3022.

14. Liu Z, Reinhardt F, Reents R: **The effective daughter contribution concept applied to multiple trait models for approximating reliability of estimated breeding values.** *Interbull Bull* 2001, **27**:41-47.

15. Edel C, Emmerling R, Götz KU: **Optimized aggregation of phenotypes for MA-BLUP evaluation in German Fleckvieh.** *Interbull Bull* 2008, **40**:178-183.

16. VanRaden PM: **Efficient methods to compute genomic predictions.** *J Dairy Sci* 2008, **91**:4414-4423.

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

18. Vitezica ZG, Aguilar I, Misztal I, Legarra A: **Bias in genomic predictions for populations under selection.** *Genet Res (Camb)* 2011, **93**:357-366.

19. Christensen OF: **Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation.** *Genet Sel Evol* 2012, **44**:37.

20. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH: **BLUPF90 and related programs (BGF90).** In *Proceedings of the 7th World Congress Applied to Livestock Production: 19-23 August 2002; Montpellier*; 2002:28-07.

21. Visscher PM: **A note on the asymptotic distribution of likelihood ratio tests to test variance components**. *Twin Res Hum Genet* 2006, **9**:490-495.

22. Geweke J: **Evaluating the accuracy of sampling-based approaches to calculating posterior moments.** In *Bayesian Statistics 4*. Edited by Bernado JM, Berger JO, Dawid AP, Smith AFM. Oxford: Clarendon Press; 1992.

23. R Development Core Team: **R: A language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing; 2011.

24. Plummer M, Best N, Cowles K, Vines K: **CODA: Convergence diagnosis and output analysis for MCMC.** *R News* 2006, **6**:7-11.

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

26. Legarra A, Robert-Granié C, Manfredi E, Elsen JM: **Performance of genomic selection in mice.** *Genetics* 2008, **180**:611-618.

27. Wilks SS: **The large-sample distribution of the likelihood ratio for testing composite hypotheses.** *Ann Math Stat* 1938, **9**:60-62.

28. Lynch M, Walsh JB: *Genetics and Analysis of Quantitative Traits.* Sunderland: Sinauer Associates, Inc.; 1998.

29. Wellmann R, Bennewitz J: **The contribution of dominance to the understanding of quantitative genetic variation.** *Genet Res (Camb)* 2011, **93**:139-154.

30. Pausch H, Aigner B, Emmerling R, Edel C, Götz KU, Fries R: **Imputation of high-density genotypes in the Fleckvieh cattle population.** *Genet Sel Evol* 2013, **45**:3.

31. Van Tassell CP, Wiggans GR, Norman HD: **Method R estimates of heritability for milk, fat, and protein yields of United States dairy cattle.** *J Dairy Sci* 1999, **82**:2231-2237.

32. Misztal I: **Estimation of variance components with large-scale dominance models.** *J Dairy Sci* 1997, **80**:965-974.

33. Varona L, Misztal I, Bertrand JK, Lawlor TJ: **Effect of full sibs on additive breeding values under the dominance model for stature in United States Holsteins.** *J Dairy Sci* 1998, **81**:1126-1135.

34. Henderson CR: **Sire evaluation and genetic trends.** In *Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush.* Edited by American Society of Animal Science, American Dairy Science Association and Poultry Science Association. Champaign; 1973:10-41.

Summary and Authors' Contributions

Dominance variance of nine milk production and conformation traits was estimated with additive and dominance models using yield deviations and SNP genotypes of 1996 Fleckvieh cows and ranged from 3.3% to 50.5% of the total genetic variance. Yield deviations of high-density genotyped Fleckvieh cows were used to assess cross-validation accuracy of genomic predictions with additive and dominance models. Cross-validation accuracy of predicted breeding values was higher with genomic models than with the pedigree model, but inclusion of dominance effects did not increase the accuracy of the predicted breeding and total genetic values. Due to small dominance effect relationships, predictions of individual dominance deviations were inaccurate and including dominance in the model did not improve prediction accuracy in cross-validation. Expected breeding values and total genetic values for putative offspring were calculated by means of SNP effects. Selection mating partners based on total genetic value instead of breeding value would result in a larger expected total genetic superiority in progeny, i.e. 14.8% for milk yield and 27.8% for protein yield and reduce the expected additive genetic gain only by 4.5% for milk yield and 2.6% for protein yield.

J. Ertl performed the analysis and drafted the manuscript. J. Ertl, A. Legarra and K.-U. Götz designed the study. A. Legarra, Z. G. Vitezica and L. Varona developed methods. C. Edel and R. Emmerling prepared phenotypic and genotypic data. A. Legarra, Z. G. Vitezica, L. Varona, C. Edel, R. Emmerling and K.-U. Götz revised the manuscript.

4th Chapter

Considering dominance in reduced single-step genomic evaluations

Johann Ertl, Christian Edel, Eduardo Pimentel, Reiner Emmerling & Kay-Uwe Götz

Bavarian State Research Centre for Agriculture, Institute of Animal Breeding, Poing-Grub, Germany

published in *Journal of Animal Breeding and Genetics* 135:151-158

Summary

Single-step models including dominance can be an enormous computational task and can even be prohibitive for practical application. In this study, we try to answer the question whether a reduced single-step model is able to estimate breeding values of bulls and breeding values, dominance deviations and total genetic values of cows with acceptable quality. Genetic values and phenotypes were simulated (500 repetitions) for a small Fleckvieh pedigree consisting of 371 bulls (180 thereof genotyped) and 553 cows (40 thereof genotyped). This pedigree was virtually extended for 2,407 non-genotyped daughters. Genetic values were estimated with the single-step model and with different reduced single-step models. Including more relatives of genotyped cows in the reduced single-step model resulted in a better agreement of results with the single-step model. Accuracies of genetic values were largest with single-step and smallest with reduced single-step when only the cows genotyped were modelled. The results indicate that a reduced single-step model is suitable to estimate breeding values of bulls and breeding values, dominance deviations and total genetic values of cows with acceptable quality.

1 | Introduction

The single-step model has been developed to estimate additive-genetic breeding values using pedigree and genomic information jointly (Aguilar et al., 2010; Christensen & Lund, 2010; Legarra, Aguilar, & Misztal, 2009; Misztal, Legarra, & Aguilar, 2009). As methods to calculate the dominance relationship matrix from SNP genotypes are now available (Vitezica, Varona, & Legarra, 2013), it is straightforward to extend the single-step model by a dominance part. Several studies have found non-negligible amounts of dominance variance in different traits of dairy cattle (Ertl et al., 2014; Miglior, Burnside, & Kennedy, 1995; Tempelman &

Burnside, 1990; Van Tassell, Misztal, & Varona, 2000). As more and more cows are genotyped, it might be interesting to apply dominance models for an optimal genetic evaluation of cows. However, inversion of the dominance relationship matrix is computationally not feasible for large systems and the direct calculation of the inverted dominance relationship matrix via sire-dam subclass effects (Hoeschele & VanRaden, 1991) can result in a very large matrix. A possible solution for computational and convergence problems with the single-step method might be to reduce the size of the model using daughter yield deviations (DYD; VanRaden & Wiggans, 1991) for bulls and yield deviations (YD; VanRaden & Wiggans, 1991) for a subset of (genotyped) cows as pseudo-phenotypes in a reduced single-step model. Several groups of authors (Gao et al., 2012; Harris, Winkelman, & Johnson, 2013; Su et al., 2012) have tested reduced (additive) single-step models, which have reduced dimension as compared to single-step models, with de-regressed proofs as pseudo-phenotypes. Unlike de-regressed proofs, YD contain dominance deviations by definition (VanRaden & Wiggans, 1991) and can therefore be used as pseudo-phenotypes of cows in a dominance model. In this study, we compared results from a full single-step evaluation containing additive and dominance effects with several reduced models combining the DYD of bulls and YD of selected cows.

2 | Material and Methods

Suppose a sex-limited trait like milk yield of dairy cattle where only cows can have observations. Following quantitative genetics theory and supposing that there is additive-genetic and dominance variation in the trait, the phenotype of a cow (y_i) can be partitioned into additive-genetic, dominance and residual components:

$$y_i = \mu_i + u_i + v_i + e_i,$$

where μ_i is the sum of fixed effects for cow i, u_i is its breeding value, v_i is its dominance deviation and e_i is the residual. Phenotypic variance (σ_Y^2) consists of additive-genetic (σ_A^2), dominance (σ_D^2) and residual variance (σ_E^2). Breeding values of animals i and j are correlated by their (additive-genetic) relationship coefficient s_{ij}. Dominance deviations of animals i and j are correlated by their dominance relationship coefficient t_{ij}. Thus, distributions of breeding values, dominance deviations and residuals follow covariance structures $\mathbf{S}\sigma_A^2$, $\mathbf{T}\sigma_D^2$ and $\mathbf{I}\sigma_E^2$, respectively, where **S** and **T** are true additive-genetic and dominance relationship matrices. True relationship matrices are never known and estimated by means of either additive-genetic (**A**) and dominance ($\mathbf{A_D}$) numerator relationship matrices (using pedigree information) or additive-genetic (**G**) and dominance ($\mathbf{G_D}$) genomic relationship matrices (typically using SNP genotypic information; e.g., VanRaden, 2008; Vitezica et al., 2013). For the case that not all animals are genotyped, pedigree and genomic information can be combined into single-step relationship matrices (Aguilar et al., 2010; Christensen & Lund, 2010; Legarra et al., 2009):

$$\mathbf{H} = \begin{bmatrix} \mathbf{H_{11}} & \mathbf{H_{12}} \\ \mathbf{H_{21}} & \mathbf{H_{22}} \end{bmatrix} = \begin{bmatrix} \mathbf{A_{11}} - \mathbf{A_{12}}\mathbf{A_{22}^{-1}}\mathbf{A_{21}} + \mathbf{A_{12}}\mathbf{A_{22}^{-1}}\mathbf{G}\mathbf{A_{22}^{-1}}\mathbf{A_{21}} & \mathbf{A_{12}}\mathbf{A_{22}^{-1}}\mathbf{G} \\ \mathbf{G}\mathbf{A_{22}^{-1}}\mathbf{A_{21}} & \mathbf{G} \end{bmatrix}$$

$$\mathbf{H_D} = \begin{bmatrix} \mathbf{H_{D11}} & \mathbf{H_{D12}} \\ \mathbf{H_{D21}} & \mathbf{H_{D22}} \end{bmatrix} = \begin{bmatrix} \mathbf{A_{D11}} - \mathbf{A_{D12}}\mathbf{A_{D22}^{-1}}\mathbf{A_{D21}} + \mathbf{A_{D12}}\mathbf{A_{D22}^{-1}}\mathbf{G_D}\mathbf{A_{D22}^{-1}}\mathbf{A_{D21}} & \mathbf{A_{D12}}\mathbf{A_{D22}^{-1}}\mathbf{G_D} \\ \mathbf{G_D}\mathbf{A_{D22}^{-1}}\mathbf{A_{D21}} & \mathbf{G_D} \end{bmatrix}$$

Subscripts 1 and 2 indicate not genotyped and genotyped animals, respectively. These equations assume no covariance between breeding values and dominance deviations, that is no inbreeding.

For this study, we simulated repeatedly breeding values, dominance deviations, total genetic values and phenotypes for a semi-real pedigree, estimated the genetic values both with a single-step model and with reduced single-step models and compared the estimates between models and with the true values. We used a small real-life pedigree structure from the Fleckvieh pedigree consisting of 371 bulls and 553 cows from the German-Austrian routine

evaluation. We extended this core pedigree with 2,407 female descendants to construct the final example pedigree. For a subset of 180 bulls and 40 cows of this core pedigree the genotypes (Illumina 50K SNP chip) were available. A subset of 10 of the 40 genotyped cows was defined to be validation cows for testing the predictive ability of the model. Each validation cow had a full-sister among the remaining 30 genotyped cows. There was one pair of full-sisters among the remaining 30 genotyped cows. Table 1 summarizes some characteristics of the example pedigree.

Table 1. Summary of the animals in the pedigree

	Genotyped	**Not genotyped**	**Total**
Core pedigree			
Bulls	180	191	371
Cows with records	30	513	543
Cows without records	10	0	10
Additional daughters (comprising DYD of bulls in reduced models)	0	2,407	2,407
Total	220	3,111	3,331

In each of 500 repetitions, simulated true breeding values (**u**), dominance deviations (**v**), total genetic values (**g**; **g** = **u** + **v**) and residuals (**e**) for the animals in the pedigree were generated by drawing a single sample from a corresponding multivariate normal distribution assuming $\mathbf{u} \sim N(0; \mathbf{H}\sigma_A^2)$, $\mathbf{v} \sim N(0; \mathbf{H_D}\sigma_D^2)$ and $\mathbf{e} \sim N(0; \mathbf{I}\sigma_E^2)$ with σ_A^2 = 0.4 and σ_D^2 = 0.2 and σ_E^2 = 0.4. **H** and $\mathbf{H_D}$ are additive and dominance single-step relationship matrices built for the pedigree structure and genotypes described above and were treated as true relationship matrices for simulation. For the purpose of comparison of single-step and reduced single-step models, this approximation should be valid even if the simulation of genetic values was based only on

estimated polygenic relationships but not on linkage disequilibrium information. Phenotypes of cows (**y**) were calculated by summing up breeding values, dominance deviations and residuals:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z_A u} + \mathbf{Z_D v} + \mathbf{e},$$

where μ is an overall mean and **1** is a vector of 1s,. $\mathbf{Z_A}$ and $\mathbf{Z_D}$ are design matrices connecting observations with the corresponding animals.

For means of comparison, breeding values and dominance deviations were estimated both with the single-step model (**1S**) as described by Legarra et al. (2009) and Aguilar et al. (2010) and with a reduced single-step model (**R1S**) following Harris et al. (2013). In 1S, genetic evaluation is based on records of cows directly:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z_A}\begin{bmatrix}\mathbf{u_1}\\ \mathbf{u_2}\end{bmatrix} + \mathbf{Z_D}\begin{bmatrix}\mathbf{v_1}\\ \mathbf{v_2}\end{bmatrix} + \mathbf{e}$$

G was calculated by the method of VanRaden (2008) using allele frequencies estimated from 11,344 Fleckvieh genotypes of the April 2014 German-Austrian routine evaluation. The pedigree-based dominance relationship matrix $\mathbf{A_D}$ was calculated as described in Mrode (2005). The genomic dominance relationship matrix $\mathbf{G_D}$ was computed by the method described by Vitezica et al. (2013). **G** and $\mathbf{G_D}$ were scaled to **A** and $\mathbf{A_D}$ using the method of Vitezica, Aguilar, Misztal, and Legarra (2011). The variance-covariance matrix of observations is given by:

$$\mathbf{V} = \mathbf{Z_A}(\mathbf{H}\sigma_A^2)\mathbf{Z_A'} + \mathbf{Z_D}(\mathbf{H_D}\sigma_D^2)\mathbf{Z_D'} + \mathbf{I}\sigma_E^2$$

The estimator of the vector of fixed effects is given by:

$$\hat{\mu} = (\mathbf{1'V^{-1}1})^{-1}\mathbf{1'V^{-1}y}$$

Breeding values and dominance deviations are estimated with the following equations:

$$\hat{\mathbf{u}} = \mathbf{H}\sigma_A^2\mathbf{Z_A'V^{-1}}(\mathbf{y} - \mathbf{1}\hat{\mu}); \hat{\mathbf{v}} = \mathbf{H_D}\sigma_D^2\mathbf{Z_D'V^{-1}}(\mathbf{y} - \mathbf{1}\hat{\mu})$$

In R1S, doubled daughter yield deviations (2DYD) of 371 bulls and yield deviations (YD) of different subsets of cows, that is

i 30 genotyped cows (with records),

ii 30 genotyped cows plus their 29 dams,

iii 30 genotyped cows plus their 36 daughters and

iv 30 genotyped cows plus their 29 dams and 36 daughters,

were used as pseudo-phenotypes instead of records. These subsets of animals included in R1S will be referred to as sub models 1-4 in the results.

A conventional additive-genetic evaluation was run first in order to obtain YD and 2DYD. YD of genotyped cows were omitted in the calculation of 2DYD of their sire in order to avoid double counting. 2DYD and YD are used jointly in R1S:

$$\begin{bmatrix}\mathbf{2DYD}\\ \mathbf{YD}\end{bmatrix} = \mathbf{1}\mu^* + \mathbf{W_A}\begin{bmatrix}\mathbf{u_1^*}\\ \mathbf{u_2^*}\end{bmatrix} + \mathbf{W_D}\begin{bmatrix}\mathbf{v_1^*}\\ \mathbf{v_2^*}\end{bmatrix} + \begin{bmatrix}\mathbf{e_{DYD}}\\ \mathbf{e_{YD}}\end{bmatrix}$$

$\mathbf{W_A}$ and $\mathbf{W_D}$ are design matrices which connect pseudo-phenotypes with breeding values and dominance deviations. Breeding values and dominance deviations as well as the respective relationship matrices are marked with an asterisk to indicate that the dimension of these vectors and matrices is reduced in R1S compared to 1S. Dominance deviations are by definition included in YD, but not in DYD. The variance of $\mathbf{e_{YD}}$ is σ_E^2 and of e_{DYD_i} (the i-th element of $\mathbf{e_{DYD}}$) is $\frac{2\sigma_A^2+4(\sigma_D^2+\sigma_E^2)}{n_{dau_i}}$ (Carillier et al., 2013). n_{dau_i} is the number of daughters' YD included in the calculation of the *i*-th DYD. The variance-covariance matrix of $\begin{bmatrix}\mathbf{2DYD}\\ \mathbf{YD}\end{bmatrix}$ is:

$$\mathbf{V}^* = \mathbf{W_A}(\mathbf{H}^*\sigma_A^2)\mathbf{W_A'} + \mathbf{W_D}(\mathbf{H_D}^*\sigma_D^2)\mathbf{W_D'} + \begin{bmatrix}\mathbf{N}(2\sigma_A^2 + 4(\sigma_D^2 + \sigma_E^2)) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma_E^2\end{bmatrix}$$

where **N** is a diagonal matrix with $\frac{1}{n_{dau_i}}$ as *i*-th element of the diagonal. No covariance was assumed between $\mathbf{e_{DYD}}$ and $\mathbf{e_{YD}}$ although in reality such covariance might exist. The estimator of the overall mean for R1S is given by:

$$\widehat{\mu^*} = (\mathbf{1'V^{*-1}1})^{-1}\mathbf{1'V^{*-1}}\begin{bmatrix}\mathbf{2DYD}\\ \mathbf{YD}\end{bmatrix}$$

Breeding values and dominance deviations were estimated with the following equations:

$$\widehat{\mathbf{u}^*} = \mathbf{H}^*\sigma_A^2\mathbf{W_A'V^{*-1}}\left(\begin{bmatrix}\mathbf{2DYD}\\ \mathbf{YD}\end{bmatrix} - \mathbf{1}\widehat{\mu^*}\right); \widehat{\mathbf{v}^*} = \mathbf{H_D}^*\sigma_D^2\mathbf{W_D'V^{*-1}}\left(\begin{bmatrix}\mathbf{2DYD}\\ \mathbf{YD}\end{bmatrix} - \mathbf{1}\widehat{\mu^*}\right)$$

Besides the two dominance models, purely additive 1S and R1S models were applied for the sake of comparison. The model equation of the additive 1S is given by:

$$\mathbf{y} = \mathbf{1}\mu_{\text{add}} + \mathbf{Z_A}\begin{bmatrix}\mathbf{u_{add1}}\\ \mathbf{u_{add2}}\end{bmatrix} + \mathbf{e_{add}}$$

The variance of $\mathbf{e_{add}}$ is $\sigma_D^2 + \sigma_E^2$. The fixed effect μ_{add} and breeding values $\mathbf{u_{add}}$ were estimated as described above just omitting the dominance part. The additive R1S was modelled as:

$$\begin{bmatrix}\mathbf{2DYD}\\ \mathbf{YD}\end{bmatrix} = \mathbf{1}\mu_{\text{add}}^* + \mathbf{W_A}\begin{bmatrix}\mathbf{u^*_{add1}}\\ \mathbf{u^*_{add2}}\end{bmatrix} + \begin{bmatrix}\mathbf{e_{addDYD}}\\ \mathbf{e_{addYD}}\end{bmatrix}$$

The variance of $\mathbf{e_{addYD}}$ is $\sigma_D^2 + \sigma_E^2$ and of e_{addDYD_i} is $\frac{2\sigma_A^2+4(\sigma_D^2+\sigma_E^2)}{n_{dau_i}}$.

Simulation and estimation of breeding values and dominance deviations were repeated 500 times. In each repetition, the correlation between estimated values from 1S and R1S models as well as the correlation between true and estimated breeding values $[r(u,\hat{u})]$, dominance deviations $[r(v,\hat{v})]$ and total genetic values $[r(g,\hat{g})]$ were calculated for different groups of animals: (i) 30 genotyped cows with YD in R1S, (ii) 10 genotyped validation cows which have no records but full-sisters with YD in R1S and (iii) 180 genotyped bulls with DYD in R1S.

Additionally, the regression coefficients of true on estimated breeding values [$b(u, \hat{u})$], dominance deviations [$b(v, \hat{v})$] and total genetic values [$b(g, \hat{g})$] were calculated.

3 | Results

3.1 | Correlations between true and estimated values

Mean $r(u, \hat{u})$, $r(v, \hat{v})$ and $r(g, \hat{g})$ for single-step and reduced single-step models with dams and/or daughters of genotyped cows either modelled in R1S or not are given in Tables 2, 3, and 4 for genotyped cows with YD, genotyped validation cows and genotyped bulls, respectively.

Compared to 1S, R1S resulted in slightly reduced accuracies of breeding values, dominance deviations and total genetic values for genotyped cows with YD. Accuracies with R1S were largest when dams and daughters were modelled in addition to genotyped cows and lowest when only genotyped cows were modelled. This decrease in accuracy was more pronounced for breeding values than for dominance deviations or total genetic values. Modelling of daughters of genotyped cows in R1S tended to result in slightly larger accuracy than modelling of dams of genotyped cows.

$r(u, \hat{u})$, $r(v, \hat{v})$ and $r(g, \hat{g})$ for validation cows were considerably smaller than for cows with records. With 1S, $r(u, \hat{u})$ and $r(g, \hat{g})$ were considerably larger and $r(v, \hat{v})$ was slightly larger than with R1S models. With R1S, accuracies were largest when dams and daughters were modelled additionally to genotyped cows but were not much smaller when only dams were modelled in addition to genotyped cows.

Compared to 1S, $r(u, \hat{u})$ and $r(g, \hat{g})$ for genotyped bulls were reduced with R1S but not affected by inclusion of dams or daughters of genotyped cows in the R1S model. $r(v, \hat{v})$ was very small for both 1S and R1S models. Accuracies of estimated values for bulls did not depend on modelling of dams and/or daughters of genotyped cows.

Table 2. Correlations between true and estimated breeding values, dominance deviations and total genetic values for genotyped cows with YD in the reduced single-step model (n = 30) obtained from single-step (1S) and reduced single-step models (R1S) with different sets of cows in the model (mean ± standard error from 500 replications)

Model	$r(u, \hat{u})$ **(additive models)**	$r(u, \hat{u})$	$r(v, \hat{v})$	$r(g, \hat{g})$
1S	0.732 ± 0.004	0.732 ± 0.004	0.483 ± 0.007	0.807 ± 0.003
R1S (sub model 1)	0.687 ± 0.005	0.686 ± 0.005	0.464 ± 0.007	0.796 ± 0.003
R1S (sub model 2)	0.702 ± 0.005	0.702 ± 0.005	0.471 ± 0.007	0.799 ± 0.003
R1S (sub model 3)	0.710 ± 0.005	0.709 ± 0.005	0.473 ± 0.007	0.801 ± 0.003
R1S (sub model 4)	0.723 ± 0.004	0.722 ± 0.004	0.479 ± 0.007	0.804 ± 0.003

Table 3. Correlations between true and estimated breeding values, dominance deviations and total genetic values for genotyped validation cows (n = 10) obtained from single-step (1S) and reduced single-step models (R1S) with different sets of cows in the model (mean ± standard error from 500 replications)

Model	$r(u, \hat{u})$ **(additive models)**	$r(u, \hat{u})$	$r(v, \hat{v})$	$r(g, \hat{g})$
1S	0.584 ± 0.010	0.584 ± 0.010	0.276 ± 0.009	0.592 ± 0.009
R1S (sub model 1)	0.475 ± 0.010	0.470 ± 0.010	0.272 ± 0.009	0.529 ± 0.009
R1S (sub model 2)	0.479 ± 0.010	0.474 ± 0.010	0.270 ± 0.009	0.530 ± 0.009
R1S (sub model 3)	0.476 ± 0.010	0.472 ± 0.010	0.271 ± 0.009	0.527 ± 0.009
R1S (sub model 4)	0.480 ± 0.010	0.476 ± 0.010	0.269 ± 0.009	0.529 ± 0.009

Table 4. Correlations between true and estimated breeding values, dominance deviations and total genetic values for genotyped bulls (*n*=180) obtained from single-step (1S) and reduced single-step models (R1S) with different sets of cows in the model (mean ± standard error from 500 replications)

Model	$r(u, \hat{u})$ **(additive models)**	$r(u, \hat{u})$	$r(v, \hat{v})$	$r(g, \hat{g})$
1S	0.771 ± 0.002	0.771 ± 0.002	0.070 ± 0.002	0.634 ± 0.003
R1S (sub model 1)	0.726 ± 0.002	0.723 ± 0.002	0.064 ± 0.002	0.596 ± 0.003
R1S (sub model 2)	0.726 ± 0.002	0.723 ± 0.002	0.064 ± 0.002	0.596 ± 0.003
R1S (sub model 3)	0.726 ± 0.002	0.723 ± 0.002	0.064 ± 0.002	0.596 ± 0.003
R1S (sub model 4)	0.726 ± 0.002	0.723 ± 0.002	0.064 ± 0.002	0.596 ± 0.003

Accuracies of breeding values estimated with the additive 1S and R1S models were comparable to, but slightly larger than the respective accuracies from 1S and R1S models including dominance in all cases. Accuracy of estimated breeding values from additive models was largest with 1S and second largest with R1S sub model 4 when cows served as validation group. There was no difference between R1S sub models in the accuracy of estimated breeding values of genotyped bulls.

3.2 | Inflation of estimated genetic values

Mean $b(u, \hat{u})$, $b(v, \hat{v})$ and $b(g, \hat{g})$ for 1S and R1S models with dams and/or daughters of genotyped cows either modelled in R1S or not are given in Table 5 for genotyped cows with YD, genotyped validation cows and genotyped bulls. As the R1S sub model 4 had turned out to be closest to 1S in all cases, results are only presented for this sub model. $b(u, \hat{u})$, $b(v, \hat{v})$ and $b(g, \hat{g})$ were very close to 1 in both models for genotyped cows with YD. For validation cows

and genotyped bulls, $b(u,\hat{u})$, $b(v,\hat{v})$ and $b(g,\hat{g})$ with 1S were not far from 1 while $b(u,\hat{u})$ and $b(g,\hat{g})$ with R1S were larger than 1 and $b(v,\hat{v})$ amounted to 0.569 (validation cows) and only 0.021 (genotyped bulls), which means that dispersion of estimated dominance deviations was far too large. In contrast to R1S, dispersion of predicted dominance deviations was approximately correct for genotyped bulls with 1S even if the predicted dominance deviations were not accurate at all and impaired accuracy of predicted total genetic values. The standard error of the mean $b(v,\hat{v})$ was quite large for validation cows (both 1S and R1S) and genotyped bulls (only 1S). For additive 1S and R1S models $b(u,\hat{u})$ was close to 1.

Table 5. Regression coefficients of true on estimated breeding values, dominance deviations and total genetic values for genotyped cows with YD in the reduced single-step model, genotyped validation cows and genotyped bulls obtained from single-step (1S) and reduced single-step model (R1S) with dams and daughters of genotyped cows in the model (mean ± standard error from 500 replications)

Model	$\boldsymbol{b(u,\hat{u})}$ **(additive models)**	$\boldsymbol{b(u,\hat{u})}$	$\boldsymbol{b(v,\hat{v})}$	$\boldsymbol{b(g,\hat{g})}$
Genotyped cows with YD (n = 30)				
1S	1.001 ± 0.008	1.008 ± 0.008	1.004 ± 0.016	0.992 ± 0.011
R1S (sub model 4)	1.000 ± 0.009	1.017 ± 0.009	1.011 ± 0.017	1.006 ± 0.012
Genotyped validation cows (n = 10)				
1S	1.008 ± 0.023	1.010 ± 0.023	0.924 ± 0.456	1.003 ± 0.034
R1S (sub model 4)	1.064 ± 0.035	1.157 ± 0.039	0.569 ± 0.302	1.180 ± 0.054
Genotyped bulls (n = 180)				
1S	0.993 ± 0.003	0.996 ± 0.003	1.007 ± 0.119	1.002 ± 0.005
R1S (sub model 4)	0.994 ± 0.004	1.143 ± 0.005	0.021 ± 0.009	1.072 ± 0.009

3.3 | Comparison between estimated values from single-step and reduced single-step models

Mean correlations between breeding values, dominance deviations and total genetic values estimated with 1S and different R1S models are shown in Table 6.

Table 6. Correlations between estimated breeding values, dominance deviations and total genetic values from single-step (1S) and different reduced single-step models (R1S) for genotyped cows with YD in the reduced single-step model, genotyped validation cows and genotyped bulls (mean ± standard error from 500 replications)

Correlation of estimated values between 1S and R1S …	Breeding value (additive models)	Breeding value	Dominance deviation	Total genetic value
Genotyped cows with YD (n = 30)				
Sub model 1	0.934 ± 0.001	0.928 ± 0.002	0.959 ± 0.001	0.984 ± 0.000
Sub model 4	0.985 ± 0.000	0.983 ± 0.000	0.989 ± 0.000	0.996 ± 0.000
Genotyped validation cows (n = 10)				
Sub model 1	0.739 ± 0.008	0.730 ± 0.008	0.612 ± 0.013	0.733 ± 0.008
Sub model 4	0.755 ± 0.008	0.746 ± 0.008	0.657 ± 0.012	0.747 ± 0.008
Genotyped bulls (n = 180)				
Sub model 1	0.941 ± 0.001	0.938 ± 0.001	0.036 ± 0.004	0.926 ± 0.001
Sub model 4	0.942 ± 0.001	0.938 ± 0.001	0.032 ± 0.004	0.927 ± 0.001

The concordance of estimated breeding values, dominance deviations and total genetic values for cows with YD was large with correlations between estimates from both models >0.92 when only genotyped cows were modelled in R1S and >0.98 when dams and daughters were modelled in addition to genotyped cows in R1S. Concordance of estimated values for validation cows with 1S and R1S models was much smaller than for cows with records.

Correlations between dominance breeding values/total genetic values of genotyped bulls estimated with 1S and R1S models were quite large while correlations between dominance deviations of genotyped bulls estimated with 1S and R1S models were close to 0. Correlations between estimated values with 1S and R1S sub models 2 and 3 always lay between the results for sub models 1 and 4 and are not reported in Table 6.

4 | Discussion

Single-step genomic evaluation has the important advantage that all available information is considered and weighted optimally (e.g., Legarra et al., 2009). However, several authors have reported difficulties in the convergence of large single-step systems (Harris et al., 2013; Liu, Goddard, Reinhardt, & Reents, 2014). The problem of convergence will be even more severe when dominance deviations are modelled in addition to breeding values in real-world systems. Another issue might be the calculation of the inverse of $\mathbf{A_D}$ because inversion of routine-size matrices is computationally not feasible. For the additive part, Henderson (1976) has presented an algorithm to compute the inverse of **A** directly. When dominance is modelled, the inverse of $\mathbf{A_D}$ is required in addition to the inverse of **A**. Although Hoeschele and VanRaden (1991) have described an algorithm to calculate this inverse directly (for non-inbred pedigrees), the computation might still be unfeasible in routine-size genetic evaluations. A reduced single-step model which uses DYD of bulls and YD of a subset of genotyped cows could therefore be a practical alternative to a full 1S model because the dimension of the model is considerably reduced with aggregation of daughters' records to DYD.

The aim of this study was to analyze whether a R1S model with DYD and YD as pseudo-phenotypes can be a practical alternative to 1S when dominance is modelled. For a practical application of R1S, it would be required that the accuracies of breeding values, dominance

deviations and total genetic values of cows (with or without own records) estimated with R1S should not be much smaller than with 1S. Dominance deviations and total genetic values of bulls are predicted inaccurately because these bulls do not have own records in female traits and have only small dominance relationships with record-tested cows, but for these bulls, dominance deviations and total genetic values are not relevant and the only relevant criterion for bulls is that $r(u, \hat{u})$ with R1S should be comparable to 1S. We additionally examined whether it is sufficient to model only the genotyped cows that one is interested in with their YD in the reduced single-step model or if it would be necessary to model the dams and/or daughters of these cows additionally in the reduced single-step model.

Although 1S performed better in the analysed criteria, accuracy and dispersion of both genetic values (breeding values, dominance deviations and total genetic values) of reference cows and breeding values of reference bulls estimated with R1S were satisfactory. Predictive ability of R1S for the additive part, however, was worse than 1S in our study. Perhaps, refined modelling techniques could improve prediction accuracy of R1S but, as every R1S model contains by definition less information than a 1S model, a certain difference will remain. The results demonstrated that an R1S model can be an acceptable alternative to a full 1S model when the purpose is to estimate breeding values, dominance deviations and total genetic values of cows with records and breeding values of bulls with daughter records.

We could not expect, of course, that R1S estimates genetic values with the same quality as a full 1S model because (especially dominance-specific) information gets lost during calculation of DYD (VanRaden & Wiggans, 1991). Other attempts of additive reduced single-step evaluation also resulted in reduced accuracy of breeding values compared to single-step evaluation (Baloche et al., 2014). Apart from the modelling of dominance, a difference of our study to other reduced single-step analyses (Harris et al., 2013; Su et al., 2012) is the use of YD and DYD as pseudo-phenotypes instead of de-regressed proofs (Garrick et al., 2009).

Either DYD/YD or de-regressed proofs are used in genomic evaluations and there is no evident advantage of one of both types of pseudo-phenotypes over the other. However, in an analysis that includes dominance deviations, YD have to be used as pseudo-phenotypes for cows because de-regressed breeding values do not contain dominance-specific information. In contrast, dominance deviations are not expected to be removed during calculation of YD because the dominance covariance structure is not similar to the covariance structure of any random environmental effect. Simulation studies with different data structures could help to enlighten this question but were beyond the scope of this investigation. Estimation quality of dominance deviations and total genetic values for bulls should not be considered when comparing 1S and R1S because for bulls these values are typically not relevant. Taking this into account, R1S is an acceptable alternative to 1S without too many compromises for relevant estimates.

R1S combines the advantage of 1S (correct weighting of pedigree, genomic, phenotypic information) with the reduced computational cost of a multiple-step approach; however, genetic values (i.e., breeding values, dominance deviations and total genetic values) are estimated only for a subset of animals whose genetic values one is interested in. Prediction quality of R1S additionally depends essentially on the quality of DYD and YD as pseudo-phenotype inputs in this model.

Provided that an investigator is interested in both breeding values of bulls (genotyped or not genotyped; with daughter records) and breeding values, dominance deviations and total genetic values of genotyped cows with records, all available bulls should be modelled in R1S with their DYD, and genotyped cows should be modelled with their YD. Depending on the computational capacities, female relatives of the genotyped cows (i.e., dam and/or daughters) should be modelled additionally in order to increase the accuracy of the estimates. From a practical point of view, it must be considered that, typically, records of the dam of a cow are

available whereas records of a daughter of a cow are only available as soon as the daughter is in her first lactation. Thus, dams of genotyped cows are expected to be able to contribute information the earliest possible to R1S genomic evaluation and should be modelled preferably. Also full- and half-sisters will contribute information, but their impact on R1S evaluation was not analysed in the present study. This should be a question of further research.

The computational effort depends essentially on the dimension of the model. Thus, the computational benefit (or, more precisely, the saving of computational capacity) is proportional to $\frac{k^2-l^2}{k^2}$, where *k* is the number of animals in the 1S model and *l* is the number of animals in the R1S model. In our example, the computational benefit amounts to 98.6%, 98.3%, 98.3% and 98.0%, respectively, for R1S sub models 1–4.

The results of this analysis are based on the infinitesimal model used for simulation. In reality, traits are defined by a usually very large, however, finite number of quantitative trait loci. It should be a matter of further research to analyze 1S and R1S with real phenotypes.

The evaluation model assumes Hardy-Weinberg equilibrium (HWE). Vitezica, Legarra, Toro, and Varona (2017) have described a non-additive model that remains orthogonal even if the population is not in HWE. In this case, the additive and dominance effects of individuals will remain orthogonal and their accuracies will keep a straightforward interpretation, but the additive effect does no longer have the meaning of a breeding value because breeding values do not exist for populations deviating from HWE.

Given the fact, that accuracy of breeding values was not larger with dominance than with additive models, and with respect to the computational costs, a dominance model should only be applied when there is interest in dominance deviations and/or total genetic values of cows.

5 | Conclusions

The results indicate that a reduced single-step dominance model which is based on DYD of bulls and YD of genotyped cows is suitable to estimate breeding values of bulls and breeding values, dominance deviations and total genetic values of cows with acceptable quality and can be an alternative to a single-step model when the single-step method cannot be applied due to computational limitations. The more cows are modelled in the reduced single-step model the closer the results are to single-step results.

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). 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 genotypes of bulls.

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genomic evaluation of Holstein final score. *Journal of Dairy Science, 93,* 743-752. https://doi.org/10.3168/jds.2009-2730

Baloche, G., Legarra, A., Sallé, G., Larroque, H., Astruc, J. M., Robert-Granié, C., & Barillet, F. (2014). Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *Journal of Dairy Science, 97,* 1107-1116. https://doi.org/10.3168/jds.2013-7135

Carillier, C., Larroque, H., Palhière, I., Clément, V., Rupp, R., & Robert-Granié, C. (2013). A first step toward genomic selection in the multi-breed French dairy goat population. *Journal of Dairy Science, 96,* 7294-7305. https://doi.org/10.3168/jds.2013-6789

Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution, 42,* 2. https://doi.org/10.1186/1297-9686-42-2

Ertl, J., Legarra, A., Vitezica, Z. G., Varona, L., Edel, C., Emmerling, R., & Götz, K.-U. (2014). Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. *Genetics Selection Evolution, 46,* 40. https://doi.org/10.1186/1297-9686-46-40

Gao, H., Christensen, O. F., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S., & Su, G. (2012). Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genetics Selection Evolution, 44,* 8. https://doi.org/10.1186/1297-9686-44-8

Garrick, D. J., Taylor, J. F., & Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution, 41,* 55. https://doi.org/10.1186/1297-9686-41-55

Harris, B. L., Winkelman, A. M., & Johnson, D. L. (2013). Impact of including a large number of female genotypes on genomic selection. *Interbull Bull, 47,* 23-27.

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. *Biometrics, 32,* 69-83. https://doi.org/10.2307/2529339

Hoeschele, I., & VanRaden, P. M. (1991). Rapid inversion of dominance relationship matrices for noninbred populations by including sire by dam subclass effects. *Journal of Dairy Science, 74,* 557-569. https://doi.org/10.3168/jds.S0022-0302(91)78203-9

Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science, 92,* 4656-4663. https://doi.org/10.3168/jds.2009-2061

Liu, Z., Goddard, M. E., Reinhardt, F., & Reents, R. (2014). A single-step genomic model with direct estimation of marker effects. *Journal of Dairy Science, 97,* 5833-5850. https://doi.org/10.3168/jds.2014-7924

Miglior, F., Burnside, E. B., & Kennedy, B.W. (2009). Production traits of Holstein cattle: Estimation of nonadditive genetic variance components and inbreeding depression. *Journal of Dairy Science, 78,* 1174-1180. https://doi.org/10.3168/jds.S0022-0302(95)76735-2

Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science, 92,* 4648-4655. https://doi.org/10.3168/jds.2009-2064

Mrode, R. A. (2005). *Linear models for the prediction of animal breeding values.* Wallingford, UK: CABI Publishing. https://doi.org/10.1079/9780851990002.0000

Su, G., Madsen, P., Nielsen, U. S., Mäntysaari, E. A., Aamand G. P., Christensen, O. F., & Lund, M. S. (2012). Genomic prediction for Nordic Red Cattle using one-step and selection index blending. *Journal of Dairy Science, 95,* 909-917. https://doi.org/10.3168/jds.2011-4804

Tempelman, R. J., & Burnside, E. B. (1990). Additive and nonadditive genetic variation for production traits in Canadian Holsteins. *Journal of Dairy Science, 73,* 2206-2213. https://doi.org/10.3168/jds.S0022-0302(90)78900-X

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. *Journal of Dairy Science, 83,* 1873-1877. https://doi.org/10.3168/jds.S0022-0302(00)75059-4

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science, 91,* 4414-4423. https://doi.org/10.3168/jds.2007-0980

VanRaden, P. M., & Wiggans, G. R. (1991). Derivation, calculation, and use of national animal model information. *Journal of Dairy Science, 74,* 2737-2746. https://doi.org/10.3168/jds.S0022-0302(91)78453-1

Vitezica, Z. G., Aguilar, I., Misztal, I., & Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genetics Research, 93,* 357-366. https://doi.org/10.1017/S001667231100022X

Vitezica, Z. G., Legarra, A., Toro, M. A., & Varona, L. (2017). Orthogonal estimates of variances for additive, dominance and epistatic effects in populations. *Genetics, 206,* 1297–1307. https://doi.org/10.1534/genetics.116.199406

Vitezica, Z. G., Varona, L., & Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics, 195,* 1223-1230. https://doi.org/10.1534/genetics.113.155176

Zeng, J., Toosi, A., Fernando, R. L., Dekkers, J. C. M., & Garrick, D. J. (2013). Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genetics Selection Evolution, 45,* 11. https://doi.org/10.1186/1297-9686-45-11

Summary and Authors' Contributions

Single-step models including dominance can be an enormous computational task and can even be prohibitive for practical application. In this study the question was addressed whether a reduced single-step model is able to estimate breeding values of bulls and breeding values, dominance deviations and total genetic values of cows with acceptable quality. Genetic values and phenotypes were simulated (500 repetitions) for a small Fleckvieh pedigree consisting of 371 bulls (180 thereof genotyped) and 553 cows (40 thereof genotyped). This pedigree was virtually extended for 2,407 non-genotyped daughters. Genetic values were estimated with the single-step model and with different reduced single-step models. Including more relatives of genotyped cows in the reduced single-step model resulted in a better agreement of results with the single-step model. Accuracies of genetic values were largest with single-step and smallest with reduced single-step when only the cows genotyped were modelled. The results indicate that a reduced single-step model is suitable to estimate breeding values of bulls and breeding values, dominance deviations and total genetic values of cows with acceptable quality.

J. Ertl performed the simulation and the analysis by means of own-written R programs and drafted the manuscript. J. Ertl, C. Edel and K.-U. Götz designed the study. C. Edel and R. Emmerling prepared phenotypic and genotypic data. C. Edel, E. Pimentel, R. Emmerling and K.-U. Götz revised the manuscript.

5th Chapter

General discussion

Potential advantages from higher SNP marker density and female genotypes in genomic analysis of Fleckvieh cattle were studied in this thesis. In the genomic analysis of female genotypes, the emphasis lay on the importance and on the potential use of dominance variance. The studies drew benefit from methodological contributions of many authors e.g. in the development of GBLUP (Meuwissen et al., 2001; VanRaden, 2008), single-step evaluation (Legarra et al., 2009; Christensen and Lund, 2010; Aguilar et al., 2010), and analysis of dominance effects (Toro and Varona, 2010; Vitezica et al., 2013). Valuable work in the application of GBLUP to practical conditions in Fleckvieh and in preparation of phenotypes and genotypes had been carried out by the team of the Institute of Animal Breeding at Bavarian Research Centre for Agriculture (Edel et al., 2011).

Although many types of statistical models have been tested for their suitability in genomic analyses, the genomic BLUP model, a linear model which is based on the infinitesimal model and assumes normal distribution of SNP effects, has turned out to perform well under the very most of practical conditions in dairy cattle populations (Hayes et al., 2009; VanRaden et al., 2009, 2013; Erbe et al., 2012; Su et al., 2012) and especially in the Fleckvieh breed (Gredler et al., 2010; Pausch et al., 2011; Pryce et al., 2011). Furthermore, the GBLUP model is compatible with conventional BLUP procedures in genetic evaluation and allows for the estimation of individual prediction reliabilities, unlike non-BLUP models. Computational demands are manageable with GBLUP, unlike many non-linear approaches. For these reasons, I have focused on BLUP models in this thesis evaluating the impacts of high-density and female genotypes. In the following general discussion I would like to comment on some aspects arising from the analyses.

Effects of marker density on the properties of genomic relationship matrices and predictive ability

Several authors have assessed the effect of higher marker density on predictive ability of genomic breeding values. Often favorable effects on prediction accuracy have been found, although gains in accuracy were not large (Harris et al, 2011; Erbe et al., 2012; Su et al., 2012; VanRaden et al., 2013). Harris et al. (2011) found in a simulation study of a trait with 50% heritability and genetic architecture close to the infinitesimal model that forward prediction accuracy with 100,000 SNP was 3.2 percentage points larger than with 20,000 SNP using ridge regression BLUP, which is equivalent to GBLUP, and with 4,799 bulls and their 15,000 genotyped daughters in the calibration data set. Erbe et al. (2012) reported losses in validation accuracy of 2, 0 and 1 percentage points for milk, fat and protein yield, respectively, and with a mixed Holstein-Jersey reference set and a Holstein validation set when 800K SNP were analyzed instead of 50K SNP. With a Jersey validation set in turn accuracy increased by 1, 2 and 4 percentage points, respectively. Su et al. (2012) found small gains in prediction reliability of 0.4, 0.9 and 0.0 percentage points in the traits protein yield, fertility and udder health, respectively, when 777K SNP were used instead of 50K SNP to predict genomic breeding values of Holstein cattle with GBLUP. In Nordic Red Dairy cattle, Su et al. (2012) found gains in prediction reliability of 1.2, 0.7, and 1.3 percentage points for the same type of comparison and the same traits. In chapter 2 of this thesis, validation reliability with imputed 777K SNP in Fleckvieh was on average 1.5 percentage points (ranging from 0.8 percentage points in fat yield to 2.3 percentage points in milkability) larger than with actually genotyped 50K SNP. We were able to show in the Appendix of chapter 2 that the potential gain in reliability with higher marker density was limited by the error arising from imputation from 50K to 777K genotypes. Other groups of authors equally reported negative impacts of imputation error on the reliability of genomic predictions (Chen et al., 2011; Dassonneville et

al., 2011; VanRaden et al., 2013). In a situation without imputation error or with the same level of imputation error in the different marker sets, the gain in prediction reliability with 777K SNP instead of 50K SNP was 2.8 percentage points, on average, and ranged from 2.1 percentage points in stature to 3.6 percentage points in milkability. In the studies of Harris et al. (2011), Erbe et al. (2012) and Su et al. (2012), it was not investigated if the reported gains in accuracy were statistically significant. Chapter 2 of this thesis proposes an approach for testing the difference in prediction accuracy with 50K and HD for significance. Reliability of genomic predictions with 777K SNP was compared with the distribution of reliabilities from different 50K samples out of 777K. All differences in prediction reliability were significant because 777K reliabilities were larger than the respective 95% quantile of the empirical 50K reliability distribution. The probability that the observed difference occurred by chance was in all traits even smaller than 1%. In addition, sampling of validation bulls showed that the difference between HD (affected by imputation error) and real 50K genotypes was significant.

The regression of pseudo-phenotypes on predicted breeding values is a common measure in validation studies. A regression coefficient < 1 indicates that the dispersion of predicted breeding values is too large. Top predicted breeding values will then be overestimated and bottom predicted breeding values will be underestimated. Such an inflation of breeding values leads to suboptimal selection decisions (Patry and Ducrocq, 2009; VanRaden et al., 2009; Mäntysaari et al., 2010). The inclusion of a polygenic effect in the model, weighted combinations of genomic and pedigree-based relationships and single-step evaluation (cf. chapter 4) have been suggested to reduce inflation (e.g. Aguilar et al., 2010; Liu et al., 2011). As reported by Su et al. (2012), the coefficient of the regression of pseudo-phenotypes on predicted breeding values was larger with 777K than with 50K and thus closer to the expected value. This means that the variance of predicted breeding values with 50K was larger than the variance of predicted breeding values with 777K. Predicted breeding values were inflated with

both marker densities but inflation was less with the higher marker density. In direct association to this observation, the model-based reliability of predicted breeding values decreased with higher marker density, in contrast to validation reliability. A plausible explanation for both observations is the finding of both Goddard et al. (2011) and Endelman and Jannink (2012) that there is a sampling error in genomic relationships calculated from markers due to finite sample size. This sampling error depends reciprocally on the number of markers (Yang et al., 2010) and causes the variance of predicted breeding values to be overestimated because genetic variance and genomic relationship coefficients (that have too large dispersion because of sampling error) are treated as true parameters in the BLUP procedure.

In direct association to the observed inflated predictions of breeding values, the model-based reliability of predicted breeding values decreased with higher marker density in chapter 2, in contrast to validation reliability. The reason is identical: The model-based reliability is overestimated because of sampling error of genomic relationships that causes too large variance in the coefficient matrix (e.g. Goddard et al., 2011).

Endelman and Jannink (2012), based on results of Ledoit and Wolf (2004), claimed that the sampling error of genomic relationship coefficient was not only reciprocally proportional to the number of markers but proportional to $\frac{n}{m*CV^2}$, where m is the number of markers, n is the number of genotyped animals and CV is the coefficient of variation of the eigenvalues of the genomic relationship matrix. They proposed a shrinkage procedure in the calculation of the genomic relationship matrix to correct for the sampling error. The shrinkage target is a diagonal matrix with diagonal elements $1+f$, where f is the individual genomic inbreeding coefficient. An equivalent approach had been already suggested by Hayes and Hill (1981) in order to render covariance matrices positively semi-definite and was called ‘bending’. Indeed,

shrinkage resulted in reduced inflation of predicted breeding values and in reduced model-based reliabilities, just as could be expected from the reflections about sampling error.

Impact of imputation error on reliability of genomic breeding values

It could be shown that the reduction of prediction reliability by about 0.013 when 50K samples from HD genotypes were used instead of original 50K genotypes was caused by imputation error in HD genotypes, by introducing an artificial genotype error - equivalent to imputation error in Fleckvieh - in original 50K genotypes. The introduction of artificial errors resulted in a decay of prediction reliability of 0.015 percentage points. Inaccuracy in the imputation of genotypes is known to hamper the predictive ability of genomic breeding values predicted by means of imputed genotypes (Chen et al., 2011; Dassonneville et al., 2011; VanRaden et al., 2013). Even though several genomic prediction studies were conducted with imputed genotypes (Erbe et al., 2012; Su et al., 2012), these authors have ignored the impact of imputation error on prediction accuracy. Pimentel et al. (2014) found in an analysis of imputed genotypes that most imputation errors occurred through replacement of a rare allele by a common allele. As a consequence, predicted breeding values of the best animals were shrunk towards the population mean, and predictions for the worst animals increased. Obviously, this phenomenon reduces the dispersion of genomic predictions, but it affects the model-based reliabilities as well. These latter impacts of imputation error on genomic predictions were not analyzed in chapter 2 of this thesis. But we have to be aware of such potential impacts of imputation error when analyses are based on imputed genotypes.

Including dominance in genomic evaluations

Genetic variance can be partitioned in additive, dominance and epistasic variance (Cockerham, 1954; Kempthorne, 1954). Additive-genetic variance results from additive (and dominance; Falconer and Mackay, 1996) effects of alleles at the different loci and is classically exploited in purebreeding schemes for livestock and plant species. Usually, additive-genetic variance covers a large part of total genetic variance, is convenient to work on because of the linear structure and has successfully been used for decades to select breeding animals by means of estimated breeding values to generate sustainable genetic progress. Non-additive genetic variance contains dominance and epistatic variance. Epistasis is the sum of interaction effects between gene loci. Non-additive genetic analysis in this thesis concentrated on dominance and its variance. Dominance arises from interaction between alleles at a locus such that the genetic value of a heterozygous genotype at a locus deviates from the sum of the respective (additive) allelic effects. Well-known examples for dominance actions are some qualitative traits like coat color: The coat color of cattle resulting from a heterozygote genotype with alleles for red and black color is not a mixture of black and red pigments but the same black color as of homozygous black genotypes because there is complete dominance of the black allele. Similarly, dominance gene action can exist also in quantitative traits and is classically exploited in crossbreeding schemes of e.g. maize, chicken and swine in order to obtain desired specific genotypes with maximal total (additive plus dominance) genetic value.

Breeding values and dominance deviations are orthogonal by definition (e.g. Falconer and Mackay, 1996) and thus no covariance exists between them. Chapters 3 and 4 of this thesis are based on this classical parametrization as implemented by Vitezica et al. (2013) for genomic analysis of dominance. However, the existence of inbreeding depression indicates that dominance can be directional (e.g. Lynch and Walsh, 1998). Wellmann and Bennewitz (2011, 2012) suggested that biologically additive and dominance effects can be interdependent. This

would result in dependence between breeding values and dominance deviations. Although the question of dependence or independence should be further investigated, the investigations of this thesis relied on the classical parametrization of independence between breeding values and dominance deviations.

It has already been known from pedigree-based estimation of variance components that there is non-negligible dominance variation in production and conformation traits of Holstein cattle (Tempelman and Burnside, 1990a, b; Miglior et al., 1995; Misztal et al.,, 1997; Van Tassell et al., 2000). In pedigree-based estimation of dominance variance, a sufficient proportion of full-sibs are required in the data set to have enough dominance relationships between individuals. Van Tassell et al. (2000) for example selected their data set such that at least 14% full-sibs were contained in the data. Tempelman and Burnside (1990a, b) obtained a data set with 15% and 16% full-sibs, respectively in conformation and production analysis, by means of selection from a larger data set. Typically, full-sib pairs are not very frequent in dairy cattle, unlike in multiparous species like pigs. In chapter 3 of this thesis, the analyzed data set was not selected for full-sib relationships and contained only 3% full-sibs, which is a typical proportion in dairy cattle. However, in addition to full-sib relationships, many genomic dominance relationships with small absolute values were present in the data set (cf. Figure 1b in chapter 3).

Albeit information content in the data set was apparently sufficient to obtain significant estimates for dominance variance in most traits, information or more specifically dominance relationships have not been sufficient to accurately predict genomic dominance deviations of validation cows in cross-validation analysis. As it is known from several studies about genomic prediction of breeding values, a sufficiently close relationship between training and validation animals is an essential factor for obtaining precise predictions (e. g. Habier et al., 2007, 2010; Pszczola et al., 2012). Analogously, larger genomic dominance relationships, i.e. more full-sib relationships, between training and validation cows would have been required to predict

dominance deviations more precisely. From a practical point of view, the full-sister has to have passed her first lactation before dominance deviations of a validation animal can be predicted with a larger accuracy. Given that a training cow is 24 months old at first calving, has finished her first lactation at the age of 34 months and has - due to use of sexed semen of a certain bull in artificial insemination - a full-sister born every 12 months after her own birth, the information available at the end of her first lactation could be used to predict dominance deviations and total genetic values of her genotyped full-sisters and to decide if (1) the oldest full-sister at the age of 22 months should be kept for dairy production, (2) the second-oldest full-sister at the age of 10 months should be raised for own replacement, (3) the third-oldest full-sister (at this time a fetus two months pre natum) should be raised for own replacement or sold for breeding/fattening purposes or (4) which out of several genotyped full-sib embryos to be conceived in five months from multiple ovulation should be carried to term (Table 1).

Table 1. Age, production stage and potential use of predicted total genetic values for different full-sisters

Age	Production stage	Use of predicted total genetic value
34 months	First lactation completed	-
22 months	First calving in 2 months	Decide whether to keep for dairy production or to sell
10 months	Insemination in 5 months	Decide whether to keep for own replacement or to sell for breeding/fattening
-2 months	Birth in 2 months	Decide whether to raise for own replacement or to sell the calf for breeding/fattening
-14 months	Conception of multiple embryos in 5 months	Decide which embryos to carry out

The earlier in the production process dominance deviations and total genetic values are available for an animal, the more the farmer can benefit from this information because he can save raising costs. On the other hand, as accurate predictions are only achievable for full-sisters, the selection candidate has to have the same parents as the reference full-sister which tends to prolong the generation interval. This can imply that the farmer has to renounce to additive-genetic gain that could have been acquired in the meantime. Positive dominance deviations have thus to be traded off against a reduction of additive-genetic gain. Furthermore, the within-herd variability would be reduced, because many cows would be full-sisters. On the other hand, this could improve the estimation of dominance variance and dominance effects. Systematic genotyping and analysis of full-sisters should be undertaken to investigate this subject. Prolonged generation interval could principally be avoided by using embryo transfer.

An even more consequent application of information about dominance effects is the selection of the optimal mating partner for a cow in order to maximize total genetic value in the offspring. Breeders and farmers choose nowadays the mating partner for a cow with respect to maximal expected breeding value of the progeny. While breeders can be indifferent to dominance deviations that occur only in a specific animal and are not transmitted to subsequent generations, farmers want to maximize the production of their cows and are therefore interested in maximal total genetic values in order to obtain maximal phenotypic values. Since the total genetic value is the sum of breeding value and dominance deviation, the optimal mating partner should be chosen to generate both excellent breeding values and positive dominance deviations in the offspring of the specific mating. Based on the approach of Toro and Varona (2010), expected breeding values, dominance deviations and total genetic values were calculated in chapter 3 for all potential matings between a certain cow and 50 bulls for the traits milk yield and protein yield. The top mating was selected for each of the 1996 cows with respect to (1) maximal breeding value or (2) maximal total genetic value. We have imposed the

restriction that one bull is not mated to more than 200 cows. The expected total genetic superiority was 14.8% (milk yield) and 27.8% (protein yield) larger when matings were selected on the expected total genetic value instead of the expected breeding value. The effect was larger for protein yield where the estimated proportion of dominance variance as compared to total genetic variance was larger. The results coincide with the finding of Toro and Varona (2010) in a simulation study that expected total genetic superiority was with selection of matings on the total genetic value 16% larger than with selection on the breeding value only. Despite the remarkable expected advantage from selection on the expected total genetic value, the expected breeding values did only decrease by 4.5% for milk yield and 2.6% for protein yield. This would not severely reduce additive genetic gain, notably because only breeding values of cows (in producer herds) but not those of AI bulls would be affected by this small reduction. While for loci where the mating partners are homozygous the genotype of the offspring can be predicted perfectly (either also homozygous when genotypes of mating partners are identical or heterozygous when mating partners are alternatively homozygous), different genotypes will be expected with varying probabilities in all other combinations. A weighted average of local breeding values, dominance deviations and total genetic values, respectively, was calculated for these loci, which resulted in values closer to the population mean than values of loci with predictable genotypes. This inaccuracy has to be accepted while the expected genetic values are still unbiased. The approach of summing up the product of values and the probability for the respective genotype has also been suggested to perform a genome-wide association study for dominance effects using the expectations of genotypes of un-genotyped cows by Boysen et al. (2013).

Potential positive dominance deviations that have been realized by means of a planned mating are not inherited to subsequent generations because the dominance deviation is associated with the individual genotype and not with the gametes. This characteristic arises from quantitative

genetic theory (e.g. Cockerham, 1954) and was confirmed by the simulation study of Toro and Varona (2010). For each generation, the optimal mating partner has to be chosen anew in order to obtain large total genetic values in the next generation.

The extension of the strategy of planned matings to multiple traits and to the aggregate genotype is straightforward even if the procedure will be much more complex then. A web-based application could be developed where a farmer provides the genotype of a cow, chooses a number of bulls as potential mating partners, defines the most preferred selection criterion and receives a mating proposal for the bull which maximizes the expected total genetic value (or breeding value) of matings.

Combining genotypic, phenotypic and pedigree information of bulls and cows in genomic analyses including dominance

In chapter 3 of this thesis, dominance effects were analyzed based on a data set of 1,996 genotypes cows. Although records of cows are valuable phenotypes with relatively large information content, it is a pity to reject to the huge amount of information that could be available from (un-genotyped) daughters of genotyped bulls. The studies on single-step genomic evaluation (Legarra et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010) have resulted in an elegant BLUP model that combines all sources of information (genotypes, phenotypes, pedigree) in an optimal way. However, the computation of a full single-step model containing additive and dominance effects is not feasible with the present technology. In order to demonstrate ways to evaluate a data set consisting of genotyped bulls and both genotyped and un-genotyped cows for breeding values and dominance deviations, a single-step model including dominance was applied to a relatively small semi-real data set adapted from the German-Austrian routine genomic evaluation.

Since its development, the single-step model has been subject of extensive investigations under a multitude of aspects. The main advantages of optimal use and weighting of all available information in single-step evaluation are increased reliability of estimates and predictions, and less bias and inflation of genomic predictions. This has been confirmed by a series of studies (e. g. Gao et al., 2012; Su et al., 2012). In practice limited computational capacities and potential convergence problems (Harris et al., 2013; Liu et al., 2014) raise the demand for a reduced single-step evaluation procedure. Given that a full single-step model contains all breeding bulls and all cows (a small proportion genotyped and the remaining large proportion not genotyped), the reduction of the model to bulls and genotyped cows is a consequent approach to save computing time and memory. For example, with the data set in chapter 4 of this thesis, the size of the model shrank from 3,331 to 411 animals. This is only 12% of the full single-step model size and would result in a 66-fold reduction of computation time for operations with quadratic increase of computation time with increasing model size. The more un-genotyped cows are included in the reduced single-step model, the more the results converge towards the full single-step model. But in this case, computing time converges towards the full single-step model requirements, too. This is not a problem when the number of genotyped cows is much smaller than the number of bulls in the evaluation as it was the case in the example of chapter 4. But in the United States already 1,187,231 genotypes of Holstein cows and heifers have been included in the genomic evaluation (Council on Dairy Cattle Breeding, 2017) and inclusion of their dams in a reduced single-step model would considerably increase the model size. The extent of model reduction and savings in computation time has to be traded off against the reduced reliability of estimated genetic values.

The finding that the accuracy of predicted breeding values was larger with full than with reduced single-step in additive analyses (e.g. Baloche et al., 2014) was confirmed by the results from additive full and reduced single-step models in chapter 4 of this thesis. The same finding

held also in general for the dominance models. The extent of benefits in accuracy of full single-step depended on the type of genetic value (breeding value, dominance deviations or total genetic value) and on the type of validation animal (cow with records, cow without records, bull with records of daughters).

Estimated breeding values showed similar properties between additive and dominance models for full and reduced single-step. This indicates that the modelling of dominance did not impair the quality of breeding value estimation – neither in the full nor in the reduced single-step model.

Predicted breeding values of validation cows were not inflated, neither with full, nor with reduced single-step models. Several papers have reported that single-step evaluation has the desired effect of reducing inflation in comparison to multi-step GBLUP (Gao et al.; 2012; Su et al., 2012). However the dispersion of predicted dominance deviations of validation cows was somewhat underestimated. Vitezica et al. (2011) outlined that scaling of genomic relationships towards the respective numerator relationships results in less bias and inflation of predictions. This scaling approach was applied in both the full and the reduced single-step model.

A reduced single-step evaluation is an acceptable alternative to the full single-step model even for estimation of dominance deviations, when computational resources are limited. Estimated dominance deviations for bulls were not valuable at all but are neither relevant for breeders. In turn, the inclusion of dominance in the single-step model did not affect the estimation of breeding values for bulls.

References

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

Christensen, O. F., and Lund, M. S. (2010): Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **42:**2.

Council on Dairy Cattle Breeding (2017): Genotypes included in evaluations by breed, chip density, presence of phenotypes (old vs. young) and evaluation year-month (cumulative). https://www.cdcb.us/Genotype/cur_density.html (accessed 2 January 2017).

Boysen, T.-J., Heuer, C., Tetens, J., Reinhardt, F., and Thaller, G. (2013): Novel Use of Derived Genotyped Probabilities to Discover Significant Dominance Effects for Milk Production Traits in Dairy Cattle. *Genetics* **193:**431-442.

Chen, J., Liu, Z., Reinhardt, F., and Reents, R. (2011): Reliability of genomic prediction using imputed genotypes for German Holsteins: Illumina 3K to 54K bovine chip. *Interbull Bulletin* **44:**51-54.

Cockerham, C. C. (1954): An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39:**859-882.

Dassonneville, R., Brøndum, R. F., Druet, T., Fritz, S., Guillaume, F., Guldbrandtsen, B., Lund, M. S., Ducrocq, V., and Su, G. (2011): Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *J. Dairy Sci.* **94:**3679-3686.

Edel, C., Schwarzenbacher, H., Hamann, H., Neuner, S., Emmerling, R., and Götz, K.-U. (2011): The German-Austrian Genomic Evaluation System for Fleckvieh (Simmental) Cattle. *Interbull Bulletin* **44:**152-156.

Endelman, J. B., and Jannink, J.-L. (2012): Shrinkage estimation of the realized relationship matrix. *G3 (Bethesda)* **2:**1405-1413.

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

Gao, H., Christensen, O. F., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S., and Su, G. (2012): Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet. Sel. Evol.* **44:**8.

Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009): Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* **41:**55.

Habier, D., Fernando, R. L., and Dekkers, J. C. M. (2007): The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics* **177:**2389-2397.

Habier, D., Tetens, J., Seefried, F.-R., Lichtner, P., and Thaller, G. (2010): The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* **42:**5.

Harris, B. L., Creagh, F. E., Winkelman, A. M., and Johnson, D. L. (2011): Experiences with the Illumina high density Bovine BeadChip. *Interbull Bulletin* **44:**3-7.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard. M. E. (2009): Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* **92:**433–443.

Hayes, J. F., and Hill, W. G. (1981): Modification of Estimates of Parameters in the Construction of Genetic Selection Indices (‘Bending’). *Biometrics* **37:**483-493.

Kempthorne, O. (1954): The Correlation between Relatives in a Random Mating Population. *Proc. R. Soc. Lond. B* **143:**103-113.

Ledoit, O., and Wolf, M. (2004): A well-conditioned estimator for large-dimensional covariance matrices. J. Multivariate Anal. 88:365-411.

Legarra, A., Aguilar, I., and Misztal, I. (2009): A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* **92:**4656-4663.

Lynch, M., and Walsh, J. B. (1998): Genetics and Analysis of Quantitative Traits. Sinauer Associates, Inc., Sunderland, 1998.

Mäntysaari, E. A., Liu, Z., VanRaden, P. M. (2010): Interbull validation test for genomic evaluations. *Interbull Bulletin* **41:**17-22.

Miglior, F., Burnside, E. B., and 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., and Gengler, N. (1997): Relationships among estimates of inbreeding depression, dominance and additive variance for linear traits in Holsteins. *Genet. Sel. Evol.* **29**:319-326.

Patry, C., and Ducrocq, V. (2009): Bias due to genomic selection. *Interbull Bulletin* **39:**77-82.

Pausch, H., Flisikowski, K., Jung, S., Emmerling, R., Edel, C., Götz, K.-U., and Fries, R. (2011): Genome-wide association study identifies two major loci affecting calving ease and growth related traits in cattle. *Genetics* **187:**289-297.

Pszczola, M., Strabel, T., Mulder, H. A., and Calus, M. P. L. (2012): Reliability of direct genomic values for animals with different relationships within and to the reference population. *J. Dairy Sci.* **95:**389–400.

Su, G., Brøndum, R. F., Ma, P., Guldbrandtsen, B., Aamand, G. P., and Lund, M. S. (2012): Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *J. Dairy Sci.* **95:**4657-4665.

Su, G., Madsen, P., Nielsen, U. S., Mäntysaari, E. A., Aamand, G. P., Christensen, O. F., and Lund, M. S. (2012): Genomic prediction for Nordic Red Cattle using one-step and selection index blending. *J. Dairy Sci.* **95:**909-917.

Tempelman, R. J., and Burnside, E. B. (1990a): Additive and nonadditive genetic variation for conformation traits in Canadian Holsteins. *J. Dairy Sci.* **73**:2214-2220.

Tempelman, R. J., and Burnside, E. B. (1990b): Additive and nonadditive genetic variation for production traits in Canadian Holsteins. *J. Dairy Sci.* **73**:2206-2213.

Thomsen, H., Reinsch, N., Xu, N., Looft, C., Grupe, S., Kühn, C., Brockmann, G. A., Schwerin, M., Leyhe-Horn, B., Hiendleder, S., Erhardt, G., Medjugorac, I., Russ, I., Förster, M., Brenig, B., Reinhardt, F., Reents, R., Blümel, J., Averdunk, G., and Kalm, E. (2001): Comparison of estimated breeding values, daughter yield deviations and de-regressed proofs within a whole genome scan for QTL. *J. Anim. Breed. Genet.* **118:**357-370.

Toro, M. A., and Varona, L. (2010): A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol.* **42**:33.

Van Tassell, C. P., Misztal, I., and 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.

VanRaden, P. M., Null, D. J., Sargolzaei, M., Wiggans, G. R., Tooker, M. E., Cole, J. B., Sonstegard, T. S., Connor, E. E., Winters, M., van Kaam, J. B. C. H. M., Valenti, A., Van Doormaal, B. J., Faust, M. A., and Doak, G. A. (2013): Genomic imputation and evaluation using high-density Holstein genotypes. *J. Dairy Sci.* **96:**668–678.

VanRaden, P. M., Tooker, M. E., and Cole, J. B. (2009): Can you believe those genomic evaluations for young bulls? *J. Dairy Sci.* **92(E-Suppl. 1):**314.

Wellmann, R., and Bennewitz, J. (2012): Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genet. Res. Camb.* **94**:21-37.

Wellmann, R., and Bennewitz, J. (2011): The contribution of dominance to the understanding of quantitative genetic variation. *Genet. Res. Camb.* **93**:139-154.

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

Summary

The objectives of this thesis were to investigate the potential benefits from genome-wide SNP genotypes at higher marker density and from female genotypes in addition to bulls' genotypes in genomic evaluation of the Fleckvieh breed. In the genomic analysis of female genotypes, the emphasis lay on the importance and on the potential use of dominance variance.

To investigate the difference in reliability of genomic predictions with medium-density (40,089; 50K) or high-density marker sets (388,951; HD) in Fleckvieh, an approximate method was developed to test differences in validation reliability for significance. The mean benefit in validation reliability of HD genotypes was 0.015 compared with real 50K genotypes and 0.028 compared with 50K samples from HD affected by imputation error and was significant for all analyzed traits. The model based reliability was, on average, 0.036 lower and the regression coefficient was 0.036 closer to the expected value with HD genotypes. Sampling error in the marker-based relationship coefficients causing overestimation of the model based reliability was smaller with HD genotypes. Inflation of the genomic predictions was reduced with HD genotypes, accordingly. Similar effects on model based reliability and inflation, but not on the validation reliability, were obtained by shrinkage estimation of the realized relationship matrix from 50K genotypes.

Variance components of nine milk production and conformation traits were estimated using yield deviations of 1996 genotyped Fleckvieh cows and estimated dominance variance ranged from 3.3% to 50.5% of the total genetic variance. Due to small dominance effect relationships between cows, predictions of individual dominance deviations were very inaccurate and including dominance in the model did not improve prediction accuracy in the cross-validation study. Additive and dominance SNP effects for milk yield and protein yield were estimated

with a BLUP model and used to calculate expectations of breeding values and total genetic values for putative offspring. Selection on total genetic value instead of breeding value would result in a larger expected total genetic superiority in progeny, i.e. 14.8% for milk yield and 27.8% for protein yield, and reduce the expected additive genetic gain by only 4.5% for milk yield and 2.6% for protein yield.

For a small real Fleckvieh pedigree consisting of 371 bulls (180 bulls thereof genotyped) and 553 cows (40 cows thereof genotyped) and extended for 2407 virtual non-genotyped daughters, genetic values were simulated and estimated with a single-step model including dominance and various reduced single-step models. Accuracies of breeding values, dominance deviations and total genetic values of genotyped cows with own performance were largest with single-step (0.726, 0.489 and 0.800, respectively) and smallest with reduced single-step when only genotyped cows were modelled (0.672, 0.470 and 0.788, respectively). For validation cows (without own performance but having a full-sister with own performance), accuracies were also largest with single-step (0.574, 0.315 and 0.609, respectively) and smallest with reduced single-step when only genotyped cows were modelled (0.481, 0.302 and 0.548, respectively). For genotyped bulls, accuracies were 0.774, 0.067 and 0.634, respectively, with single-step and 0.725-0.726, 0.066, and 0.597-0.598, respectively, with different reduced single-step models. A reduced single-step dominance model is suitable to estimate breeding values of bulls and breeding values, dominance deviations and total genetic values of cows with acceptable quality and can be an alternative when running a full single-step genomic model including dominance is not feasible.

Zusammenfassung

Ziele dieser Arbeit waren es, den möglichen Nutzen von genomweiten SNP-Genotypen mit höherer Markerdichte zu untersuchen sowie den möglichen Vorteil durch die Nutzung von weiblichen Genotypen zusätzlich zu den Genotypen von Bullen in der genomischen Zuchtwertschätzung beim Fleckvieh abzuschätzen.

Eine näherungsweise Methode wurde entwickelt um beobachtete Unterschiede in der Sicherheit der genomischen Vorhersage beim Fleckvieh zwischen Genotypen mit mittlerer (50 K; 40 089) und hoher Markerdichte (HD; 388 951) auf Signifikanz zu testen. Der mittlere Zugewinn in der Validierungssicherheit mit HD-Genotypen betrug 0,015 im Vergleich mit wirklichen 50K-Genotypen und 0,028 im Vergleich mit 50K-Stichproben aus den HD-Genotypen, die von einem Imputationsfehler betroffen waren, und war signifikant bei allen untersuchten Merkmalen. Die aus dem Modell berechnete Sicherheit war mit HD-Genotypen im Durchschnitt um 0,036 geringer und der Regressionskoeffizient um 0,036 näher am erwarteten Wert. Der Stichprobenfehler in den aus der Markerinformation errechneten Verwandtschaftskoeffizienten, der die Überschätzung der aus dem Modell berechneten Sicherheit bedingte, war mit HD-Genotypen kleiner. Damit übereinstimmend war die Streuung der genomischen Zuchtwerte mit HD-Genotypen weniger überhöht. Durch eine Manipulation („Shrinkage“) der mit 50K-Genotypen berechneten Verwandtschaftsmatrix wurden ähnliche Auswirkungen auf die aus dem Modell berechnete Sicherheit und die Streuung der genomischen Zuchtwerte, aber nicht auf die Validierungssicherheit, erreicht wie mit HD-Genotypen.

Aus den Leistungsabweichungen von 1996 genotypisierten Fleckviehkühen wurden die additiven und Dominanz-Varianzkomponenten von neun Milchleistungs- und

Exterieurmerkmalen geschätzt. Die geschätzte Dominanzvarianz betrug zwischen 3,3% und 50,5% der gesamten genetischen Varianz. Wegen der kleinen Dominanz-Verwandtschaftskoeffizienten zwischen den Kühen waren die individuellen Dominanzzuchtwerte sehr ungenau und die Aufnahme von Dominanz in das Modell verbesserte nicht die Vorhersagegenauigkeit in der Kreuzvalidierung. Additive und Dominanzeffekte von SNP für Milchmenge und Eiweißmenge wurden mit einem BLUP-Modell geschätzt und zur Berechnung von erwarteten Zuchtwerten und Genotypwerten von möglichen Nachkommen verwendet. Die Auswahl des Paarungspartners nach dem Genotypwert statt nach dem Zuchtwert würde zu einer zusätzlichen Verbesserung im Genotypwert der Nachkommen um 14,8% in der Milchmenge und um 27,8% in der Eiweißmenge führen und im Gegenzug den erwarteten additiven Zuchtfortschritt nur um 4,5% in der Milchmenge und 2,6% in der Eiweißmenge reduzieren.

Für ein kleines echtes Fleckvieh-Pedigree mit 371 Bullen (davon 180 genotypisiert) und 553 Kühen (davon 40 genotypisiert), das um 2407 nicht genotypisierte Töchter erweitert wurde, wurden genetische Werte simuliert und Zuchtwerte mit einem Single-Step-Modell inklusive Dominanz sowie verschiedenen reduzierten Single-Step-Modellen geschätzt. Die Genauigkeiten der Zuchtwerte, Dominanzabweichungen und Genotypwerte von genotypisierten Kühen mit eigener Leistung waren mit dem Single-Step-Modell am größten (0,726, 0,489 und 0,800) und am kleinsten mit einem reduzierten Single-Step-Modell, in dem nur genotypisierte Kühe enthalten waren (0,672, 0,470 und 0,788). Für die Validierungskühe (ohne Eigenleistung, aber Vollschwester mit Leistung) waren die Genauigkeiten ebenfalls mit dem Single-Step-Modell am größten (0,574, 0,315 und 0,609) und am kleinsten mit dem auf genotypisierte Kühe reduzierten Single-Step-Modell (0,481, 0,302 und 0,548). Für genotypisierte Bullen betrugen die Genauigkeiten mit dem Single-Step-Modell jeweils 0,774, 0,067 und 0,634 und für die verschiedenen reduzierten Single-Step-Modelle jeweils 0,725-

0,726, 0,066 und 0,597-0,598. Ein reduziertes Single-Step-Modell mit Dominanz eignet sich dafür, Zuchtwerte von Bullen und Zuchtwerte, Dominanzabweichungen und Genotypwerte von Kühen mit akzeptabler Qualität zu schätzen und kann als Alternative dienen, wenn ein volles Single-Step-Modell mit Dominanz aus Kapazitätsgründen nicht gerechnet werden kann.

Acknowledgements

I would like to thank

Prof. Dr. Kay-Uwe Götz, who has given me the opportunity to work in his team at the Institute of Animal Breeding in Grub and to learn a lot about animal breeding

Prof. Dr. Ruedi Fries for the opportunity to write this thesis and for numerous interesting talks

Prof. Dr. Henner Simianer for acting as co-referee of this thesis

Prof. Dr. Wilhelm Windisch for taking over the chair of the examination committee

the **German Federal Ministry of Education and Research** (Bonn, Germany) for funding this thesis within the AgroClustEr "**Synbreed** – Synergistic plant and animal breeding" (Grant no. 0315628 H)

the **Förderverein Biotechnologieforschung e.V.** (Bonn, Germany) and the **organizations contributing to the German-Austrian pool of Fleckvieh genotypes** permitting the use of genotypes

Prof. Dr. Chris-Carolin Schön for the great opportunities of training and exchange within the Synbreed project, including funding a very interesting research stay at INRA

Dr. Andrés Legarra together with **Dr. Zulma G. Vitezica** and **their colleagues at INRA in Auzeville-Tolosane** for the warm accommodation at their institute during my three-month research stay and for supervising a part of my thesis.

my colleagues at the Institute of Animal Breeding in Grub for the enjoyable teamwork. Special thanks to **Dr. Christian Edel**, **Dr. Reiner Emmerling**, **Dr. Stefan Neuner**, **Dr. Anne Haberland Pimentel**, **Dr. Eduardo Pimentel**, **Dr. Dieter Krogmeier** and **Dr. Jörg Dodenhoff** for constant motivation and sharing their expertise with me. Warm thanks to **Dr. Laura Steib** for fruitful discussions and a nice time in the office. Many thanks also to **Bernd Luntz** and the members of his team **Hubert Anzenberger**, **Wilhelm Heinrichs**, **Arnold Krämer**, **Rudolf Schnagl**, **Leonhard Schweiger**, **Stefan Schweiger**, **Heinrich Strasser (†)** and **Herbert Trager** for collecting tissue samples from cows and scoring them.

Prof. Dr. Hubert Pausch, **Dr. Michal Wysocki**, and the whole **team at the Chair of Animal Breeding** at TUM for various support and notably the genotyping of cows.

my colleagues at Bayerischer Bauernverband for constant motivation to finish this thesis

my parents as well as **all relatives and friends** who have supported and encouraged me

my beloved wife Sonja and our children Johannes, Magdalena and Elisabeth for their motivation, their understanding, their constant support and their love.

Curriculum vitae

Name:	Johann Ertl
Date of birth:	5th March 1986
Place of birth:	Bad Tölz
Nationality:	German
Family status:	married, three children

Professional experience

Since Feb 2014	**Bavarian Farmers' Association**, Munich
	Referee of Animal Husbandry and Animal Welfare
Aug 2011 – July 2014	**Bavarian State Research Centre for Agriculture – Institute of Animal Breeding**, Poing-Grub
	Research assistant
	Genomic selection in cattle within the "Synbreed" project (funded by the Federal Ministry of Education and Research)
	Ph.D. project "Investigations on genomic evaluations using high-density genotypes in Fleckvieh with consideration of dominance effects"
Feb 2013 – May 2013	**INRA – Station d'amélioration génétique des animaux**, Auzeville-Tolosane, France
	Research stay (supervisor: Dr. Andrés Legarra)

University education

Oct 2009 – July 2011	**Technical University of Munich – TUM School of Life Sciences Weihenstephan**, Freising-Weihenstephan
	Agricultural Sciences (Major field of study: Animal Sciences)
	Degree: Master of Science
	Master's Thesis: „Experimentelle Untersuchungen zum Einsatz unterschiedlichen Grundfutters (Heu bzw. Heu und Maissilage) in der Kälberaufzucht"
Oct 2006 – Oct 2009	**Technical University of Munich – TUM School of Life Sciences Weihenstephan**, Freising-Weihenstephan
	Agricultural and Horticultural Sciences
	Degree: Bachelor of Science

	Bachelor's Thesis: „Zur Energie- und Nährstoffversorgung trockenstehender und frischlaktierender Milchkühe unter besonderer Berücksichtigung der Mengenelemente Calcium, Phosphor und Magnesium“

Military service

Oct 2005 – June 2006	**Concert Band of the German Armed Forces**, Siegburg

School education

Sep 2003 – June 2005	**Gabriel-von-Seidl-Gymnasium Bad Tölz**
	General qualification for university entrance
Sep 2002 – Aug 2003	**Asam-Gymnasium München**
Sep 1998 – Aug 2002	**Staatliche Realschule Bad Tölz**
	General certificate of secondary education
Sep 1996 – Aug 1998	**Hauptschule Reichersbeuern**
Sep 1992 – Aug 1996	**Grundschule Gaißach**

Lebenslauf

Name:	Johann Ertl
Geburtsdatum:	5. März 1986
Geburtsort:	Bad Tölz
Staatsangehörigkeit:	deutsch
Familienstatus:	verheiratet, drei Kinder

Beruf

seit Febr. 2014	**Bayerischer Bauernverband**, München Referent für Tierhaltung und Tierschutz
Aug. 2011 – Juli 2014	**Bayerische Landesanstalt für Landwirtschaft – Institut für Tierzucht**, Poing-Grub Wissenschaftlicher Mitarbeiter Genomische Selektion beim Rind im Rahmen des BMBF-Projekts „Synbreed“ Promotion (Dr. agr.) „Investigations on genomic evaluations using high-density genotypes in Fleckvieh with consideration of dominance effects”
Febr. 2013 – Mai 2013	**INRA – Station d’amélioration génétique des animaux**, Auzeville-Tolosane, Frankreich Forschungsaufenthalt unter der Betreuung von Dr. Andrés Legarra

Studium

Okt. 2009 – Juli 2011	**Technische Universität München – Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt**, Freising-Weihenstephan Masterstudiengang Agrarwissenschaften (Schwerpunkt: Agrobiowissenschaften Tier) Abschluss: Master of Science Masterarbeit: „Experimentelle Untersuchungen zum Einsatz unterschiedlichen Grundfutters (Heu bzw. Heu und Maissilage) in der Kälberaufzucht“

Okt. 2006 – Okt. 2009	**Technische Universität München – Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt**, Freising-Weihenstephan
	Bachelorstudiengang Agrarwissenschaften und Gartenbauwissenschaften
	Abschluss: Bachelor of Science
	Bachelorarbeit: „Zur Energie- und Nährstoffversorgung trockenstehender und frischlaktierender Milchkühe unter besonderer Berücksichtigung der Mengenelemente Calcium, Phosphor und Magnesium“

Wehrdienst

Okt. 2005 – Juni 2006	**Musikkorps der Bundeswehr**, Siegburg

Schule

Sept. 2003 – Juni 2005	**Gabriel-von-Seidl-Gymnasium Bad Tölz**
	Abschluss: Allgemeine Hochschulreife
Sept. 2002 – Aug. 2003	**Asam-Gymnasium München**
Sept. 1998 – Aug. 2002	**Staatliche Realschule Bad Tölz**
	Abschluss: Mittlere Reife
Sept. 1996 – Aug. 1998	**Hauptschule Reichersbeuern**
Sept. 1992 – Aug. 1996	**Grundschule Gaißach**