

Bayesian models with dominance effects for genomic evaluation of quantitative traits

ROBIN WELLMANN* AND JÖRN BENNEWITZ

Department of Animal Husbandry and Animal Breeding, University of Hohenheim, D-70599 Stuttgart, Germany

(Received 11 July 2011; revised 2 November 2011 and 7 December 2011; accepted 12 December 2011)

Summary

Genomic selection refers to the use of dense, genome-wide markers for the prediction of breeding values (BV) and subsequent selection of breeding individuals. It has become a standard tool in livestock and plant breeding for accelerating genetic gain. The core of genomic selection is the prediction of a large number of marker effects from a limited number of observations. Various Bayesian methods that successfully cope with this challenge are known. Until now, the main research emphasis has been on additive genetic effects. Dominance coefficients of quantitative trait loci (QTLs), however, can also be large, even if dominance variance and inbreeding depression are relatively small. Considering dominance might contribute to the accuracy of genomic selection and serve as a guide for choosing mating pairs with good combining abilities. A general hierarchical Bayesian model for genomic selection that can realistically account for dominance is introduced. Several submodels are proposed and compared with respect to their ability to predict genomic BV, dominance deviations and genotypic values (GV) by stochastic simulation. These submodels differ in the way the dependency between additive and dominance effects is modelled. Depending on the marker panel, the inclusion of dominance effects increased the accuracy of GV by about 17% and the accuracy of genomic BV by 2% in the offspring. Furthermore, it slowed down the decrease of the accuracies in subsequent generations. It was possible to obtain accurate estimates of GV, which enables mate selection programmes.

1. Introduction

Genomic selection refers to the use of genome wide and dense markers for the prediction of breeding values (BV) and subsequent selection of individuals (Meuwissen *et al.*, 2001). Results from simulations and from real validation studies revealed that the accuracy of predicted BV of individuals without own records or without progeny records can be remarkably high (Meuwissen *et al.*, 2001; Calus *et al.*, 2008; Luan *et al.*, 2009; Habier *et al.*, 2010; Hayes *et al.*, 2009), which offers the opportunity to accurately select individuals at an early stage of their life as parents of the next generation. This technique has become a standard tool in dairy cattle breeding (Hayes *et al.*, 2009), and its implementation in other livestock species is foreseen, e.g. in poultry breeding (Wolc *et al.*,

2011), in pig breeding, and also in plant breeding (Piepho, 2009; Heffner *et al.*, 2009).

The core of genomic selection is the prediction of BV from massive marker data. Most influential methods were already proposed by Meuwissen *et al.* (2001). These are genomic best linear unbiased prediction (G-BLUP), BayesA and BayesB, which differ in their assumptions about the distribution of marker effects. G-BLUP assumes a normal distribution of marker effects, whereas BayesA assumes a more heavy tailed Student *t*-distribution (Gianola *et al.*, 2009). Since the Student *t*-distribution approximates the normal distribution when the degree of freedom ν increases, G-BLUP can be considered as a limiting case of BayesA. Although these methods perform well in simulation studies and applications, many markers are not needed to capture the effects of quantitative trait loci (QTLs) because they are either redundant or not in linkage disequilibrium (LD) with a QTL. This is accounted for by BayesB, which assumes that marker effects come from a Student *t*-distribution with a certain probability and take the

* Corresponding author: Department of Animal Husbandry and Animal Breeding, University of Hohenheim, D-70599 Stuttgart, Germany. Tel: +711 459 23008. Fax: +711 459 23101. e-mail: r.wellmann@uni-hohenheim.de

value 0 otherwise. A marker effect comes from the Student t -distribution if the marker is needed to capture a QTL effect. BayesB contains BayesA as the special case where the prior probability of a marker to be needed equals one. An alternative is the stochastic search variable selection (SSVS) model (George & McCulloch, 1993) that was introduced to QTL mapping by Yi *et al.* (2003) and was applied by Meuwissen (2009) to genomic selection. It assumes that effects of single nucleotide polymorphisms (SNPs) come from a mixture of two normal distributions with different variances, and SNPs with negligible effects come from the distribution with little variance. Bayesian SSVS is similar, but assumes prior distributions for the variances of SNP effects. As a consequence, SNP effects are from a mixture of two Student t -distributions with different variances. Bayesian SSVS is also known as BayesC (Verbyla *et al.*, 2009, 2010). It contains BayesB as the limiting case, in which the variance of one mixing distribution approaches zero. Further Bayesian models that have been proposed are the Bayesian Lasso that assumes Laplace priors (Park & Casella, 2008; Legarra *et al.*, 2011), and a Bayesian model that uses identity by descent (IBD) probabilities (Meuwissen & Goddard, 2004). However, Calus *et al.* (2008) found that a marker effect-based model that does not make use of IBD probabilities provides similar accuracies for high-density marker panels.

Nearly all published models only include additive effects (Calus, 2010) and little has been done to generalize these models for the prediction of genomic BV to explicitly account for dominance. One reason is probably that estimated BV or deregressed BV obtained from routine evaluations (Garrick *et al.*, 2009) are used as observations in most applications of genomic selection, so dominance deviations of individuals are absent in the data. However, if individual phenotypes are available, the inclusion of dominance effects could not only increase the accuracy of genomic selection, but predicted dominance effects could also be used to find mating pairs with good combining abilities by recovering inbreeding depression and utilizing possible overdominance. Toro & Varona (2010) demonstrated that the inclusion of dominance into Bayesian models can indeed increase the accuracy of genomic selection. They assumed independent additive and dominance effects which were in accordance with their simulation protocol. Reproducing kernel Hilbert space regression (RKHS regression) is proposed as an alternative method for the estimation of genotypic values (GV), especially if non-additive effects such as dominance or epistasis are included in the phenotypes (Gianola *et al.*, 2006). It assumes that for each pair of genotypes, the covariance of its GV is defined by a covariance function (de los Campos *et al.*, 2009). The covariance function is the reproducing

kernel of a Hilbert space. RKHS regression yields an estimate of the function g that maps genotypes to GV. This estimate is optimal in a well-defined sense among all functions that belong to the Hilbert space (Gianola & van Kaam, 2008). RKHS regression does not make assumptions about the mechanism that makes GV random since any symmetric and finitely positive-semidefinite function K is the reproducing kernel of a Hilbert Space (Shawe-Taylor & Cristianini, 2004).

The aim of this paper is to introduce Bayesian linear regression models for genomic evaluation of quantitative traits that account for dominance effects of QTLs. These models are generalizations of Bayesian SSVS that includes only additive effects. The proposed models differ in the way the dependency between additive effects, dominance effects and allele frequencies is modelled. Plausible informative priors are chosen which are in agreement with the genetic architectures of quantitative traits suggested in the literature. We call these generalizations the BayesD models, where D stands for dominance. The paper is organized as follows. We start with a brief literature review on the dependencies between additive effects, dominance effects and allele frequencies in real populations and discuss how Bayesian models could account for these dependencies. Statistical models that realistically account for these dependencies are defined thereafter. The joint posterior distribution of unknown model parameters is given and used to derive a Markov chain for the prediction of these parameters. Moments of the random effects are derived and are used for the calculation of the hyperparameters. We also demonstrate that kernels for RKHS regression can be derived from the assumptions of our models. The resulting RKHS estimate is the BLUP for the underlying Bayesian model. The proposed methods are applied to a simulated population. The models are compared with respect to their ability to predict dominance deviations, GV and BV.

2. Theory

(i) Possibilities to model the genetics of dominance

Since the aim of this paper is to account for dominance in genomic evaluations of quantitative traits, we present different possibilities to model the joint distribution of additive and dominance effects of the markers and show how these models are able to account for the genetic architectures that are suggested in the literature. The mathematical definitions of our models are given in the next section. Note that the proofs of all numbered equations can be found in the electronic appendix. Table 1 summarizes symbols used in this paper.

Consider biallelic QTLs with alleles 0 and 1. In this section, a_j and d_j denote the additive and dominance

Table 1. Table of symbols

\tilde{a}_j, \tilde{d}_j	Additive effect and dominance effect of marker j
a_j, d_j, δ_j	Putative additive effect, dominance effect and dominance coefficient of marker j (or additive effect, dominance effect and dominance coefficient of QTL j)
$\tilde{\theta}_j$	Effect of marker j , where $\tilde{\theta}_j = (\tilde{a}_j, \tilde{d}_j)$ or $\tilde{\theta}_j = \tilde{a}_j$
θ_j	Putative effect of marker j , where $\theta_j = (a_j, d_j)$ or $\theta_j = a_j$
v_j	$v_j = 1$ if the additive effect of marker j is positive and $v_j = 0$ otherwise
γ_j	$\gamma_j = 1$ if marker j is needed and $\gamma_j = 0$ otherwise
κ_j	Conditional scale of the effect of marker j , given γ_j
τ_j^2	Forces the distribution of the additive effect to have the desired shape
β	Vector of fixed effects
u	Vector of normally distributed random effects with covariance matrix Σ
σ^2	Variance of the errors
$\text{Pos}_j(d_j)$	Probability that a_j is positive, given d_j
$\mu_d(a_j)$	Conditional expectation of d_j , given $ a_j $
$\sigma_d^2(a_j , \tau_j^2)$	Conditional variance of d_j , given $ a_j $ and τ_j^2
$\mu_\Delta(\frac{ a_j }{s})$	Conditional expectation of the dominance coefficient, given $ a_j $
σ_Δ^2	Conditional variance of the dominance coefficient, given $ a_j $
w_j	The parameter satisfies $\frac{1-w_j}{2} = P(\text{sign}(\tilde{a}_j) = \text{sign}(\tilde{d}_j))$
$K(x, \tilde{x})$	For RKHS regression, $K(x, \tilde{x})$ denotes the kernel
$K(x)$	For BayesD, $K(x)$ denotes the function defined after equation (3)

effect of QTL j . The 1-allele has frequency p_j and the 0-allele has frequency $q_j = 1 - p_j$. The three possible genotypes 00, 01 and 11 have GV 0, $a_j + d_j$ and $2a_j$. Dominance effects can cause inbreeding depression. Inbreeding depression \mathcal{I} of a trait is the expected decrease of the phenotypic value when the inbreeding coefficient increases from 0 to 1. $\mathcal{I} > 0$ requires that the expectation of dominance effects is positive, provided that allelic effects are assumed to be random realisations from some distribution. This means that the effect of the heterozygous genotype is usually above the average effect of the two homozygous genotypes. Analogously, for a trait with outbreeding depression we have $\mathcal{I} < 0$ and the expectation of dominance effects is negative. Dominance effects also cause dominance variance. Dominance variances of up to 50% of the additive variance have been reported for traits in dairy cattle (van Tassell *et al.*, 2000). Substantially larger dominance variances are unlikely to occur due to the U-shaped distribution of

allele frequencies (Hill *et al.*, 2008). Dominance variances have also been estimated for some traits in beef cattle (Duangjinda *et al.*, 2001) and pigs (Serenius *et al.*, 2006). In general, estimates of dominance variance must be interpreted with caution because they could be confounded e.g. by maternal effects or environmental covariance of full sibs. Moreover, an accurate estimation of dominance variance requires at least 20 times as much data as the estimation of additive variance (Misztal, 1997).

Figure 1 gives an overview of different possibilities to model the joint distribution of additive effects and dominance effects. The figure shows samples drawn from the proposed joint prior distribution of additive and dominance effects of markers with allele frequency $q_j = p_j = 0.5$ for different scenarios, where additive effects are assumed to be Student t -distributed with $\nu = 2.5$ degrees of freedom. Note that the models presented in the next section are more general than this illustrative example.

The simplest possibility is to assume independence of additive and dominance effects, where dominance effects have the same distribution as additive effects. It can be seen in Fig. 1 that this prior assumes that for QTL with large dominance effect the additive effect is likely small in magnitude. This is because the t -distribution is heavy tailed, so it allows ‘large’ effects to occur. But large effects are rare events, so under independence, the probability is small that for a large dominance effect, the additive effect is also large. Thus, dominance is mainly due to overdominant alleles. But this is out of line with standard genetics theory (Kacser & Burns, 1981; Charlesworth & Willis, 2009) because theory predicts that recessive deleterious alleles rather than overdominant alleles are the primary cause of inbreeding depression. This model is therefore not further considered.

A second possibility (BayesD1) is to assume that the joint distribution of additive and dominance effects is elliptical with Student t -distributed margins. Roughly speaking, this model assumes that additive and dominance effects of a QTL are of the same magnitude. Figure 1 shows that overdominance is quite common under this assumption. The importance of overdominance cannot yet be accurately quantified (Charlesworth & Willis, 2009), so this model may be appropriate for populations in which many overdominant QTLs are expected to exist. For most applications, however, the prior possibly allows for too much overdominance.

Therefore, we consider as a third possibility (BayesD2) the case that absolute additive effects and dominance coefficients are independent. The dominance coefficient $\delta_j = \frac{d_j}{|a_j|}$ of a QTL is the ratio between the dominance effect and the absolute additive effect. If $\delta_j > 1$ or $\delta_j < -1$ then the QTL is overdominant (or underdominant). If $\delta_j = 1$ or $\delta_j = -1$ then the QTL

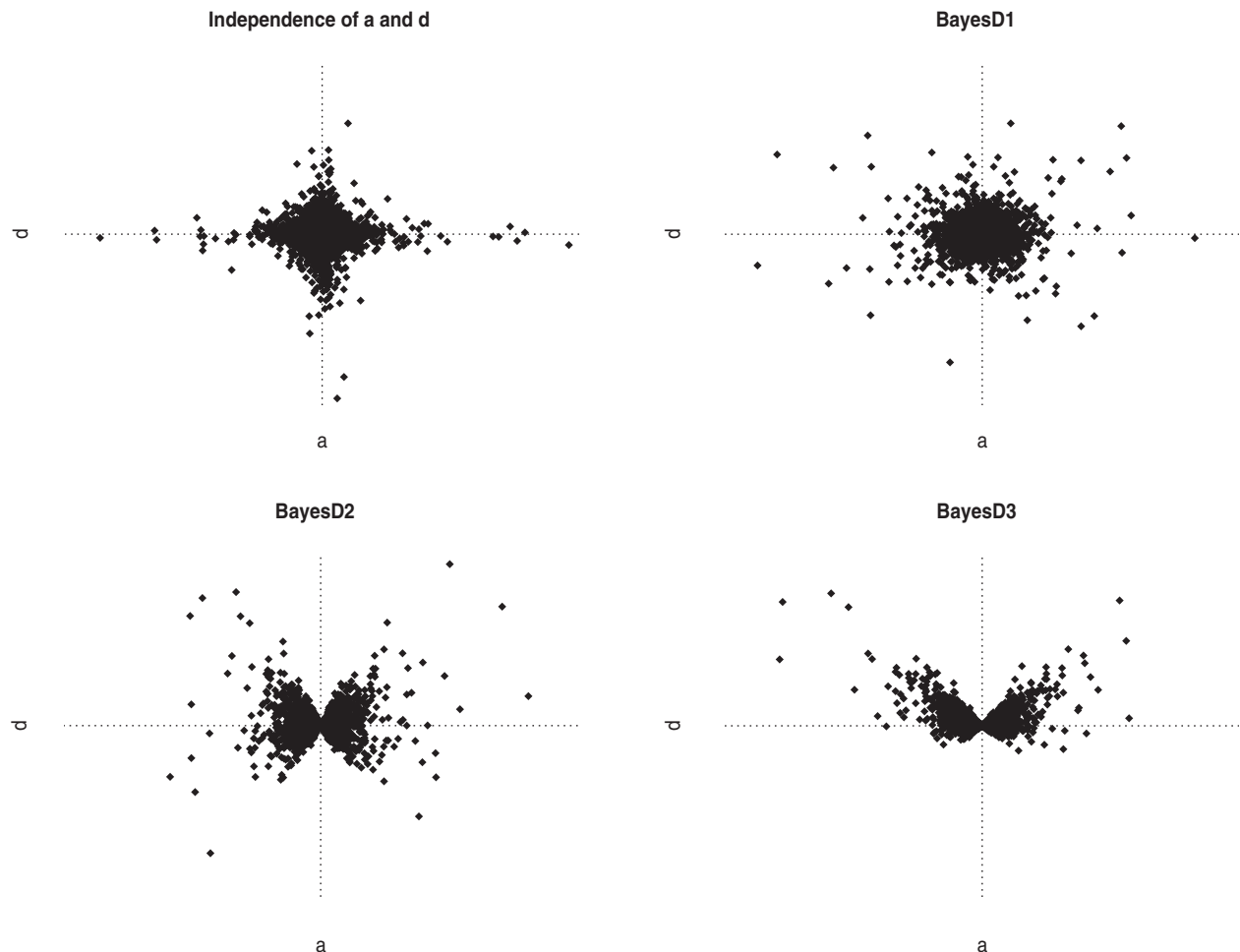


Fig. 1. Samples drawn from the joint prior distribution of additive and dominance effects of markers with allele frequency $q_j = p_j = 0.5$, where additive effects are Student t -distributed with $\nu = 2.5$ degrees of freedom. The distribution specifications of BayesD1–BayesD3 are given in Section 2(ii).

is dominant or recessive. For $\delta_j = 0$, the QTL is completely additive and a QTL with $-1 < \delta_j < 0$ or $0 < \delta_j < 1$ shows incomplete dominance or recessivity. As demonstrated in Fig. 1, the model is able to make the prior assumption that the probability is small that a dominance effect is much larger in magnitude than the additive effect, so overdominance is a rare but not negligible event. This is in accordance with Bennewitz & Meuwissen (2010) who used results from QTL experiments and found that dominance coefficients were normally distributed with small but positive mean. Here, we also assume a normal distribution for the dominance coefficient.

Caballero & Keightley (1994) showed that the dominance effect depends on the additive effect such that QTLs with large absolute additive effect likely have large dominance coefficients. Moreover, they concluded tentatively that mutations of small effect show highly variable degrees of dominance with an average dominance coefficient close to zero. Deleterious alleles with large effect are usually close to recessivity due to the hyperbolic relationship between enzyme

activity and flux (Kacser & Burns, 1981). The prior distribution of BayesD2 cannot account for this. Therefore, we consider as a fourth possibility (BayesD3) the case that markers with large absolute additive effects tend to be associated with large dominance coefficients. That is, for additive effects of markers that are small in magnitude, the average dominance coefficient is close to 0, and for additive effects of markers that are large in magnitude the average dominance coefficient is close to 1. This is also in accordance with García-Dorado *et al.* (1999), who suggested an average dominance coefficient of 0.8 for non-severe deleterious mutations and of 0.94–0.98 for new lethal mutants.

In addition to the dependency between dominance effects and absolute additive effects, the dependency between the allele frequencies and the signs of the allelic effects could also be considered. They are dependent because selection has likely shifted allele frequencies away from values for which the expected allele-frequency change per one generation of selection is high. From this argument, a characterization

of the dependency can be derived as follows: since selection for (or against) a recessive allele is inefficient if its frequency is low (Falconer & Mackay, 1996; Fig. 2.2), recessive alleles likely have low frequencies. Recessiveness of the 1-allele implies that $d_j < 0$ if $a_j > 0$ and $d_j > 0$ if $a_j < 0$. Since $p_j < 0.5$ is likely to hold, we have $\text{sign}(a_j) = -\text{sign}((q_j - p_j)d_j)$ in both cases. Now consider selection for (or against) a dominant allele. If the 1-allele is dominant, then the other allele is recessive. From the above argument it follows that the other allele likely has a low frequency ($q_j < 0.5$). It follows that $p_j > 0.5$. Thus, dominant alleles likely have high frequencies. Dominance of the 1-allele implies that $d_j > 0$ if $a_j > 0$ and $d_j < 0$ if $a_j < 0$. Since $p_j > 0.5$ is likely to hold, we have again $\text{sign}(a_j) = -\text{sign}((q_j - p_j)d_j)$ with high probability. The models BayesD2 and BayesD3 account for this. Alternatively, the equation could be derived from the contribution $2p_jq_j(a_j + (q_j - p_j)d_j)^2$ of the QTL to the additive variance because it is unlikely that a QTL has a frequency p_j for which the contribution of the QTL to the additive variance is large. The contribution is small if $a_j \approx -(q_j - p_j)d_j$. This also shows that $\text{sign}(a_j) = \text{sign}(-(q_j - p_j)d_j)$ likely holds for the majority of the alleles. Wellmann & Bennewitz (2011) obtained plausible estimates for parameters of the joint distribution of additive effects and dominance effects for the trait productive life (PL) in dairy cattle only if the model accounts for this. From this argument it also follows that alleles with large effect are not likely to have intermediate frequencies, except for overdominant alleles (Falconer & Mackay, 1996; p. 27ff.) and some pleiotropic alleles (e.g. *DGAT1* in Holstein cattle, see Grisart *et al.*, 2004).

(ii) *The linear regression model*

In this section, the general regression model and various submodels are defined. The submodels differ in the joint prior distributions of the additive and dominance effects of the markers as motivated in the previous section. We consider a linear model of the form

$$y = X\beta + Z_A\tilde{a} + Z_D\tilde{d} + Zu + E,$$

where the vector y consists of n observations, β is a vector of fixed effects and would usually include the intercept of the model. The matrix X has full column rank. The vector $u \sim \mathcal{N}_p(0, \Sigma)$ is normally distributed and independent from the marker effects with covariance matrix Σ and Z is a known $n \times p$ matrix. The vector $\tilde{a} = (\tilde{a}_1, \dots, \tilde{a}_M)^T$ contains the additive effects of the markers and $\tilde{d} = (\tilde{d}_1, \dots, \tilde{d}_M)^T$ consists of the dominance effects of the markers, where M is the number of markers. The errors E_1, \dots, E_n are independent and normally distributed with variance σ^2 , i.e. $E|\sigma^2 \sim \mathcal{N}_n(0, \sigma^2 I)$.

Additive effects and dominance effects of the markers are random variables. Randomness of allelic effects is most easily understood by imagining that the population is a random sample from all hypothetical populations to which the method could be applied, and the trait is randomly chosen among all traits with similar genetic architecture that could be analysed. That is, for each hypothetical population the effect of a marker at a particular position in the genome is a random realization from some distribution. We assume biallelic markers with alleles 0 and 1. Take $\tilde{\theta}_j$ to be the effect of marker j . We have $\tilde{\theta}_j = (\tilde{a}_j, \tilde{d}_j)$, if dominance effects are included in the model and $\tilde{\theta}_j = \tilde{a}_j$ otherwise. It is assumed that the distribution of $\tilde{\theta}_j$ is a mixture of two distributions that differ only by a scaling factor ε , so conditionally on a Bernoulli distributed indicator variable $\gamma_j \sim \mathcal{B}(1, p_{LD})$ we can write

$$\tilde{\theta}_j | \gamma_j \sim \gamma_j \mathcal{F} + (1 - \gamma_j) \varepsilon \mathcal{F},$$

where the parameter $0 \leq \varepsilon \ll 1$ is chosen small and the distribution \mathcal{F} is specified below. Thus, if $\gamma_j = 1$, then the marker effect comes from the distribution with large variance. This occurs with probability $p_{LD} = E(\gamma_j)$. Markers j with $\gamma_j = 1$ are those needed to capture the effects of QTLs. The *a priori* expected number of markers that are needed is Mp_{LD} .

We can write $\tilde{\theta}_j = \kappa_j \theta_j$, where $\kappa_j = (1 - \gamma_j)\varepsilon + \gamma_j$ and $\theta_j \sim \mathcal{F}$ is called the ‘putative’ effect of marker j . In the remaining part of the paper, $(a_j, d_j) = \theta_j$, (or $a_j = \theta_j$) denote the putative additive effect and the putative dominance effect of marker j . Thus, if marker j is needed to capture a QTL effect ($\gamma_j = 1$), then $\tilde{a}_j = a_j$ and $\tilde{d}_j = d_j$. Otherwise, the putative marker effect is regressed towards zero and we have $\tilde{a}_j = \varepsilon a_j$ and $\tilde{d}_j = \varepsilon d_j$.

The distribution of the putative marker effect θ_j is specified next. It is the distribution of the effects of markers needed to capture the effects of QTLs. As mentioned in the previous section, the sign of the additive effect $\text{sign}(a_j)$, the absolute additive effect $|a_j|$ and the dominance effect d_j depend on each other in a complicated way. The model assumes that the absolute additive effect $|a_j|$ has a folded t -distribution, since conditionally on an inverse chi-square distributed parameter τ_j^2 it has a half-normal distribution. The prior is therefore

$$\tau_j^2 | \gamma_j \sim \text{Inv} - \chi^2(v, s^2),$$

$$|a_j| | \tau_j^2, \gamma_j \sim |\mathcal{N}(0, \tau_j^2)|.$$

The putative dominance effect d_j may depend on the absolute additive effect $|a_j|$ in order to allow for a prior for which overdominance is a rare event. It may also depend on τ_j^2 in order to account for the fact that additive and dominance effects are similar in magnitude. Conditionally on $|a_j|$ and τ_j^2 , the dominance effect is normally distributed with mean $\mu_d(|a_j|) = E(d_j | |a_j|)$ and variance $\sigma_d^2(|a_j|, \tau_j^2) = \text{Var}(d_j | |a_j|, \tau_j^2)$. That

is, we define the prior distribution of the putative dominance effects as

$$d_j| |a_j|, \tau_j^2, \gamma_j \sim \mathcal{N}(\mu_d(|a_j|), \sigma_d^2(|a_j|, \tau_j^2)).$$

The probability $\text{pos}_j(d_j) = P(a_j > 0|d_j)$ that the additive effect a_j is positive, given d_j , may be different for each marker and depends on the frequency of the 1-allele and thus on the coding of the marker. Let $v_j = 1$ if $a_j > 0$ and $v_j = 0$ otherwise. Thus, $\text{sign}(a_j) = 2v_j - 1$ and v_j has the prior

$$v_j|d_j, |a_j|, \tau_j^2, \gamma_j \sim \mathcal{B}(1, \text{pos}_j(d_j)).$$

As motivated in the previous section, the sign of the additive effect depends on the dominance effect and the allele frequency such that

$$\text{sign}(a_j) = -\text{sign}(q_j - p_j)\text{sign}(d_j),$$

likely holds for the majority of QTL. It is convenient to assume that this equation also holds with high probability for the marker effects. Therefore, we assume that the probability that a_j is positive, given d_j , equals

$$P(a_j > 0|d_j) = \text{pos}_j(d_j) = \frac{1 - w_j \text{sign}(d_j)}{2},$$

where $w_j \in (-1, 1)$ may depend on the frequency of marker j . For example, we may choose $w_j = 0$, $w_j = 0.9 \text{sign}(q_j - p_j)$ or $w_j = q_j - p_j$. This parameter has the interpretation

$$w_j = 1 - 2P(\text{sign}(a_j) = \text{sign}(d_j)). \tag{1}$$

For the variance of the errors, we use the prior

$$\sigma^2 \sim \text{Inv} - \chi^2(v^*, s^{*2}) \text{ or } p(\sigma^2) \propto 1 \text{ or } p(\sigma^2) \propto \frac{1}{\sigma^2}.$$

For the improper uniform prior let $v^* = -2$, $s^{*2} = 0$, and for the improper prior $p(\sigma^2) \propto \frac{1}{\sigma^2}$ let $v^* = 0$. Random effects are *a priori* independent, i.e. $u, \sigma^2, (\tilde{a}_1, \tilde{d}_1, \gamma_1, \tau_1^2), \dots, (\tilde{a}_M, \tilde{d}_M, \gamma_M, \tau_M^2)$ are independent.

We consider the following submodels. The first submodel (BayesD0) contains only additive effects, so $\mu_d(|a_j|) = 0$, $\sigma_d^2(|a_j|, \tau_j^2) = 0$, and $w_j = 0$. Since $w_j = 0$, the putative additive effects a_j have a t distribution with v degrees of freedom, mean 0 and variance $\text{var}(a_j) = E(\tau_j^2) = s^2 \frac{v}{v-2}$ if $v > 2$. For $v \leq 2$, the variance does not exist. If v is large, then the putative additive effects are approximately normally distributed. Note that BayesC of Verbyla *et al.* (2010) appears as the special case where $\varepsilon > 0$. Most of the Bayesian models mentioned in the introduction are also special cases or limiting cases of BayesD0.

In model BayesD1, additive effects and dominance effects are conditionally independent given τ_j^2 with $w_j = 0$, $\mu_d(|a_j|) = \mu_D$ and $\sigma_d^2(|a_j|, \tau_j^2) = s_D^2 \tau_j^2$, where $s_D > 0$. As a consequence, the distribution of the putative marker effect $\theta_j = (a_j, d_j)$ is elliptical.

In model BayesD2, absolute additive effects $|a_j|$ and dominance coefficients $\delta_j = \frac{d_j}{|a_j|} \sim \mathcal{N}(\mu_\Delta, \sigma_\Delta^2)$ are independent. Thus, $\mu_d(|a_j|) = |a_j| \mu_\Delta$ and $\sigma_d^2(|a_j|, \tau_j^2) = a_j^2 \sigma_\Delta^2$.

In model BayesD3, additive effects and dominance coefficients are dependent such that large additive effects are associated with large dominance coefficients. More precisely, we assume $\delta_j| |a_j| \sim \mathcal{N}(\mu_\Delta \left(\frac{|a_j|}{s}\right), \sigma_\Delta^2)$, where $\mu_\Delta(x) = \frac{x}{s_\Delta + x}$ with $s_\Delta > 0$. Note that the function μ_Δ increases with $\mu_\Delta(0) = 0$ and $\lim_{x \rightarrow \infty} \mu_\Delta(x) = 1$, so dominance coefficients of marker effects that are small in magnitude are centred at zero and dominance coefficients of marker effects that are large in magnitude are centred at one. Thus, $\mu_d(|a_j|) = |a_j| \mu_\Delta \left(\frac{|a_j|}{s}\right)$, and $\sigma_d^2(|a_j|, \tau_j^2) = a_j^2 \sigma_\Delta^2$.

(iii) *The joint posterior distribution and the Markov chain*

In this section, we present the joint posterior distribution for construction of the Markov chain. Since the Markov chain cannot be used for $\varepsilon = 0$, we assume $\varepsilon > 0$ in this section. In applications, the parameter ε would usually be chosen as small as possible in order to approximate a BayesB-type model. Alternatively, it could be chosen such that the accuracies of predicted BV and GV are maximized.

Let $\xi = (\beta, u, \sigma^2)$. We have $p(\xi) \propto p(u)p(\sigma^2)$ because the assumption that β is a fixed effect is equivalent to the assumption that it has the flat prior $p(\beta) \propto 1$. The joint posterior distribution is

$$p(\tilde{\theta}, \xi, \gamma, \tau^2|y) \propto p(y|\tilde{\theta}, \xi)p(u)p(\sigma^2) \times \prod_{j=1}^M p(\tilde{\theta}_j|\tau_j^2, \gamma_j)p(\tau_j^2)p(\gamma_j),$$

where the likelihood function is

$$p(y|\tilde{\theta}, \xi) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{E^T E}{2\sigma^2}\right),$$

with $E = y - X\beta - Zu - Z_A \tilde{a} - Z_D \tilde{d}$. An explicit representation of the conditional prior distribution $p(\tilde{\theta}_j|\tau_j^2, \gamma_j)$ of the marker effects is needed in order to derive a Markov chain that uses this parameterization. We have

$$p(\tilde{\theta}_j|\tau_j^2, \gamma_j) = \frac{2g_j(\tilde{a}_j, \tilde{d}_j)}{\kappa_j \sqrt{2\pi\tau_j^2}} \exp\left(-\frac{\tilde{a}_j^2}{2\kappa_j^2 \tau_j^2}\right) \psi_{\kappa_j, \tau_j^2}(\tilde{a}_j, \tilde{d}_j), \tag{2}$$

where $\kappa_j = (1 - \gamma_j)\varepsilon + \gamma_j$,

$$g_j(\tilde{a}_j, \tilde{d}_j) = \begin{cases} \frac{1}{2}, & \text{for BayesD0 - BayesD1,} \\ \frac{1 - w_j \text{sign}(\tilde{a}_j) \text{sign}(\tilde{d}_j)}{2}, & \text{for BayesD2 - BayesD3,} \end{cases}$$

and

$$\psi_{\kappa_j, \tau_j^2}(\tilde{a}_j, \tilde{d}_j) = \begin{cases} 1, & \text{for BayesD0,} \\ \frac{1}{\kappa_j \sqrt{2\pi\sigma_d^2}} \exp\left(-\frac{\left(\frac{\tilde{d}_j - \mu_d}{\kappa_j}\right)^2}{2\sigma_d^2}\right), & \text{for BayesD1} \\ & \text{-BayesD3.} \end{cases}$$

Here, we used the abbreviation

$$\sigma_d^2 = \sigma_d^2 \left(\frac{|\tilde{a}_j|}{\kappa_j}, \tau_j^2 \right), \quad \mu_d = \mu_d \left(\frac{|\tilde{a}_j|}{\kappa_j} \right).$$

The Markov chain for the prediction of the model parameters is generated by Gibbs sampling with a Metropolis–Hastings step. The algorithm is described in Appendix A. Starting with initial values, the parameters are sampled from their full conditional posterior distributions. The lengthy but straightforward proofs of the full conditional posterior distributions are given in the electronic appendix.

(iv) Moments of the random effects

Moments of the random effects given in this section are needed for the calculation of hyper-parameters. They are also needed for the calculation of the RKHS-kernel which is defined in the next section. For the second moment of a_j to exist, we assume $v > 2$. Since the putative marker effect θ_j and γ_j are independent, we have $E(\tilde{a}_j^k) = E(\kappa_j^k)E(a_j^k)$ and $E(\tilde{d}_j^k) = E(\kappa_j^k)E(d_j^k)$ for $k \in \{1, 2\}$. Moreover, we have $E(\tilde{a}_j \tilde{d}_j) = E(\kappa_j^2)E(a_j d_j)$ and $E(|\tilde{a}_j|) = E(\kappa_j)E(|a_j|)$, where $E(\kappa_j^k) = (1 - p_{LD})\epsilon^k + p_{LD}$. These formulae depend on moments of the putative marker effects. They can be calculated as

$$E(a_j) = -w_j E\left(|a_j| \left(1 - 2\phi\left(\frac{-\mu_d(|a_j|)}{\sigma_d(|a_j|, \tau_j^2)}\right)\right)\right), \quad (3)$$

$$E(|a_j|) = \lambda \sqrt{E(a_j^2)},$$

$$E(a_j^2) = s^2 \frac{v}{v-2},$$

$$E(d_j) = E(\mu_d(|a_j|)),$$

$$E(d_j^2) = E(\sigma_d^2(|a_j|, \tau_j^2)) + E(\mu_d(|a_j|)^2),$$

$$E(a_j d_j) = -w_j E(|a_j| |d_j|),$$

$$E(|a_j| |d_j|) = E\left(|a_j| \mu_d(|a_j|) K\left(\frac{\sigma_d(|a_j|, \tau_j^2)}{\mu_d(|a_j|)}\right)\right),$$

where (Psarakis & Panaretos, 1990):

$$\lambda = 2\sqrt{\frac{v-2}{\pi}} \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})(v-1)},$$

$$K(x) = x\sqrt{\frac{2}{\pi}} \exp\left(-\frac{1}{2x^2}\right) + \left(1 - 2\phi\left(-\frac{1}{x}\right)\right),$$

ϕ is the cumulative distribution function of a standard normal distribution, and Γ is the Gamma function. The formulae can be further simplified for the considered submodels. Simplified formulae are given in Table 2, where $t \sim t_v$ has a t -distribution with v degrees of freedom.

(v) Prediction of genotypic values

Take $\Omega = \{0, 1, 2\}^M$ to be the set of all possible multilocus genotypes at the M markers. For $x \in \Omega$, x_j is the number of 1-alleles at a particular marker j . According to the regression model described above, an individual with genotype x is assumed to have GV

$$g_{GV}(x) = \sum_{j=1}^M (\tilde{a}_j + (2 - x_j)\tilde{d}_j)x_j,$$

provided that Z_A is the gene content matrix with entries 0, 1, 2 and Z_D is the indicator matrix for heterozygosity. That is, the GV is the sum of all additive effects and dominance effects that are carried by the individual. According to Falconer & Mackay (1996), the GV can be partitioned into a BV, a dominance deviation, and a contribution to the overall mean that is equal for all individuals. The BV is

$$g_{BV}(x) = \sum_{j=1}^M (\tilde{a}_j + (q_j - p_j)\tilde{d}_j)(x_j - 2p_j),$$

and dominance deviation is

$$g_{DV}(x) = \sum_{j=1}^M -\tilde{d}_j x_j (x_j - 1 - 2p_j) - 2p_j^2 \tilde{d}_j.$$

These are the formulae of Falconer & Mackay (1996, Table 7.3), except that summation is over the markers rather than over the QTLs.

Different methods are considered for the prediction of BV, dominance deviations and GV: the Bayesian methods explained above and an RKHS method. For the Bayesian methods, additive and dominance effects of the markers are predicted as posterior means from the MCMC algorithm. The predicted values are then inserted into the above equations to get the estimated genomic genotypic value $EGV(x)$, the estimated genomic breeding value $EBV(x)$ and the estimated genomic dominance deviation $EDV(x)$.

In our paper, an RKHS method is used only for the prediction of GV. The kernel is derived from the

Table 2. Model-specific expectations

Model	$E(d_j)$	$E(d_j^2)$	$E(a_j d_j)$
BayesD0	0	0	0
BayesD1	μ_D	$E(a_j^2)s_D^2 + \mu_D^2$	0
BayesD2	$\mu_\Delta E(a_j)$	$E(a_j^2)\sigma_\Delta^2 + E(a_j^2)\mu_\Delta^2$	$-w_j E(a_j^2)\mu_\Delta K\left(\frac{\sigma_\Delta}{\mu_\Delta}\right)$
BayesD3	$sE(t \mu_\Delta(t))$	$E(a_j^2)\sigma_\Delta^2 + s^2 E(t^2\mu_\Delta(t)^2)$	$-w_j s^2 E\left(t^2\mu_\Delta(t)K\left(\frac{\sigma_\Delta}{\mu_\Delta(t)}\right)\right)$

assumptions of the regression model. Recall that the known genotypes are fixed explanatory variables. Randomness of GV arises from the randomness of allelic effects, so the function $g_{GV}(\cdot)$ is random. Since RKHS regression assumes that GV are normally distributed with mean zero (Gianola & van Kaam, 2008), but $E(g_{GV}(x)) = \sum_{j=1}^M (E(\tilde{a}_j) + (2 - x_j)E(\tilde{d}_j))x_j \neq 0$, we cannot estimate $g_{GV}(\cdot)$ directly. Instead, we estimate $g(x) = g_{GV}(x) - E(g_{GV}(x))$ with RKHS regression from the observations that are diminished by the expected GV. Then, an estimate of the GV is obtained as $\hat{g}_{GV}(x) = \hat{g}(x) + E(g_{GV}(x))$. For convenience, $g(x)$ is also said to be a GV. The model assumptions imply that the covariance between GV of individuals with genotypes x and \tilde{x} is

$$\begin{aligned} \text{Cov}(g(x), g(\tilde{x})) &= \sum_{j=1}^M x_j \tilde{x}_j \text{Var}(\tilde{a}_j) + \sum_{j=1}^M x_j \tilde{x}_j (2 - \tilde{x}_j) \\ &\times (2 - x_j) \text{Var}(\tilde{d}_j) + \sum_{j=1}^M x_j \tilde{x}_j (4 - x_j - \tilde{x}_j) \text{Cov}(\tilde{a}_j, \tilde{d}_j). \end{aligned} \tag{4}$$

This equation defines the covariance matrix K for the GV of the phenotyped and genotyped individuals in the estimation set. Formulae for the calculation of the variances and covariances that appear on the right-hand side of the equation are given in the previous section. Then, the covariance matrix K can be calculated and used as a genomic relationship matrix for BLUP-prediction of GV. The resulting estimates are RKHS estimates (de los Campos *et al.*, 2009). Take $\hat{\mathbf{g}}$ to be the vector of predicted GV for individuals in the estimation set. It is used for the prediction of GV of non-phenotyped individuals as follows. For an individual with genotype x_0 , the vector K_0 of covariances between the GV of this individual and the GV of all individuals in the estimation set can be calculated from equation (4). The estimated GV of the new individual is $\hat{g}(x_0) = K_0^T K^{-1} \hat{\mathbf{g}}$. Since x_0 was arbitrary, this equation defines an estimate \hat{g} of g . Note that the function $K(x, \tilde{x}) = \text{Cov}(g(x), g(\tilde{x}))$ is symmetric and finitely positive-semidefinite because it is defined via a covariance function. The name RKHS regression results from the fact that a Hilbert space \mathcal{H} exists for which the covariance function $K(x, \tilde{x})$ is a reproducing kernel. The Hilbert space is defined in

Appendix C. The estimated function \hat{g} belongs to this Hilbert space and is optimal in a well-defined sense among all functions that belong to this Hilbert space (Gianola & van Kaam, 2008).

This RKHS estimate is also the BLUP. Since all RKHS estimates are linear, it is the best estimate that can be obtained with RKHS regression, provided that the model assumptions are satisfied by the data and that the first two moments of the random effects are assumed known. In the applications, we calculated the kernel from the assumptions of BayesD2, so the RKHS estimate is nothing but the BLUP of BayesD2 (as opposed to the best predictor).

(vi) Calculation of hyper-parameters

For the calculation of the constant hyper-parameters, we assume whenever possible that additive variance, dominance variance and inbreeding depression of a random mating population are completely explained by the markers. This is not optimal because in fact, markers from low-density panels are not able to explain the total dominance variance. Thus, higher accuracies could be achieved when the parameters are chosen by a grid search or by cross validation. But we think that this is the natural way to choose the parameters. If the contribution of LD to the additive variance and the dominance variance are neglected, then expected additive variance V_{AM} , dominance variance V_{DM} and inbreeding depression \mathcal{I}_M explained by markers are

$$\begin{aligned} V_{AM} &= E\left(\sum_{j \in \mathcal{M}} h_j (\tilde{a}_j + (q_j - p_j) \tilde{d}_j)^2\right), \\ V_{DM} &= E\left(\sum_{j \in \mathcal{M}} h_j^2 \tilde{d}_j^2\right), \\ \mathcal{I}_M &= E\left(\sum_{j \in \mathcal{M}} h_j \tilde{d}_j\right), \end{aligned}$$

where $h_j = 2p_j q_j$ is the heterozygosity of marker j in the case of Hardy Weinberg equilibrium. Note that these formulae assume that the matrix Z_A in the model specification denotes the gene content matrix and Z_D is the indicator matrix for heterozygosity.

The expectations can be calculated, which gives

$$\begin{aligned}
 V_{AM} &= ME(\kappa_j^2)(\bar{h}_o E(a_j^2) - 2E(|a_j||d_j|)\tilde{\gamma}_M + \gamma_M E(d_j^2)), \\
 V_{DM} &= ME(\kappa_j^2)\bar{h}_o^2 E(d_j^2), \\
 \mathcal{I}_M &= ME(\kappa_j) \bar{h}_o E(d_j),
 \end{aligned}
 \tag{5}$$

where \bar{h}_o is the average heterozygosity, \bar{h}_o^2 is the average squared heterozygosity, $\gamma_M = \frac{1}{M} \sum_{j \in \mathcal{M}} h_j (q_j - p_j)^2$ and $\tilde{\gamma}_M = \frac{1}{M} \sum_{j \in \mathcal{M}} h_j (q_j - p_j) w_j$.

For the model without dominance effects (BayesD0), the parameter s^2 is obtained from the condition $V_{AM} = V_A$, which gives

$$s^2 = \frac{V_A}{M \bar{h}_o E(\kappa_j^2)} \frac{v-2}{v}.
 \tag{6}$$

For the calculation of the fixed hyper-parameters for models that include dominance effects it is assumed that V_A, V_D and \mathcal{I} are known or have been estimated. Three parameters are chosen such that the conditions $V_{AM} = V_A$, $V_{DM} = V_D$ and $\mathcal{I}_M = \mathcal{I}$ hold. These are s^2, μ_D and s_D^2 for BayesD1, $s^2, \mu_\Delta, \sigma_\Delta^2$ for BayesD2, and $s^2, s_\Delta, \sigma_\Delta^2$ for BayesD3. The formulae for the calculation of these parameters are lengthy and are given in Appendix B.

3. Application

(i) Simulation

A Fisher–Wright diploid population with population size $N = 1000$ was simulated by sampling individuals for breeding with replacement for 5000 generations. Thereafter, the effective population size decreased for 400 generations from 1000 to 100 with a fast decrease in the most recent generations according to $N_{e,t-400} = 100 + 900 \frac{1 - e^{0.005t-2}}{1 - e^{-2}}$. This formula was chosen in order to reproduce the LD-pattern that is observed in cattle (compare with the estimated historic N_e of cattle breeds, given in Villa-Angulo *et al.*, 2009). The total population size remained constant. This was achieved by reducing the number n_{mt} of males and increasing the number n_{ft} of females such that $N_{e,t} = \frac{4n_{mt}n_{ft}}{n_{mt} + n_{ft}}$.

The genome consisted of one chromosome of 1 Morgan with a mutation rate of 5×10^{-8} . We simulated ten populations. For each population, ten traits with the same characteristics were simulated, so in total there were 100 replicates. In generation 0, 50 SNP for each trait were randomly selected among the about 84 000 segregating SNP to become a QTL. This corresponds to 1500 QTLs in a 30 Morgan genome. Additive and dominance effect of each QTL was sampled according to Scenario 3 fitted to PL in Wellmann & Bennewitz (2011). Thus, the QTL effects were sampled from the same distribution for all traits. Alleles with large effect tend to be partially recessive

or dominant with heterozygous effect above the average effect of the two homozygotes. But alleles of small effect show highly variable dominance coefficients. The sign of the additive effect a_j was chosen dependent on the allele frequency such that the contribution of the QTL to the additive variance is small, i.e. $\text{sign}(a_j) = -\text{sign}((q_j - p_j)d_j)$. After additive and dominance effects had been sampled, they were rescaled by the same factor to obtain a heritability of $h^2 = V_A = 0.15$ for each trait. Realized dominance variance and inbreeding depression varied considerably between traits because only one chromosome was simulated. Average dominance variance and inbreeding depression were $\bar{V}_D = 0.07$ and $\bar{\mathcal{I}} = 0.6$.

Starting with generation 1, the population was maintained without selection for five generations with $N_{e,t} = 100$ but $N_t = 1000$. The markers were identified in generation 1 based on the minor allele frequency (MAF) and on the distance to neighbouring markers. We considered three marker sets with 1500, 3000 and 6000 markers. This corresponds to 45 000, 90 000 and 180 000 markers in a 30 Morgan genome. Markers with high MAF were favoured. Average r^2 -values between adjacent markers were similar for all marker sets. They ranged from 0.36 to 0.40. The marker effects were predicted once in generation 1 from 1000 individuals. Expected dominance variance $V_D = 0.072$ and inbreeding depression $\mathcal{I} = 0.59$ were used as parameters to predict marker effects. Thus, the same hyper-parameters were used for all traits. Predicted marker effects were used to calculate estimated BV, dominance values and GV in generations 1–5. According to the scaling argument introduced by Meuwissen (2009), the results can be extended to a population with a 30 Morgan genome and 30 000 individuals in the estimation set.

(ii) Analysis of the simulated data sets

We compared the accuracies that are obtained by the BayesD methods with G-BLUP, BayesA, BayesC and RKHS regression. Recall that BayesC does not contain dominance effects, BayesD1 assumes conditionally independent additive effect and dominance effects, BayesD2 assumes independent absolute additive effects and dominance coefficients, and BayesD3 assumes dependent absolute additive effects and dominance coefficients such that large additive effects are associated with large dominance coefficients.

G-BLUP and BayesA assume that all markers have a non-negligible effect, so $p_{LD} = 1$. For all other methods, we have chosen the scaling factor $\varepsilon = 0.01$ and p_{LD} depending on the marker panel. For marker panel $k = 1, 2, 3$ with $M = 750 \times 2^k$ markers, we used $p_{LD} = 0.8 \times 0.35^k$. That is, we expected a priori that on average $p_{LD} M / 50 = 12 \times 0.7^k$ markers are needed to

Table 3. Accuracies of predicted BV, dominance deviations (DV) and GV in generations 2 and 5, regressions b_{BV} , b_{DV} and b_{GV} and of true on predicted values, and computation times per cycle relative to BayesA

Generation	Accuracy BV		Accuracy DV		Accuracy GV		b_{BV}	b_{DV}	b_{GV}	Time
	2	5	2	5	2	5	2	2	2	
G-BLUP	0.735	0.665	–	–	0.621	0.564	0.999	–	0.994	(1.0)
RKHS	–	–	–	–	0.648	0.593	–	–	0.954	–
BayesA	0.783	0.730	–	–	0.661	0.618	1.110	–	1.106	1.0
BayesC	0.790	0.740	–	–	0.667	0.626	1.028	–	1.024	1.0
BayesD1	0.798	0.753	0.615	0.604	0.762	0.725	1.017	0.824	0.979	2.3
BayesD2	0.805	0.764	0.662	0.654	0.777	0.744	1.022	0.884	0.993	2.5
BayesD3	0.807	0.767	0.678	0.670	0.782	0.750	1.028	0.932	1.005	2.5

capture the effect of one QTL for marker panel k . The formula for the calculation of p_{LD} was chosen such that the expected number of required markers approaches the number of QTL when the size of the marker panel approaches the total number of SNP in the genome. For BayesD2 and BayesD3, we used $w_j = q_j - p_j$. This parameter determines the probability that additive effect and dominance effect have the same sign for marker j . The degrees of freedom also had to be specified for each method. We used $\nu = 100$ for G-BLUP, $\nu = 2.1$ for BayesA, and $\nu = 2.5$ for all other methods. The prior for the error variance was $p(\sigma^2) \propto 1/\sigma^2$. For simplicity, $X\beta = 1\mu$ and $Zu = 0$ were assumed. For BayesD1, the MCMC chains were run for 50 000 cycles and for all other methods, they were run for 10 000 cycles. The first 50% of the cycles were discarded. We used ten cycles in the Metropolis–Hastings step to sample the additive effects. Marker effects were estimated by the posterior means. The RKHS-regression estimate of the GV was calculated as described in Section (v). The reproducing kernel is defined by equation (4) and the underlying regression model is the same as for BayesD2, so the RKHS estimate has the property that it is the BLUP of BayesD2.

(iii) Results

Table 3 shows the accuracies of predicted BV, dominance deviations (DV) and GV for generation 2 and 5, where 3000 markers per chromosome are used. It also shows the regressions b_{BV} , b_{DV} and b_{GV} of true BV, dominance deviations and GV on estimated values. Among all BayesD methods, BayesD3 yielded the highest accuracies and the regressions of true on predicted values are close to one. BayesD2 performed very similarly and was only slightly worse. In generation 2, BayesC yielded a 7% higher accuracy of predicted BV than G-BLUP and BayesD3 resulted in 10% higher accuracy. The accuracy of G-BLUP decreased by 10% from generation 2 to generation 5, whereas the accuracy of BayesD3 decreased only by

5%. As a consequence, BayesC had a 11% higher accuracy in generation 5, and BayesD3 had a 15% higher accuracy than G-BLUP. The methods differ very much in their ability to predict GV. The accuracy of GV in generation 2 was 15% less than the accuracy of BV if G-BLUP was used for the prediction, but it was only 3% less if BayesD3 was used. The accuracy of GV of BayesD3 exceeded that of G-BLUP by 26% in generation 2 and by 33% in generation 5. The accuracy of GV obtained from RKHS regression, exceeded that of G-BLUP by 4% in generation 2 and by 5% in generation 5. The choice of p_{LD} and ν did not affect the RKHS estimate (not shown). The results are visualized in Fig. 2. The figure shows the accuracies of predicted BV, dominance deviations and GV in generations 1–5 for the set with 3000 markers. It can be seen that the accuracy of the dominance deviation is below the accuracy of the BV. Interestingly, the accuracy of the dominance deviation decreases only very little from generation 1 to generation 5. This suggests that markers that capture the dominance effect of a QTL must be in high LD and therefore in close proximity of the QTL.

Table 3 also shows that the BayesD methods required about twice as much computation time per cycle as the methods without dominance effects. This was expected because twice as many effects need to be predicted. The Metropolis–Hastings step that was needed to sample the additive effects in methods BayesD2 and BayesD3 increased the computation time only slightly. Figure 3 shows the mean accuracy of predicted GV in generations 1–5 when the sampler was stopped after 10, 100, 1000 or 10 000 cycles, and the first 50% of the cycles were discarded. GV were predicted from 3000 markers per chromosome. It can be seen that BayesA, BayesC, BayesD2 and BayesD3 approximately reached convergence after 1000 iterations, whereas BayesD1 may not have reached convergence, even after 10 000 iterations. The same applies for BV and dominance deviations.

Figure 4 shows the mean accuracy of predicted BV, dominance deviations and GV in generations 1–5 for

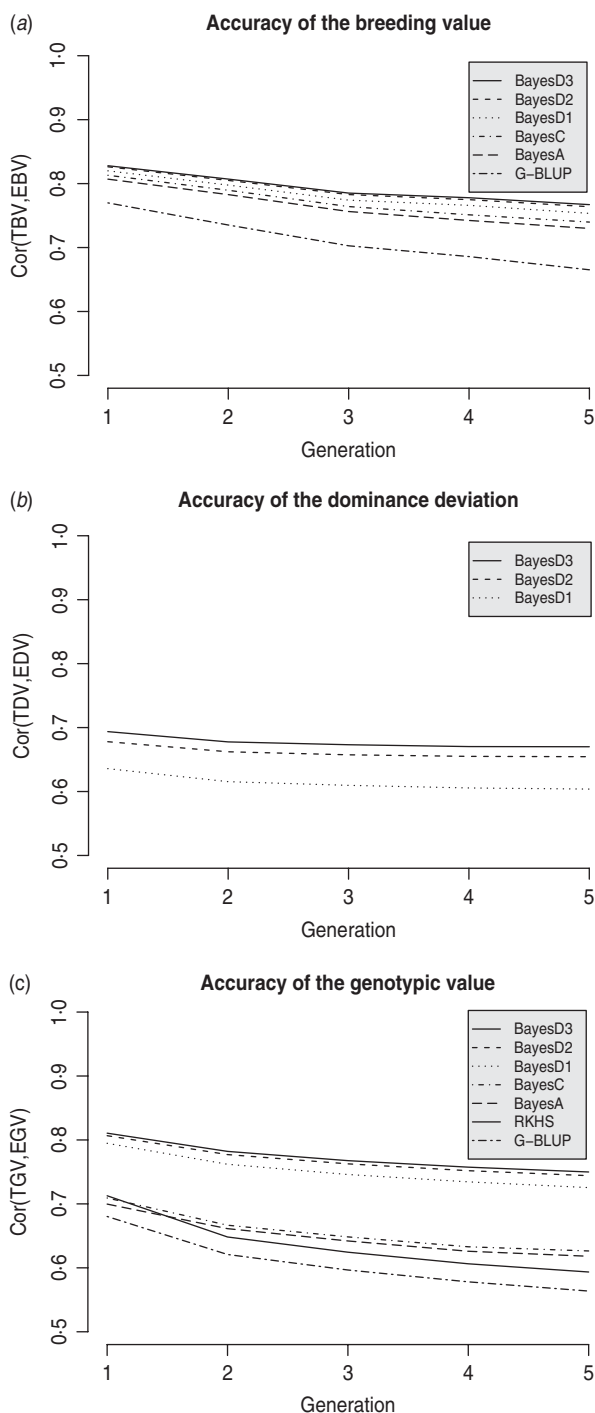


Fig. 2. Accuracy of predicted BV (a), dominance deviations (b) and GV (c) for generations 1–5 and 3000 markers per chromosome.

the different marker panels. Markers from high-density panels had a smaller MAF on average than markers from low-density panels, so they were less informative. This explains why the accuracies of G-BLUP, BayesA and BayesC increase only slightly for high-density panels. It can be seen that for an accurate prediction of dominance deviations and GV, high-density marker panels are needed. The likely

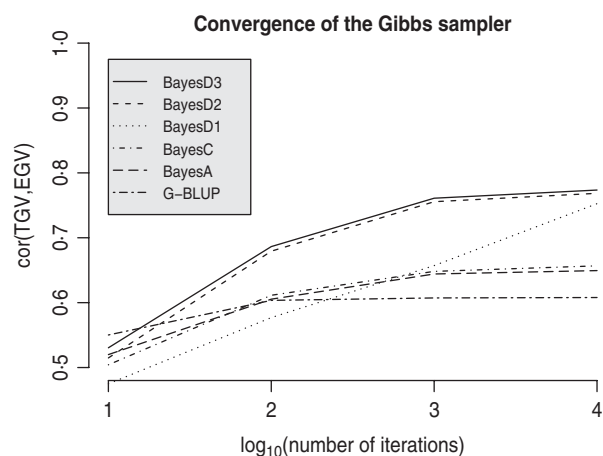


Fig. 3. Mean accuracy of predicted GV in generations 1–5 calculated from 10^x iterations of the sampler for 3000 markers per chromosome.

reason for the increased accuracy of dominance deviations for high-density marker panels is, that the QTLs are on average in higher LD with a marker. For each QTL, we calculated the maximum r^2 -value between the QTL and a marker. The average maximum r^2 -values were 0.49, 0.64, and 0.80 for the different marker panels.

4. Discussion

New Bayesian models for the prediction of genomic BV, dominance deviations and GV have been introduced. The BayesD models outperformed BayesA and BayesC for the simulated data. BayesD3 and BayesD2, which both assume dependent additive and dominance effects, performed best. We showed that these methods not only enable an accurate prediction of BV, dominance deviations and GV, but they also lead to a smaller decrease of the accuracy of genomic BV in subsequent generations. Accuracies of genomic BV of BayesD methods were larger than the accuracies of methods that do not account for dominance. However, for an accurate prediction of dominance deviations and GV, high-density marker panels are needed. Computation time per predicted random effect was similar to BayesA. Interestingly, BayesD1 produced rather high accuracies even though the assumptions of this model were violated in the simulated data. For example, it did not take into account that dominance effects decreased rather than increased the additive variance of the population. However, it yielded smaller accuracies than the other BayesD methods and it showed a slower convergence of the Markov chain.

G-BLUP showed the strongest decrease of the accuracies in subsequent generations. This was expected because G-BLUP assumes normally distributed additive marker effects. This distribution is not heavy

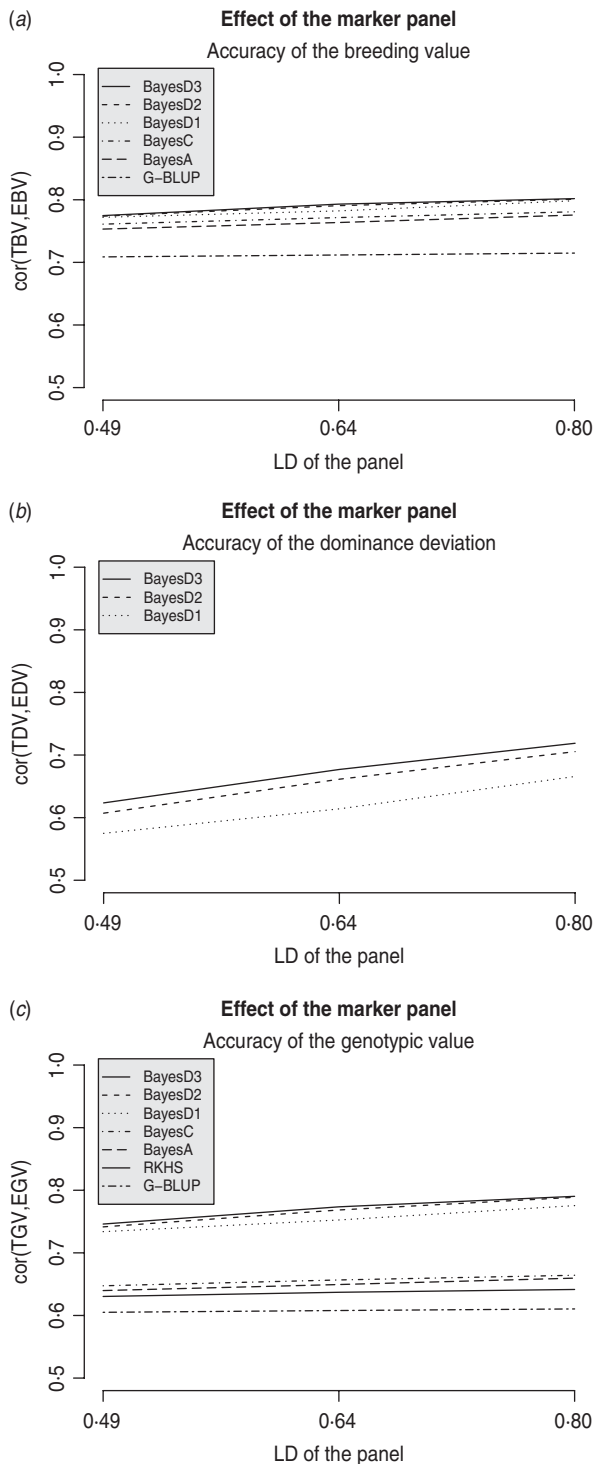


Fig. 4. Accuracy of predicted BV (a), dominance deviations (b), and GV (c) for marker panels with 1500, 3000 and 6000 markers per chromosome. The average maximum r^2 values of a QTL with a marker are shown on the x-axis for the different panels.

tailed. Therefore, more markers would be needed to capture the effect of one large QTL. These additional markers partly have a greater distance to the QTL, so recombinations can cause a greater drop in accuracy. BayesD3 showed the smallest decrease. Moreover, for

the BayesD methods, the accuracy of the dominance deviations decreased only very little in subsequent generations. This suggests that a marker must be in strong LD with a QTL in order to capture the dominance effect of the QTL because otherwise, recombinations are likely to cause a faster decrease of the accuracies. As additive and dominance effects are dependent, and since both of them affect the BV of an individual, this could explain why the inclusion of dominance effects also slows the decrease in accuracy of genomic BV in subsequent generations, as shown in Fig. 2).

Since the BayesD2 estimate and the RKHS estimate (which is the BLUP of BayesD2) were derived from the same statistical model, it was expected that both methods provide similar accuracies. However, the increase in accuracies of GV over G-BLUP was much smaller for RKHS regression than for the Bayesian models. The reason is probably that RKHS regression could provide the best predictor if the GV are normally distributed, but this assumption was violated even though a relatively large number of QTLs was simulated. For RKHS regression, the increase in accuracy over G-BLUP was larger than the increases reported by other authors. Ober *et al.* (2011) found that the accuracy in the validation set obtained with universal kriging was 0.013 larger than the accuracy of a genomic BLUP method. These authors chose the kernel from the family of Matérn covariance functions and additive variance and dominance variance were equal in the simulation. In our study, RKHS regression yielded a 0.027 higher accuracy than G-BLUP, although the dominance variance was only about half as large as the additive variance in our simulation. This suggests that a model-based definition of the kernel can increase the accuracies of RKHS regression estimates. A kernel that accounts for non-additive effects and uses SNP information was also proposed by Gianola & de los Campos (2008) in analogy to the model of Henderson (1985) which relies on the assumptions of Cockerham (1954) and Kempthorne (1954). In contrast to our model that assumes that the known genotypes are fixed and randomness of GV is due to randomness of the allelic effects, Cockerham and Kempthorne assumed implicitly that randomness of GV arises from randomness of the genotypes, and the function g that maps genotypes to GV is unknown and fixed. The joint distribution of the genotypes induces a covariance between GV that depends on the genetic architecture of the trait, i.e. on the unknown function g . Although it is often stated in the literature that g can be arbitrary, from these arguments it follows that RKHS regression makes strong assumptions about the genetic architecture of the trait because GV are inferred from the covariance structure of the GV, and the covariance structure depends on the genetic

architecture. Gianola & de los Campos (2008) stated that the choice of the kernel is indeed absolutely critical for attaining good predictions in RKHS regression.

If the number of markers is large and the prior probability p_{LD} of a marker for being needed to capture a QTL effect is small, then the cumulative effect of all SNP drawn from the distribution with small variance can explain a considerable part of the variance even if $\varepsilon=0.01$. But it is desirable that the model assumes only a small polygenic component if the marker density is large, so ε should be small in this case. However, a small $\varepsilon \ll 0.01$ considerably slows convergence of the Markov chain, so one has to compromise. Alternatively, a BayesB-type algorithm that allows for $\varepsilon=0$ may be used for this model.

The conditional variances τ_j^2 were introduced only in order to obtain a hierarchical model that facilitates estimation of the marker effects with an MCMC algorithm. As noted by Gianola *et al.* (2009) for BayesA, these parameters cannot be estimated precisely from the data because the model does not allow Bayesian learning on these parameters. However, since they have no biological interpretation, estimates of these parameters are not needed.

There are several possibilities to further generalize our model. For example, the parameter s^2 could be different for each marker. That is, each marker could have its own variance. This could make sense in order to account for prior knowledge about a QTL that is in LD with the marker, or in order to account for the joint distribution of additive effects, dominance effects and allele frequencies. In the latter case, s_j^2 would depend on the allele frequency of marker j . We assumed a folded t -distribution for the absolute values of the putative additive effects. As a consequence, we had to choose $\nu > 2$ because otherwise the additive effects would have infinite variance. Since the degree of freedom ν controls the thickness of the tails of a t -distribution, the choice of ν could have a large effect on the accuracies. In this paper, this parameter was chosen by a grid search using cross validation. Alternatively, ν could be treated as random and sampled with a Metropolis–Hastings step. Priors are proposed that assign a small probability to large values of ν , but exclude $\nu < 2$ (e.g. $p(\nu) \propto \frac{1}{\nu^2}$, see Rosa *et al.*, 2004). The t -distribution could be modified to become a generalized hyperbolic distribution in order to force the variance of the additive effects to exist even for $\nu < 2$.

Different models may be appropriate for application depending on the objective of a study. If the aim of a study is the exploration of the genetic architecture of a quantitative trait then allelic effects may be predicted with different models and the assumptions of the model with the best predictive ability are

likely to give a good description of the genetic architecture. However, it is unknown as to how epistasis would affect the predictive ability of the models, so inclusion of epistasis would be the logical next step. Alternatively, the genetic architecture could be explored by evaluation of the posterior distribution. In this case, BayesD1 may be the method of choice because it makes weak prior assumptions. Xu (2003) demonstrated that improper priors with heavy tails for additive and dominance effects produce clearer signals of QTL than the normal distribution. Therefore, small values of ν and p_{LD} could be preferable for QTL detection because this results in a more heavy-tailed distribution. If the aim of a study is the prediction of genomic BV or GV, then the model with best predictive ability should be chosen, provided that the computation time is acceptable. This is likely to be BayesD2 or BayesD3 because these models give the best fit to the genetic architectures that are suggested in the literature. Our simulation study confirms the superiority of BayesD2 and BayesD3. Similar joint distributions of additive effects and dominance effects were assumed for the simulation protocol and for BayesD3. If, contrary to our assumptions, the true joint distribution is very different for a trait (e.g. a trait with many overdominant alleles), then of course BayesD2 and BayesD3 would be not superior.

New methods have been introduced that enable computationally feasible, simultaneous and accurate prediction of BV, dominance deviations and GV for high-density marker sets. The number of females genotyped with a low-density marker set in dairy cattle breeds is increasing rapidly. High density marker genotypes or even whole genome sequences of sires and grandsires are becoming available for imputation. That is, high-density marker sets can be used for the prediction of BV and GV even though most individuals are only genotyped with a low-density marker set (Meuwissen & Goddard, 2010a, 2010b). Thus, the data needed to estimate dominance effects are becoming available. The conclusions drawn in this study are based on simulation experiments. The simulation protocol was designed to realistically model the dependencies between additive and dominance effects of QTLs for quantitative traits following the suggestions of Wellmann & Bennewitz (2011). Once real genomic data from traits with precise information on the genetic architecture including dominance effects become available, it should be used to validate the proposed models in the spirit of Hayes *et al.* (2010).

R. W. was supported by a grant from the Deutsche Forschungsgemeinschaft, DFG. The authors thank Christine Baes for language correction. The manuscript has benefited from the critical comments of the anonymous reviewers.

Appendix A: The Markov Chain

In this section, the Markov chain used for the prediction of the model parameters is described. In each cycle, the fixed effect β , the random effect u , the marker effects $\tilde{\theta}_1, \dots, \tilde{\theta}_M$, the indicator variables $\gamma_1, \dots, \gamma_M$, the conditional variances $\tau_1^2, \dots, \tau_M^2$, and the error variance σ^2 are sampled (in this order). Details are described below.

A.1. Sampling of the fixed effect β

The full conditional posterior distribution of β is

$$\beta | \tilde{\theta}, u, \sigma^2, \gamma, \tau^2, y \sim \mathcal{N}(\hat{\beta}, \sigma^2(X^T X)^{-1}), \tag{7}$$

where

$$\hat{\beta} = (X^T X)^{-1} X^T (y - Zu - Z_A \tilde{a} - Z_D \tilde{d}).$$

In the special case where $\beta = \mu$ contains only the intercept, i.e. $X\beta = 1\mu$, the full conditional posterior simplifies to

$$\mu | \tilde{\theta}, u, \sigma^2, \gamma, \tau^2, y \sim \mathcal{N}\left(\frac{1}{n} 1^T (y - Zu - Z_A \tilde{a} - Z_D \tilde{d}), \frac{\sigma^2}{n}\right).$$

A.2. Sampling of the random effect u

The full conditional posterior distribution of u is

$$u | \tilde{\theta}, \beta, \sigma^2, \gamma, \tau^2, y \sim \mathcal{N}_p(\tilde{u}, \sigma^2(Z^T Z + \sigma^2 \Sigma^{-1})^{-1}), \tag{8}$$

where

$$\tilde{u} = (Z^T Z + \sigma^2 \Sigma^{-1})^{-1} Z^T (y - X\beta - Z_A \tilde{a} - Z_D \tilde{d}).$$

A.3. Sampling of the marker effects $\tilde{\theta}_1, \dots, \tilde{\theta}_M$

We have

$$p(\tilde{a}_j | \tilde{a}_{-j}, \tilde{d}, \xi, \gamma, \tau^2, y) \propto f(\tilde{a}_j) g_j(\tilde{a}_j, \tilde{d}_j) \psi(\tilde{a}_j), \tag{9}$$

where

$$\psi(\tilde{a}_j) = \psi_{\kappa_j, \tau_j^2}(\tilde{a}_j, \tilde{d}_j) \propto \begin{cases} 1 & \text{for BayesD0} \\ 1 & \text{for BayesD1,} \\ \frac{1}{|\tilde{a}_j|} \exp\left(-\frac{(\tilde{d}_j - |\tilde{a}_j| \mu_\Delta)^2}{2\sigma_\Delta^2 \tilde{a}_j^2}\right) & \text{for BayesD2,} \\ \frac{1}{|\tilde{a}_j|} \exp\left(-\frac{(\tilde{d}_j - |\tilde{a}_j| \mu_\Delta \left(\frac{|\tilde{a}_j|}{\kappa_j s}\right))^2}{2\sigma_\Delta^2 \tilde{a}_j^2}\right) & \text{for BayesD3,} \end{cases}$$

and $f(\tilde{a}_j) = f_{\mathcal{N}(\mu_f, \sigma_f^2)}(\tilde{a}_j)$ is the density of a normal distribution with mean and variance

$$\mu_f = \frac{y^T Z_{A(j)}}{Z_{A(j)}^T Z_{A(j)} + \sigma^2 / (\kappa_j^2 \tau_j^2)},$$

$$\sigma_f^2 = \frac{\sigma^2}{Z_{A(j)}^T Z_{A(j)} + \sigma^2 / (\kappa_j^2 \tau_j^2)},$$

where

$$y' = y - X\beta - Zu - Z_{A(-j)} \tilde{a}_{-j} - Z_D \tilde{d}.$$

Since $g_j(\tilde{a}_j, \tilde{d}_j) = \text{pos}_j(\tilde{d}_j)$ if $\tilde{a}_j > 0$ and $g_j(\tilde{a}_j, \tilde{d}_j) = 1 - \text{pos}_j(\tilde{d}_j)$ if $\tilde{a}_j < 0$, a distribution with density proportional to $h(\tilde{a}_j) = f(\tilde{a}_j) g_j(\tilde{a}_j, \tilde{d}_j)$ is a mixture of two truncated normal distributions. Random numbers \tilde{a}_{cand} from h are needed as candidate values for the Metropolis–Hastings step that samples \tilde{a}_j . The probability p_{pos} that a random variable with this distribution is positive equals

$$p_{\text{pos}} = \frac{g_j(1, \tilde{d}_j) \int_0^\infty f(\tilde{a}_j) d\tilde{a}_j}{g_j(1, \tilde{d}_j) \int_0^\infty f(\tilde{a}_j) d\tilde{a}_j + g_j(-1, \tilde{d}_j) \int_{-\infty}^0 f(\tilde{a}_j) d\tilde{a}_j} = \frac{g_j(1, \tilde{d}_j)(1 - F(0))}{g_j(1, \tilde{d}_j)(1 - F(0)) + g_j(-1, \tilde{d}_j)F(0)},$$

where F is the cumulative distribution function with density f . Sampling of \tilde{a}_{cand} from h proceeds as follows:

Sample I_{pos} from $\mathcal{B}(1, p_{\text{pos}})$.
 If $I_{\text{pos}} = 0$ then sample U from $\mathcal{U}[0, F(0)]$, else sample U from $\mathcal{U}[F(0), 1]$.
 Return $\tilde{a}_{\text{cand}} = F^{-1}(U)$.

Here, $\mathcal{U}[a, b]$ denotes the uniform distribution on the interval $[a, b]$. The Metropolis–Hastings algorithm can be used to sample \tilde{a}_j from the full conditional posterior as follows (Chib & Greenberg, 1995):

Sample \tilde{a}_j from h
 For (i in 1:maxIt){
 Sample \tilde{a}_{cand} from h
 Let $\alpha = \min\left(\frac{\psi(\tilde{a}_{\text{cand}})}{\psi(\tilde{a}_j)}, 1\right)$
 With probability α let $\tilde{a}_j = \tilde{a}_{\text{cand}}$
 }
 Return \tilde{a}_j

Note that the Metropolis–Hastings step is not needed if $\psi(\tilde{a}_j) \propto 1$.

It remains to be shown how \tilde{d}_j is sampled for models that include dominance effects. The full conditional posterior distribution of \tilde{d}_j is

$$p(\tilde{d}_j | \tilde{a}, \tilde{a}_{-j}, \xi, \gamma, \tau^2, y) \propto \tilde{f}(\tilde{d}_j) g_j(\tilde{a}_j, \tilde{d}_j), \tag{10}$$

where $\tilde{f}(\tilde{d}_j) = f_{\mathcal{N}(\mu_f, \sigma_f^2)}(\tilde{d}_j)$ is the density of a normal distribution with mean and variance

$$\mu_f = \frac{y'^T Z_{D(j)} + \sigma^2 \mu_d / (\sigma_d^2 \kappa_j)}{Z_{D(j)}^T Z_{D(j)} + \sigma^2 / (\sigma_d^2 \kappa_j^2)},$$

$$\sigma_f^2 = \frac{\sigma^2}{Z_{D(j)}^T Z_{D(j)} + \sigma^2 / (\sigma_d^2 \kappa_j^2)},$$

where

$$y' = y - X\beta - Zu - Z_A \tilde{a} - Z_{D(-j)} \tilde{d}_{-j}.$$

Since $g_j(\tilde{a}_j, \tilde{d}_j) = (1 - w_j \text{sign}(\tilde{a}_j))/2$ if $\tilde{d}_j > 0$ and $g_j(\tilde{a}_j, \tilde{d}_j) = (1 + w_j \text{sign}(\tilde{a}_j))/2$ if $\tilde{d}_j < 0$, the full conditional posterior distribution of \tilde{d}_j is a mixture of two truncated normal distributions. We have

$$\begin{aligned} \tilde{p}_{\text{pos}} &= P(\tilde{d}_j > 0 | \text{ELSE}) \\ &= \frac{g_j(\tilde{a}_j, 1)(1 - \tilde{F}(0))}{g_j(\tilde{a}_j, 1)(1 - \tilde{F}(0)) + g_j(\tilde{a}_j, -1)\tilde{F}(0)}, \end{aligned}$$

where \tilde{F} is the cumulative distribution function with density \tilde{f} . Sampling of \tilde{d}_j proceeds as follows:

Sample I_{pos} from $\mathcal{B}(1, \tilde{p}_{\text{pos}})$.
 If $I_{\text{pos}} = 0$ then sample U from $\mathcal{U}[0, \tilde{F}(0)]$, else sample U from $\mathcal{U}[\tilde{F}(0), 1]$.
 Return $\tilde{d}_j = \tilde{F}^{-1}(U)$.

A.4. Sampling of $\gamma_1, \dots, \gamma_M$

The full conditional posterior distribution of γ_j is

$$\gamma_j | \tilde{\theta}, \xi, \gamma_{-j}, \tau^2, y \sim \mathcal{B}\left(1, \frac{\omega_1 p_{\text{LD}}}{\omega_1 p_{\text{LD}} + \omega_0(1 - p_{\text{LD}})}\right), \quad (11)$$

where

$$\begin{aligned} \omega_0 &= \frac{1}{\varepsilon} \exp\left(-\frac{\tilde{a}_j^2}{2\varepsilon^2\tau_j^2}\right) \psi_{\varepsilon, \tau_j^2}(\tilde{a}_j, \tilde{d}_j), \\ \omega_1 &= \exp\left(-\frac{\tilde{a}_j^2}{2\tau_j^2}\right) \psi_{1, \tau_j^2}(\tilde{a}_j, \tilde{d}_j). \end{aligned}$$

Updating of $\gamma_1, \dots, \gamma_M$ is only needed if $p_{\text{LD}} < 1$. After sampling of γ_j , κ_j needs to be updated.

A.5. Sampling of $\tau_1^2, \dots, \tau_M^2$

For BayesD1, the full conditional posterior distribution of τ_j^2 is

$$\begin{aligned} \tau_j^2 | \tilde{\theta}, \xi, \gamma, \tau_{-j}^2, y \sim \\ \text{Inv-}\chi^2\left(v+2, \frac{\frac{\tilde{a}_j^2}{\kappa_j^2} + \left(\frac{\tilde{d}_j - \mu_D}{s_D}\right)^2 + v s^2}{v+2}\right). \end{aligned} \quad (12)$$

Otherwise, the full conditional posterior distribution of τ_j^2 is

$$\tau_j^2 | \tilde{\theta}, \xi, \gamma, \tau_{-j}^2, y \sim \text{Inv-}\chi^2\left(v+1, \frac{\frac{\tilde{a}_j^2}{\kappa_j^2} + v s^2}{v+1}\right). \quad (13)$$

A.6. Sampling of the error variance σ^2

The full conditional posterior distribution of σ^2 is

$$\begin{aligned} \sigma^2 | \tilde{\theta}, \beta, u, \gamma, \tau^2, y \sim \\ \text{Inv-}\chi^2\left(\frac{n + v^*, E^T E + v^* s^{*2}}{n + v^*}\right), \end{aligned} \quad (14)$$

where $E = y - X\beta - Zu - Z_A \tilde{a} - Z_D \tilde{d}$.

Appendix B: Calculation of hyper-parameters

For BayesD1 with conditionally independent additive effect and dominance effects the parameters s^2, μ_D and s_D need to be specified. We have

$$\begin{aligned} \mu_D &= \frac{\mathcal{I}}{M \bar{h}_o E(\kappa_j)}, \\ s^2 &= \frac{V_A - \frac{\gamma_M}{h_o^2} V_D}{M \bar{h}_o E(\kappa_j^2)} \frac{v-2}{v}, \\ s_D^2 &= \frac{\frac{V_D}{M \bar{h}_o^2 E(\kappa_j^2)} - \mu_D^2}{s^2} \frac{v-2}{v}. \end{aligned} \quad (15)$$

For BayesD2 with independent absolute additive effect and dominance coefficients the parameters s^2, μ_Δ and σ_Δ^2 are obtained as follows. Since

$$\frac{\sigma_\Delta^2}{\mu_\Delta^2} = \frac{V_D}{\mathcal{I}^2} \frac{M \bar{h}_o^2 E(\kappa_j)^2 \lambda^2}{h_o^2 E(\kappa_j^2)} - 1, \quad (16)$$

we can calculate $K(\sigma_\Delta/\mu_\Delta)$. We have

$$\mu_\Delta = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \quad (17)$$

where

$$\begin{aligned} a &= V_A - \frac{\gamma_M V_D}{h_o^2}, \\ b &= \frac{2E(\kappa_j^2) \mathcal{I}^2 K\left(\frac{\sigma_\Delta}{\mu_\Delta}\right) \tilde{\gamma}_M}{\lambda^2 M \bar{h}_o^2 E(\kappa_j)^2}, \\ c &= \frac{-\mathcal{I}^2 E(\kappa_j^2)}{\lambda^2 M \bar{h}_o E(\kappa_j)^2}. \end{aligned}$$

Then $\sigma_\Delta^2 = \mu_\Delta^2 \frac{\sigma_\Delta^2}{\mu_\Delta^2}$ can be calculated. Finally, s^2 is obtained from

$$s^2 = \frac{V_D}{(\sigma_\Delta^2 + \mu_\Delta^2) M \bar{h}_o^2 E(\kappa_j^2)} \frac{v-2}{v}.$$

For BayesD3 with dependent absolute additive effect and dominance coefficients, the parameters s^2, s_Δ and σ_Δ^2 are calculated as follows. At first, the parameter s_Δ is determined such that the following

equation holds:

$$\begin{aligned} \bar{h}_o \frac{v}{v-2} &= \frac{V_A \bar{h}_o^2 - V_D \gamma_M}{I^2} \frac{E(\kappa_j)^2 \bar{h}_o^2}{E(\kappa_j^2) \bar{h}_o^2} ME(|t| \mu_\Delta(|t|))^2 \\ &+ 2 \tilde{\gamma}_M E\left(t^2 \mu_\Delta(|t|) K\left(\frac{\sigma_\Delta}{\mu_\Delta(|t|)}\right)\right), \end{aligned} \tag{18}$$

where

$$\begin{aligned} \sigma_\Delta^2 &= \left(\frac{V_D}{I^2} \frac{E(\kappa_j)^2 \bar{h}_o^2}{E(\kappa_j^2) \bar{h}_o^2} ME(|t| \mu_\Delta(|t|))^2 - E(t^2 \mu_\Delta(|t|)^2) \right) \\ &\times \frac{v-2}{v}. \end{aligned}$$

This was done by a grid search, where the expectations were estimated from simulated random variables $t \sim t_v$. Then σ_Δ^2 can be calculated and s^2 is obtained from

$$s^2 = \frac{I^2}{M^2 E(\kappa_j)^2 \bar{h}_o^2 E(|t| \mu_\Delta(|t|))^2}.$$

Appendix C: Hilbert space for RKHS regression

This section defines the Hilbert space \mathcal{H} for which a symmetric and positive semidefinite kernel K is reproducing. Consider the linear space S that consists of all linear combinations of functions $K(x, \cdot): \Omega \rightarrow \mathbb{R}$ with $x \in \Omega$. Each function $f \in S$ can be written as $f = \sum_{j=1}^N a_j K(x_j, \cdot)$ with $N \in \mathbb{N}$, pairwise different $x_1, \dots, x_N \in \Omega$ and $a \in \mathbb{R}^N$. For two functions $g_1 = \sum_{j=1}^{N_1} a_j K(x_j, \cdot)$, $g_2 = \sum_{k=1}^{N_2} b_k K(\tilde{x}_k, \cdot)$, the inner product is defined as

$$\langle g_1, g_2 \rangle = \sum_{j=1}^{N_1} \sum_{k=1}^{N_2} a_j b_k K(x_j, \tilde{x}_k).$$

S is a pre-Hilbert space with inner product $\langle \cdot, \cdot \rangle$. The closure of S under the inner product $\langle \cdot, \cdot \rangle$ is a Hilbert space \mathcal{H} . It is called the native space of the kernel K . Note that \mathcal{H} can be larger than S because it also includes all limiting functions of Cauchy sequences. The kernel K is a reproducing kernel of \mathcal{H} . For proofs see the literature on Hilbert spaces, e.g. Shawe-Taylor & Cristianini (2004).

References

Bennewitz, J. & Meuwissen, T. H. E. (2010). The distribution of QTL additive and dominance effects in porcine F2 crosses. *Journal of Animal Breeding and Genetics* **127**, 171–179.
 Caballero, A. & Keightley, P. D. (1994). A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* **138**, 883–900.
 Calus, M. P. L. (2010). Genomic breeding value prediction: methods and procedures. *Animal* **4**, 157–164.

Calus, M. P. L., Meuwissen, T. H. E., de Roos, A. P. W. & Veerkamp, R. F. (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics* **178**, 553–561.
 Charlesworth, D. & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics* **10**, 783–796.
 Chib, S. & Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *American Statistical Association* **49**, 327–335.
 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.
 de los Campos, G., Gianola, D. & Rosa, G. J. M. (2009). Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. *Journal of Animal Science* **87**, 1883–1887.
 Duangjinda, M., Bertrand, J. K., Misztal, I. & Druet, T. (2001). Estimation of additive and nonadditive genetic variances in Hereford, Gelbvieh, and Charolais by Method R. *Journal of Animal Science* **79**, 2997–3001.
 Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. London, UK: Longman.
 Garcia-Dorado, A., López-Fanjul, C. & Caballero, A. (1999). Properties of spontaneous mutations affecting quantitative traits. *Genetic Research Cambridge* **74**, 341–350.
 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.
 George, E. I. & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881–889.
 Gianola, D. & de los Campos, G. (2008). Inferring genetic values for quantitative traits non-parametrically. *Genetic Research* **90**, 525–540.
 Gianola, D., Fernando, R. L. & Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* **173**, 1761–1776.
 Gianola, D. & van Kaam, J. B. C. H. M. (2008). Reproducing Kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* **178**, 2289–2303.
 Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E. & Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* **183**, 347–363.
 Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J. J., Kvasz, A., Mni, M., Simon, P., Frère, J.-M., Coppieters, W. & Georges, M. (2004). Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2398–2403.
 Habier, D., Tetens, J., Seefried, F.-R., Lichtner, P. & Thaller, G. (2010). The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution* **42**, 5.
 Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. (2009). Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* **92**, 433–443.
 Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J. & Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in holstein

- cattle as contrasting model traits. *PLoS Genetics* **6**, 31001139.
- Heffner, E. L., Sorrells, M. E. & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Science* **49**, 1–12.
- Henderson, C. R. (1985). Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *Journal of Animal Science* **60**, 111–117.
- Hill, W. G., Goddard, M. E. & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* **4**, e1000008.
- Kacser, H. & Burns, J. A. (1981). The molecular basis of dominance. *Genetics* **97**, 639–666.
- Kempthorne, O. (1954). The correlation between relatives in a random mating population. *Proceedings of the Royal Society of London. Series B* **143**, 103–113.
- Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F. & Fritz, S. (2011). Improved Lasso for genomic selection. *Genetic Research Cambridge* **93**, 77–87.
- Luan, T., Woolliams, J. A., Lien, S., Kent, M., Svendsen, M. & Meuwissen, T. H. E. (2009). The accuracy of genomic selection in norwegian red cattle assessed by cross-validation. *Genetics* **183**, 1119–1126.
- Meuwissen, T. H. E. (2009). Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genetics Selection Evolution* **41**, 35.
- Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829.
- Meuwissen, T. H. E. & Goddard, M. E. (2004). Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genetics, Selection, Evolution* **36**, 261–279.
- Meuwissen, T. H. E. & Goddard, M. E. (2010a). Accurate prediction of genetic values for complex traits by whole genome resequencing. *Genetics* **185**, 623–631.
- Meuwissen, T. H. E. & Goddard, M. E. (2010b). The use of family relationships and linkage disequilibrium to impute phase and missing genotypes in up to whole-genome sequence density genotypic data. *Genetics* **185**, 1441–1449.
- Misztal, I. (1997). Estimation of variance components with large-scale dominance models. *Journal of Dairy Science* **80**, 965–974.
- Ober, U., Erbe, M., Long, N., Porcu, E., Schlather, M. & Simianer, H. (2011). Predicting genetic values: a Kernel-based best linear unbiased prediction with genomic data. *Genetics* **188**, 695–708.
- Park, T. & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association* **103**, 681–686.
- Piepho, H. P. (2009). Ridge regression and extensions for genomewide selection in maize. *Crop Science* **49**, 1165–1176.
- Psarakis, S. & Panaretos, J. (1990). The folded *t* distribution. *Communications in Statistics – Theory and Methods* **19**, 2717–2734.
- Rosa, G. J. M., Gianola, D. & Padovani, C. R. (2004). Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC. *Journal of Applied Statistics* **31**, 855–873.
- Serenius, T., Stalder, K. J. & Puonti, M. (2006). Impact of dominance effects on sow longevity. *Journal of Animal Breeding and Genetics* **123**, 355–361.
- Shawe-Taylor, J. & Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge: Cambridge University Press.
- Toro, M. A. & Varona, L. (2010). A note on mate allocation for dominance handling in genomic selection. *Genetics Selection Evolution* **42**, 33.
- 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.
- Verbyla, K. L., Hayes, B. J., Bowman, P. J. & Goddard, M. E. (2009). Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genetics Research Cambridge* **91**, 307–311.
- Verbyla, K. L., Bowman, P. J., Hayes, B. J. & Goddard, M. E. (2010). Sensitivity of genomic selection to using different prior distributions. *BMC Proceedings* **4** (Suppl. 1), S5.
- Villa-Angulo, R., Matukumalli, L. K., Gill, C. A., Choi, J., Van Tassell, C. P. & Grefenstette, J. J. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genetics* **10**, 19.
- Wellmann, R. & Bennewitz, J. (2011). The contribution of dominance to the understanding of quantitative genetic variation. *Genetics Research Cambridge* **93**, 139–154.
- Wolc, A., Stricker, C., Arango, J., Settar, P., Fulton, J. E., O’Sullivan, N. P., Preisinger, R., Habier, D., Fernando, R., Garrick, D. J., Lamont, S. J. & Deckers, J. C. M. (2011). Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution* **43**, 5.
- Xu, S. (2003). Estimating polygenic effects using markers of the entire genome. *Genetics* **163**, 789–801.
- Yi, N., George, V. & Allison, D. B. (2003). Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics* **164**, 1129–1138.