Genomic Model with Correlation Between Additive and Dominance Effects

Tao Xiang,[*,‡,§,1] **Ole Fredslund Christensen,**[§] **Zulma Gladis Vitezica,**[**] **and Andres Legarra**[‡]

*Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education and Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, P. R. China, ‡INRA, UMR 1388 GenPhySE, F-31326 Castanet-Tolosan, France, §Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark, and **Université de Toulouse, UMR 1388 GenPhySE, F-31326 Castanet-Tolosan, France

ORCID ID: 0000-0001-8893-7620 (A.L.)

ABSTRACT Dominance genetic effects are rarely included in pedigree-based genetic evaluation. With the availability of single nucleotide polymorphism markers and the development of genomic evaluation, estimates of dominance genetic effects have become feasible using genomic best linear unbiased prediction (GBLUP). Usually, studies involving additive and dominance genetic effects ignore possible relationships between them. It has been often suggested that the magnitude of functional additive and dominance effects at the quantitative trait loci are related, but there is no existing GBLUP-like approach accounting for such correlation. Wellmann and Bennewitz (2012) showed two ways of considering directional relationships between additive and dominance effects, which they estimated in a Bayesian framework. However, these relationships cannot be fitted at the level of individuals instead of loci in a mixed model, and are not compatible with standard animal or plant breeding software. This comes from a fundamental ambiguity in assigning the reference allele at a given locus. We show that, if there has been selection, assigning the most frequent as the reference allele orients the correlation between functional additive and dominance effects. As a consequence, the most frequent reference allele is expected to have a positive value. We also demonstrate that selection creates negative covariance between genotypic additive and dominance genetic values. For parameter estimation, it is possible to use a combined additive and dominance relationship matrix computed from marker genotypes, and to use standard restricted maximum likelihood algorithms based on an equivalent model. Through a simulation study, we show that such correlations can easily be estimated by mixed model software and that the accuracy of prediction for genetic values is slightly improved if such correlations are used in GBLUP. However, a model assuming uncorrelated effects and fitting orthogonal breeding values and dominant deviations performed similarly for prediction.

KEYWORDS genomic model; additive genetic effects; dominance genetic effects; correlation; shared data resource; GenPred; Genomic Selection

FROM quantitative genetics theory, statistical additive genetic values (also called breeding values) of individuals are obtained from average allele substitution effects (α), which are functions of functional additive and dominant gene/marker effects (a and d) (Falconer and Mackay 1996). Dominance deviations are differences between genotypic values and breeding values, and only include a part of the dominant effects of the genes/markers (Falconer and Mackay 1996). Additive genetic variance ($2pq\alpha^2$) includes variation due to the functional additive and dominant effects, and dominance genetic variance ($(2pqd)^2$) involves only the functional dominant effects.

The inclusion of dominance in genomic evaluation models has been proposed by several authors (Su *et al.* 2012; Vitezica *et al.* 2013, 2016; Ertl *et al.* 2014; Muñoz *et al.* 2014; Aliloo *et al.* 2016, 2017; Xiang *et al.* 2016). In those studies, additive and dominant marker effects (a and d) are considered uncorrelated. However, QTL analyses show that the *magnitudes* of these effects are dependent (*e.g.*, Bennewitz and Meuwissen 2010).

In the genomic era, the relationship between magnitudes of additive and dominant gene/marker effects has scarcely been modeled (Wellmann and Bennewitz 2011, 2012; Bennewitz *et al.* 2017). Wellmann and Bennewitz (2011) reviewed evidence for association of magnitudes of additive and dominant effects across QTL. These magnitudes have been shown to be related through the dominance coefficients $\delta = d/|a|$. QTL with large absolute additive effect are likely to be associated with large dominance coefficients, while small additive effects tend to be associated with small dominance coefficients (Caballero and Keightley 1994). These results suggest that across QTL loci it is possible to construct joint distributions of additive and dominant effects. Wellmann and Bennewitz (2012) suggested a general hierarchical Bayesian model where absolute additive QTL effects and dominance coefficients were assumed to be dependent with $cor(|\boldsymbol{a}|, \boldsymbol{\delta}) > 0$.

Recently, the dependencies between additive and dominant gene effects were considered in a Bayesian model for association analysis (Bennewitz *et al.* 2017). For genomic prediction, methods commonly used are linear mixed models and best linear unbiased prediction (BLUP), *i.e.*, genomic BLUP (GBLUP) methods (Su *et al.* 2012; Vitezica *et al.* 2013). In these models, the relationship between additive and dominant marker effects is ignored. Examining the relationship between a and d and including it in such genomic models could improve the accuracy of predictions.

Even though Bayesian models can take into account the dependencies between additive and dominant gene effects, their implementation by Markov Chain Monte Carlo methods is not straightforward. These models need customary implementations and would be computationally slow for large data sets. In addition, if the relationship between functional additive and dominant marker effects cannot be described as covariance structure of additive and dominance effects on individuals, then the standard mixed model animal breeding software, *e.g.*, DMU (Madsen and Jensen 2013), BLUPf90 family (Misztal *et al.* 2002), or ASREML (Gilmour *et al.* 2009), cannot be used for estimating such relationships.

In this study, we present a novel method to quantify the importance of the relationship between additive and dominant effects of QTL using a GBLUP-like method. This method relies on the fact that allele substitution effects contain functional additive and dominant effects that, after phenotypic selection, tend to offset each other. Our approach is based on fixing the most frequent allele as the reference allele. By simulation, we evaluate the benefit of accounting for this relationship (between a and d) explicitly in the genomic model for genetic parameter estimation and the prediction of genetic values, comparing it with the model ignoring the relationship between a and d, and with the classical orthogonal model based on breeding values and dominance deviations (Vitezica *et al.* 2013).

Materials and Methods

Theory

Consider a quantitative trait that is determined by biallelic quantitative trait loci. For each locus, let the midpoint of the genotypic values of the two homozygotes be the origin (0). Relative to this origin, the genotypic value of a homozygote is defined as either a or $-a$, and the genotypic value of the heterozygote is defined as d, which is the amount the heterozygote deviates from the origin. Note that the functional value a can be either positive or negative (a point that is rarely explicit in many textbooks), and the magnitude of the difference between the two homozygotes is therefore the absolute value $|2a|$. For many traits and species, an advantage of heterozygosity is observed, known as heterosis or the related phenomenon inbreeding depression (*i.e.*, Lynch and Walsh 1998, chapter 10). This phenomenon is typically modeled as a regression of the phenotype on the degree of heterozygosity, for which several metrics exist (*i.e.*, Silió *et al.* 2016; Xiang *et al.* 2016).

Using a genotypic model (*i.e.*, Su *et al.* 2012; Vitezica *et al.* 2016) for additive and dominant genotypic effects $\boldsymbol{u}$ and $\boldsymbol{v}$ on individuals we use the notation $\boldsymbol{u} = \boldsymbol{Za}$ and $\boldsymbol{v} = \boldsymbol{Wd}$, respectively, where $\boldsymbol{a}$ and $\boldsymbol{d}$ are vectors of functional additive and dominant QTL effects across individuals, respectively, and $\boldsymbol{Z}$ and $\boldsymbol{W}$ are the respective incidence matrices. Considering one QTL, the matrix $\boldsymbol{W}$ has entries 0, 1, and 0 for genotypes BB, Bb, and bb, respectively. For the matrix $\boldsymbol{Z}$, there are two ways of coding genotypes depending on the selected reference allele: if allele B is the selected reference allele, then matrix $\boldsymbol{Z}$ has entries 1, 0, and -1 for BB, Bb, and bb, respectively (case 1 in Table 1); if allele b is selected to be the reference allele, matrix $\boldsymbol{Z}$ is coded as -1, 0, and 1 for genotypes BB, Bb, and bb, respectively (case 2 in Table 1). Thus, in case 1, additive genotypic effects $\boldsymbol{u}$ for BB, Bb, and bb are a, 0, and $-a$, respectively, and in case 2, additive genotypic effects $\boldsymbol{u}$ for BB, Bb, and bb are $-a$, 0, and a, respectively. The dominant genotypic effects $\boldsymbol{v}$ are always 0, d, and 0 for genotypes BB, Bb, and bb, respectively. These two cases can be found in Table 1.

Phenotypic (directional) selection, either artificial by breeding or natural, operates to change the phenotype in a desired direction; throughout this paper we will use the shorter term "selection" for this, and without loss of generality we will assume that the direction of selection is upwards. Selection acts mainly on the statistical additive component of genetic values. Thus, considering a segregating QTL, then selection will act on the allele substitution effect

$$\alpha = a + (1 - 2p)d.$$

Therefore, one would expect that in a long-term selected population, QTL with large allele substitution effects are fixed and only QTL with small allele substitution effects segregate.

Table 1 Different ways of allele coding for incidence matrices for additive and dominance effects

Genotypes Case 1	BB	Bb	bb
Matrix $\mathbf{Z}$	1	0	-1
$\boldsymbol{u} = \mathbf{Z}\boldsymbol{a}$	a	0	$-a$
Matrix $\mathbf{W}$	0	1	0
$\boldsymbol{v} = \mathbf{W}\boldsymbol{d}$	0	d	0
Frequencies	p^2	$2pq$	q^2
Case 2			
Matrix $\mathbf{Z}$	-1	0	1
$\boldsymbol{u} = \mathbf{Z}\boldsymbol{a}$	$-a$	0	a
Matrix $\mathbf{W}$	0	1	0
$\boldsymbol{v} = \mathbf{W}\boldsymbol{d}$	0	d	0
Frequencies	q^2	$2pq$	p^2

For one biallelic QTL locus, there are three genotypes: BB, Bb, and bb. The genotypic value of a homozygote is defined as a or $-a$ and the genotypic value of the heterozygote is defined as d. a refers to an arbitrary allele whose frequency is p. In a genotypic model, genotypic additive ($\boldsymbol{u}$) and dominance effects ($\boldsymbol{v}$) on individuals are $\mathbf{Z}\boldsymbol{a}$ and $\mathbf{W}\boldsymbol{d}$, respectively. Incidence matrix $\mathbf{Z}$ can be coded in two cases: (1) $\mathbf{Z}$ is coded as 1, 0, and -1 for BB, Bb, and bb, respectively, and (2) $\mathbf{Z}$ is coded as -1, 0, and 1 for genotypes BB, Bb, and bb, respectively. In both two cases, matrix $\mathbf{W}$ is coded as 0, 1, and 0 for genotypes BB, Bb, and bb, respectively.

Thus, if there is selection, it is expected that no segregating QTL will have a very large allele substitution effect after some generations of selection, leading to:

$$\alpha = a + (1 - 2p)d = \varepsilon \qquad (1)$$

where ε is small. From this we can derive $a. = (2p - 1)d + \varepsilon$, and we see that a and $(2p - 1)d$ will tend to have the same sign and magnitudes that are positively associated. At this point, a refers to an arbitrary allele (B or b) whose frequency is p, and both a and d can have positive or negative values. The allelic frequency $0 < p < 1$ and therefore $2p - 1$ can also be positive or negative, and we see from $a = (2p - 1)d + \varepsilon$ that there is no association between sign of a and signs of either $(2p - 1)$ or d. We will call this "random allele coding" (RAC).

An alternative presentation of the same idea is considering allelic frequency at equilibrium after several rounds of selection (Crow and Kimura 2009; Charlesworth and Charlesworth 2010). If the fitnesses at the locus are $-a$, d, and a (the actual values are functions of a and d that depend on the selection intensity and the part of genetic variance explained by the locus; Charlesworth and Charlesworth 2010, box 3.7), the frequency at equilibrium is $p = 0.5 + a/2d$, which is a rewriting of Equation (1). Thus, $2p - 1$ and a/d have the same sign.

Consider now the equilibria. If $-|a| < d < |a|$, loci tend to fixation toward the favorable homozygote. If $|d| > |a|$ (overdominance or underdominance), equilibria are *stable* when $d > |a|$ (overdominance) and there is maintenance of polymorphisms in the population, but if $d < -|a|$ (underdominance) the equilibrium is unstable and any random event will lead loci to fixation (Crow and Kimura 2009; Charlesworth

and Charlesworth 2010). Thus, selection tends to maintain primarily overdominant alleles and therefore, after a selection process, it is expected that $d > 0$ (either across segregating loci or across repeated evolutionary histories of the same locus).

Now we extend the reasoning to several loci with random effects drawn from some distributions. Consider a collection of loci with elements a_i, d_i, and p_i, respectively, and (a_i, d_i, p_i), $i = 1, \ldots, N$ are treated independent and identically distributed across loci. We can formalize the above finding to the covariance between a and $(2p - 1)d$ as $cov(a_i, (2p_i - 1)d_i) > 0$, with $E(a_i) = 0$ and $E(d_i) > 0$. For RAC we see from $a_i = (2p_i - 1)d_i + \varepsilon_i$ that there is no association between sign of a_i and signs of either $(2p_i - 1)$ or d_i. This we formalize as:

$$\begin{aligned} \text{cov}(a_i, d_i) &= E(a_i d_i) - E(a_i)E(d_i) \\ &= E\big(((2p_i - 1)d_i + \varepsilon_i)d_i\big) - 0 \\ &\approx E\big((2p_i - 1)d_i^2\big) = 0 \end{aligned} \qquad (2)$$

The last identity holds because $2p_i - 1$ has zero mean due to p_i having a symmetric distribution with mean 0.5 [could be a uniform distribution between 0 and 1 or a U-shaped distribution (Hill *et al.* 2008)] for the RAC, and assuming that conditional mean and variance of d_i do not depend on p_i. This means that postulating in the model a covariance between a and d has no meaning (or interest) because this covariance is 0 under these assumptions.

However, returning to the one-locus case, we can measure its allelic frequency p, and we can arbitrarily fix the most frequent allele as the reference allele such that $p > 0.5$. We will indicate this hereinafter as "major allele coding" (MAC). In such case, allelic frequency p is denoted as $p^{MAC}(p = p^{MAC} > 0.5)$. Additive effect a still refers to the allele whose frequency is p (actually now is p^{MAC}) and is denoted as a^{MAC}. Still, a^{MAC} can have positive or negative values. The functional dominance effect d is invariant to the reference allele. Equation (1) still holds:

$$\alpha^{MAC} = a^{MAC} + (1 - 2p^{MAC})d = \varepsilon.$$

By the restriction $p^{MAC} > 0.5$, the term $2p^{MAC} - 1$ in $a^{MAC} \approx (2p^{MAC} - 1)d + \varepsilon$ becomes strictly positive and we see that a^{MAC} and d will tend to have the same sign. In addition, because $E(d) > 0$ and ε is small, then it turns out that $E(a^{MAC}) > 0$.

Consider now the covariance, across several loci, between functional additive (a^{MAC}) and dominant (d) effects; then we formalize the above finding to

$$\begin{aligned} cov(a_i^{MAC}, d_i) &= cov\big((2p_i^{MAC} - 1)d_i + \varepsilon, d_i\big) \\ &\approx cov\big((2p_i^{MAC} - 1) + d_i, d_i\big) > 0 \end{aligned} \qquad (3)$$

where the last inequality follows from

$$cov\left(\left(2p_i^{MAC} - 1\right)d_i, d_i\right)$$
$$= E\left(cov\left(\left(2p_i^{MAC} - 1\right)d_i, d_i | p_i^{MAC}\right)\right)$$
$$+ cov\left(E\left(\left(2p_i^{MAC} - 1\right)d_i | p_i^{MAC}\right), E\left(d_i | p_i^{MAC}\right)\right),$$

where the first term equals $E((2p_i^{MAC} - 1)var(d_i|p_i^{MAC})) > 0$, and the second term is zero since $E(d_i|p_i) = E(d_i)$ for segregating loci. If the conditional variance of d_i is not depending on p_i^{MAC}, then $((2p_i^{MAC} - 1)d_i, d_i) = Var(d_i)E(2p_i^{MAC} - 1)$ where both terms are larger than zero.

Thus, we have shown that, by a careful coding of the model, it is possible to express an *after selection* dependency between functional additive and dominance effects at biallelic QTL as a covariance.

So far, we did not consider the dominant coefficient δ. If, across loci, there is a biological relationship between the magnitudes of a and d (*e.g.*, Bennewitz and Meuwissen 2010), this mechanism will reinforce the previously mentioned dependency between functional additive and dominance effects. In Appendix A, we consider $d_i = \delta_i|a_i|$, and we argue that it often holds that $cov(a_i, d_i) > 0$ when $cov(\delta_i, |a_i|) > 0$ [*BayesD3* model in Wellmann and Bennewitz (2012)] and $cov(a_i, d_i) = 0$ when $cov(\delta_i, |a_i|) = 0$ [*BayesD2* model in Wellmann and Bennewitz (2012)].

Covariance between genotypic additive and dominant effects

In Hardy–Weinberg equilibrium (see Table 1), the covariance between genotypic additive effects u_j and dominance effects v_j (for a random individual j drawn from a population) for one QTL can be derived as:

$$cov(u_j, v_j) = E(u_jv_j) - E(u_j)E(v_j),$$

where the expectations can be computed from expectations of conditional expectations $E(u_jv_j|a, d, p) = 0$ (because the cross product is always zero across the three possible genotypes), $E(u_j|a, d, p) = (2p - 1)a$, and $E(v_j|a, d, p) = 2pqd$, respectively. Therefore,

$$cov(u_j, v_j) = -E((2p - 1)a)E(2pqd).$$

As we have already noted, $E(d) > 0$ (there will be tendency that only overdominant mutations eventually remain in heterozygous state) and therefore the term $E(2pqd)$ is always positive, irrelevant of the coding. When coding genotypes as in the RAC, $E(p) = 0.5$ and thus $E(u_j) = E((2p - 1)a) = 0$. Therefore, $cov(u_j, v_j) = 0$ holds for RAC. When coding genotypes as in the MAC, $E(2p^{MAC} - 1) > 0$, and there will be a tendency that a^{MAC} and d have the same sign [see Equation (2)], so that $E(a^{MAC}) > 0$. In other words, after long-term selection there will be a tendency that the allele with the positive effect will be most frequent, and hence a^{MAC} is positive. To conclude, there will be tendency that both a^{MAC} and d are positive.

Therefore, $E((2p^{MAC} - 1)a^{MAC})$ and $E(2pqd)$ are both positive, and hence for the MAC coding

$$cov(u_j, v_j) = -E((2p^{MAC} - 1)a^{MAC})E(2p^{MAC}q^{MAC}d) < 0 \quad (4)$$

The intuition behind this is that after long-term selection, individual genotypic additive and dominance effects tend to offset each other in polymorphic loci (otherwise there would be fixation).

The expression in Equation (4) extends, assuming linkage equilibrium, to $cov(u_j, v_j) = -\sum E((2p_i^{MAC} - 1)a_i^{MAC})E(2p_iq_id_i) < 0$ for the case of multiple QTL.

For a statistical analysis based on mixed models and genomic data, the assumption is $Var\begin{pmatrix} a \\ d \end{pmatrix} = \begin{pmatrix} \sigma_a^2 & \sigma_{ad} \\ \sigma_{ad} & \sigma_d^2 \end{pmatrix}$, and matrices Z and W are known. From the α being small property, we have that signs of a and d are uncorrelated for RAC but positively correlated for MAC. Therefore, we obtain $\sigma_{ad} = 0$ for RAC [Equation (2)] and $\sigma_{ad} > 0$ for MAC [Equation (3)].

Variance component estimation

According to the theory sketched before, the variance–covariance structure between genotypic additive and dominant effects is:

$$var\begin{pmatrix} u \\ v \end{pmatrix} = var\begin{pmatrix} Za \\ Wd \end{pmatrix} = var\left[\begin{pmatrix} Z & 0 \\ 0 & W \end{pmatrix}\begin{pmatrix} a \\ d \end{pmatrix}\right]$$
$$= \begin{pmatrix} Z & 0 \\ 0 & W \end{pmatrix}var\begin{pmatrix} a \\ d \end{pmatrix}\begin{pmatrix} Z & 0 \\ 0 & W \end{pmatrix}'$$
$$= \begin{pmatrix} Z & 0 \\ 0 & W \end{pmatrix}\begin{pmatrix} I\sigma_a^2 & I\sigma_{ad} \\ I\sigma_{ad} & I\sigma_d^2 \end{pmatrix}\begin{pmatrix} Z' & 0 \\ 0 & W' \end{pmatrix} \quad (5)$$
$$= \begin{pmatrix} ZZ'\sigma_a^2 & ZW'\sigma_{ad} \\ WZ'\sigma_{ad} & WW'\sigma_d^2 \end{pmatrix},$$

where Z and W contain genotypic codings, a and d are additive and dominant SNP effects, σ_a^2 is the additive variance for SNP effects, σ_d^2 is the dominance variance for SNP effects, and σ_{ad} is covariance between additive and dominance SNP effects. For different analyses, elements in matrix Z will be coded using RAC (the reference allele is chosen at random for each locus) or using MAC (the reference allele is the most frequent for each locus).

This (co)variance structure in Equation (5) cannot be fit in usual BLUP or restricted maximum likelihood (REML) software directly, because the covariance matrix is not factorizable as a Kronecker product of a relationship structure times a covariance matrix. A solution for this issue is to use an equivalent model with two additional unknown random effects u^* and v^* (Fernández *et al.* 2017). Let $u^* = Wa$ and $v^* = Zd$; these effects have no biological meaning *per se*. The variance structures for u^* and v^* are $var(u^*) = WW'\sigma_a^2$ and $var(v^*) = ZZ'\sigma_d^2$, respectively. Then, the (co)variance structure for all the four random effects is:

$$var\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{u^*} \\ \boldsymbol{v^*} \\ \boldsymbol{v} \end{pmatrix} = var\begin{pmatrix} \boldsymbol{Za} \\ \boldsymbol{Wa} \\ \boldsymbol{Zd} \\ \boldsymbol{Wd} \end{pmatrix} = var\left[\begin{pmatrix} \boldsymbol{Z} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{W} \end{pmatrix}\begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{a} \\ \boldsymbol{d} \\ \boldsymbol{d} \end{pmatrix}\right]$$

$$= \begin{pmatrix} \boldsymbol{Z} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{W} \end{pmatrix}\begin{pmatrix} \boldsymbol{I}\sigma_a^2 & \boldsymbol{I}\sigma_a^2 & \boldsymbol{I}\sigma_{ad} & \boldsymbol{I}\sigma_{ad} \\ \boldsymbol{I}\sigma_a^2 & \boldsymbol{I}\sigma_a^2 & \boldsymbol{I}\sigma_{ad} & \boldsymbol{I}\sigma_{ad} \\ \boldsymbol{I}\sigma_{ad} & \boldsymbol{I}\sigma_{ad} & \boldsymbol{I}\sigma_d^2 & \boldsymbol{I}\sigma_d^2 \\ \boldsymbol{I}\sigma_{ad} & \boldsymbol{I}\sigma_{ad} & \boldsymbol{I}\sigma_d^2 & \boldsymbol{I}\sigma_d^2 \end{pmatrix}\begin{pmatrix} \boldsymbol{Z'} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W'} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z'} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{W'} \end{pmatrix} \qquad (6)$$

$$= \begin{pmatrix} \boldsymbol{ZZ'}\sigma_a^2 & \boldsymbol{ZW'}\sigma_a^2 & \boldsymbol{ZZ'}\sigma_{ad} & \boldsymbol{ZW'}\sigma_{ad} \\ \boldsymbol{WZ'}\sigma_a^2 & \boldsymbol{WW'}\sigma_a^2 & \boldsymbol{WZ'}\sigma_{ad} & \boldsymbol{WW'}\sigma_{ad} \\ \boldsymbol{ZZ'}\sigma_{ad} & \boldsymbol{ZW'}\sigma_{ad} & \boldsymbol{ZZ'}\sigma_d^2 & \boldsymbol{ZW'}\sigma_d^2 \\ \boldsymbol{WZ'}\sigma_{ad} & \boldsymbol{WW'}\sigma_{ad} & \boldsymbol{WZ'}\sigma_d^2 & \boldsymbol{WW'}\sigma_d^2 \end{pmatrix} = \begin{pmatrix} \sigma_a^2 & \sigma_{ad} \\ \sigma_{ad} & \sigma_d^2 \end{pmatrix} \otimes \begin{pmatrix} \boldsymbol{ZZ'} & \boldsymbol{ZW'} \\ \boldsymbol{WZ'} & \boldsymbol{WW'} \end{pmatrix} = \boldsymbol{K_0} \otimes \boldsymbol{K},$$

where $\boldsymbol{K}$ is an additive-dominance unscaled relationship matrix $\begin{pmatrix} \boldsymbol{ZZ'} & \boldsymbol{ZW'} \\ \boldsymbol{WZ'} & \boldsymbol{WW'} \end{pmatrix}$; $\boldsymbol{K_0}$ is a covariance matrix $\begin{pmatrix} \sigma_a^2 & \sigma_{ad} \\ \sigma_{ad} & \sigma_d^2 \end{pmatrix}$ that associates to SNP effects. Equation (6) is a typical correlated random effects structure, and such structures can be fit using mixed model software.

Nevertheless, variance components in $\boldsymbol{K_0}$ are associated to the scale of SNP effects. To adjust variance components in the $\boldsymbol{K_0}$ matrix to be associated to the scale of individuals, similar to Vitezica *et al.* (2016) and Xiang *et al.* (2016), a genomic relationship matrix $\boldsymbol{G} = \boldsymbol{K}/\{tr(\boldsymbol{K})\}/2n$ is introduced, where $tr(\boldsymbol{K})$ is the trace of the relationship matrix $\boldsymbol{K}$ and $2n$ is twice the number of individuals involved in the matrix $\boldsymbol{K}$. The $\{tr(\boldsymbol{K})\}/2n$ is the average of the diagonal elements in the relationship matrix $\boldsymbol{K}$. Let a constant $k = \{tr(\boldsymbol{K})\}/2n$, then matrix $\boldsymbol{G} = \boldsymbol{K}/k = \begin{pmatrix} \boldsymbol{ZZ'}/k & \boldsymbol{ZW'}/k \\ \boldsymbol{WZ'}/k & \boldsymbol{WW'}/k \end{pmatrix}$. Then, the Equation (6) will change to:

$$var\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{u^*} \\ \boldsymbol{v^*} \\ \boldsymbol{v} \end{pmatrix} = \boldsymbol{K_0} \otimes \boldsymbol{K} = \boldsymbol{K_0} \otimes (\boldsymbol{G} \times k)$$

$$= \begin{pmatrix} k\sigma_a^2 & k\sigma_{ad} \\ k\sigma_{ad} & k\sigma_d^2 \end{pmatrix} \otimes \boldsymbol{G} \qquad (7)$$

$$= \begin{pmatrix} \sigma_A^{*2} & \sigma_{AD}^* \\ \sigma_{AD}^* & \sigma_D^{*2} \end{pmatrix} \otimes \begin{pmatrix} \boldsymbol{ZZ'}/k & \boldsymbol{ZW'}/k \\ \boldsymbol{WZ'}/k & \boldsymbol{WW'}/k \end{pmatrix},$$

where σ_A^{*2}, σ_D^{*2} and σ_{AD}^* are estimated genotypic additive variance, genotypic dominance variance, and covariance between genotypic additive and dominance effects at the level of individuals, and these can be estimated using REML implemented in standard animal breeding software.

Still, these estimated variance components cannot be interpreted as the genetic variances in the population (Legarra 2016). The estimated variance components in Equation (7) should be scaled to the expected variance components of a population as in Legarra (2016) (see Appendix B), as follows:

$$\sigma_u^2 = D_{\boldsymbol{ZZ'}/k}\sigma_A^{*2}, \quad \sigma_v^2 = D_{\boldsymbol{WW'}/k}\sigma_D^{*2}, \quad \sigma_{u,v} = D_{\boldsymbol{ZW'}/k}\sigma_{AD}^*, \qquad (8)$$

where σ_u^2 is the expected genotypic additive variance, σ_v^2 is the expected genotypic dominance variance, $\sigma_{u,v}$ is the covariance between expected genotypic additive and dominance effects, and statistics $D_{\boldsymbol{M}} = mean(diag(\boldsymbol{M})) - mean(\boldsymbol{M})$ for $\boldsymbol{M} = \boldsymbol{ZZ'}/k, \boldsymbol{ZW'}/k, \boldsymbol{WW'}/k$, respectively. Thus, the total genetic variance is $\sigma_{(u+v)}^2 = \sigma_u^2 + \sigma_v^2 + 2\sigma_{u,v} = D_{\boldsymbol{ZZ'}/k}\sigma_A^{*2} + D_{\boldsymbol{WW'}/k}\sigma_D^{*2} + 2D_{\boldsymbol{ZW'}/k}\sigma_{AD}^*$; and the expected correlation between genotypic additive effects $\boldsymbol{u}$ and genotypic dominance effects $\boldsymbol{v}$ is:

$$r_{u,v} = \frac{\sigma_{u,v}}{\sigma_u \sigma_v} = \frac{D_{\boldsymbol{ZW'}/k}\sigma_{AD}^*}{\sqrt{D_{\boldsymbol{ZZ'}/k}\sigma_A^{*2}D_{\boldsymbol{WW'}/k}\sigma_D^{*2}}}, \qquad (9)$$

Replacing $D_{\boldsymbol{ZW'}/k}$ in the formula above by its expectation $-\sum_i (2p_i - 1)(2p_i q_i)/k$ (see Appendix B), which is negative, and using Equations (3) and (7) to see that $\sigma_{AD}^* > 0$, we obtain that $r_{u,v}$ is negative, as also shown in Equation (4).

We derived formulae that can be used to estimate the covariance between genotypic additive and dominant effects. Based on this, four hypothesis can be proposed: (1) association between additive and dominant effects is captured in genomic model by a covariance if the MAC is used, but cannot be captured if the RAC is used; (2) based on Equation (9), the covariance $\sigma_{u,v}$ is negative when the absolute additive effects and dominance coefficients of QTL are positively correlated [*e.g.*, *BayesD3* in Wellmann and Bennewitz (2012)]; (3) a long-term directional selection is a possible cause for $\sigma_{u,v}$ different from 0; and (4) the predictive ability of a

genomic model could be improved if the $\sigma_{u,v}$ is included. A simulation study was used to test these hypotheses.

Genomic models

To obtain mixed models with centered a and d (equivalently, centered u and v) two regressions are needed: in MAC and RAC, a regression on the proportion of heterozygotes [or its counterpart the genomic inbreeding, see Xiang *et al.* (2016)] and in MAC, a regression on the proportion of major alleles.

In this study, the full genomic model (M1) can be written as

$$\text{M1}: \mathbf{y} = \mathbf{1}\mu + \mathbf{m}A + \mathbf{f}b + (\mathbf{I} \quad \mathbf{0})\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{u}^* \end{pmatrix} + (\mathbf{0} \quad \mathbf{I})\begin{pmatrix} \boldsymbol{v}^* \\ \boldsymbol{v} \end{pmatrix} + \mathbf{e},$$

where $\mathbf{y}$ is a vector of user defined phenotypic values, μ is the overall mean, and $\mathbf{m}A$ (only included in MAC, not for RAC) models the regression of phenotype on proportion of most frequent alleles, *e.g.*, $E(\boldsymbol{u}) > 0$; $\mathbf{m}$ is a vector with elements $m_j = \sum_{i=1}^{N} Z_{ji}/N$, matrix Z_{ji} is the element in the incidence matrix for the additive effects for random individual j with MAC, N is the number of SNP markers used, and A is the regression coefficient, which needs to be estimated; $\mathbf{f}b$ models the inbreeding depression (Vitezica *et al.* 2016; Xiang *et al.* 2016) (*e.g.*, $E(\boldsymbol{v}) > 0$), where $\mathbf{f}$ is inbreeding coefficient and b is the inbreeding depression parameter per unit of inbreeding, which needs to be estimated. Vectors $\boldsymbol{u}$, $\boldsymbol{u}^*$, $\boldsymbol{v}^*$, and $\boldsymbol{v}$ are genotypic additive and dominance individual effects as in Equation (7), matrix $\mathbf{I}$ is an identity matrix to assign genotypic effects $\boldsymbol{u}$ and $\boldsymbol{v}$ to the corresponding phenotypic records, and $\mathbf{e}$ is the overall residual. The expectation of both $\boldsymbol{u}$ and $\boldsymbol{v}$ are zero after inclusion of $\mathbf{m}A$ and $\mathbf{f}b$ in the model. The covariance structure of random effects $\boldsymbol{u}$, $\boldsymbol{u}^*$, $\boldsymbol{v}^*$, and $\boldsymbol{v}$ are as in the Equation (7). If $\boldsymbol{u}$ has n levels, then an additional n levels for $\boldsymbol{u}^*$ (from $n + 1$ to $2n$) need to be declared to achieve the factorizable structure of the covariance matrix. Similarly, declare $\boldsymbol{v}^*$ variables with levels 1 to n, and then levels of $\boldsymbol{v}$ are from $n + 1$ to $2n$.

The M1 model was compared to a submodel with covariance between $\boldsymbol{u}$ and $\boldsymbol{v}$ equal to zero ($\sigma_{u,v} = 0$) (Vitezica *et al.* 2016; Xiang *et al.* 2016), as

$$\text{M2}: \quad \mathbf{y} = \mathbf{1}\mu + \mathbf{m}A + \mathbf{f}b + \boldsymbol{u} + \boldsymbol{v} + \mathbf{e},$$

where $\mathbf{m}A$ was only included for MAC, but not for RAC; $var(\boldsymbol{u}) = \mathbf{Z}\mathbf{Z}'\sigma_a^2$ and $var(\boldsymbol{v}) = \mathbf{W}\mathbf{W}'\sigma_d^2$. For both M1 and M2, variance components and associated SE were estimated by GREML using AIREMLf90 (Misztal *et al.* 2002) in different scenarios. Estimated genetic parameters were scaled as in Equation (8).

In addition, a model (M3) with orthogonal breeding values ($\boldsymbol{u}^o$) and dominance deviations ($\boldsymbol{v}^o$) (Vitezica *et al.* 2013) including genomic inbreeding depression was used. For M3, only RAC was investigated.

$$\text{M3}: \mathbf{y} = \mathbf{1}\mu + \mathbf{f}b + \boldsymbol{u}^o + \boldsymbol{v}^o + \mathbf{e},$$

where $\boldsymbol{u}^o = \mathbf{Z}^o\boldsymbol{\alpha}$ and $\boldsymbol{v}^o = \mathbf{W}^o\mathbf{d}$. Orthogonal incidence matrices $\mathbf{Z}^o$ and $\mathbf{W}^o$ were coded as follows:

$$\mathbf{Z}^o = \begin{cases} 2 - 2p \\ 1 - 2p \\ 0 - 2p \end{cases} \text{ for genotypes } \begin{cases} BB \\ Bb \\ bb \end{cases}; \quad \mathbf{W}^o = \begin{cases} -2q^2 \\ 2pq \\ -2p^2 \end{cases} \text{ for }$$

genotypes $\begin{cases} BB \\ Bb \\ bb \end{cases}$ and p is the allele frequency of the second allele for each locus. Note that M3-RAC is strictly equivalent to M3-MAC as the cross products $\mathbf{Z}^o\mathbf{Z}^{o'}$ and $\mathbf{W}^o\mathbf{W}^{o'}$ are invariant to MAC or RAC coding, and therefore only M3-RAC is investigated.

The variance proportions of genotypic additive variance (h_u^2) was calculated as $h_u^2 = \sigma_u^2/\sigma_p^2$, where σ_u^2 was the variance of genotypic additive effects; σ_p^2 was the phenotypic variance, equal to the sum of variance of total genotypic effects ($\sigma_{(u+v)}^2$) and residual variance (σ_e^2). The dominant variance proportion (h_v^2) was calculated as $h_v^2 = \sigma_v^2/\sigma_p^2$, where σ_v^2 was variance of genotypic dominance effects. The broad sense heritability was calculated as $H^2 = \sigma_{(u+v)}^2/\sigma_p^2$. Note that variance proportions of total genotypic variance (H^2) is not the sum of h_u^2 and h_v^2 because the covariance between the genotypic additive and dominant effects is not equal to zero.

The goodness of fit of the models was measured by the $-2log(likelihood)$. The superiority of M1 over M2 was tested by a likelihood ratio test (LRT), which was calculated as $-2log(likelihood\ for\ M1) - (-2log(likelihood\ for\ M2)$. The differences were assumed to follow a χ^2 distribution with 1 d.f.

Simulation

Phenotypic and genotypic data sets were simulated by the software QMSim, version 1.10 (Sargolzaei and Schenkel 2009). The trait was designed to be controlled by both additive and dominant gene actions, and the population mimicked a pig population. Heritability in the narrow sense was 0.38 and the additive genotypic variance σ_u^2 was 0.66. The phenotypic variance σ_p^2 was 1.74 and the dominance genotypic variance σ_v^2 was set to 0.174. No polygenic effect was simulated.

The simulation steps are presented in Figure 1. In the first simulation step for creating the historical population (HP), 2500 discrete generations with a constant population size of 500 were simulated. To mimic the bottleneck of a pig population, from generation 2501 onwards, the population size was gradually reduced to 65 at generation 2530. Then, from the generation 2531 to 2535, the population size gradually increased from 65 to 220. At the last generation of the HP (generation 2535), the number of males was 20 and the number of females was 200. The LD decay of chromosome 1 was checked and it was in line with that in a real Danish Landrace/Yorkshire pig population (Wang *et al.* 2013). In the second simulation step, a recent population 1 (RP1) was generated by randomly mating the 20 males and 200 females from the last generation in the HP. Each female had 10 offspring and the sex proportions for progeny were fixed to 0.5. The goal of the RP1 was to expand the population size. The RP1 only had one generation (generation 2536), and there were 1000 males and 1000 females existing in RP1. In the next simulation step, 100 males and 500 females were
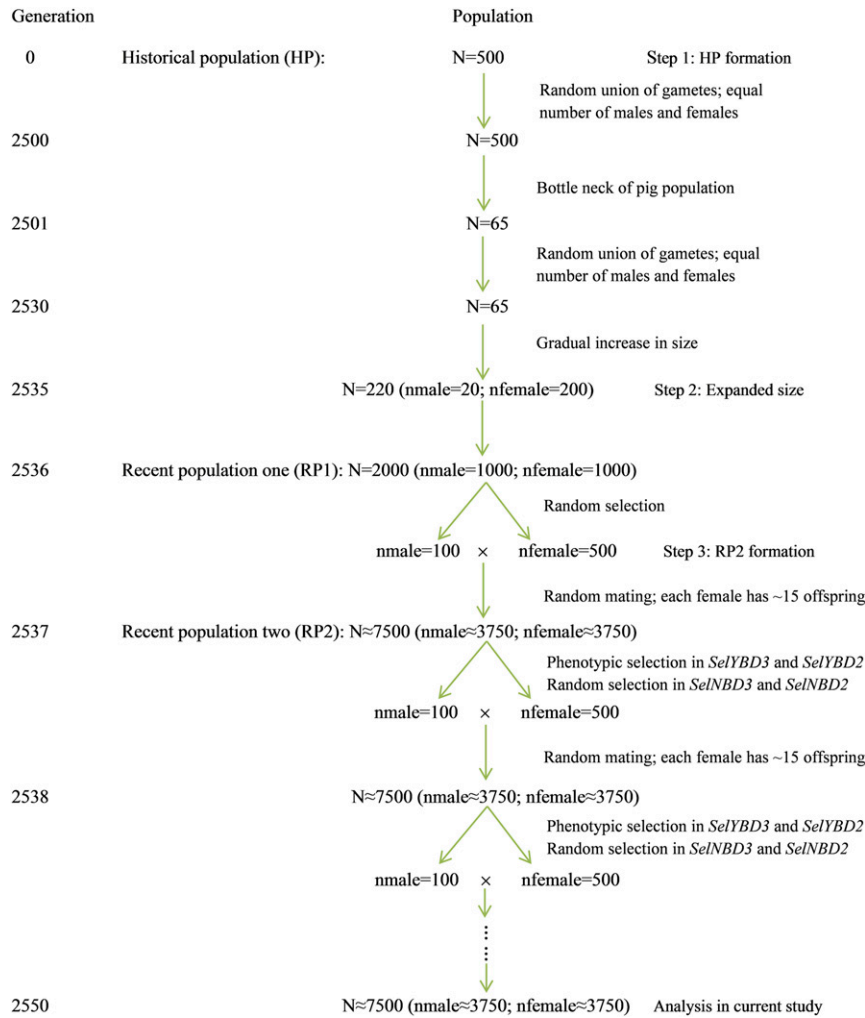
Figure 1 Simulation steps for creating a pig populations. HP, historical population; RP, recent population.

randomly selected from the RP1 as the founders of recent population 2 (RP2). The RP2 had 13 generations (generation 2537–2550). For each of these generations, 100 males and 500 females who either had the highest phenotypic values (for a scenario with phenotypic selection) or were randomly selected (for a scenario with no selection) from the former generation were kept as parents for the next generation. Each female had ~15 progeny (from 7 to 22) to mimic a real pig population. The sex ratio for the progeny was fixed as 0.5. For the last generation in RP2, there were ~7500 individuals in total. All the simulation steps were repeated 10 times to create 10 independent data sets for the further analysis.

The genome consisted of 18 autosomes of 120 cM each. For the first generation in the HP, each chromosome had 200 segregating biallelic QTL loci and 18,200 biallelic marker loci randomly located (thus, in total, 3600 QTL loci and 327,600 marker loci). In the first generation, allele frequencies for QTL and markers loci were 0.5, and the recurrent mutation rate was 2.5×10^{-4}.

Since software QMSim cannot simulate dominance effects, in the RP2 (from generation 2537 onwards), the option of "ebv_est = external_bv" in QMSim was used to base selection decisions on a user provided file. In this study, selection decisions were made according to phenotypic values (phenotypic selection) and thus, the provided file contained user-defined phenotypes for each individual.

These phenotypic values were simulated as follows. For each QTL, additive effects a and dominance effects d were constant across generations. The additive effects a of 3600 QTL loci were drawn from a Student's t-test distribution with 2.5 d.f. [$a = rt(3600, 2.5)$ in R] (Wellmann and Bennewitz 2012). Then, dominance functional effects d were drawn in two different ways: *BayesD2* and *BayesD3*, as is proposed in Wellmann and Bennewitz (2012). For *BayesD2*, the magnitudes of a and d were related as $d = \delta|a|$ and $cor(\delta_i, |a_i|) = 0$. Bennewitz and Meuwissen (2010) showed that dominance coefficients δ follow a normal distribution with mean 0.193 and SD 0.312. Therefore, the dominance coefficients were simulated as $\delta = rnorm(3600, mean = 0.193, sd = 0.312)$ and $d = \delta|a|$ (element-wise multiplication) in R. For the *BayesD3*, the magnitudes of a and d were related as $d = \delta|a|$ and $cor(\delta_i, |a_i|) > 0$. Wellmann and Bennewitz (2012) assumed dominance coefficients δ follow a conditional normal distribution as $\delta||a| \sim N(\mu_\Delta(|a|/s), \sigma_\Delta^2)$, where s is a scaling parameter and $\mu_\Delta(x) = x/(s_\Delta + x)$ with $s_\Delta > 0$. To generate

Table 2 Summary for different scenarios

Scenario	*SelYBD3*		*SelYBD2*		*SelNBD3*		*SelNBD2*	
Model	$\mathbf{y} = \mathbf{1}\mu + (\mathbf{m}A) + \mathbf{f}b + (\mathbf{I} \quad \mathbf{0})\begin{pmatrix} \mathbf{u} \\ \mathbf{u}^* \end{pmatrix} + (\mathbf{0} \quad \mathbf{I})\begin{pmatrix} \mathbf{v}^* \\ \mathbf{v} \end{pmatrix} + \mathbf{e},\ var\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{ZZ}'\sigma_a^2 & \mathbf{ZW}'\sigma_{ad} \\ \mathbf{WZ}'\sigma_{ad} & \mathbf{WW}'\sigma_d^2 \end{pmatrix}$							
Model 2	$\mathbf{y} = \mathbf{1}\mu + (\mathbf{m}A) + \mathbf{f}b + \mathbf{u} + \mathbf{v} + \mathbf{e},\ var(\mathbf{u}) = \mathbf{ZZ}'\sigma_a^2,\ var(\mathbf{v}) = \mathbf{WW}'\sigma_d^2$							
Selection	Phenotypic selection		Phenotypic selection		Random selection		Random selection	
a, d relation	$d = \delta|a|;\ cor(\delta,|a|) > 0$		$d = \delta|a|;\ cor(\delta,|a|) = 0$		$d = \delta|a|;\ cor(\delta,|a|) > 0$		$d = \delta|a|;\ cor(\delta,|a|) = 0$	
Allele coding	MAC	RAC	MAC	RAC	MAC	RAC	MAC	RAC

Four scenarios—*SelYBD3*, *SelYBD2*, *SelNBD3*, and *SelNBD2*—were outcomes of cross combinations: phenotypic selection (selection Yes) or random selection (selection No) in combination with two ways of relating functional additive effects *a* and dominance effects $d = \delta|a|$: BayesD3 with $cor(\delta,|a|) > 0$; BayesD2 with $cor(\delta,|a|) = 0$. Two genotypic models were investigated: model 1 considers the covariance between genotypic additive and dominance effects, and model 2 considers independent additive and dominance effects. Genotypes were coded in both MAC and RAC ways. MAC, major allele coding; RAC, random allele coding; *SelNB*, Selection No Bayes; *SelYB*, Selection Yes Bayes.

the dominance coefficients $\boldsymbol{\delta}$ following a distribution similar to that in *BayesD2*, $s_\Delta = 4$ was chosen and $\boldsymbol{\delta} = |\boldsymbol{a}|/(4 + |\boldsymbol{a}|) + 0.3*rnorm(3600)$ was used in R for *BayesD3*. The bivariate plots between $\boldsymbol{a}$ and $\boldsymbol{d}$ in the 10 simulated data sets are shown in the supplemental files (Supplemental Material, Figure S1 for *BayesD2* and Figure S2 for *BayesD3*). The mean empirical correlation (s.e.) between $|\boldsymbol{a}|$ and $\boldsymbol{\delta}$ over 10 repetitions was, -0.000013 (0.0015) for *BayesD2* and $= 0.355$ (0.039) for *BayesD3*.

Then, for each individual, based on the genotypes of 3600 QTL loci and the corresponding QTL effects, "true" genotypic additive effects $\boldsymbol{u}$ and dominance effects $\boldsymbol{v}$ of each individual were calculated. Residual effects were sampled from a normal distribution. Afterward, the phenotype for each individual was calculated as the sum of an overall mean, true additive effects, dominance effects, and residual effects. Internally, QMSim used these phenotypes as the selection criteria for the next generation in scenarios with selection. In scenarios with no selection, replacement at each generation was done at random.

At the last generation of RP2, on average ∼650 QTL loci were segregating. Once the simulation was finished, each SNP marker simulated by QMSim was retained for the subsequent analyses if the minor allele frequency was larger than 0.05. In total, there were ∼50K segregating markers retaining. The parameter file for QMSim can be found in the supplemental material.

Four scenarios were studied (Table 2). Two scenarios had phenotypic selection: *SelYBD3* ("Selection Yes BayesD3"), where $\boldsymbol{a}$ and $\boldsymbol{d}$ were related as in *BayesD3*; and *SelYBD2* ("Selection Yes BayesD2"), where $\boldsymbol{a}$ and $\boldsymbol{d}$ were related as in *BayesD2*. Two scenarios had no selection: *SelNBD3* ("Selection No BayesD3") and *SelNBD2* ("Selection No BayesD2"), with $\boldsymbol{a}$ and $\boldsymbol{d}$ related as in BayesD3 and BayesD2, respectively. Within each scenario, MAC (most frequent allele as reference) in combination with M1 and M2 (M1-MAC and M2-MAC) and RAC (random allele as reference) in combination with M1, M2, and M3 (M1-RAC, M2-RAC, and M3-RAC) were applied. For each scenario, 10 replicates were run.

Predictive abilities

The effect on predictive abilities of including the covariance between the genotypic additive and dominance effects in the genomic model was investigated. Prediction was performed by using M1, M2, and M3 in the scenarios with selection. In each replicate, 20% of individuals in the last generation of RP2 were randomly masked and put into the validation population; the remaining 80% of individuals were used as the training population. The predictive ability was measured as the correlation, in the validation population, between true total genetic values g that were known from the simulation and the estimated total genetic values $\hat{g} = \mathbf{m}\hat{A} + \mathbf{f}\hat{b} + \hat{u} + \hat{v}$ for MAC and $\hat{g} = \mathbf{f}\hat{b} + \hat{u} + \hat{v}$ for RAC. Bias was measured as the regression coefficient of total g on estimated genetic values $\hat{g}$. The Hotelling–Williams *t*-test at a 95% confidence level was applied to evaluate the significance for the differences of predictive abilities between M1, M2, and M3 within different scenarios. Table 2 summarizes the analyses for the different scenarios.

Data availability

The authors state that all data necessary for confirming the conclusions presented in the manuscript are represented fully within the manuscript. Supplemental material available at Figshare: https://doi.org/10.25386/genetics.6111068.

Results and Discussion

Regressions on proportion of most frequent alleles and on genomic inbreeding in M1

Average regression coefficients (s.e.) of phenotypes on the proportion of the most frequent alleles (*A*) in the model were 12.61 (1.32) for *SelYBD3*, 12.10 (1.97) for *SelYBD2*, 0.02 (2.32) for *SelNBD3*, and 0.02 (2.51) for *SelNBD2*. The positive *A* for scenarios with selection is in agreement with the tendency that after long-term selection, additive effects for most frequent alleles are positive. The average (s.e.) inbreeding depression parameters per unit of inbreeding (*b*) were -6.36 (1.73) for *SelYBD3*, -6.20 (1.45) for *SelYBD2*, -6.63 (1.34) for *SelNBD3*, and -6.18 (1.52) for *SelNBD2*. These negative inbreeding depression parameters confirm the phenomenon of inbreeding depression.

Variance components

Variance components were calculated using QTL (to obtain the true values) and also, using markers, for the four scenarios combining models M1 (with correlated additive and
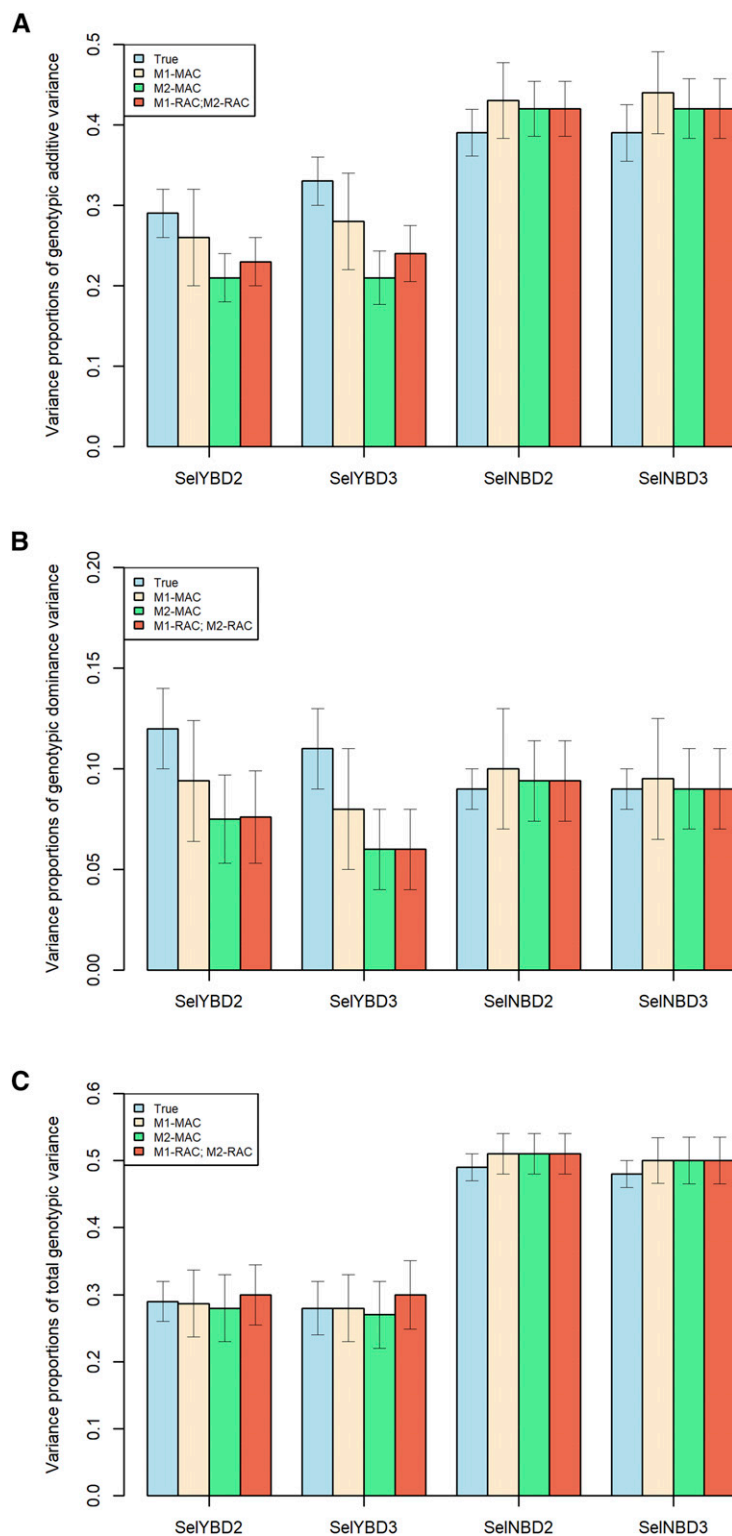
Figure 2 Variance proportions of genotypic additive variance (a) genotypic dominance variance (b) and total genotypic variance (c) over total phenotypic variance. "True" represents the results that were calculated based on the QTL loci in the last generation (generation 2550); "M1-MAC" indicates MAC in combination with M1 was applied; "M2-MAC" indicates MAC in combination with M2 was applied; "M1-RAC; M2-RAC" represents one of the following combinations: RAC in combination with M1 and RAC in combination with M2 was applied in the respective scenario; *SelYBD2*, *SelYBD3*, *SelYBD2*, and *SelYBD3* were the four studied scenarios. M1, model 1; M2, model 2; MAC, major allele coding; RAC, random allele coding; *SelYB*, Selection Yes Bayes.

dominant genotypic effects) or M2 (with uncorrelated effects) with MAC (most frequent allele as reference) or RAC (random allele as reference). This gives four combinations (M1-MAC, M2-MAC, M1-RAC, and M2-RAC). Results were the mean of 10 replicates. Variance proportions of genotypic additive (h_u^2), genotypic dominance (h_v^2), and total genotypic (H^2) variance

over phenotypic variance are presented in Figure 2. All the other variance components are in Table S1 in the supplemental material.

For scenarios with phenotypic selection (*SelYBD2* and *SelYBD3*), estimated variance components in M1-MAC were not statistically significantly different from the true ones,
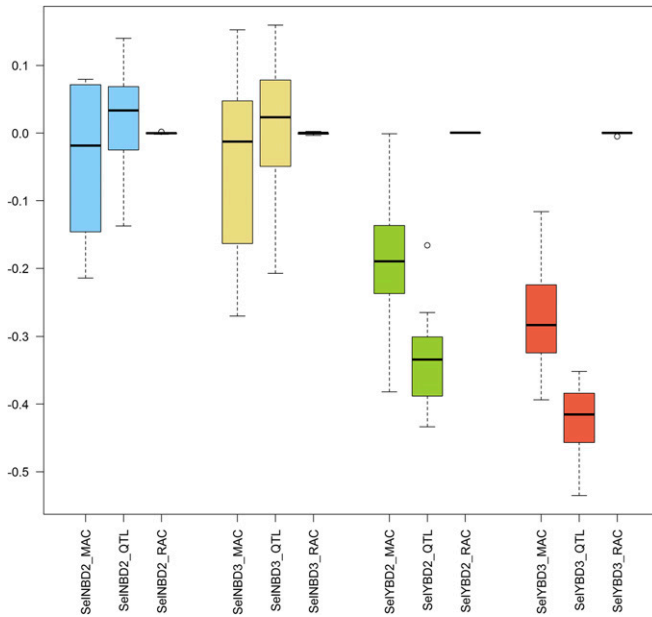
Figure 3 Genetic correlations between genotypic additive and dominance effects based on model 1 in different scenarios across 10 repetitions. QTL indicate the true correlations that were calculated based on QTL; either MAC and RAC was applied to code genotypes. *SelNBD2*, *SelNBD3*, *SelYBD2*, and *SelYBD3* were the four studied scenarios. MAC, major allele coding; RAC, random allele coding; *SelNB*, Selection No Bayes; *SelYB*, Selection Yes Bayes.

Table 3 Average goodness of fit and $-2log(likelihood)$ of M 1 and M2 across 10 repetitions, and likelihood ratio test between M1 and M2 in different scenarios

Scenario	Matrix Z coding	M1	M2	χ^2 value	*P*-value
SelYBD3	MAC	−5463.49	−5437.12	26.37	1.41e−7
	RAC	−5385.44	−5384.91	0.88	0.1741
SelYBD2	MAC	−5412.52	−5398.66	13.90	9.64e−5
	RAC	−5305.91	−5303.72	2.19	0.0695
SelNBD3	MAC	−4643.10	−4640.99	2.10	0.0736
	RAC	−4637.95	−4636.43	1.52	0.1095
SelNBD2	MAC	−4524.30	−4523.15	1.16	0.1407
	RAC	−4520.11	−4518.41	1.70	0.0961

MAC and RAC represent allele coding applied to code genotypes; *SelNBD2*, *SelNBD3*, *SelYBD2*, and *SelYBD3* were the four studied scenarios. M1, model 1; M2, model 2; MAC, major allele coding; RAC, random allele coding; *SelNB*, Selection No Bayes; *SelYB*, Selection Yes Bayes.

while results in the other cases (M1-RAC, M2-MAC, and M2-RAC) were slightly but statistically different from the true ones. Similarly, all the calculated genetic parameters (h_u^2, h_v^2 and H^2) in M1-MAC were not significantly different from the true values, while values in the other cases (M1-RAC, M2-MAC, and M2-RAC) were slightly different from the true ones in these two scenarios.

For scenarios with random selection (*SelNBD2* and *SelNBD3*), estimated genetic parameters in different cases did not dramatically deviate from the true ones. In all cases, σ_u^2, h_u^2 and H^2 values were slightly overestimated. Estimates from M1-MAC did not show advantages over those from other combinations.

Overall, the variance components estimated with model M1 were close to the true values in different scenarios. In animal and plant breeding, most traits have experienced long-term selection. Thus, to estimate the genetic parameters more precisely, a genomic model considering covariances between genotypic additive and dominance effects (like M1) seems appropriate.

Genetic correlations

True and estimated genetic correlations between genotypic additive effects and genotypic dominance effects ($r_{u,v}$) for combinations M1-MAC and M1-RAC in different scenarios are shown in Figure 3. When RAC was used, the $r_{u,v}$ was almost 0 in any scenario. However, when the most frequent allele was the reference (MAC), scenarios with phenotypic selection (*SelYBD3* and *SelYBD2*) yielded negative estimates

of $r_{u,v}$. In such two scenarios, the absolute values of estimated $r_{u,v}$ were slightly lower than the absolute values of true $r_{u,v}$. For scenarios with random selection (*SelNBD3* and *SelNBD2*), both true and estimated $r_{u,v}$ were around zero.

Results showed that when a and d are related (Bennewitz and Meuwissen 2010; Wellmann and Bennewitz 2012), a negative correlation may be generated between the genotypic additive and dominance effects after long-term selection. The long-term directional selection seems to be a precondition for producing such correlation. M1-MAC can capture part of such correlation.

LRT

The goodness of fit of M1 and M2 in different scenarios is shown as $-2log(likelihood)$ in the third and fourth columns in Table 3 for both MAC and RAC, respectively. For each allele coding way, M1 always had smaller numeric values of $-2log(likelihood)$ than M2 within different scenarios, which indicated that the M1 fitted the data set better than the M2.

For all the four scenarios, LRT showed no significant differences in fitting the data set between M1 and M2 ($p > 0.05$) when RAC was applied. However, when MAC was used, M1 fitted the data set significantly better than M2 in scenarios with phenotypic selection (*SelYBD3* and *SelYBD2*). For scenarios with random selection (*SelNBD3* and *SelNBD2*), there were no significant differences of goodness of fit between M1 and M2, no matter which allele coding was applied. However, it can be seen that within each scenario, there was a tendency that M1 fitted the data set better than M2, which suggested that a genomic model including the relationships between genotypic additive and dominance effects fits the simulated data better than a model without considering such relationships.

Predictive abilities

Based on the results of LRT, M1 only provided a better fit for the data set than M2 in scenarios with phenotypic selection (*SelYBD3* and *SelYBD2*) when MAC was used for coding SNP markers. Besides, in animal breeding, most traits have experienced long-term selection. Therefore, predictive abilities

Table 4 Average predictive abilities and regression coefficients of total genetic values on predicted total genetic values with the respective SE in scenarios *SelYBD3* and *SelYBD2*

Scenario	*SelYBD3*				*SelYBD2*			
Coding	MAC	MAC	RAC	RAC	MAC	MAC	RAC	RAC
Model	M1	M2	M2	M3	M1	M2	M2	M3
$cor(\hat{g}, g)$	0.746[a] (0.031)	0.739[b] (0.031)	0.734[b] (0.034)	0.741[a] (0.033)	0.732[a] (0.027)	0.729[b] (0.028)	0.727[b] (0.028)	0.730[a] (0.027)
Regression coefficients	1.014 (0.046)	1.013 (0.050)	1.013 (0.051)	1.017 (0.046)	1.014 (0.046)	1.018 (0.045)	1.024 (0.039)	1.025 (0.039)

SelYBD2 and *SelYBD3* were the two scenarios with phenotypic selection; M1, M2, and M3 indicate models 1, 2, and 3, respectively. Numbers in the parenthesis are the SE. MAC, major allele coding; RAC, random allele coding; *SelYB*, Selection Yes Bayes.
[a,b] Significant differences ($P < 0.05$) by Hotelling–Williams *t*-test.

obtained from using M1 and M2 were only compared in these two scenarios (M1-MAC and M2-MAC). Predictive abilities were also computed for M2-RAC, because this model is commonly used in other dominance studies (Su *et al.* 2012; Vitezica *et al.* 2016; Xiang *et al.* 2016). Furthermore, M3-RAC was also applied so that the predictive abilities from coding genotypes in a classical (orthogonal) manner (Falconer and Mackay 1996; Vitezica *et al.* 2013) can be compared to that from coding genotypes in a genotypic manner directly.

Predictive abilities and unbiasedness in different scenarios are compared in Table 4. Overall, the Hotelling–Williams *t*-test indicated that predictive abilities from M1 were significantly higher than those from M2 ($p < 0.05$), but similar to those from M3 ($p > 0.05$). When MAC was applied, predictive abilities from M1 (M1-MAC) were ~1% and 0.4% higher than those from M2 (M2-MAC) for *SelYBD3* and *SelYBD2*, respectively. This result is in agreement with LRT showing that goodness of fit of M1 was significantly superior to that of M2. In addition, it also implies that the relationships between additive and dominant effects were appropriately captured. The differences of predictive abilities between M1-MAC and M2-RAC increased to 1.6% and 0.7% for *SelYBD3* and *SelYBD2*, respectively. These differences indicated that consideration of the covariance between genotypic additive and dominance effects via using MAC would possibly generate a small extra genetic gain. For M2, when MAC was applied (M2-MAC), the predictive abilities were ~0.7% and 0.3% higher than using RAC (M2-RAC). This result indicated that predictive abilities benefited from the inclusion of regression of phenotype on the proportion of the most frequent alleles in the model, even if the covariance between additive and dominance effects were not considered in the M2. This result again recommends the use of MAC in other genomic evaluation models to enhance their predictive abilities, which is in line with the smaller numeric values of $-2log(likelihood)$ for M1 than for M2. In terms of unbiasedness, the regression coefficients observed from M1 were slightly closer to 1 than those from M2 in scenario *SelYBD2*, but there was no clear trend in scenario *SelYBD3*.

However, when comparing the predictive abilities from M3-RAC with other combinations, except for M1-MAC, M3-RAC performed slightly better than M2-MAC and M2-RAC. This phenomenon indicates that when the covariance

between additive and dominant effects is ignored, genomic evaluation via coding genotypes in the genotypic manner may work worse than coding genotypes in the classical, orthogonal manner (in terms of breeding values and dominance deviations). However, when considering the covariance between additive and dominance effects, coding genotypes in the MAC (M1-MAC) increased predictive abilities ~0.7% and 0.3% compared to M3-RAC, which suggested the use of MAC in combination with models considering the covariance between additive and dominant effects. In terms of unbiasedness, the regression coefficients observed from M3 were slightly further from 1 than those from M1 and M2.

Overall, the comparisons of the different models and allele codings confirmed that the predictive ability of a genomic model could slightly improve if the $\sigma_{u,v}$ is included (M1-MAC). The inclusion of such covariance does not need extra computing time. The computing time for estimating variance components is similar for M1-MAC and M2-MAC, which is slightly shorter than that for M3-RAC/M3-MAC.

Wellmann and Bennewitz (2012) also derived an reproducing kernel Hilbert space (RKHS) model for estimating total genetic values that assumed a correlation between additive and dominance effects, which was also a genomic BLUP model. This model was derived on the basis of *BayesD2*. A property in the RKHS model is that the covariance among two genotypic values depends on the assumed *a priori* joint distribution of a and d, *i.e.*, $var(a)$, $var(d)$, and $cov(a, d)$, which were obtained from theoretical arguments and used for computing the Kernel matrix. We prefer instead to estimate these parameters explicitly from the data sets. Even though standard software can be used for prediction in the RKHS model, it cannot be used for the estimation of genetic parameters.

Conclusions

We first presented a novel and simple way of incorporating, in populations under selection, the correlation between genotypic additive and dominance effects in a genomic model that can be implemented using standard mixed model software. Our study showed that: if the population is in HW equilibrium and the absolute additive effects and dominance coefficients of QTL are positively correlated [*e.g.*, *BayesD3* in Wellmann and Bennewitz (2012)], after directional selection, a negative correlation between genotypic additive and dominance effects is expected. A genomic model based on a reference

allele that is the most frequent one at each locus can capture part of the negative genetic correlation between genotypic additive and dominant effects, but this cannot be captured if the RAC is used. This new approach is applied to a directional selected trait controlled by both additive and dominant gene actions. When such correlation is taken into account in the model, accuracies of estimated total genetic values can be improved by up to 1.5% in genomic evaluation while bias is slightly reduced.

Acknowledgments

Literature Cited

Aliloo, H., J. E. Pryce, O. González-Recio, B. G. Cocks, and B. J. Hayes, 2016 Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. Genet. Sel. Evol. 48: 8. https://doi.org/10.1186/s12711-016-0186-0

Aliloo, H., J. Pryce, O. González-Recio, B. Cocks, M. Goddard et al., 2017 Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. J. Dairy Sci. 100: 1203–1222. https://doi.org/10.3168/jds.2016-11261

Bennewitz, J., and T. Meuwissen, 2010 The distribution of QTL additive and dominance effects in porcine F2 crosses. J. Anim. Breed. Genet. 127: 171–179. https://doi.org/10.1111/j.1439-0388.2009.00847.x

Bennewitz, J., C. Edel, R. Fries, T. H. Meuwissen, and R. Wellmann, 2017 Application of a Bayesian dominance model improves power in quantitative trait genome-wide association analysis. Genet. Sel. Evol. 49: 7. https://doi.org/10.1186/s12711-017-0284-7

Caballero, A., and P. D. Keightley, 1994 A pleiotropic nonadditive model of variation in quantitative traits. Genetics 138: 883–900.

Charlesworth, B., and D. Charlesworth, 2010 Elements of Evolutionary Genetics. Roberts and Company Publishers, Greenwood Village, CO.

Crow, J. F., and M. Kimura, 2009 An Introduction to Population Genetics Theory. Blackburn Press, Caldwell, NJ.

Ertl, J., A. Legarra, Z. G. Vitezica, L. Varona, C. Edel et al., 2014 Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. Genet. Sel. Evol. 46: 40. https://doi.org/10.1186/1297-9686-46-40

Falconer, D. S., and T. F. C. Mackay, 1996 Introduction to Quantitative Genetics, Ed. 4. Pearson Education Limited, LONGMAN, Harlow, United Kingdom.

Fernández, E., A. Legarra, R. Martínez, J. P. Sánchez, and M. Baselga, 2017 Pedigree-based estimation of covariance between dominance deviations and additive genetic effects in closed rabbit lines considering inbreeding and using a computationally simpler equivalent model. J. Anim. Breed. Genet. 134: 184–195. https://doi.org/10.1111/jbg.12267

Gilmour, A. R., B. Gogel, B. Cullis, R. Thompson, and D. Butler, 2009 ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, United Kingdom.

Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4: e1000008. https://doi.org/10.1371/journal.pgen.1000008

Legarra, A., 2016 Comparing estimates of genetic variance across different relationship models. Theor. Popul. Biol. 107: 26–30. https://doi.org/10.1016/j.tpb.2015.08.005

Lynch, M., and B. Walsh, 1998 Genetics and Analysis of Quantitative Traits. Sinauer Associates Inc, Sunderland, MA.

Madsen, P., and J. Jensen, 2013 A User's Guide to DMU. Version 6, Release 5.2. Center for Quantitative Genetics and Genomics. Dept. of Molecular Biology and Genetics. Aarhus University, Tjele, Denmark.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet et al., 2002 BLUPF90 and related programs (BGF90). Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France, pp. 21–22.

Muñoz, P. R., M. F. Resende, S. A. Gezan, M. D. V. Resende, G. de los Campos et al., 2014 Unraveling additive from nonadditive effects using genomic relationship matrices. Genetics 198: 1759–1768. https://doi.org/10.1534/genetics.114.171322

Sargolzaei, M., and F. S. Schenkel, 2009 QMSim: a large-scale genome simulator for livestock. Bioinformatics 25: 680–681. https://doi.org/10.1093/bioinformatics/btp045

Silió, L., C. Barragán, A. I. Fernández, J. García-Casco, and M. C. Rodríguez, 2016 Assessing effective population size, coancestry and inbreeding effects on litter size using the pedigree and SNP data in closed lines of the Iberian pig breed. J. Anim. Breed. Genet. 133: 145–154. https://doi.org/10.1111/jbg.12168

Sorensen, D., R. Fernando, and D. Gianola, 2001 Inferring the trajectory of genetic variance in the course of artificial selection. Genet. Res. 77: 83–94 (erratum: Genet. Res. 77: 297).

Speed, D., and D. J. Balding, 2015 Relatedness in the post-genomic era: is it still useful? Nat. Rev. Genet. 16: 33–44. https://doi.org/10.1038/nrg3821

Su, G., O. F. Christensen, T. Ostersen, M. Henryon, and M. S. Lund, 2012 Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One 7: e45293. https://doi.org/10.1371/journal.pone.0045293

Vitezica, Z. G., L. Varona, and A. Legarra, 2013 On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genet. 194: 1223–1230. https://doi.org/10.1534/genetics.113.155176

Vitezica, Z. G., L. Varona, M. J. Elsen, I. Misztal, W. Herring et al., 2016 Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. Genet. Sel. Evol. 48: 6. https://doi.org/10.1186/s12711-016-0185-1

Wang, L., P. Sørensen, L. Janss, T. Ostersen, and D. Edwards, 2013 Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds. BMC Genet. 14: 115. https://doi.org/10.1186/1471-2156-14-115

Wellmann, R., and J. Bennewitz, 2011 The contribution of dominance to the understanding of quantitative genetic variation. Genet. Res. 93: 139–154. https://doi.org/10.1017/S0016672310000649

Wellmann, R., and J. Bennewitz, 2012 Bayesian models with dominance effects for genomic evaluation of quantitative traits. Genet. Res. 94: 21–37. https://doi.org/10.1017/S0016672312000018

Xiang, T., O. F. Christensen, Z. G. Vitezica, and A. Legarra, 2016 Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. Genet. Sel. Evol. 48: 98.

Communicating editor: M. Calus

Appendix A

Here, we consider $d_i = \delta_i|a_i|$ and argue that it often holds that $cov(a_i, d_i) > 0$ when $cov(\delta_i, |a_i|) > 0$ [*BayesD3* model in Wellmann and Bennewitz (2012)] and $cov(a_i, d_i) = 0$ when $cov(\delta_i, |a_i|) = 0$ [*BayesD2* model in Wellmann and Bennewitz (2012)].

In the main part of the paper, we have derived that the covariance between additive and dominance effects is determined by $Var(d_i)$, and since $E(d_i^2) = E(\delta_i^2 a_i^2) = E(\delta_i^2)E(a_i^2) + cov(\delta_i^2, a_i^2)$, and $E(d_i) = E(\delta_i|a_i|) = E(|a_i|)E(\delta_i) + cov(\delta_i, |a_i|)$, we see that the result holds when $cov(\delta_i^2, a_i^2) - cov(\delta_i, |a_i|)^2 - 2E(|a_i|)E(\delta_i)cov(\delta_i, |a_i|) > 0$ when $cov(\delta_i, |a_i|) > 0$, and $cov(\delta_i^2, a_i^2) = 0$ when $cov(\delta_i, |a_i|) = 0$. This latter property often holds; for example it holds when $cov(\delta_i^2, a_i^2) = 2cov(\delta_i, |a_i|)^2 + 4E(|a_i|)E(\delta_i)cov(\delta_i, |a_i|)$, which we show below to hold for the bivariate normal case.

Here, we derive an expression for $cov(a^2, \delta^2)$ when assuming a bivariate normal distribution on $(\delta, |a|)$. For simplicity of notation assume that $\mu_1 = E(|a|)$, $\mu_2 = E(\delta)$, $\sigma_1^2 = var(|a|)$, $\sigma_2^2 = var(\delta)$ and $\sigma_{12} = cov(\delta, |a|)$, and define the centered random variables $X_1 = |a| - \mu_1$ and $X_2 = \delta - \mu_2$. First, we write $cov(a^2, \delta^2) = cov(X_1^2, X_2^2) + 2\mu_1 cov(X_1, X_2^2) + 2\mu_2 cov(X_1^2, X_2)$, which reduces to $cov(a^2, \delta^2) = E(X_1^2 X_2^2) - \sigma_1^2\sigma_2^2 + 2\mu_1 E(X_1 X_2^2) + 2\mu_2 E(X_1^2 X_2) + 4\mu_1\mu_2 E(X_1 X_2)$. For the bivariate normal distribution the central moments are $E(X_1^2 X_2^2) = \sigma_1^2\sigma_2^2 + 2\sigma_{12}^2$, $E(X_1 X_2^2) = E(X_1^2 X_2) = 0$ and $E(X_1 X_2) = \sigma_{12}$, and from this we obtain that $cov(a^2, \delta^2) = 2\sigma_{12}^2 + 4\mu_1\mu_2\sigma_{12}$.

Finally, we note that, strictly speaking, $|a|$ cannot follow a normal distribution, so the above expression is only an approximation.

Appendix B

Consider a set of individuals with genetic values (in a broad sense, these can be understood as genotypic values) $\boldsymbol{g}$, and these genetic values are assumed drawn from a certain distribution, *i.e.*, $E(\boldsymbol{g}) = \boldsymbol{0}$ and $Var(\boldsymbol{g}) = \mathbf{G}\sigma_g^2$. Since the genetic values are unknown and drawn from a sampling distribution, the variances of these genetic values will also have a certain distribution (Sorensen *et al.* 2001; Legarra 2016). Legarra (2016) showed that, on average, the expectation of the variance for these genetic values is $V_G = (\overline{diag(\mathbf{G})} - \bar{\mathbf{G}})\sigma_g^2 = D_g\sigma_g^2$, where $\overline{diag(\mathbf{G})}$ is the average of the diagonal of $\mathbf{G}$ and $\bar{\mathbf{G}}$ is the average of matrix $\mathbf{G}$, and D_g is the difference between the two. Thus, the expected variance V_G is associated with relationship matrix $\mathbf{G}$. Only if D_g is equal to 1, the expectation of variance for genetic values V_G can be identical to the estimated variance component σ_g^2 (Speed and Balding 2015). In this study, the submatrices $\boldsymbol{ZZ}'/k$, $\boldsymbol{ZW}'/k$, and $\boldsymbol{WW}'/k$ do not yield respective $D_{ZZ'/k}$, $D_{ZW'/k}$, and $D_{WW'/k}$ equal to 1. Thus, the estimated variance components were scaled to the expected variance components of a population using, for instance, $D_{ZW'/k} = \overline{diag(\boldsymbol{ZW}'/k)} - \overline{(\boldsymbol{ZW}'/k)}$ and $Cov_{uv} = D_{ZW'/k}\sigma_{ad}^*$.

The values of the statistics can be derived analytically. For one SNP locus, if the population is in Hardy–Weinberg equilibrium and p is the frequency of the allele whose homozygote has the genotypic value a, the frequencies of different genotypes in the $\boldsymbol{Z}$ and $\boldsymbol{W}$ matrices for one locus are:

$$
\begin{array}{c c c c}
 & p^2 & 2pq & q^2 \\
\boldsymbol{Z} & 1 & 0 & -1 \\
\boldsymbol{W} & 0 & 1 & 0
\end{array}
$$

Therefore, the elements in the matrix $\boldsymbol{ZW}'/k$ with corresponding frequencies are:

$$
\begin{array}{c c c c}
 & p^2 & 2pq & q^2 \\
p^2 & 0 & 1/k & 0 \\
2pq & 0 & 0 & 0 \\
q^2 & 0 & -1/k & 0
\end{array}
$$

For instance, in a one-locus model, value $1/k$ appears with frequency $2pq$ (the number of animals heterozygote) times p^2 (the number of homozygotes) in matrix $\boldsymbol{ZW}'/k$. Extending the reasoning to all loci, from the above table, $\overline{diag(\boldsymbol{ZW}'/k)} = 0$ (because an individual cannot be homozygote and heterozygote at the same time); $\overline{(\boldsymbol{ZW}'/k)} = \frac{1}{k}\sum 2p_iq_i(p_i^2 - q_i^2) = \frac{1}{k}\sum 2p_iq_i(2p_i - 1)$, where $k = \{tr(\mathbf{K})\}/2n = (p^4 + q^4 + 4p^2q^2)/2 > 0$. Thus, $D_{ZW'/k} = \overline{diag(\boldsymbol{ZW}'/k)} - \overline{(\boldsymbol{ZW}'/k)} = -\frac{1}{k}\sum 2p_iq_i(2p_i - 1)$. If $p > 0.5$ (MAC), $D_{ZW'/k}$ is negative because $\sum(2p_i - 1) > 0$, but if RAC is used then $D_{ZW'/k} = 0$ because $\sum(2p_i - 1) = 0$.