

On the Additive and Dominant Variance and Covariance of Individuals Within the Genomic Selection Scope

Zulma G. Vitezica,^{*,1} Luis Varona,^{*,†} and Andres Legarra[§]

^{*}Université de Toulouse, Unité Mixte de Recherche 1289, Institut National de la Recherche Agronomique/Institut National Polytechnique de Toulouse, F-31326 Castanet-Tolosan, France, [†]Departamento de Anatomía, Embriología y Genética, Universidad de Zaragoza, 50013 Zaragoza, Spain, [‡]Instituto de Biocomputación y Física de los Sistemas Complejos, Universidad de Zaragoza, 50018 Zaragoza, Spain, and [§]Institut National de la Recherche Agronomique, UR 631 Station d'Amélioration Génétique des Animaux, F-31326 Castanet-Tolosan, France

ABSTRACT Genomic evaluation models can fit additive and dominant SNP effects. Under quantitative genetics theory, additive or “breeding” values of individuals are generated by substitution effects, which involve both “biological” additive and dominant effects of the markers. Dominance deviations include only a portion of the biological dominant effects of the markers. Additive variance includes variation due to the additive and dominant effects of the markers. We describe a matrix of dominant genomic relationships across individuals, **D**, which is similar to the **G** matrix used in genomic best linear unbiased prediction. This matrix can be used in a mixed-model context for genomic evaluations or to estimate dominant and additive variances in the population. From the “genotypic” value of individuals, an alternative parameterization defines additive and dominance as the parts attributable to the additive and dominant effect of the markers. This approach underestimates the additive genetic variance and overestimates the dominance variance. Transforming the variances from one model into the other is trivial if the distribution of allelic frequencies is known. We illustrate these results with mouse data (four traits, 1884 mice, and 10,946 markers) and simulated data (2100 individuals and 10,000 markers). Variance components were estimated correctly in the model, considering breeding values and dominance deviations. For the model considering genotypic values, the inclusion of dominant effects biased the estimate of additive variance. Genomic models were more accurate for the estimation of variance components than their pedigree-based counterparts.

GENOMIC evaluation models typically fit only marker or haplotypic additive effects, either explicitly, estimating the effect of each marker (Meuwissen *et al.* 2001; VanRaden 2008; De Los Campos *et al.* 2009), or implicitly through the so-called “genomic” relationship matrix (VanRaden 2008; Goddard 2009), which uses an equivalent model from which the marker effects can be inferred by backsolving. The additive substitution effect represents the average change in genotypic value that results when an A_1 allele is randomly substituted for an A_2 allele in a population (Falconer 1981; Lynch and Walsh 1998). The substitution effects in quantitative genetics

are of primal interest, as they capture a large part of dominant and higher-order interactions across genes and alleles (*i.e.*, epistasis) (*e.g.*, Cockerham 1954; Kempthorn 1954; Falconer 1981). In addition, selection acts on additive substitution effects because alleles, not genotypes, are passed from parents to offspring. However, dominance is of theoretical and practical interest, because it is heavily used in crosses of animal breeds and plant lines (*e.g.*, in pigs, poultry, or corn). In principle, assortative mating or mate allocation (when the breeder chooses specific couples of individuals as parents for the next generation, for instance, favorable dominant combinations) can boost the field performances of livestock and crops (Varona and Misztal 1999; Toro and Varona 2010).

In livestock populations, one of the main reasons why dominance effects have not been widely used or estimated *per individual* is that pedigree relationships are not enough informative, as large full-sib families are typically needed for any accurate estimate. In prolific species such as chickens and

Copyright © 2013 by the Genetics Society of America
doi: 10.1534/genetics.113.155176

Manuscript received July 21, 2013; accepted for publication October 4, 2013;
published Early Online October 11, 2013.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.155176/-DC1>.

¹Corresponding author: Unité Mixte de Recherche 1289 TANDEM, ENSAT, Avenue de l'Agrobiopole, Postal Box 32607, 31326 Auzeville Tolosane, France. E-mail: zulma.vitezica@ensat.fr

pigs, the litter effect is highly confounded with family. In addition, the prediction of dominant values is typically very cumbersome because it involves complex computations (e.g., Misztal *et al.* 1998; Mrode and Thompson 2005).

Recently, genomic evaluations have renewed the interest in the prediction of nonadditive genetic effects (e.g., Toro and Varona 2010; Su *et al.* 2012; Wellmann and Bennewitz 2012). One of the reasons is that it is much easier to work with dominance, knowing for each evaluated locus which animals are heterozygotes, but also that prediction of the genotypic value of future matings is straightforward (Toro and Varona 2010).

An example of the richness of parametric genomic prediction methods for the estimation of marker effects is to have equivalent models and interpretations through the genetic covariance between individuals [*i.e.*, genomic relationships (VanRaden 2008; Goddard 2009; Yang *et al.* 2010)] and estimation of the base population variances (Forni *et al.* 2011; Legarra *et al.* 2011b; Sillanpää 2011). These equivalences are not quite completely described for the case of dominance effects in genomic evaluations (e.g., Toro and Varona 2010; Su *et al.* 2012; Wellmann and Bennewitz 2012). These works are either incomplete or, in part, induce erroneous interpretations of correct models. The aim of this study is to show the equivalences between additive and dominant effects at the marker and the population levels and to present how to compute from genotypes the covariances between individuals due to dominant deviations, e.g., “dominant genomic relationships”. Real data and a simulated example are used to illustrate the principles.

Theory

A model including additive and dominant effects of the SNP markers can be written as

$$y_i = \mu + \sum_{j=1}^n t_{ij}a_j + \sum_{j=1}^n x_{ij}d_j + e_i \quad (1)$$

(Toro and Varona 2010) or, in matrix form for a set of individuals,

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{T}\mathbf{a} + \mathbf{X}\mathbf{d} + \mathbf{e},$$

where y_i is the phenotypic value of the individual i and μ is the population mean. An additive effect a_j and a dominant effect d_j are included for each of the n SNP markers. The covariate $t_{ij} = 1, 0$, and -1 , for SNP genotypes A_1A_1 , A_1A_2 , and A_2A_2 , respectively. For the dominant component, $x_{ij} = 0, 1$, and 0 for SNP genotypes A_1A_1 , A_1A_2 , and A_2A_2 , respectively.

In model (1), the meaning of “dominant” and “additive” SNP effects is clear from the formulation of the model and basically parallels the biological intuitive meaning. However its transposition to classical, pedigree-based concepts (in particular, partitioning of the variance) is not that obvious.

In the following, and to avoid ambiguities, we talk either about (*biological*) “genotypic” additive and dominant values or about (*statistical*) “breeding” values and dominance deviations. Quoting Hill *et al.* (2008, p. 1) the “biological” value is “the observations of dominance [...] at the level of gene action at individual loci, exemplified by a table of genotypic values,” and the “statistical” values are “the observations of variance due to these components in analysis of data from a population.” Traditional treatment of quantitative genetics distinguish carefully between biological effects and statistical effects (e.g., Crow and Kimura 1970; Falconer 1981; Hill *et al.* 2008). For instance, even when the biological action of genes is mostly dominant or epistatic, most of the genetic variation transforms into additive in the sense of transmission to offspring or of a “substitution” effect of a gene in a population (e.g., Hill *et al.* 2008).

For instance, a recent study by Su *et al.* (2012) partitions genetic variance into additive and dominant, in such a way that these estimated variances are not directly comparable to pedigree-based estimates. In this work we explain why this is so and how partitioning should proceed. We also present a dominant genomic relationship matrix that can be introduced in genomic evaluation linear mixed models similar to genomic best linear unbiased prediction (GBLUP) (VanRaden 2008). The first part of this work is just a short review of classical treatments (e.g., Crow and Kimura 1970; Falconer 1981).

Consider one locus. Following model (1) the genotypic value G of an individual is

$$G_{A_1A_1} = a \quad G_{A_1A_2} = d \quad G_{A_2A_2} = -a,$$

where the values a and d are deviations from the midpoint of the two homozygous genotypes (Falconer 1981). The genetic mean of the population is therefore

$$E(G) = (p - q)a + 2pqd,$$

where p is the frequency of A_1 and $q = 1 - p$.

Classical parameterization

Classically, the genotypic values are split in additive (or breeding) values and dominant deviations, as in Falconer (1981),

$$G = E(G) + u + v,$$

where u and v stand for breeding values and dominant deviations. The breeding value for an individual is

$$\begin{aligned} u_{A_1A_1} &= (2 - 2p)(a + d(q - p)) = (2 - 2p)\alpha \\ u_{A_1A_2} &= (1 - 2p)(a + d(q - p)) = (1 - 2p)\alpha \\ u_{A_2A_2} &= (-2p)(a + d(q - p)) = (-2p)\alpha, \end{aligned}$$

where $\alpha = a + d(q - p)$ is the substitution effect of the gene. So, the breeding values of a set of individuals are $\mathbf{u} = \mathbf{z}\alpha$ (with \mathbf{z} coded as in VanRaden 2008),

$$z_i = \begin{cases} (2-2p) \\ (1-2p) \\ -2p \end{cases} \text{ for genotypes } \begin{cases} A_1A_1 \\ A_1A_2 \\ A_2A_2 \end{cases} \quad (2)$$

in individual i . Matrix $\mathbf{Z} = (\mathbf{z}_1 \dots \mathbf{z}_n)$ including all markers is identical to matrix \mathbf{T} in (1) but centered.

Also, the dominant deviation of an individual is

$$\begin{aligned} v_{A_1A_1} &= -2q^2d \\ v_{A_1A_2} &= 2pqd \\ v_{A_2A_2} &= -2p^2d. \end{aligned}$$

So $\mathbf{v} = \mathbf{w}d$ with

$$w_i = \begin{cases} -2q^2 \\ 2pq \\ -2p^2 \end{cases} \text{ for genotypes } \begin{cases} A_1A_1 \\ A_1A_2 \\ A_2A_2 \end{cases}.$$

Note that $\mathbf{W} = (\mathbf{w}_1 \dots \mathbf{w}_n) \neq \mathbf{X}$ in model (1).

The genetic variance is

$$\begin{aligned} \sigma_G^2 &= 2pq[a + d(q-p)]^2 + [2pqd]^2 \\ &= 2pq\alpha^2 + [2pqd]^2. \end{aligned}$$

We can partition the genetic variance σ_G^2 into components due to individual additive value (breeding values, u), and dominance deviations (v). In fact, $E(u) = 0$ and the *additive genetic variance* is

$$\sigma_A^2 = 2pq[a + d(q-p)]^2 = 2pq\alpha^2.$$

Also like breeding values, dominance deviation sums to zero ($E(v) = 0$) and the *dominance genetic variance* is

$$\sigma_D^2 = [2pqd]^2.$$

Partitioning the total genetic variance, we can see that $\sigma_G^2 = \sigma_A^2 + \sigma_D^2$, where the first term is the additive genetic variance and the second term corresponds to the dominance genetic variance. Note that this parameterization implies that, in a noninbred population with Hardy–Weinberg equilibrium (*i.e.*, without inbreeding) the covariance between the breeding value and the dominance deviation is zero (Falconer 1981). Without inbreeding, the inclusion of dominance in the model is easy (Henderson 1985). However, its inclusion is more complicated with inbreeding, first, because there may be inbreeding depression and, second, because the covariance between inbred individuals with dominance is no longer a function of just σ_A^2 and σ_D^2 (De Boer and Hoeschele 1993; Lynch and Walsh 1998).

If a and d effects are considered to be random and covariance between them equal to zero, the covariance of additive individual effects, \mathbf{u} , is

$$\text{Cov}(\mathbf{u}) = \mathbf{z}\mathbf{z}'\sigma_a^2 + \mathbf{z}\mathbf{z}'(q-p)^2\sigma_d^2,$$

where σ_a^2 and σ_d^2 are the SNP variances for additive and dominant components, respectively.

Now, we have that $\sigma_A^2 = 2pq\sigma_a^2 + 2pq(q-p)^2\sigma_d^2$, assuming that the covariance between a and d is zero. So in fact

$$\text{Cov}(\mathbf{u}) = \frac{\mathbf{z}\mathbf{z}'}{2pq}\sigma_A^2 = \mathbf{G}\sigma_A^2 \quad (3)$$

as expected, which is the classical \mathbf{G} matrix of GBLUP (VanRaden 2008). It is well known that in a base population with Hardy–Weinberg equilibrium, the average of the diagonal of \mathbf{G} is 1, whereas the average off-diagonal is 0.

As for the dominant deviations, its covariance is

$$\text{Cov}(\mathbf{v}) = \mathbf{w}\mathbf{w}'\sigma_d^2.$$

If we take $\sigma_D^2 = (2pq)^2\sigma_d^2$, then

$$\text{Cov}(\mathbf{v}) = \frac{\mathbf{w}\mathbf{w}'}{(2pq)^2}\sigma_D^2 = \mathbf{D}\sigma_D^2. \quad (4)$$

In a base population with Hardy–Weinberg conditions, it turns out that the diagonal of \mathbf{D} sums to $[p^2(-2q^2)^2 + 2pq(2pq)^2 + q^2(-2p^2)^2]/(2pq)^2$, which is equal to 1. In addition, the sum of off-diagonal elements of \mathbf{D} is 0. In effect, this sum can be written as

$$\frac{(p^2 \quad 2pq \quad q^2) \begin{pmatrix} -2q^2 \\ 2pq \\ -2p^2 \end{pmatrix} \begin{pmatrix} -2q^2 \\ 2pq \\ -2p^2 \end{pmatrix}' (p^2 \quad 2pq \quad q^2)'}{(2pq)^2},$$

which sums to 0. Both features correspond to proper definitions of dominant relationships in a base population (Hayes *et al.* 2009).

Alternative (genotypic) parameterization

From the genotypic value, we can also define u^* and v^* as the genotypic additive and dominant effects, *i.e.*, the parts attributable to the additive and dominance biological effects of the markers. Note that this u^* is not a breeding value.

Thus, we have $G = E(G) + u^* + v^*$. The mean of genotypic values of individuals due to additive effects is $(p - q)a$, and that due to dominant effects is $2pqd$.

Therefore

$$\begin{aligned} u_{A_1A_1}^* &= (2 - 2p)a \\ u_{A_1A_2}^* &= (1 - 2p)a \\ u_{A_2A_2}^* &= (-2p)a, \end{aligned}$$

which is as in (2) and as in VanRaden (2008). Also,

$$\begin{aligned} v_{A_1A_1}^* &= -2pqd \\ v_{A_1A_2}^* &= (1 - 2pq)d \\ v_{A_2A_2}^* &= -2pqd \end{aligned}$$

or $\mathbf{v} = \mathbf{h}d$, with \mathbf{h} as

$$h_i = \begin{cases} 0 - 2pq \\ 1 - 2pq \\ 0 - 2pq \end{cases} \text{ for genotypes } \begin{cases} A_1A_1 \\ A_1A_2 \\ A_2A_2. \end{cases}$$

Note that incidence matrix $\mathbf{H} = (\mathbf{h}_1 \dots \mathbf{h}_n)$ is simply \mathbf{X} in (1) but centered. Remember that $\sigma_{A^*}^2$ is the variance of genotypic additive values, and $\sigma_{D^*}^2$ is the variance of genotypic dominant values. Then

$$\sigma_{A^*}^2 = 2pqa^2$$

and

$$\sigma_{D^*}^2 = 2pq(1 - 2pq)d^2.$$

The covariances across genotypic additive values, u^* , are

$$\text{Cov}(\mathbf{u}^*) = \mathbf{z}\mathbf{z}'\sigma_a^2$$

and introducing $\sigma_{A^*}^2 = 2pq\sigma_a^2$,

$$\text{Cov}(\mathbf{u}^*) = \frac{\mathbf{z}\mathbf{z}'}{2pq}\sigma_{A^*}^2,$$

which is also the classical \mathbf{G} matrix of genomic BLUP (VanRaden 2008), but with a different variance component.

As for the covariance in genotypic values due to dominance, this is

$$\text{Cov}(\mathbf{v}^*) = \mathbf{h}\mathbf{h}'\sigma_d^2.$$

Introducing $\sigma_{D^*}^2 = 2pq(1 - 2pq)\sigma_d^2$, then

$$\text{Cov}(\mathbf{v}^*) = \frac{\mathbf{h}\mathbf{h}'}{2pq(1 - 2pq)}\sigma_{D^*}^2.$$

This is as in Su *et al.* (2012). For model (1), \mathbf{H} corresponds to the incidence matrix \mathbf{X} (shifted by a constant). However, note that $\mathbf{W} \neq \mathbf{H}$.

Therefore, the breeding (or classical) and the genotypic models are equivalent models to explain the data (\mathbf{y}) but their interpretation must be different. The first model is termed in breeding values and dominant deviations and the second model in genotypic additive and dominant effects of markers. In particular, the first model provides direct estimates of breeding values, whereas the second model is more obviously interpretable in terms of dominant and additive effects of the markers. The interpretation of the variances is also different; for instance, $\sigma_{A^*}^2$ is the variance useful in selection or in prediction of potential selection response. In particular, to compare with pedigree-based estimates σ_A^2 should be used, not $\sigma_{A^*}^2$. Also, the dominant relationship matrices are different. For instance, the covariance matrices for three individuals with the three possible genotypes, each one for one locus and assuming $\sigma_d^2 = 1$ and $p = 0.7$ (the frequency of A_1), are for the breeding model

$$\text{Cov}(\mathbf{v}) = \frac{\mathbf{w}\mathbf{w}'}{(2pq)^2}\sigma_D^2 = \begin{pmatrix} 0.03 & -0.08 & 0.18 \\ -0.08 & 0.18 & -0.41 \\ 0.18 & -0.41 & 0.96 \end{pmatrix}$$

with $\sigma_D^2 = (2pq)^2\sigma_d^2$ and

$$\mathbf{w} = \begin{pmatrix} -2q^2 \\ 2pq \\ -2p^2 \end{pmatrix} = \begin{pmatrix} -0.18 \\ 0.42 \\ -0.98 \end{pmatrix}$$

and for the genotypic model

$$\text{Cov}(\mathbf{v}^*) = \frac{\mathbf{h}\mathbf{h}'}{2pq(1 - 2pq)}\sigma_{D^*}^2 = \begin{pmatrix} 0.18 & -0.24 & 0.18 \\ -0.24 & 0.34 & -0.24 \\ 0.18 & -0.24 & 0.18 \end{pmatrix}$$

with $\sigma_{D^*}^2 = 2pq(1 - 2pq)\sigma_d^2$ and

$$\mathbf{h} = \begin{pmatrix} 0 - 2pq \\ 1 - 2pq \\ 0 - 2pq \end{pmatrix} = \begin{pmatrix} -0.42 \\ 0.58 \\ -0.42 \end{pmatrix}.$$

Thus, the dominant deviations across individuals have stronger covariances when two individuals share a rare allele (-0.41) in the classical parameterization.

Going from genotypic to breeding additive and dominance variances

Assume that we have variances estimated in either model. Then we have the identities

$$\sigma_A^2 = 2pq\sigma_a^2 + 2pq(q-p)^2\sigma_d^2$$

$$\sigma_{A^*}^2 = 2pq\sigma_a^2$$

$$\sigma_D^2 = (2pq)^2\sigma_d^2$$

$$\sigma_{D^*}^2 = 2pq(1 - 2pq)\sigma_d^2.$$

It can be verified that $\sigma_{A^*}^2 + \sigma_{D^*}^2 = \sigma_A^2 + \sigma_D^2$. Also, if $p = q = 0.5$, all variances are identical, and if $d = 0$, $\sigma_{A^*}^2 = \sigma_A^2$. From these identities, conversion across models is immediate.

Extension to multiple loci

The extension to multiple loci assumes linkage equilibrium and uncorrelated marker effects, which are ordinary assumptions (VanRaden 2008; Gianola *et al.* 2009).

Then

$$\sigma_A^2 = \sum (2p_i q_i) \sigma_a^2 + \sum (2p_i q_i (q_i - p_i)^2) \sigma_d^2$$

$$\sigma_{A^*}^2 = \sum (2p_i q_i) \sigma_a^2$$

$$\sigma_D^2 = \sum (2^2 p_i^2 q_i^2) \sigma_d^2$$

Table 1 Heritabilities and variance estimates (standard deviations) under genotypic and breeding models for different mouse traits

Heritability/variance	Weight	Growth speed	Body length	Body mass index
$h_{A^*}^2$ ^a	0.192 (0.028)	0.020 (0.013)	0.111 (0.028)	0.053 (0.021)
$h_{D^*}^2$ ^a	0.022 (0.011)	0.015 (0.011)	0.017 (0.013)	0.033 (0.018)
$\sigma_{D^*}^2/(\sigma_{D^*}^2 + \sigma_{A^*}^2)$ ^a	0.105 (0.051)	0.433 (0.222)	0.137 (0.093)	0.382 (0.171)
h_A^2 ^b	0.199 (0.028)	0.025 (0.012)	0.117 (0.027)	0.064 (0.020)
h_D^2 ^b	0.015 (0.008)	0.011 (0.007)	0.012 (0.009)	0.022 (0.013)
$\sigma_D^2/(\sigma_D^2 + \sigma_A^2)$ ^b	0.072 (0.035)	0.296 (0.151)	0.094 (0.064)	0.261 (0.117)
h_A^2 ^c	0.197 (0.027)	0.016 (0.012)	0.113 (0.028)	0.058 (0.021)

^a Model based on additive breeding value and dominance deviations (breeding model).^b Model based on additive and dominant effects of markers (genotypic model).^c Model including additive effects only.

$$\sigma_{D^*}^2 = \sum (2p_i q_i (1 - 2p_i q_i)) \sigma_d^2.$$

Also, the covariance matrices are simple extensions of (3) and (4):

$$\text{Cov}(\mathbf{u}) = \frac{\mathbf{ZZ}'}{2 \sum p_i q_i} \sigma_{A^*}^2 = \mathbf{G} \sigma_{A^*}^2 \quad (5)$$

$$\text{Cov}(\mathbf{v}) = \frac{\mathbf{WW}'}{\sum (2p_i q_i)^2} \sigma_D^2 = \mathbf{D} \sigma_D^2 \quad (6)$$

$$\text{Cov}(\mathbf{u}^*) = \frac{\mathbf{ZZ}'}{2 \sum p_i q_i} \sigma_{A^*}^2 = \mathbf{G} \sigma_{A^*}^2 \quad (7)$$

$$\text{Cov}(\mathbf{v}^*) = \frac{\mathbf{HH}'}{2 \sum p_i q_i (1 - 2p_i q_i)} \sigma_{D^*}^2. \quad (8)$$

For instance, Su *et al.* (2012) estimated several variance components using markers. These are parameterized in term of genotypic effects; *i.e.*, $G = E(G) + u^* + v^*$. Assuming, for instance, that their allelic frequencies p_i are drawn from a $\beta(2, 2)$ distribution, the terms $\sum (2p_i q_i)$, $\sum (2^2 p_i^2 q_i^2)$, etc., can be computed (by Monte Carlo or algebraic integration). Then, the results of Su *et al.* (2012) could be converted as follows:

$$\sigma_{D^*}^2 = 309 \rightarrow \sigma_d^2 = 0.052 \rightarrow \sigma_D^2 = 232$$

$$\sigma_{A^*}^2 = 2081 \rightarrow \sigma_A^2 = 2158.$$

In this example, the dominance variance (σ_D^2) is not 13% of the genetic variance (σ_G^2), but 9.7%. The genotypic model overestimates the dominance genetic variance and, consequently, underestimates additive genetic variance.

Inbreeding

It is not clear whether the inbreeding is taken automatically into account in the models. The model in Equation 1 makes no hypothesis of Hardy–Weinberg equilibrium. Inbreeding (deviations from Hardy–Weinberg) is accounted for because genotypes are observed. However, properties of the breeding

and genotypic models are no longer well defined; for instance, covariances between \mathbf{u} and \mathbf{v} appear. But because the genotypic and breeding models are equivalent to model (1), the genotypic value can always be correctly inferred. However, model (1) does not account for inbreeding depression. This needs further investigation.

Examples

Mouse data analysis

Legarra *et al.* (2008) analyzed phenotypes and genotypes of mouse data (Valdar *et al.* 2006), including 1884 individuals and 10,946 markers. For illustration, we have reanalyzed the data and estimated σ_a^2 and σ_d^2 , using the same data set and model (1) (plus an extra “cage” effect, as in the original work). Variance components under genotypic and breeding models were estimated for four traits (weight, growth speed, body length, and body mass index). Estimation was by Bayesian inference, using flat priors for the variances and a Gibbs sampling algorithm. We used the publicly available software GS3 (Legarra *et al.* 2011a) for estimation. Both the software and the mouse data are available at <http://genoweb.toulouse.inra.fr/~alegarra/g3dist.tar.gz>.

Simulated data

We simulated a quantitative trait with additive and dominant QTL action, so that the parts of total variance attributable to additive and dominance effects (in terms of breeding values and dominant deviations) were 0.20 and 0.10, respectively. Simulation was as in Toro and Varona (2010), including 10 chromosomes of 1 M and 1000 markers each. A population was generated by mutation and drift over 1000 generations with an effective population size of 100. After 1000 generations, the population was expanded to 1000 individuals (500 per sex) and remained at 1000 with random mating for 2 discrete generations. Therefore, the data set consisted of 2100 individuals. These 2100 individuals were genotyped and phenotyped and then used to estimate variance components. No inbreeding occurred within the 2100 individuals.

In variance components estimation four different models were used. To verify whether the correct genetic parameters could be estimated using markers, the genomic

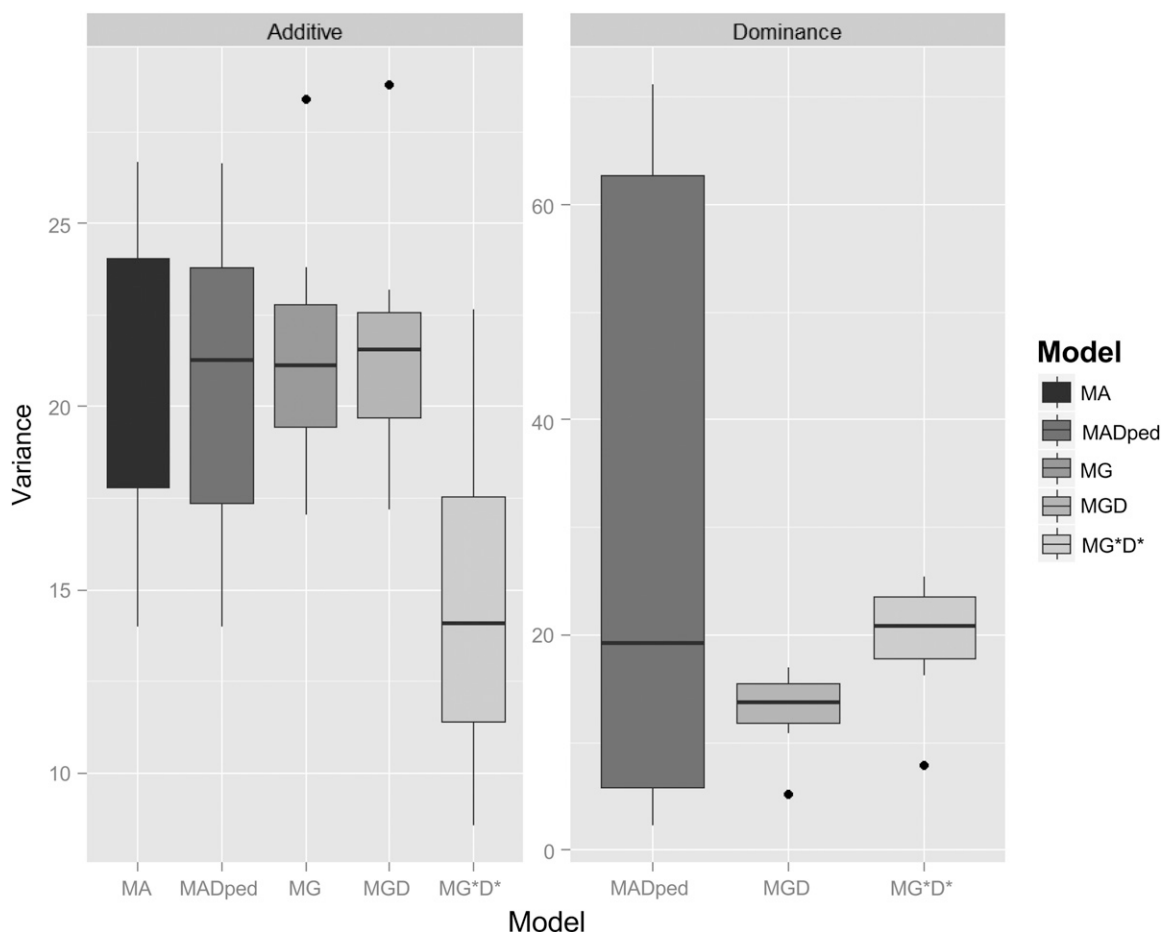


Figure 1 Estimations of variance components under different models. Prediction models were based on best linear unbiased prediction: a mixed model based on the pedigree additive relationship matrix and phenotypes (MA), an alternative model including pedigree-based additive and dominant relationship matrices (MADped), a genomic model including additive genomic relationship matrix (MG), and a genomic model involving breeding values and dominance deviations (MGD). Variance components under genotypic parameterization (MG*D*) as in Su *et al.* (2012) were calculated using the allele frequencies and the variances of the additive and dominant effects of the markers.

model involving breeding values and dominance deviations (MGD) was

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

$$\text{Var}(\mathbf{u}) = \mathbf{G}\sigma_A^2; \quad \text{Var}(\mathbf{v}) = \mathbf{D}\sigma_D^2; \quad \text{Var}(\mathbf{e}) = \mathbf{I}\sigma_e^2,$$

where the covariance matrices \mathbf{G} and \mathbf{D} were constructed as in (5) and (6).

A pedigree-based alternative model including additive and dominant relationship matrices (MADped) used

$$\text{Var}(\mathbf{u}) = \mathbf{A}\sigma_A^2; \quad \text{Var}(\mathbf{v}) = \mathbf{D}_{\text{ped}}\sigma_D^2; \quad \text{Var}(\mathbf{e}) = \mathbf{I}\sigma_e^2,$$

where \mathbf{A} is a pedigree-based relationship matrix and \mathbf{D}_{ped} is a pedigree-based dominant relationship matrix (Cockerham 1954; Henderson 1985). A simplified model $\mathbf{y} = \mathbf{1}\mu + \mathbf{u} + \mathbf{e}$ with two different variance structures, either $\text{Var}(\mathbf{u}) = \mathbf{G}\sigma_A^2$ (MG) or $\text{Var}(\mathbf{u}) = \mathbf{A}\sigma_A^2$ (MA), was also tested. These two variance structures define a genomic model with an additive genomic relationship matrix (MG) and a pedigree-based

model (MA) respectively. Genetic (σ_A^2 , σ_D^2) and residual (σ_e^2) variances were computed by restricted maximum likelihood, using the software remlf90 (Misztal *et al.* 2002) (available at <http://nce.ads.uga.edu/wiki/doku.php>).

Variance components under genotypic parameterization (σ_A^2 , σ_D^2) as in Su *et al.* (2012) were also calculated using the model

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{u}^* + \mathbf{v}^* + \mathbf{e},$$

with covariance structures as in (7) and (8), estimating thus variances of the additive and dominant effects of the markers (MG*D*).

In addition, we verified the ability of each model to correctly predict breeding values and dominant deviations for each individual in the last generation. Regression of true breeding value (TBV) on estimated breeding value (EBV) was used as a measure of the inflation of the prediction model, where a regression coefficient of 1 denotes no inflation. The predictive ability of model was also evaluated by accuracy computed as the correlation between TBV

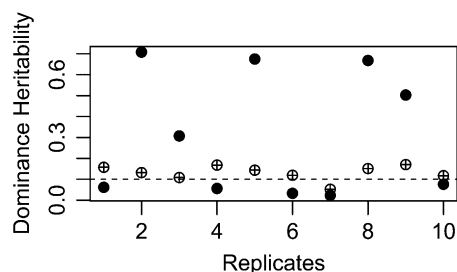


Figure 2 Dominance heritability (proportion of dominance variance to phenotypic variance) for MGD (circled plus) and for MADped (solid circle) for each replicate. The dashed horizontal line is the simulated value for the dominance heritability. MADped is the model including pedigree-based additive and dominant relationship matrices. MGD is a genomic model involving breeding values and dominance deviations.

and EBV. Results were the mean of the 10 replicates for each model.

Simulated data sets (phenotypes, genotypes, and pedigree) are available at http://genoweb.toulouse.inra.fr/~zvitezic/simu_for_genetics.tar.gz. Programs for simulation of the data sets and for construction of the different relationship matrices are available on request from the authors.

Results and Discussion

Mouse data analysis

The results for variance components are shown in Table 1.

Estimates of the partitioning of the genetic variance do clearly differ; for instance, for growth the proportion of genetic variance attributed to dominance effects halves if one uses the classical parameterization. In addition, broad-sense heritability is sometimes much higher than its additive estimate (e.g., in growth speed, additive heritability is 0.025 whereas broad-sense heritability ($h_a^2 + h_d^2$) is 0.036). The ratio between h_D^2 and h_a^2 ranges from 1.36 to 1.50 for all the traits and depends only on the distribution of the allelic frequencies of markers; the reason it is not constant is because of Monte Carlo noise in the estimates. The correlations between estimated (\mathbf{u}, \mathbf{u}^*) and (\mathbf{v}, \mathbf{v}^*) in this data set were 0.99 and 0.86. This illustrates that estimated individual effects are different for the different parameterizations, in particular for dominance effects. We stress that the correlation between total estimated genetic values ($\mathbf{u} + \mathbf{v}$) and ($\mathbf{u}^* + \mathbf{v}^*$) is 1, because both models are equivalent.

Simulated data

Estimates of additive and dominance variances for the four models are presented in Figure 1. Estimates of additive genetic variance were 20.7 ± 4.1 , 21.4 ± 3.3 , 21.4 ± 3.3 , and 20.5 ± 4.2 for the models MA, MG, MGD and MADped, respectively. As was expected, the SD of estimates was greater in MA (MADped) than in MG (MGD). This means that the amount of statistical information to estimate variances is greater with markers than with pedigree. The additive genetic variance was well estimated also when models

Table 2 Accuracy and inflation estimates (standard deviations) for different prediction models including additive and dominant matrices using simulated data

Effect	Accuracy ^a	Inflation ^b
Additive		
MA	0.58 (0.04)	0.98 (0.15)
MG	0.68 (0.02)	0.96 (0.08)
MGD	0.69 (0.02)	0.96 (0.08)
MADped	0.58 (0.03)	0.99 (0.15)
Dominant		
MGD	0.44 (0.03)	0.91 (0.26)
MADped	0.32 (0.03)	1.23 (1.28)

Prediction models were based on best linear unbiased prediction (BLUP): a mixed model based on the pedigree additive relationship matrix and phenotypes (MA), an alternative model including pedigree-based additive and dominant relationship matrices (MADped), a genomic model including the additive genomic relationship matrix (MG), and a genomic model involving breeding values and dominance deviations (MGD).

^a The correlation between true and estimated breeding values.

^b The coefficient of regression of true on estimated breeding values.

included dominant effects. These results do not agree with Su *et al.* (2012) who found lower estimates of additive genetic variance when the model included dominance and epistasis effects. The expanded model of Su *et al.* (2012) does not handle properly the fact that the (additive) substitution effects α include part of biological dominance. For instance, when only additive effects are included in their data analysis, part of biological dominance is included; but when dominance is fitted, this part goes to the dominance variation. The additive variance estimated by the increasingly complex models (with dominance and epistasis) in Su *et al.* (2012) has not the same meaning throughout all models, whereas in the breeding parameterization it is always the same, irrespective of whether additive only or additive and dominance effects are fitted.

The dominance variance estimated with the MGD model was 13.2 ± 3.5 , which agrees with the simulated dominance variance of 10. Estimates of the proportion of dominance variance to phenotypic variance (“dominance heritability”) are shown in Figure 2. Dominance variance estimation had a large standard error in MADped (Figure 1 and Figure 2). These results illustrate the difficulties in obtaining a good estimate of dominance variance from pedigree information. In addition, the results show how genomic information allows one to obtain an accurate estimation of dominant deviations. Further, in practice, use of dominance through the implementation of mate allocations using markers is straightforward, contrary to pedigree-based methods (Varona and Misztal 1999).

Under the genotypic MG*D* model, the additive and dominance variances were 14.7 ± 4.4 and 19.9 ± 5.3 , respectively. The dominance variance was overestimated; it was 57.5% instead 38% of the genetic variance (Figure 1). Consequently, this genotypic parameterization underestimates additive genetic variance. These results show that estimates from the genotypic model are not directly comparable with pedigree-based estimates, although comparable values can be derived from them using allele frequencies.

The degree of inflation from the prediction methods is indicated by the coefficient of regression of TBV on EBV (Table 2). The optimal prediction model for genetic merit of young individuals would have a regression coefficient close to 1. For each effect, the differences among the models were small, and the approaches achieved very similar inflation.

The accuracies for each model (correlations between TBV and EBV) are shown in Table 2. Compared with the MA and MADped models, all genomic prediction methods (MG and MGD) increased accuracy by ~18% for additive effects. With dominant effects, accuracy from MGD was higher (32.5%) than from MADped (Table 2). In addition, the genomic models including dominance have the advantage that they provide a simple framework, compared with pedigree models, to estimate dominance effects (*i.e.*, Su *et al.* 2012; Ertl *et al.* 2013).

Conclusions

Models using genomic additive and dominant relationships can recover information correctly. The parameterization is largely a matter of convenience, but the parameterization in terms of breeding values and substitution effects is more adequate for selection (both for ranking animals and for predicting genetic improvement). The genotypic model allows simpler estimates of the results of mate allocations. Translating estimates from one model to the other one is simple if the distribution of the allelic frequencies is available. We note, however, that estimates from the genotypic model are not comparable with pedigree-based estimates, although the transposition is fairly easy using the identities that we provide.

Acknowledgments

We are grateful to members of the X-GEN project, Miguel Toro, and Luis Alberto Garcia-Cortes for their helpful and constructive comments. This work has been financed by INRA SELGEN metaprogram, project X-GEN (Z.G.V. and A.L.), as well as the Comisión Interministerial de Ciencia y Tecnología (CICYT) of Spain, project AGL2010-15903 (L.V.). This project was partly supported by the Toulouse Midi-Pyrénées bioinformatic platform.

Literature Cited

- 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.
- Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- De Boer, I. J. M., and I. Hoeschele, 1993 Genetic evaluation methods for populations with dominance and inbreeding. *Theor. Appl. Genet.* 86: 245–258.
- de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra *et al.*, 2009 Predicting quantitative traits with regression models for dense molecular markers and pedigrees. *Genetics* 182: 375–385.
- Ertl, J., A. Legarra, Z.G. Vitezica, L. Varona, C. Edel *et al.*, 2013 Genomic analysis of dominance effects in milk production and conformation traits of Fleckvieh cattle. *Interbull Bull.* 47 (in press).
- Falconer, D. S., 1981 *Introduction to Quantitative Genetics*. Longman, New York.
- Forni, S., I. Aguilar, and I. Misztal, 2011 Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet. Sel. Evol.* 43: 1.
- Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando, 2009 Additive genetic variability and the Bayesian alphabet. *Genetics* 183: 347–363.
- Goddard, M. E., 2009 Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136: 245–257.
- Hayes, B. J., P. M. Visscher, and M. E. Goddard, 2009 Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91: 47–60.
- Henderson, C. R., 1985 Best linear unbiased prediction of non-additive genetic merits. *J. Anim. Sci.* 60: 111–117.
- 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(2): e1000008.
- Kempthorn, O., 1954 The correlation between relatives in a random mating population. *Proc. R. Soc. Lond. B* 143: 103–113.
- Legarra, A., C. Robert-Granié, E. Manfredi, and J. M. Elsen, 2008 Performance of genomic selection in mice. *Genetics* 180: 611–618.
- Legarra, A., A. Ricard, and O. Filangi, 2011a GS3: Genomic Selection, Gibbs Sampling, Gauss-Seidel (and BayesCpi). Available at: <http://genoweb.toulouse.inra.fr/~alegarra/>.
- Legarra, A., C. Robert-Granié, P. Croiseau, F. Guillaume, and S. Fritz, 2011b Improved Lasso for genomic selection. *Genet. Res.* 93: 77–87.
- Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Misztal, I., L. Varona, L. Culbertson, J. K. Bertrand, J. W. Wabry *et al.*, 1998 Studies on the value of incorporating the effect of dominance in genetic evaluations of dairy cattle, beef cattle and swine. *Biotechnol. Agron. Soc. Environ.* 2: 227–233.
- 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, Communication 28-07.
- Mrode, R. A., and R. Thompson, 2005 *Linear Models for the Prediction of Animal Breeding Values*. CABI Publishing, Cambridge, MA.
- Sillanpää, M. J., 2011 On statistical methods for estimating heritability in wild populations. *Mol. Ecol.* 20: 1324–1332.
- Su, G., O. F. Christensen, T. Ostensen, 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(9): e45293.
- Toro, M. A., and L. Varona, 2010 A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol.* 42: 33.
- Valdar, W., L. C. Solberg, D. Gauguier, S. Burnett, P. Klennerman *et al.*, 2006 Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* 38: 879–887.
- VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91: 4414–4423.
- Varona, L., and I. Misztal, 1999 Prediction of parental dominance combinations for planned matings, methodology, and simulation results. *J. Dairy Sci.* 82(10): 2186–2191.
- Wellmann, R., and J. Bennewitz, 2012 Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genet. Res.* 94: 21–37.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42: 565–569.

Communicating editor: D.-J. De Koning