

The contribution of dominance to the understanding of quantitative genetic variation

ROBIN WELLMANN* AND JÖRN BENNEWITZ

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

(Received 6 May 2010 and in revised form 10 November 2010; first published online 12 April 2011)

Summary

Knowledge of the genetic architecture of a quantitative trait is useful to adjust methods for the prediction of genomic breeding values and to discover the extent to which common assumptions in quantitative trait locus (QTL) mapping experiments and breeding value estimation are violated. It also affects our ability to predict the long-term response of selection. In this paper, we focus on additive and dominance effects of QTL. We derive formulae that can be used to estimate the number of QTLs that affect a quantitative trait and parameters of the distribution of their additive and dominance effects from variance components, inbreeding depression and results from QTL mapping experiments. It is shown that a lower bound for the number of QTLs depends on the ratio of squared inbreeding depression to dominance variance. That is, high inbreeding depression must be due to a sufficient number of QTLs because otherwise the dominance variance would exceed the true value. Moreover, the second moment of the dominance coefficient depends only on the ratio of dominance variance to additive variance and on the dependency between additive effects and dominance coefficients. This has implications on the relative frequency of overdominant alleles. It is also demonstrated how the expected number of large QTLs determines the shape of the distribution of additive effects. The formulae are applied to milk yield and productive life in Holstein cattle. Possible sources for a potential bias of the results are discussed.

1. Introduction

Quantitative traits can be controlled by many genes and environmental factors. One main concept to deal with these traits is the infinitesimal model (Fisher, 1918; Goddard, 2001) which assumes an infinite number of genes each with infinitesimal effects. But in fact, the number of genes is finite. Even though the assumptions of this model are violated, the use of this concept has yielded substantial genetic gain for many quantitative traits in livestock and plant breeding over the past few decades (Dekkers & Hospital, 2002). Research on the genetic architecture of quantitative traits and its deviation from the infinitesimal model has a long history, see e.g. Lynch & Walsh (1998), Mackay (2001) or Hill (2010) and references therein. Many results are trait specific and controversial. Having knowledge about the number of quantitative

trait loci (QTLs) as well as the distribution of their effects for a particular trait would contribute to a deeper understanding of its genetic architecture. Furthermore, in animal breeding, genetic evaluation methods that use massive genetic marker data have become common in practice (Hayes *et al.*, 2009). For the development of these models as well as for the assessment of long-term genetic progress using this technology it would be helpful to have knowledge about the number of QTLs and the distribution of QTL effects (Goddard, 2009; Daetwyler *et al.*, 2010).

Estimators of the number of QTLs have been proposed by Hayes & Goddard (2001) and Chamberlain *et al.* (2007). Hayes & Goddard (2001) estimated 50–100 segregating QTLs per trait in dairy cattle. Chamberlain *et al.* (2007) concluded that at least 30 QTLs are segregating. However, both authors could not rule out the possibility that there are many more QTLs with small effects, because they used results from QTL mapping studies. It was frequently shown that mapped QTLs were separated into a series of

* Corresponding author: Department of Animal Husbandry and Animal Breeding, University of Hohenheim, D-70599 Stuttgart, Germany. e-mail: r.wellmann@uni-hohenheim.de

additional QTLs by fine mapping experiments (e.g. Mackay & Anholt, 2006), pointing towards larger numbers of QTLs. Support for a higher number of QTLs comes also from human height data (Visscher, 2008) and from genomic selection experiments. Luan *et al.* (2009) using 20 k single nucleotide polymorphisms (SNPs) for genomic breeding value estimation in dairy cattle found that the method called G-BLUP for the estimation of genomic breeding values produced similar or better than Bayesian methods. This is usually only the case if many QTLs are segregating and the ratio of marker to QTL is low (Meuwissen & Goddard, 2010). Several estimators are known that are not applicable to outbred populations, e.g. the Castle–Wright estimator or the estimator proposed by Otto & Jones (2000). Hill *et al.* (2008) noted that for a given rate of inbreeding depression, as the number of loci increases and the gene frequencies move towards 0 or 1, the dominance variance decreases towards zero. Consequently, the relationship between dominance variance and inbreeding depression, which is the mean decrease of the trait value when the inbreeding coefficient increases from 0 to 1, could be used to estimate the number of QTLs. Such an estimator would not rely on QTL mapping results but would need knowledge about inbreeding depression and dominance variance, which have been estimated for some populations and traits (Misztal, 1997).

A second main issue in the genetic architecture of quantitative traits is the distribution of QTL effects. In a meta-analysis of information from QTL mapping experiments, Hayes & Goddard (2001) found the distribution to be moderately leptocurtic, consistent with many genes of small effect and few of large effect. Some studies have reported double exponential distributions (Eyre-Walker & Keightley, 2007; Bennewitz & Meuwissen, 2010). Traits are known with segregating QTLs of large effect (e.g. the pleiotropic *DGATI* in Holstein cattle, see Grisart *et al.*, 2004) even though the trait has been under selection for a long time. On the other hand, recent large-scale association studies conducted for human height revealed 54 variants which collectively explained only a few percent of the genetic variance of this highly heritable trait (Visscher, 2008), indicating that there are many QTLs with small effects and none with an exceptional large effect for this trait. This suggests that the distribution of QTL effects is trait specific. In QTL mapping studies, it is often estimated how many QTLs with a ‘large’ effect are segregating. Naturally, these ‘large’ QTLs determine the distribution of the effects and could therefore be used to derive parameters of the distribution. However, the true number of segregating large QTLs may be even smaller than reported by QTL mapping studies due to publication bias, as negative results tend not to be published.

The distribution of the dominance coefficient was derived by Bennewitz & Meuwissen (2010) for meat production traits in pigs. The dominance coefficient $\delta_n = d_n/|a_n|$ depends on the dominance effect d_n and on the absolute additive effect $|a_n|$, which is half the difference between the homozygous genotypic values. They postulated a normal distribution with a mean of $\mu_\Delta = 0.193$ and a standard deviation of $\sigma_\Delta = 0.312$ for segregating alleles. Thus, overdominance is a rare but not negligible event for these traits. However, if dominance is more important, e.g. for fitness-related traits, these figures might not hold and overdominance might play a more important role. For example, Ishikawa (2009) mapped an overdominant allele affecting body weight in mice at 6 weeks of age with a dominance coefficient of up to 6.6. Luo *et al.* (2001) found that overdominance is an important property of most loci associated with heterosis in rice. Rocha *et al.* (2004) found more directional dominance in fitness-related traits compared to growth or body composition traits in mice. García-Dorado *et al.* (1999) suggested in their review an average dominance coefficient of $0.94 < \bar{\delta} < 0.98$ ($0.01 < \bar{h} = (1 - \bar{\delta})/2 < 0.03$) for new lethal mutants and argued that the typical value $\bar{\delta} = 0.2$ ($\bar{h} = 0.4$) can be questioned for new non-severe deleterious mutations, but suggested $\bar{\delta} = 0.8$ ($\bar{h} = 0.1$). However, even if dominance is common, it does not necessarily cause much dominance variance due to the U-shaped distribution of allele frequencies (Hill *et al.*, 2008).

Also important is the joint distribution of the additive effects a_n of the mutant alleles, the dominance coefficients δ_n and allele frequencies p_n . Several studies assume that they are independent, e.g. Bennewitz & Meuwissen (2010). But this is likely not true. Kacser & Burns (1981) argued that the relationship between enzyme activity and end-product is hyperbolic. Thus, if a high enzyme activity is needed to produce a large trait value, then it is expected that the allele that increases the trait value shows incomplete dominance (i.e. $\delta_n > 0$). Moreover, since a differentiable function is locally approximately linear, one would expect that alleles with small effect show little dominance (i.e. $\delta_n \approx 0$), although this is not always confirmed empirically, e.g. Caballero & Keightley (1994), Bennewitz & Meuwissen (2010). Thus, $|a_n|$ and δ_n should be positively correlated. Not only arguments based on metabolic pathways (Keightley, 1996) but also arguments based on selection point into this direction. New mutations that affect a trait that was under selection are most often deleterious and recessive. Recessivity of mutants arises as a side effect of the margin of safety built into most metabolic pathways. But the extent to which this safety margin results from natural selection remains controversial (Bourguet, 1999). Since selection on modifier alleles that act on heterozygote

mutant alleles would be much stronger than that acting on homozygotes (Fisher, 1928) and since advantageous dominant alleles less likely become lost by random drift than recessive alleles with the same function, alleles that have been fixed are likely to be dominant over new mutant alleles. However, selection coefficients on modifier alleles would be extremely small (Wright, 1934), so modifier alleles may be of little importance. More important for the joint distribution of allelic effects and allele frequencies is the change of an allele frequency by one generation of selection, as it depends on the current allele frequency (Falconer & Mackay, 1996, p. 28). Since selection against a recessive deleterious allele becomes inefficient when the allele frequency p_n is small, and since most mutant alleles are recessive and deleterious, one would expect that segregating deleterious alleles ($a_n < 0$) with small frequency ($p_n \approx 0$) tend to be recessive ($\delta_n \approx 1$). For the same reason, segregating advantageous alleles ($a_n > 0$) with high frequency ($p_n \approx 1$) tend to be dominant ($\delta_n \approx 1$). Due to the U-shaped distribution of allele frequencies, most allele frequencies are close to 0 or 1, so these arguments are valid for the majority of the alleles. Therefore, we have likely $p_n > 0.5$ if $\text{sign}(a_n) = \text{sign}(\delta_n)$ and $p_n < 0.5$ if $\text{sign}(a_n) \neq \text{sign}(\delta_n)$. All these arguments also confirm that $|a_n|$ and δ_n are likely positively correlated.

Allelic effects and allele frequencies affect the additive variance, the dominance variance and the inbreeding depression of a trait. However, dominance variance is not only determined by these quantities, but may be increased quite substantially by linkage disequilibrium as shown by Avery & Hill (1979). Given that alleles are coded as 0 and 1, the coefficient of linkage disequilibrium $D_{n,m}$ between loci n and m equals the covariance of the alleles, if they are chosen at random from the infinite gametic pool.

The objective of this paper was to find parameter settings for the simulation of a particular trait that account for dominance realistically and are consistent with literature reports. A further concern is the identification of genetic architectures for a particular trait that are consistent with available estimates of variance components and inbreeding depression. First, formulae were developed that quantify the contributions of different sources that affect the phenotypic variance. The minimum number of QTLs is calculated that is needed such that inbreeding depression on average does not cause more dominance variance than the true dominance variance of the trait. This novel estimator predicts a lower bound for the number of QTLs. Using the developed formulae, it is shown how the number of 'large' QTLs can be used to derive parameters of the distribution of QTL effects. Finally, the role of overdominance for cases where the ratio of dominance variance to additive genetic variance is high is investigated. The developed

formulae are based on the availability of additive genetic variance as well as dominance variance components. They were applied using published variance components estimates for milk yield and productive life (PL) in Holstein dairy cattle.

2. Theory

This section is divided into several parts. In (i), formulae for the variance components are derived that account for linkage disequilibrium. Their expectations under a wide range of assumptions are derived in (ii). Part (iii) discusses the distribution of allele frequencies and linkage disequilibrium of alleles. In (iv), parameters that describe the joint distribution of additive effect and dominance coefficients are derived for different scenarios. Parameters that are estimated in (iii) and (iv) are needed to evaluate the formulae that are derived in later parts. In (v), formulae are derived to estimate the contributions of different sources to the additive and dominance variance. Lower bounds for the number of QTLs are presented in (vi). Implications on the importance of overdominance can be found in (vii). Upper bounds for the number of large QTLs are given in (viii). Part (ix) shows how the formulae can be used to estimate the number of QTLs, the variance of the additive effects, and the mean and variance of the dominance coefficients. For a better readability, the proofs of the formulae are given in the electronic appendix (available at <http://journals.cambridge.org/GRH>). A list of symbols used in the paper is given in Table 1.

(i) Variance components and inbreeding depression

According to Falconer & Mackay (1996), the breeding value of an individual is the sum of all substitution effects that are carried by this individual, i.e. up to an additive constant the breeding value is

$$BV = \sum_{n \in Q} \alpha_n (v_n + m_n) \quad \text{with } \alpha_n = a_n + d_n (q_n - p_n),$$

where $v_n \in \{0, 1\}$ is the paternal and $m_n \in \{0, 1\}$ is the maternal allele of the individual at QTL n and Q consists of all polymorphic QTLs. Here, it is assumed that all QTLs are biallelic with alleles 0 and 1. Moreover, we assume throughout the paper the absence of genetic interactions. The mutant allele at QTL n has frequency p_n , additive effect a_n and dominance effect d_n . The other allele has frequency $q_n = 1 - p_n$. The dominance deviation of the individual is

$$DV = \sum_{n \in Q} -2d_n (v_n - p_n)(m_n - p_n),$$

see Falconer & Mackay (1996, Table 7.3). The additive variance is the variance of the breeding value and the dominance variance is the variance of the dominance deviation of an individual whose parents are

Table 1. Table of symbols

a_n	Additive effect of the mutant allele at QTL n
a'_n	Additive effect of an arbitrarily coded allele at QTL n
d_n	Dominance effect of QTL n
$\delta_n = d_n/ a_n $	Dominance coefficient of QTL n
p_n	Frequency of allele 1 (the mutant allele) at locus n
$q_n = 1 - p_n$	Frequency of allele 0 at locus n
$\alpha_n = a_n + d_n(q_n - p_n)$	Allele substitution effect of QTL n
$h_n = 2p_nq_n$	Heterozygosity at locus n in the case of HWE
N, N_e	Population size and effective population size
V_A, V_D, \mathcal{I}	Additive and dominance variance and inbreeding depression
$\bar{V}_A, \bar{V}_D, \bar{\mathcal{I}}$	Expected additive and dominance variance and inbreeding depression
$V_A^a, V_A^{ad}, V_A^d, V_A^{LD}$	Contributions of different sources to the additive variance
$\bar{V}_A^a, \bar{V}_A^{ad}, \bar{V}_A^d, \bar{V}_A^{LD}$	Expected contributions of different sources to the additive variance
V_D^d, V_D^{LD}	Contributions of different sources to the dominance variance
$\bar{V}_D^d, \bar{V}_D^{LD}$	Expected contributions of different sources to the dominance variance
$\mu_A = E(a_n), \sigma_A^2 = \text{Var}(a_n)$	Mean and variance of the additive effect of mutant alleles of QTL
$\mu_D = E(d_n), \sigma_D^2 = \text{Var}(d_n)$	Mean and variance of the dominance effect of QTL
$\mu_\Delta = E(\delta_n), \sigma_\Delta^2 = \text{Var}(\delta_n)$	Mean and variance of the dominance coefficient of QTL
$\mu_{A'} = 0, \sigma_{A'}^2 = E(a_n'^2)$	Mean and variance of the additive effect of arbitrarily coded alleles
F	Distribution function of $a'_n/\sigma_{A'}$
$\lambda = E(a'_n)/\sigma_{A'}$	
$K = E(\delta_n)/E(\delta_n)$	
$c_v = \sigma_\Delta/\mu_\Delta$	Coefficient of variation of the dominance coefficient
$\gamma = E(h_n(q_n - p_n)^2)/E(h_n)$	
$D_{LD}^{0,0}$	Average expected coefficient of LD of a chromosome
$D_{LD}^{i,j,k}$	Generalization of the average expected coefficient of LD
$r_1 = \text{Cov}(a_n , \delta_n)/E(a_n)E(\delta_n)$	Describes the dependency between a_n and δ_n
$r_2 = \text{Cov}(a_n^2, \delta_n^2)/E(a_n^2)E(\delta_n^2)$	Describes the dependency between a_n and δ_n
$r_{2,1} = \text{Cov}(a_n^2, \delta_n)/E(a_n^2)E(\delta_n)$	Describes the dependency between a_n and δ_n
p_k	Probability that a segregating QTL has MAF larger than k
$s > 0$	QTL with $ a_n > s$ are considered as 'large'
\mathcal{Q}, \mathcal{Q}	Set of all polymorphic QTLs, number of polymorphic QTLs
$\mathcal{Q}_c, \mathcal{N}_c$	Set of all polymorphic QTLs and set of nucleotides at chromosome c
\mathcal{C}, \mathcal{C}	Set of chromosomes, number of chromosomes
$\bar{Q} = p_k \mathcal{Q}$	Expected number of QTLs with MAF $> k$
\bar{n}	Expected number of large QTLs with MAF $> k$

randomly chosen from the population. Thereby, the additive effects, the dominance effects and the allele frequencies are assumed to be fixed parameters and randomness arises from random sampling of the genotypes. That is, only the paternal and maternal alleles are random. Since maternal and paternal alleles are independent and identically distributed with $\text{Var}(v_n) = \text{Var}(m_n) = p_nq_n$ and $E(v_n) = E(m_n) = p_n$, we obtain (see the proofs in the electronic appendix)

$$V_A = \text{Var}(\text{BV}) = \sum_{n \in \mathcal{Q}} h_n \alpha_n^2 + 2 \sum_{n < m \in \mathcal{Q}} 2\alpha_n \alpha_m D_{n,m}, \quad (1)$$

where $h_n = 2p_nq_n$ is the heterozygosity at locus n in the case of Hardy-Weinberg Equilibrium (HWE) and $D_{n,m} = \text{Cov}(v_n, v_m) = \text{Cov}(m_n, m_m)$ is the coefficient of linkage disequilibrium between locus n and m . For the dominance variance, we obtain

$$V_D = \text{Var}(\text{DV}) = \sum_{n \in \mathcal{Q}} h_n^2 d_n^2 + 2 \sum_{n < m \in \mathcal{Q}} 4d_n d_m D_{n,m}^2. \quad (2)$$

Note that $V_A + V_D$ equals the genotypic variance within a line given by Avery & Hill (1979). If linkage disequilibrium is neglected, i.e. if $D_{n,m} = 0$ is assumed, then only the left summands in eqns (1) and (2) remain, which is equal to the well-known formulae given in Falconer & Mackay (1996). We assume that loci from different chromosomes are not in linkage disequilibrium, but loci from the same chromosome may be linked. Take \mathcal{C} to be the set of chromosomes and for $c \in \mathcal{C}$ let \mathcal{Q}_c denote the polymorphic QTL at chromosome c . Then we have

$$V_A = \sum_{n \in \mathcal{Q}} h_n \alpha_n^2 + 2 \sum_{c \in \mathcal{C}} \sum_{n < m \in \mathcal{Q}_c} 2\alpha_n \alpha_m D_{n,m} \quad (3)$$

and

$$V_D = \sum_{n \in \mathcal{Q}} h_n^2 d_n^2 + 2 \sum_{c \in \mathcal{C}} \sum_{n < m \in \mathcal{Q}_c} 4d_n d_m D_{n,m}^2. \quad (4)$$

The first summand in eqn (3) equals

$$\sum_{n \in Q} h_n a_n^2 = \sum_{n \in Q} h_n a_n^2 + 2 \sum_{n \in Q} h_n (q_n - p_n) a_n d_n + \sum_{n \in Q} h_n (q_n - p_n)^2 d_n^2.$$

Therefore, we can write

$$V_A = V_A^a + V_A^{ad} + V_A^d + V_A^{LD},$$

where $V_A^a = \sum_{n \in Q} h_n a_n^2$ and $V_A^d = \sum_{n \in Q} h_n (q_n - p_n)^2 d_n^2$ are the contributions of single additive effects and dominance effects to V_A , respectively. $V_A^{ad} = \sum_{n \in Q} 2 \times h_n (q_n - p_n) a_n d_n$ is a correction term that can be negative. Finally, $V_A^{LD} = 2 \sum_{c \in C} \sum_{m < m \in Q_c} 2 a_n a_m D_{n,m}$ is the contribution that comes from covariances between linked loci. Similarly, we can write

$$V_D = V_D^d + V_D^{LD},$$

where $V_D^d = \sum_{n \in Q} h_n^2 d_n^2$ is the contribution of single dominance effects to V_D and $V_D^{LD} = 2 \sum_{c \in C} \sum_{m < m \in Q_c} 4 d_n d_m D_{n,m}^2$ is the contribution that comes from covariances between linked loci.

The expected genotypic value of an individual that is randomly chosen from the population in HWE is

$$G_{HWE} = \mu + \sum_{n \in Q} 0 q_n^2 + (a_n + d_n) 2 p_n q_n + 2 a_n p_n^2.$$

Assume that an inbred line is established from two individuals that are randomly chosen from the population. Individuals from this line carry the genotype 00 at locus n with probability q_n and the genotype 11 with probability p_n . Thus, the expected genotypic value of a completely inbred individual is

$$G_{inbred} = \mu + \sum_{n \in Q} p_n 2 a_n + q_n 0.$$

The inbreeding depression, i.e. the expected decrease of the genotypic value when the inbreeding coefficient increases from 0 to 1 is therefore

$$\mathcal{I} = G_{HWE} - G_{inbred} = \sum_{n \in Q} h_n d_n.$$

(ii) *Expectations of variance components and inbreeding depression*

In this paper, we treat a population as if it would be a random outcome of a simulation study, i.e. a population is viewed as a realization from all hypothetical populations. Therefore, not only the genotypes are random but also the additive effects, the dominance effects, the allele frequencies and the coefficients of linkage disequilibrium for each pair of loci. Alternatively, randomness of the additive and dominance effects could be due to uncertainty about the true effects (Gianola *et al.*, 2009). In both cases, V_A and V_D are random variables. In our setting,

V_A and V_D are treated as random since they would attain different values in each hypothetical replicated population. Their expectations are of interest in order to relate parameters of the distributions of additive effects and dominance coefficients to the expected variance components.

For $n \in Q$, the additive effect a_n of the mutant allele has mean μ_A and variance σ_A^2 , the dominance effect d_n has mean μ_D and variance σ_D^2 . The dominance coefficient $\delta_n = d_n/|a_n|$ has mean μ_Δ and variance σ_Δ^2 . Take \mathcal{N}_c to be the set of nucleotides at chromosome c . The probability that a nucleotide is a segregating QTL is equal for all nucleotides and nucleotides become QTL independent from each other. For simplicity, we assume that all chromosomes have equal length. The expected number of segregating QTLs at one chromosome is therefore Q/C , where $Q = E(\#Q)$ is the expected total number of segregating QTLs and $C = \#C$ is the number of chromosomes.

We derive expectations of the variance components under a wide range of assumptions. That is, we first derive equations using only a small number of assumptions and successive include more of them. The assumptions are

- (A1) (p_n, a_n, d_n) and (p_m, a_m, d_m) are identically distributed,
- (A2) allelic effects are independent from the allelic effects and the allele frequencies at other loci and linkage disequilibrium does not depend on the allelic effects (see below),
- (A3) $D_{LD}^{1,0} = D_{LD}^{1,1} = D_{LD}^{1,0} = 0$ (see below),
- (A4) dominance effect d_n and heterozygosity h_n are independent,
- (A5) $(|a_n|, d_n)$ and heterozygosity h_n are independent,
- (A6) (a) $a_n|\delta_n, p_n$ has a symmetrical distribution with mean $\mu_A = 0$, or
 (b) $p_n > 0.5$ if $\text{sign}(a_n) = \text{sign}(\delta_n)$ and $p_n < 0.5$ if $\text{sign}(a_n) \neq \text{sign}(\delta_n)$, i.e. $\text{sign}(a_n) = -\text{sign}((q_n - p_n)\delta_n)$ (see the introduction),

given $n \neq m \in Q$. Note that additive effects and dominance coefficients may be dependent. By using only (A1) we obtain

$$\tilde{V}_A = E(V_A) = \tilde{V}_A^a + \tilde{V}_A^d + \tilde{V}_A^{ad} + \tilde{V}_A^{LD}, \tag{5}$$

$$\tilde{V}_D = E(V_D) = \tilde{V}_D^d + \tilde{V}_D^{LD},$$

$$\tilde{\mathcal{I}} = E(\mathcal{I}) = QE(h_n d_n | n \in Q),$$

where

$$\tilde{V}_A^a = QE(h_n a_n^2 | n \in Q), \tag{6}$$

$$\tilde{V}_A^d = QE(h_n (q_n - p_n)^2 d_n^2 | n \in Q),$$

$$\tilde{V}_A^{ad} = 2QE(h_n (q_n - p_n) a_n d_n | n \in Q),$$

$$\tilde{V}_D^d = QE(h_n^2 d_n^2 | n \in \mathcal{Q}),$$

$$\tilde{V}_A^{LD} = \frac{2Q^2}{C} \frac{1}{\#\mathcal{N}_c^2} \sum_{n \neq m \in \mathcal{N}_c} E(\alpha_n \alpha_m D_{n,m} | n, m \in \mathcal{Q})$$

and

$$\tilde{V}_D^{LD} = \frac{4Q^2}{C} \frac{1}{\#\mathcal{N}_c^2} \sum_{n \neq m \in \mathcal{N}_c} E(d_n d_m D_{n,m}^2 | n, m \in \mathcal{Q}).$$

In most of the following formulae, we suppress the condition that the SNPs are polymorphic QTLs in order to improve the readability.

Note that a_n denotes the effect of the mutant allele. However, in real populations it is unknown which allele the mutant allele is. One would rather use an arbitrary coding of the alleles. Take a'_n to be the additive effect of the arbitrarily coded allele. We have $a'_n = a_n$ or $a'_n = -a_n$, each with probability 0.5. Although $|a'_n| = |a_n|$, the distributions of both random variables are different. Whereas the effect of the mutant allele may have a non-zero mean μ_A , the effect of the arbitrary coded allele a'_n has a symmetrical distribution with mean $\mu_{A'} = 0$ and variance $\sigma_{A'}^2 = E(a_n^2)$. Since a'_n has mean 0, we can write $E(|a_n|) = E(|a'_n|) = \lambda \sigma_{A'}$, where the parameter λ depends on the standardized distribution of a'_n .

Let $D_{n,m}^{i,j,k} = D_{n,m}^i (q_n - p_n)^j (q_m - p_m)^k$, which is needed to simplify \tilde{V}_A^{LD} . Now we assume additionally that linkage disequilibrium does not depend on the allelic effects and that allelic effects at different loci are independent, i.e. a_n is independent from $(a_m, d_m, p_m, D_{n,m})$ and d_n is independent from $(a_m, d_m, p_m, D_{n,m}, D_{n,m}^{1,1,1})$, given $n \neq m \in \mathcal{Q}$. With (A1) and (A2) we obtain

$$\tilde{V}_A^{LD} = 2 \frac{Q^2}{C} (\mu_A^2 D_{LD}^{1,0,0} + 2\mu_A \mu_D D_{LD}^{1,0,1} + \mu_D^2 D_{LD}^{1,1,1}), \quad (7)$$

$$\tilde{V}_D^{LD} = 4 \frac{Q^2}{C} \mu_D^2 D_{LD}^{2,0,0},$$

where

$$D_{LD}^{i,j,k} = \frac{1}{\#\mathcal{N}_c^2} \sum_{n \neq m \in \mathcal{N}_c} E(D_{n,m}^{i,j,k} | n, m \in \mathcal{Q}).$$

Roughly speaking, $E(D_{n,m}^{2,0,0} | n, m \in \mathcal{Q})$ would be the mean D^2 -value for loci n and m , averaged over many hypothetical populations for which n and m are polymorphic QTLs. If these expected D^2 -values are averaged over all pairs of loci at one chromosome, then $D_{LD}^{2,0,0}$ is obtained. Simulations of neutral alleles showed that likely $D_{LD}^{1,0,0} = D_{LD}^{1,0,1} = D_{LD}^{1,1,0} = 0$ but $D_{LD}^{1,1,1} \neq 0$ (A3), so this will be assumed in the rest of this paper. Thus, linkage disequilibrium not only contributes to the dominance variance but also to the additive variance. If there are many QTLs per chromosome that affect the trait, the contributions from

single loci can in principle be much smaller than the contributions that arise from covariances of linked loci, since the latter increases quadratically with the number of QTLs. If additionally dominance effect and heterozygosity are independent (A4), then

$$\tilde{V}_A^d = QE(h_n)(\mu_D^2 + \sigma_D^2)\gamma, \quad (8)$$

$$\tilde{V}_D^d = QE(h_n)(\mu_D^2 + \sigma_D^2) \frac{1-\gamma}{2},$$

$$\tilde{\mathcal{I}} = QE(h_n)\mu_D,$$

where $\gamma = E(h_n(q_n - p_n)^2) / E(h_n) = 1 - 2E(h_n^2) / E(h_n)$. The last equation in (8) shows that inbreeding depression occurs only if dominance effects do not have mean zero. It may be more convenient to express the formulae for \tilde{V}_A^d , \tilde{V}_D^d and $\tilde{\mathcal{I}}$ using moments of the dominance coefficients rather than moments of the dominance effects. These can be obtained from eqn (8) with

$$\mu_D = \mu_A \lambda \sigma_{A'} (1 + r_1), \quad (9)$$

$$\mu_D^2 + \sigma_D^2 = (\mu_A^2 + \sigma_A^2) \sigma_{A'}^2 (1 + r_2),$$

where

$$r_j = \frac{\text{Cov}(\delta_n^j, |a_n|^j)}{E(\delta_n^j)E(|a_n|^j)} \quad \text{for } j=1, 2.$$

The scale invariant parameters r_1 and r_2 characterize the dependency between additive effects and dominance coefficients. Scale invariance means in particular that r_1 and r_2 do not depend on the variance of the additive effects. For biological reasons, we expect that $r_1, r_2 \geq 0$. If additive effects and dominance coefficients are independent, then $r_1 = r_2 = 0$.

Now we assume additionally that the absolute additive effect and the heterozygosity are independent (A5). We have

$$\tilde{V}_A^a = Q\sigma_A^2 E(h_n). \quad (10)$$

In simulation studies, the additive effect a_n often has a symmetrical distribution with mean 0 given δ_n and p_n (A6a). In this case, we have

$$\tilde{V}_A^{ad} = 0. \quad (11)$$

However, we believe that (A6a) is an unrealistic assumption for real populations for several reasons. Most important, selection acts mainly on the additive component of genotypic values. Therefore, one would expect that in a long-term selected population alleles with small contribution to the additive variance are overrepresented. The contribution of QTL n to the additive variance depends on $|a_n| = |a_n + (q_n - p_n)d_n|$ which is small if a_n and $(q_n - p_n)d_n$ have opposite signs.

Although this is violated in real populations as well for some alleles, it likely holds for the majority of the alleles as pointed out in the introduction and it provides not only more realistic results as (A6a), but it can also easily be accounted for in simulation studies by simply choosing the sign of a_n as $\text{sign}(a_n) = -\text{sign}((q_n - p_n)\delta_n)$. Under (A6b) we have

$$\tilde{V}_A^{ad} = -2\tilde{V}_A^a E(|\delta_n|)(1 + r_{2,1}) \frac{E(h_n \sqrt{1 - 2h_n})}{E(h_n)}, \quad (12)$$

where

$$r_{2,1} = \frac{\text{Cov}(a_n^2, |\delta_n|)}{\sigma_A^2 E(|\delta_n|)}.$$

We can write $E(|\delta_n|) = \mu_\Delta K$, where K depends on the coefficient of variation $c_v = \sigma_\Delta / \mu_\Delta$.

Several parameters appeared in these equations that need to be estimated for the application of formulae. The next part provides estimates of parameters that are related to heterozygosity and linkage disequilibrium. In Part (iv), parameters that characterize the joint distribution of additive effect and dominance coefficient are calculated for different scenarios.

(iii) *Heterozygosity and linkage disequilibrium*

In accordance with the diffusion approximation (Crow & Kimura, 1970) and Hill *et al.* (2008), the density of a segregating neutral mutant allele is $f(p) = 1/(\kappa p)$ on the interval $[1/(2N), 1 - 1/(2N)]$ if mutations are rare and the population is in mutation drift equilibrium. Since $f(p)$ is a density, we have $\kappa = \ln(2N - 1)$, where N is the population size. Take $p_k = P(k < p_n < 1 - k)$ to be the probability that a segregating allele has minor allele frequency (MAF) larger than k , where $1/(2N) \leq k \leq 0.5$. We have

$$\gamma = \frac{1}{3} \frac{(N - 1)^2}{N^2}, \quad E(h_n) = \frac{1}{\kappa} \frac{N - 1}{N}, \quad \text{and} \quad (13)$$

$$p_k = \frac{\ln((1/k) - 1)}{\kappa}.$$

Note that the population size is usually so large that the factors $(N - 1)^2/N^2$ and $(N - 1)/N$ can be ignored. The effective population size is assumed to be constant. But for most livestock populations, the effective population size decreased in the past. In a small bottlenecked population, more alleles with extreme frequencies become lost than new mutations arise and the remaining alleles more likely have intermediate frequencies. Thus, the mean heterozygosity of segregating alleles would be larger. But on the other hand, the population size is usually much larger than the effective population size, so many new mutations arise each generation and these new mutations decrease the mean heterozygosity of segregating alleles.

In order to get estimates of γ , $E(h_n)$ and p_k for a bottlenecked population, we simulated a population for which the effective population size decreased from $N_e = 1000$ to $N_e = 100$ within 400 generations in accordance with the results of Villa-Angulo *et al.* (2009). The total population size remained constant with $N = 1000$, which was achieved by an unequal number of males and females. We obtained the estimates $\hat{E}(h_n) = 0.19$, $\hat{E}(h_n \sqrt{1 - 2h_n}) = 0.087$, $\hat{\gamma} = 0.28$, and $\hat{p}_k = 0.71$ of the segregating alleles had an MAF larger than $k = 0.01$. We also obtained estimates for the LD from the simulated population resulting in $\hat{D}_{LD}^{1,1,1} = 2.4 \times 10^{-5}$ and $\hat{D}_{LD}^{2,0,0} = 1.3 \times 10^{-4}$. These values are used in the examples. They are needed to apply the formulae that were derived in the previous section. For a population in mutation–selection equilibrium a more extreme L- or U-shaped distribution may be expected. The shape of this distribution would affect these parameters, but not the validity of the formulae.

(iv) *Dependency between additive effect and dominance coefficient*

The formulae that are derived could be applied by assuming $r_1 = r_2 = r_{2,1} = 0$ which holds if additive effect and dominance coefficient are independent. However, the literature suggests that they are not independent. Therefore, we consider different possibilities to model dependent effects and for these scenarios we give estimates for the scale invariant parameters c_v , r_1 , r_2 , $r_{2,1}$, λ and K . In all scenarios, $|a_n|$ and δ_n are positively correlated.

Because of the scale invariance of all relevant parameters, the distributions of a'_n and δ_n need to be specified only up to constant factors that will be estimated in the next sections. That is, we only need to specify the joint distribution of random variables $\tilde{a}_n, \tilde{\delta}_n$. These are multiplied by constant factors to obtain a'_n and δ_n . The parameters $c_v, r_1, r_2, r_{2,1}, \lambda$ and K for (a_n, δ_n) are exactly the same as for $(\tilde{a}_n, \tilde{\delta}_n)$. In scenarios (1), (3) and (4), mean and standard deviation of $\tilde{\delta}_n$ were chosen as 0.2 and 0.3, as suggested by Bennewitz & Meuwissen (2010).

Scenario (1) assumes that δ_n and a'_n are independent with

$$\tilde{\delta}_n \sim \mathcal{N}(0.2, 0.3^2) \quad \text{and} \quad \tilde{a}_n \sim \mathcal{L}(0, 1),$$

where $\mathcal{L}(0, 1)$ denotes the Laplace distribution with mean 0 and variance 1.

In all other scenarios there is a lack of segregating additive alleles with large effect. This could arise in real populations because these alleles have been fixed due to selection or because of a hyperbolic relationship between enzyme activity and flux (Kacser & Burns, 1981).

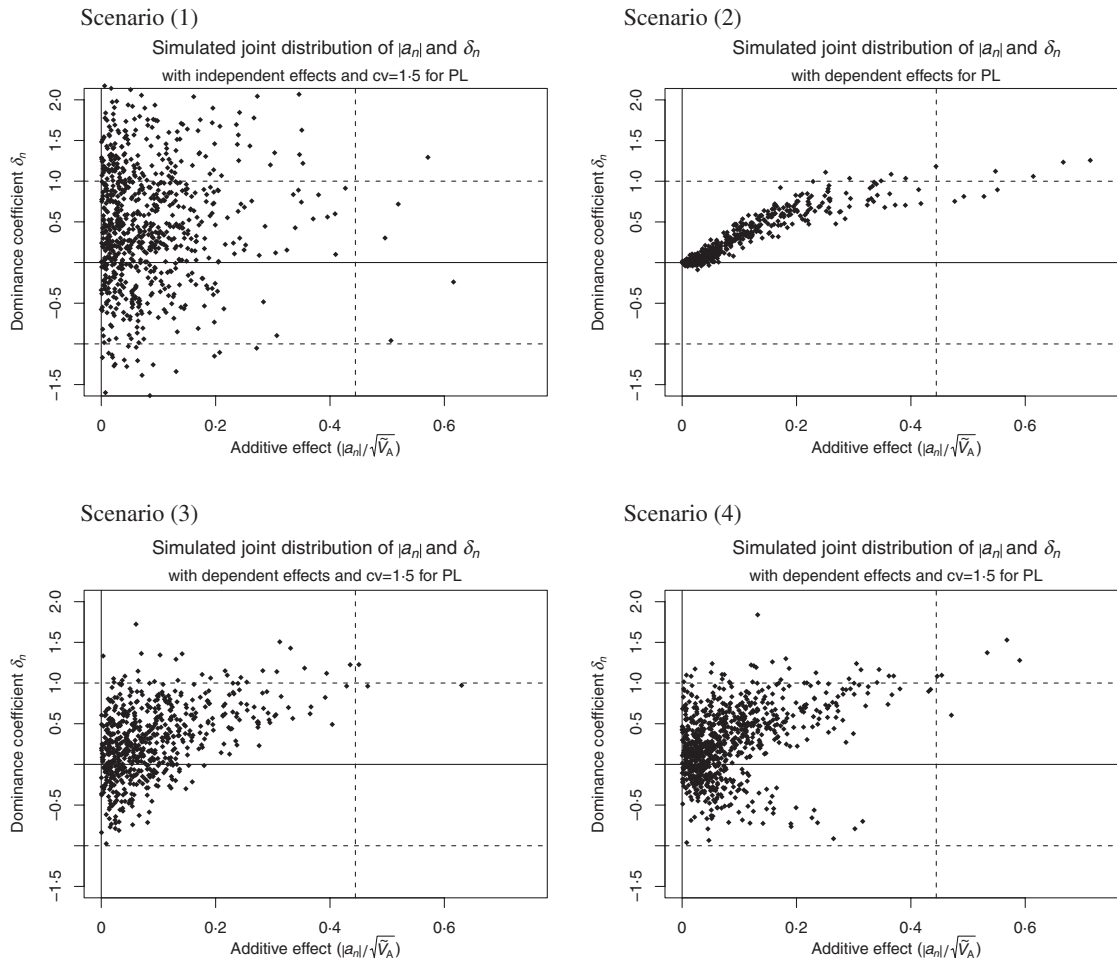


Fig. 1. Scatter plots of the considered joint distributions of $|a_n|$ and δ_n from Part (iv), fitted to the trait productive life (PL) under assumption (6b), so the signs of the additive effects are such that the alleles contribute little to the additive variance. (1) Absolute additive effect and dominance coefficient are independent. (2) Alleles with large effect show directional dominance, but alleles of small effect are additive. (3) Alleles with small effects show highly variable dominance coefficients, but for alleles with large effects, heterozygous effect is above the average effect of the two homozygotes. (4) Alleles with small effects show highly variable dominance coefficients, but alleles with large effect are incomplete recessive or dominant.

Scenario (2) reflects the conclusions drawn from the analysis of enzyme networks. Genes of large effect show directional dominance, but for genes of small effect the dominance coefficient is close to 0. That is,

$$\tilde{a}_n \sim \mathcal{L}(0, 1) \text{ and } \tilde{\delta}_n | \tilde{a}_n \sim \mathcal{N}\left(\frac{\tilde{a}_n^2}{1 + \tilde{a}_n^2}, 0.01 | \tilde{a}_n\right).$$

Scenario (3) reflects the results reported by Caballero & Keightley (1994), i.e. alleles with large effect tend to be partially recessive or even over-dominant and the heterozygous effect is above the average effect of the two homozygotes. But alleles of small effect show highly variable dominance coefficients. More precisely,

$$\tilde{\delta}_n \sim \mathcal{N}(0.2, 0.3^2) \text{ and } \tilde{a}_n | \tilde{\delta}_n \sim \mathcal{N}(0, \exp(3\tilde{\delta}_n)).$$

Scenario (4) is similar to scenario (3), but for alleles with large effect, the heterozygous effect could also be

below the average effect of the two homozygotes. This may be more suitable for traits where mutant alleles with large effect could also be recessive and advantageous. We have

$$\tilde{\delta}_n \sim \mathcal{N}(0.2, 0.3^2) \text{ and } \tilde{a}_n | \tilde{\delta}_n \sim \mathcal{N}(0, (0.5 + |\tilde{\delta}_n|)^4).$$

Scatter plots of the distributions are shown in Fig. 1. This figure was fitted for the trait PL in Holstein cattle (see Examples section). Table 2 shows the parameter values for all scenarios.

(v) Contributions of different sources to variance components

We have under assumptions (A1) and (A2)

$$\tilde{V}_D^{LD} = 4 \frac{(\tilde{I} - QCov(h_n, d_n))^2}{CE(h_n)^2} D_{LD}^{2,0,0}. \tag{14}$$

Table 2. Characteristics of the considered joint distributions of $|a_n|$ and δ_n that are described in Part (iv)

Parameter	Scenario			
	1	2	3	4
c_v	1.50	1.00	1.50	1.50
r_1	0.00	0.90	0.68	0.60
r_2	0.00	2.55	1.40	1.75
$r_{2,1}$	0.00	1.47	0.67	0.86
λ	0.71	0.71	0.72	0.69
K	1.45	1.02	1.45	1.45

Unfortunately, the covariance is unknown, so this equation cannot be used directly to estimate the contribution of linkage disequilibrium to the dominance variance. But it shows that if falsely $\text{Cov}(h_n, d_n) = 0$ is assumed, although alleles with intermediate frequencies tend to have larger dominance effects (i.e. $\text{Cov}(h_n, d_n) > 0$ due to overdominance) then \tilde{V}_D^{LD} would become overestimated. Similarly, we obtain under (A1)–(A3) that

$$\tilde{V}_A^{\text{LD}} = 2 \frac{(\tilde{I} - Q \text{Cov}(h_n, d_n))^2}{CE(h_n)^2} D_{\text{LD}}^{1,1,1}. \tag{15}$$

Again we have to assume that $\text{Cov}(h_n, d_n) = 0$ for being able to evaluate this formula. Then the contributions of dominance effects of linked loci to the additive and dominance variance depend only on the squared inbreeding depression, which can be estimated, but not directly on the unknown number of QTLs.

Since \tilde{V}_D can be estimated from the population and \tilde{V}_D^{LD} can be estimated with eqn (14), the contribution of single loci to the dominance variance can be obtained from

$$\tilde{V}_D^d = \tilde{V}_D - \tilde{V}_D^{\text{LD}}.$$

This contribution would become underestimated if falsely $\text{Cov}(h_n, d_n) = 0$ is assumed in (14) although $\text{Cov}(h_n, d_n) > 0$. If additionally h_n and d_n are independent (A4), then the contribution of dominance effects of single loci to the additive variance can be obtained from

$$\tilde{V}_A^d = \frac{2\gamma}{1-\gamma} \tilde{V}_D^d. \tag{16}$$

Under (A6a) we have

$$\tilde{V}_A^{\text{ad}} = 0,$$

whereas under (A6b) we have

$$\tilde{V}_A^{\text{ad}} = -(\tilde{V}_A - \tilde{V}_A^d - \tilde{V}_A^{\text{LD}}) \times \frac{2E(|\delta_n|)(1+r_{2,1})E(h_n\sqrt{1-2h_n})}{E(h_n) - 2E(|\delta_n|)(1+r_{2,1})E(h_n\sqrt{1-2h_n})}. \tag{17}$$

The latter formula depends, however, on $E(|\delta_n|)$ and $r_{2,1}$. The contribution of additive effects of single loci to the additive variance can then be obtained from

$$\tilde{V}_A^a = \tilde{V}_A - \tilde{V}_A^d - \tilde{V}_A^{\text{LD}} - \tilde{V}_A^{\text{ad}}$$

In summary, these formulae enable to estimate the contributions of different sources to the additive and dominance variance under (A1)–(A6a).

(vi) Lower bounds for the number of QTLs

In this section, we derive lower bounds for the expected number of QTLs. We obtain from eqns (5) and (6)

$$\frac{\tilde{I}^2}{\tilde{V}_D - \tilde{V}_D^{\text{LD}}} = \frac{Q^2 E(h_n d_n)^2}{QE(h_n^2 d_n^2)} = Q \frac{E(h_n d_n)^2}{E(h_n^2 d_n^2)}.$$

Thus,

$$Q = \frac{\tilde{I}^2}{\tilde{V}_D - \tilde{V}_D^{\text{LD}}} \frac{E(h_n^2 d_n^2)}{E(h_n d_n)^2} \geq \frac{\tilde{I}^2}{\tilde{V}_D - \tilde{V}_D^{\text{LD}}} \geq \frac{\tilde{I}^2}{\tilde{V}_D}. \tag{18}$$

Note that the proof only used (A1) and $\tilde{V}_D^{\text{LD}} \geq 0$, so the bounds hold regardless of how intense the selection is and how deleterious new mutations tend to be. Under (A1)–(A4) the contribution of linkage disequilibrium \tilde{V}_D^{LD} can be estimated as described in the previous part if dominance variance and inbreeding depression have been estimated for the population. Note that these bounds depend neither on the distribution of the additive effects nor on the correlation between $|a_n|$ and δ_n . However, they can be further improved under (A1)–(A4). Since

$$\frac{E(h_n^2 d_n^2)}{E(h_n d_n)^2} = \frac{1-\gamma}{2E(h_n)} \frac{(1+r_2)}{\lambda^2(1+r_1)^2} \frac{\mu_\Delta^2 + \sigma_\Delta^2}{\mu_\Delta^2} \tag{19}$$

and

$$\frac{\mu_\Delta^2 + \sigma_\Delta^2}{\mu_\Delta^2} \geq 1 + r_1^2 \frac{\lambda^2}{1-\lambda^2}, \tag{20}$$

we have

$$Q \geq \frac{\tilde{I}^2}{\tilde{V}_D - \tilde{V}_D^{\text{LD}}} \frac{1-\gamma}{2E(h_n)} \frac{1+r_1^2 \lambda^2 / (1-\lambda^2)}{\lambda^2(1+r_1)^2} (1+r_2). \tag{21}$$

The factor $(1+r_2)$ on the right-hand side is typically larger than 1 (see Table 2). Moreover, we have

$$\frac{1-\gamma}{2E(h_n)} = \frac{E(h_n^2)}{E(h_n)^2} \geq 1$$

and

$$\frac{1+r_1^2 \lambda^2 / (1-\lambda^2)}{\lambda^2(1+r_1)^2} \geq 1. \tag{22}$$

Thus, if any of these factors is unknown, then it can be skipped in eqn (21) and the inequality still holds. In particular, we have

$$Q \geq \frac{\tilde{I}^2}{\tilde{V}_D - \tilde{V}_D^{LD}} \frac{1-\gamma}{2E(h_n)} \tag{23}$$

(vii) *The role of overdominance*

Using (A1)–(A5) we obtain

$$\tilde{V}_D^d \tag{8} \quad QE(h_n)(\mu_D^2 + \sigma_D^2) \frac{1-\gamma}{2}$$

$$\tag{9} \quad QE(h_n)\sigma_{A'}^2 (\mu_\Delta^2 + \sigma_\Delta^2)(1+r_2) \frac{1-\gamma}{2}$$

$$\tag{10} \quad \tilde{V}_A^a (\mu_\Delta^2 + \sigma_\Delta^2)(1+r_2) \frac{1-\gamma}{2}.$$

Thus,

$$E(\delta_n^2) = \mu_\Delta^2 + \sigma_\Delta^2 = \frac{1}{1+r_2} \frac{\tilde{V}_D^d}{\tilde{V}_A^a} \frac{2}{1-\gamma}. \tag{24}$$

That is, the second moment of the dominance coefficients depends only on the ratio $\tilde{V}_D^d/\tilde{V}_A^a$ that can be estimated under (A1)–(A6a), but not on the unknown number of QTLs.

A lower bound for the probability of an allele to be under- or overdominant can be derived if the dominance coefficient is normally distributed and the ratio of dominance variance to additive variance is sufficient large. If $\tilde{V}_D^d/\tilde{V}_A^a > d_{\max} = (1+r_2)(1-\gamma)/2$, then eqn (24) shows that $\mu_\Delta^2 + \sigma_\Delta^2 \geq 1$. But then $P(|\delta_n| > 1)$ is minimized if $\mu_\Delta = 0$ and $\sigma_\Delta = 1$. It follows that

$$P(|\delta_n| > 1) \geq 1 - \frac{1}{\sqrt{2\pi}} \int_{-1}^1 e^{-t^2/2} dt = 0.317, \tag{25}$$

$$\text{if } \frac{\tilde{V}_D^d}{\tilde{V}_A^a} > d_{\max}.$$

Such a large probability appears to be unrealistic in real populations for almost any trait, see for example Charlesworth *et al.* (2009). Possible reasons for $\tilde{V}_D^d/\tilde{V}_A^a > d_{\max}$ to occur are the violation of one of the assumptions (A1)–(A6) or a wrong value for r_2 . If $E(\delta_n^2) > 1$ is obtained under (A6a), then (A6b) should be used instead. That is, \tilde{V}_A^{ad} and \tilde{V}_A^a should be calculated again using eqn (17) with some realistic values for $E(|\delta_n|)$ and $r_{2,1}$.

(viii) *Upper bounds for the number of large QTLs*

The contribution \tilde{V}_A^a of additive effects of single loci to the additive variance is what remains if the contributions of all other factors are subtracted from the additive variance. On the other hand, we have $\tilde{V}_A^a = \sigma_A^2 QE(h_n)$. Unfortunately, this equation cannot

be used directly to estimate the number of QTLs as σ_A^2 is unknown. But for many traits the number of large QTLs which are the QTLs whose absolute additive effect $|a_n|$ exceeds a given threshold value s and whose MAF is larger than (say) $k=0.01$ is known from QTL mapping studies and could be used to estimate the expected number of large QTLs

$$\tilde{n} = E(\#\{n: |a_n| > s, p_n \in I\}),$$

where $I = (0.01, 0.99)$. Since $\tilde{V}_A^a = \sigma_A^2 QE(h_n)$ we obtain

$$\tilde{n} = 2p_k QF\left(-v\sqrt{Q}\right), \tag{26}$$

where F is the distribution function of a'_n/σ_A , and $v = s\sqrt{E(h_n)/\tilde{V}_A^a}$. If a particular distribution function F (e.g. the normal distribution or the Laplace distribution) is assumed for the standardized additive effects and an estimate of \tilde{n} is given, then eqn (26) can be solved for the number Q of QTLs. Unfortunately, estimates for \tilde{n} obtained from real data are often rather poor and F is also not known with certainty. But eqn (26) can also be used to derive upper bounds for \tilde{n} . For a given distribution function F of the standardized additive effects an upper bound for the expected number of large QTLs is (see eqn (26))

$$\tilde{n} \leq \max_{Q \geq Q_{\min}} 2p_k QF(-v\sqrt{Q}),$$

where Q_{\min} is a lower bound for the number of QTLs. From eqn (26) it follows with the Tschebyscheff inequality that $\tilde{n} \leq p_k/v^2$. The proof is similar to the proof of eqn (27). This bound for the expected number of large QTLs depends neither on the total number of QTLs and the distribution of the QTL effects nor on the correlation between $|a_n|$ and δ_n . The Vysochanskij–Petunin inequality refines the Tschebyscheff inequality for unimodal distributions. That is, if $Q > 8/(3v^2)$ and an unimodal distribution is assumed for the additive effects, then the Vysochanskij–Petunin inequality gives the better bound

$$\tilde{n} \leq \frac{4p_k}{9} \frac{\tilde{V}_A^a}{s^2 E(h_n)}. \tag{27}$$

The bound depends on the contribution \tilde{V}_A^a of single additive effects to the additive variance. As shown in section (v), this can be estimated under (A6a). The bound holds in principle also under (A6b). However, then \tilde{V}_A^a depends on the unknown joint distribution of the additive effects and dominance coefficients and could be even larger than the additive variance \tilde{V}_A since \tilde{V}_A^{ad} is negative. In the scenarios considered so far, the bound may increase by a factor of 3 under (A6b), but then the QTLs with large effect are close to dominance or recessivity, so that they contribute little to the additive variance.

(ix) *Parameter estimation*

In this section, we show how the presented equations can be used, e.g. to design a simulation experiment for the assessment of genome-wide evaluations of quantitative traits in outbred populations. Estimates for \tilde{V}_A , \tilde{V}_D and \tilde{I} must be available. The goal is to estimate μ_Δ , σ_Δ^2 , $\sigma_{A'}^2$, Q and the contributions of the different sources to the additive and dominance variance by the formulae such that the simulated populations have on average the desired additive variance, dominance variance and inbreeding depression. The procedure is as follows:

- Estimate the parameters p_k , $E(h_n)$, γ and $E(h_n \times \sqrt{1-2h_n})$ that depend on the distribution of allele frequencies and the parameters $D_{LD}^{0,0}$ and $D_{LD}^{1,1}$ that describe the effect of linkage disequilibrium. This is demonstrated in section (iii). Ensure that the simulation reproduces these values on average.
- Then \tilde{V}_D^{LD} , \tilde{V}_A^{LD} , \tilde{V}_D^d and \tilde{V}_A^d can be calculated as described in section (v).
- A joint distribution for the scaled additive effects and the scaled dominance coefficients must be postulated. For this distribution, the parameters c_v , r_1 , r_2 , $r_{2,1}$, λ and K are to be calculated. Alternatively, the values from Table 2 can be used (see section (iv)) for the proposed joint distributions.
- The expected number of QTLs can then be calculated as

$$Q = \frac{\tilde{I}^2}{\tilde{V}_D - \tilde{V}_D^{LD}} \frac{1-\gamma}{2E(h_n)} \frac{(1+r_2)}{\lambda^2(1+r_1)^2} (1+c_v^2). \quad (28)$$

- We have

$$\begin{aligned} \mu_\Delta &= \pm \sqrt{\alpha} && \text{under (A6a),} \\ \mu_\Delta &= \pm \sqrt{\alpha + (\alpha\beta)^2} - \alpha\beta && \text{under (A6b),} \end{aligned} \quad (29)$$

and $\sigma_\Delta = \mu_\Delta c_v$, where

$$\begin{aligned} \alpha &= \frac{1}{1+c_v^2} \frac{1}{1+r_2} \frac{\tilde{V}_D^d}{\tilde{V}_A - \tilde{V}_A^d - \tilde{V}_A^{LD}} \frac{2}{1-\gamma}, \\ \beta &= \frac{K(1+r_{2,1})E(h_n\sqrt{1-2h_n})}{E(h_n)}. \end{aligned}$$

- Under (A6a) we have $\tilde{V}_A^{ad} = 0$ and under (A6b), \tilde{V}_A^{ad} can be obtained from eqn (17) since $E(|\delta_n|) = \mu_\Delta K$.
- $\tilde{V}_A^a = \tilde{V}_A - \tilde{V}_A^d - \tilde{V}_A^{ad} - \tilde{V}_A^{LD}$ and $\sigma_{A'}^2 = \tilde{V}_A^a / QE(h_n)$.

3. Examples

In this section we use published data for the application of the derived equations. Van Tassell *et al.* (2000) estimated fractions of variance accounted for by additive variance (h^2) and dominance variance (d^2) as well as inbreeding depression (\mathcal{I}) for milk yield, PL

Table 3. *Estimates for milk yield and PL in Holstein cattle from literature*

	Milk yield	PL
h^2	0.342	0.118
d^2	0.053	0.057
d^2/h^2	0.15	0.48
\mathcal{I}	3170 kg	24.2 months
V_p	1216 ²	13.1 ²
V_A	711 ²	4.5 ²
V_D	280 ²	3.1 ²

and other traits from Holstein data. The estimates are given in Table 3. They are based on milk records and PL records of more than 730 000 cows. Data were selected to maximize the number of full sibs in the analysis. Variance components were estimated by Method R. Properties of this method were examined by Duangjinda *et al.* (2001). Unfortunately, the phenotypic variances were not reported. Therefore, literature estimates were used which are $V_p = 1216^2$ for preadjusted 305-day milk yield during the first lactation (Miglior *et al.*, 1995) and $V_p = 13.1^2$ for total month in milk by 84 months of age (Van Raden *et al.*, 1993).

Clearly, the conclusions drawn in this section may hold only if these parameters have been estimated with sufficient precision and if the model assumptions hold. The precision of the estimates may be not very high for two reasons. At first, we cannot rule out the possibility of a bias due to confounding factors, and secondly, the variance components were estimated and not their expectations in the sense of this paper. But the variance components are close to their expectations if the number of QTLs is sufficiently large and therefore, the estimates should give realistic values to deal with.

We obtained the bounds shown in Table 4, where $\tilde{Q} = p_k Q$ with $p_k = 0.71$ is the expected number of QTLs with MAF > 0.01. We expect at least 187 QTLs for milk and 84 QTLs for PL (obtained from eqn (23)).

We chose the threshold values for a QTL to be large as $s = 200$ kg for milk yield and $s = 2$ months for PL. More than 18.9 large QTLs for milk can be excluded if a unimodal distribution of the additive effects is assumed. Furthermore, if \tilde{n} is assumed to be larger than 9.6, then neither a normal distribution nor a Laplace distribution of the QTL effects is possible. If a normal distribution is assumed for the additive effects for PL then the expected number of large QTLs is 0.7 at most. These bounds have been derived under (A6a).

Figure 1 shows scatter plots of the joint distributions of additive effects and dominance coefficients that were fitted to PL under assumption (6b).

Table 4 Estimated parameters for Milk yield and PL

	Milk yield scenario				PL scenario			
	1	2	3	4	1	2	3	4
\tilde{Q}	1216	736	998	1359	548	332	450	612
μ_{Δ}	0.293	0.190	0.186	0.172	0.480	0.301	0.299	0.274
σ_{Δ}	0.439	0.191	0.278	0.257	0.720	0.302	0.449	0.411
$E(\delta_n^2)$	0.279	0.073	0.112	0.096	0.749	0.182	0.291	0.244
$\sigma_{A'}$	47.7	63.9	53.7	46.5	0.494	0.684	0.565	0.494
\tilde{n}	3.3	8.8	5.5	5.3	1.8	5.3	3.1	3.1
$\tilde{V}_A^q/\tilde{V}_G$	1.252	1.354	1.300	1.327	1.171	1.361	1.26	1.309
$\tilde{V}_A^{ad}/\tilde{V}_G$	-0.485	-0.587	-0.533	-0.560	-0.743	-0.933	-0.832	-0.881
$\tilde{V}_A^{LD}/\tilde{V}_G$	0.098	0.098	0.098	0.098	0.245	0.245	0.245	0.245
$\tilde{V}_A^{LD}/\tilde{V}_G$	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
$\tilde{V}_B^{LD}/\tilde{V}_G$	0.009	0.009	0.009	0.009	0.010	0.010	0.010	0.010
$\tilde{V}_B^{LD}/\tilde{V}_G$	0.126	0.126	0.126	0.126	0.316	0.316	0.316	0.316

The shape of the distributions were given in advance for each scenario. Only the scales were adjusted. It can be seen in the most realistic Scenarios (2)–(4) that QTLs with large additive effect are close to recessivity or dominance, so that they contribute little to the additive variance. Druet *et al.* (2001) support the hypothesis that large additive QTLs for PL are absent since the fertility, which determines length of PL, has a heritability of only 1%. The QTL on the right-hand side of the vertical lines in Fig. 1 are those whose absolute additive effect exceeds $s=2$ months.

Table 5 shows the parameter estimates for all scenarios for milk yield and PL. For scenarios (2)–(4), the estimated number of QTLs was between 736 and 1359 and the expected number of large QTLs was about 5.3–8.8. For PL, the estimated number of QTLs was between 332 and 612 and the expected number of large QTLs was about 3.1–5.3. The table also shows for the different sources their estimated contribution to the expected genotypic variance $\tilde{V}_G = \tilde{V}_A + \tilde{V}_D$. It can be seen that the genotypic variance is affected only little by linkage disequilibrium. Moreover, we have in all scenarios $\tilde{V}_A^d + \tilde{V}_A^{ad} + \tilde{V}_A^{LD} < 0$, so dominance effects diminish rather than increase the additive variance.

Equation (28) suggests that the number of QTLs depends heavily on the coefficient of variation of the dominance coefficient c_v . In scenarios (1), (3) and (4), this coefficient was equal to 1.5 as it was suggested by Bennewitz & Meuwissen (2010). However, the true value may be different for some traits. In order to see how the parameter values depend on c_v we calculated them also for other values. Results are shown in Figs 2 and 3 for PL (left) and milk yield (right). The dotted line shows the results obtained under assumption (6a) for independent Laplace distributed additive effects and normally distributed dominance coefficients. The dashed line shows the results obtained under

Table 5 Bounds for the parameters under different assumptions

	Milk yield	PL	Assumption
$\tilde{n} \leq$	18.9	5.4	(6a) Unimodal distribution of additive effects
$\tilde{n} \leq$	9.6	2.0	(6a) Laplace distribution of additive effect
$\tilde{n} \leq$	6.7	0.7	(6a) Normal distribution of additive effects
$\tilde{Q} \geq$	187	84	

assumption (6a) for dependent additive effects and dominance coefficients. Here, $\delta_n \sim \mathcal{N}(0.2, (0.2c_v)^2)$ was used and \tilde{a}_n was defined as in scenario (3). This choice results for large c_v in a more heavy tailed distribution of a'_n . The solid line shows the results obtained under assumption (6b) for the same joint distribution of additive effects and dominance coefficients.

It can be seen that under (A6b) with dependent effects (solid lines) the smallest mean and standard deviation of the dominance coefficients and therefore the smallest fraction of overdominant alleles was obtained. It also provided the highest number of large QTLs. The number of QTLs depends heavily on the coefficient of variation, i.e. if c_v is small then μ_{Δ} and $\sigma_{A'}$ are small as well, so inbreeding depression must be caused by many genes. The number of large QTLs changes only little although the total number of QTLs strongly increases for increasing c_v which is due to the more heavy tailed distribution of a'_n for large c_v .

4. Discussion

Novel methods to estimate the number of QTLs and to quantify the importance of linkage disequilibrium

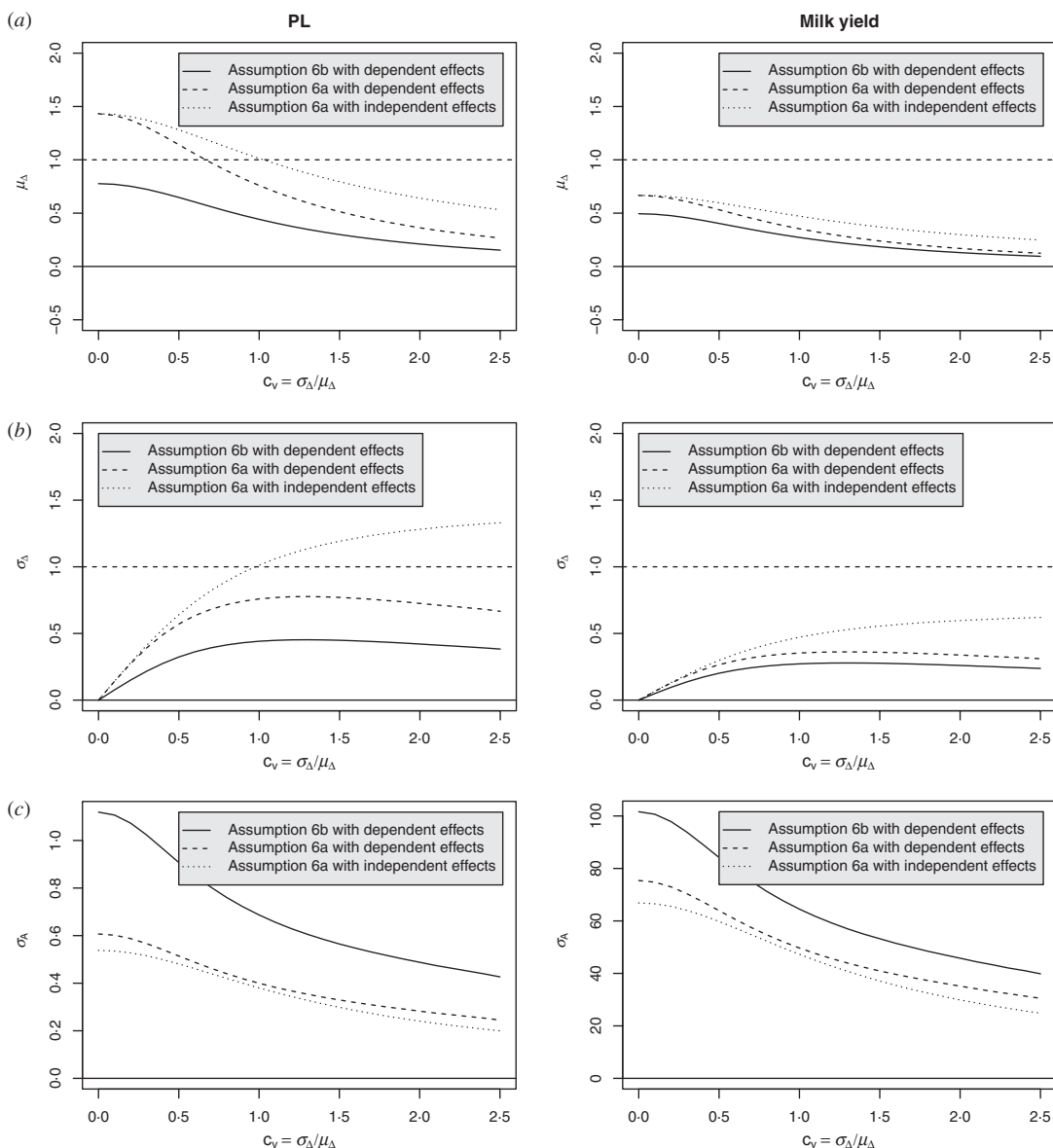


Fig. 2. Different parameters for PL and milk yield as a function of the coefficient of variation of the dominance coefficient: (a) expectation of the dominance coefficient, (b) standard deviation of the dominance coefficient and (c) standard deviation of the additive effects.

on the additive and dominance variance have been derived. It is demonstrated how parameters of the distribution of the QTLs effects can be estimated if estimates for the variance components and inbreeding depression are available. In the following, several aspects of the approach are critically discussed.

(i) *Bounds for the number of QTLs*

Lower bounds for the expected number of QTLs have been derived. The bound given by eqn (18) equals the ratio of squared inbreeding depression to dominance variance. It could be derived under weak assumptions, but may be poor if the trait shows little directional dominance. The bound could be improved if dominance effects and heterozygosity are independent

(see eqn (21)). This equation gives the largest bound, but the calculation requires knowledge on the distribution of the additive effects and on the dependency between additive effects and dominance coefficients. This is not needed to calculate the bound given in eqn (23). Since this bound holds even if $c_v = 0$ although, we are likely to have $c_v > 0$, the bound is rather conservative. Larger bounds could be derived under more stringent assumptions. There could be less QTLs in a population, but the average number of QTLs (averaged over many populations with the same characteristics) would not fall below the bound.

This novel method to estimate the number of QTLs affecting a trait gives a higher number than estimates obtained from the QTL mapping experiments mentioned in the Introduction section. The reason

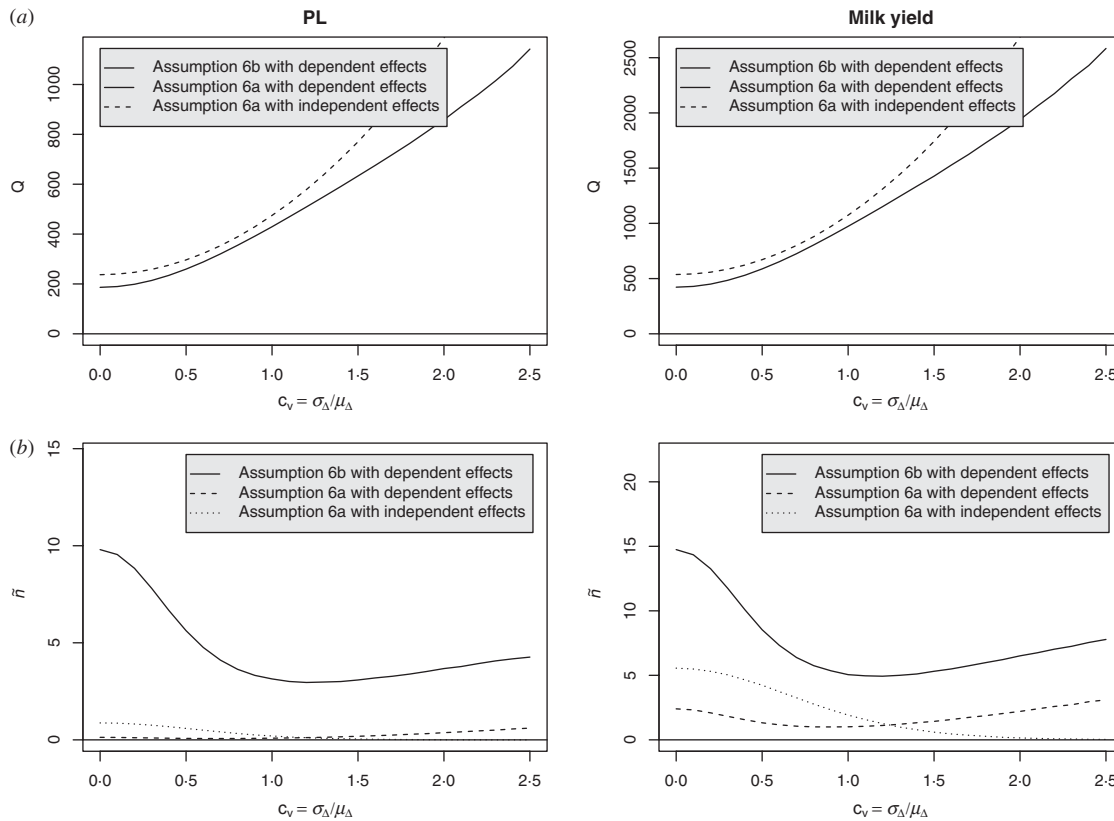


Fig. 3. Different parameters for PL and milk yield as a function of the coefficient of variation of the dominance coefficient: (a) Number of QTLs, (b) number of large QTLs.

is that inbreeding depression must be due to a sufficient number of QTLs, because otherwise it would cause more than the true dominance variance of the trait. Therefore, our estimates are larger than the estimates of other authors who neglected dominance and could not well account for QTLs with small effects. This supports the hypothesis of a large number of genes affecting a quantitative trait. An upper bound for the expected number of large QTLs is also derived.

(ii) *Shape of the distribution*

A critical issue is the shape of the distribution of the additive effects. It is demonstrated that the distribution of QTL effects must be heavy tailed if many QTLs with large effect are expected to segregate in a population, provided that the trait is affected by directional dominance. This means that a distribution function for standardized additive effects is admissible, only if eqn (26) can be solved for Q under side condition (21).

(iii) *Quality of the estimates*

Parameter estimates and bounds for the number of QTLs that are obtained as described above may be imprecise for several reasons. The estimation of

dominance variance requires large amounts of data with a high proportion of full sibs. For equivalent accuracy, the estimation of dominance variance requires at least 20 times as much data than the estimation of additive variance (Miszta, 1997). In particular, estimates of the ratio of squared inbreeding depression to dominance variance, which determines the bound for the number of QTLs, may be poor for traits where both of them are small. Dominance variance is difficult to estimate as it could be confounded e.g. by maternal effects, environmental covariance of fullsibs and variation in relationship. We cannot rule out the possibility that the dominance variance has been overestimated, but see Duangjinda *et al.* (2001). Moreover, our approach treats the population as a random realization. Realized additive variance, dominance variance and inbreeding depression deviate by chance from their expectations.

(iv) *Assumptions that may be violated in the presence of selection*

Our model could account for effects of selection on the distribution of allele frequencies and on the dependency between a_n and δ_n , but it could not account for some other effects such as a dependency between

the dominance effects and the heterozygosity. Since overdominant alleles likely have intermediate frequencies because of selection, they contribute more to dominance variance, so a smaller proportion of overdominant alleles would be sufficient to explain a large dominance variance in a long-term selected population. Independence of a_n and $D_{n,m}$, which was assumed to calculate \tilde{V}_A^{LD} , may not hold in the presence of selection because then allelic effects and allele frequencies are dependent and allele frequencies affect $D_{n,m}$. This was neglected in the present study. In the examples we used estimates for LD and for parameters of the distribution of allele frequencies that were obtained for neutral alleles although the traits under examination are not neutral. Further research is needed in order to generalize the model or to find out how robust our results are when the assumptions that have been made are violated.

(v) *The role of overdominance*

We showed that the second moment of the dominance coefficient is determined by the ratio of dominance variance due to single dominance effects \tilde{V}_D^d to the additive variance due to single additive effects \tilde{V}_A^a , but does not depend on the number of QTLs. If the second moment is larger than one, then the probability of an allele to be overdominant becomes large (see eqn (25)). This may occur in computer simulations when populations are simulated without selection (according to assumption (A6a)) and the trait shows much dominance variance. But it is unlikely to occur in real populations. A large second moment does not occur if additive effects and dominance coefficients are dependent ($r_2 \gg 0$) and a_n and $(q_n - p_n) d_n$ have opposite signs (A6b). Frankham (2009) argued that in long-term artificially selected populations the proportion of variation due to overdominant alleles is likely much higher than in wild populations at equilibrium. Overdominance of an allele that affects PL could arise as follows: an individual has a long PL if it is good enough for a large number of traits. Consider a pleiotropic gene with alleles a and A . The allele a improves one trait (e.g. mastitis resistance) but reduces another trait (e.g. milking speed). Suppose that one a -allele is sufficient to prevent the individual from being culled because of the first trait. But two a -alleles decrease the second trait so much that the individual would likely be culled because of the second trait. In this hypothetical example, genotype AA would be culled early because of little mastitis resistance and genotype aa would be culled early because of slow milking speed. Only genotype Aa is likely to survive several lactations. Thus, the gene would be overdominant for PL. This is called antagonistic pleiotropy. The concept of antagonistic pleiotropy was popularized by Rose (1982) and critically analysed by

Curtsinger (1994) and Hedrick (1999). Curtsinger argued that when many loci are involved, the conditions for intermediate equilibrium frequencies are difficult to satisfy. But he did not take into account that the number of affected fitness components likely also increases when the number of loci increases that affect total fitness. He also ignored the recurrent introduction of new alleles by mutations. Our approach also could not clarify how important overdominance for PL is, since some scenarios yielded a non-negligible portion of overdominant alleles, whereas others did not. However, the most reliable scenarios yielded little overdominance.

(vi) *Conclusion*

In conclusion, formulae are developed that can be used to obtain estimates for the number and distribution of QTL effects. The examples demonstrated the need to estimate them for each trait separately. Our method needs estimates of dominance variance and inbreeding depression. But if reliable estimates are available, then estimates for the number and distribution of QTL effects can be obtained that are very helpful in the choice of the appropriate method for the prediction of genomic breeding values, even though the dependency between marker effects and QTL effects needs clarification. Additionally they can be used as input parameters for stochastic simulations that mimic the real situation closely.

R.W. was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG). The manuscript has benefited from critical and helpful comments of the anonymous reviewers. The authors also thank Professor W. G. Hill for constructive critique and encouraging comments on an earlier version of the manuscript.

References

- Avery, P. J. & Hill, W. G. (1979). Variance in quantitative traits due to linked dominant genes and variance in heterozygosity in small populations. *Genetics* **91**, 817–844.
- Bennewitz, J. & Meuwissen, T. H. E. (2010). The distribution of QTL additive and dominance effects in porcine F2 crosses. *Journal of Animal Breeding and Genetics* **127**, 171–179.
- Bourguet, D. (1999). The evolution of dominance. *Heredity* **83**, 1–4.
- Caballero, A. & Keightley, P. D. (1994). A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* **138**, 883–900.
- Chamberlain, A. J., McPartlan, H. C. & Goddard, M. E. (2007). The number of loci that affect milk production traits in dairy cattle. *Genetics* **177**, 1117–1123.
- Charlesworth, D. & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics* **10**, 783–796.
- Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. New York, NY: Harper and Row.

- Curtis, J. W., Service, P. M. & Prout, T. (1994). Antagonistic pleiotropy, reversal of dominance, and genetic polymorphism. *American Naturalist* **144**, 210–228.
- Daetwyler, H. D., Pong-Wong, R., Villanueva, B. & Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* **185**, 1021–1031.
- Dekkers, J. C. M. & Hospital, F. (2002). Utilization of molecular genetics in genetic improvement of plants and animals. *Nature Reviews Genetics* **3**, 22–32.
- Druet, T., Sölkner, J., Groen, A. F. & Gengler, N. (2001). Additive and dominance genetic variance of fertility by method R and preconditioned conjugate gradient. *Journal of Dairy Science* **84** (Online).
- Duangjinda, M., Misztal, I., Bertrand, J. K. & Tsuruta, S. (2001). The empirical bias of estimates by restricted maximum likelihood, Bayesian method, and method R under selection for additive, maternal, and dominance models. *Journal of Animal Science* **79**, 2991–2996.
- Eyre-Walker, A. & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics* **8**, 610–618.
- Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. London: Longman.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399–433.
- Fisher, R. A. (1928). The possible modification of the response of the wild type to recurrent mutations. *American Naturalist* **62**, 115–126.
- Frankham, R. (2009). Genetic architecture of reproductive fitness and its consequences. In *Adaption and Fitness in Animal Populations, Evolutionary and Breeding Perspectives on Genetic Resource Management* (ed. J van der Werf, H.-U. Graser, R. Frankham & C. Gondro), pp. 15–39. Springer Science + Business Media B. V.
- García-Dorado, A., López-Fanjul, C. & Caballero, A. (1999). Properties of spontaneous mutations affecting quantitative traits. *Genetical Research Cambridge* **74**, 341–350.
- Gianola, D., des los Campos, G., Hill, W. G., Manfredi, E. & Fernando, R. (2009). Additive genetic variability and the bayesian alphabet. *Genetics* **183**, 347–363.
- Goddard, M. E. (2001). The validity of genetic models underlying quantitative traits. *Livestock Production Science* **72**, 117–127.
- Goddard, M. E. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257. doi:10.1007/s10709-008-9308-0.
- Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J. J., Kvasz, A., Mni, M., Simon, P., Frère, J.-M., Coppieters, W. & Georges, M. (2004). Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences of the USA* **101**, 2398–2403.
- Hayes, B. & Goddard, M. E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution* **33**, 209–229.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* **92**, 433–443.
- Hedrick, P. W. (1999). Antagonistic pleiotropy and genetic polymorphism: a perspective. *Heredity* **82**, 126–133.
- Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B* **365**, 73–85.
- Hill, W. G., Goddard, M. E. & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* **4**, e1000008.
- Ishikawa, A. (2009). Mapping an overdominant quantitative trait locus for heterosis of body weight in mice. *Journal of Heredity* **100** (4), 501–504.
- Kacser, H. & Burns, J. A. (1981). The molecular basis of dominance. *Genetics* **97**, 639–666.
- Keightley, P. D. (1996). A metabolic basis for dominance and recessivity. *Genetics* **143**, 621–625.
- Luan, T., Woolliams, J. A., Lien, S., Kent, M., Svendsen, M. & Meuwissen, T. H. E. (2009). The accuracy of genomic selection in norwegian red cattle assessed by cross-validation. *Genetics* **183**, 1119–1126.
- Luo, L. J., Li, Z.-K., Mei, H. W., Shu, Q. Y., Tabien, R., Zhong, D. B., Ying, C. S., Stansel, J. W., Khush, G. S. & Paterson, A. H. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. grain yield components. *Genetics* **158**, 1755–1771.
- Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer.
- Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**, 303–339.
- Mackay, T. F. C. & Anholt, R. R. (2006). Of flies and man: *Drosophila* as a model for human complex traits. *Annual Review of Genomics and Human Genetics* **7**, 339–367.
- Meuwissen, T. & Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole genome resequencing. *Genetics* **185**, 623–631.
- Miglior, F., Burnside, E. B. & Kennedy, B. W. (1995). Production traits of Holstein cattle: estimation of non-additive genetic variance components and inbreeding depression. *Journal of Dairy Science* **78**, 1174–1180.
- Misztal, I. (1997). Estimation of variance components with large-scale dominance models. *Journal of Dairy Science* **80**, 965–974.
- Otto, S. P. & Jones, C. D. (2000). Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics* **156**, 2093–2107.
- Rocha, J. L., Eisen, E. J., Siewerdt, F., Van Vleck, L. D. & Pomp, D. (2004). A large-sample QTL study in mice: III. Reproduction. *Mammalian Genome* **15**, 878–886.
- Rose, M. R. (1982). Antagonistic pleiotropy, dominance and genetic variation. *Heredity* **48**, 63–78.
- Van Raden, P. M. & Klaaskate, E. J. H. (1993). Genetic evaluation of length of productive life including predicted longevity of live cows. *Journal of Dairy Science* **76**, 2758–2764.
- van Tassell, C. P., Misztal, I. & Varona, L. (2000). Method R estimates of additive genetic, dominance genetic, and permanent environmental fraction of variance for yield and health traits of Holsteins. *Journal of Dairy Science* **83**, 1873–1877.
- Villa-Angulo, R., Matukumalli, L. K., Gill, C. A., Choi, C., Van Tassell, C. P. & Grefenstette, J. J. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genetics* **10**, 19.
- Visscher, P. M. (2008). Sizing up human height variation. *Nature Genetics* **40**, 489–490.
- Wright, S. (1934). Molecular and evolutionary theories of dominance. *American Naturalist* **68**, 24–53.