OXFORD GENETICS

GENETICS, 2021, 219(4), iyab138
DOI: 10.1093/genetics/iyab138
Advance Access Publication Date: 27 August 2021
Genomic Prediction
Highlighted Article

The correlation of substitution effects across populations and generations in the presence of nonadditive functional gene action

Andres Legarra ,[1,*] Carolina A. Garcia-Baccino,[1,2,3] Yvonne C. J. Wientjes ,[4] and Zulma G. Vitezica[1]

[1]INRAE/INP, UMR 1388 GenPhySE, Castanet-Tolosan 31326, France,
[2]Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires C1417DSQ, Argentina,
[3]SAS NUCLEUS, Le Rheu 35650, France, and
[4]Wageningen University & Research, Animal Breeding and Genomics, Wageningen 6700 AH, the Netherlands

*Corresponding author: INRAE, UMR1388 GenPhySE, CS 52627, 31326 Castanet Tolosan, France. Email: andres.legarra@inrae.fr

Abstract

Allele substitution effects at quantitative trait loci (QTL) are part of the basis of quantitative genetics theory and applications such as association analysis and genomic prediction. In the presence of nonadditive functional gene action, substitution effects are not constant across populations. We develop an original approach to model the difference in substitution effects across populations as a first order Taylor series expansion from a "focal" population. This expansion involves the difference in allele frequencies and second-order statistical effects (additive by additive and dominance). The change in allele frequencies is a function of relationships (or genetic distances) across populations. As a result, it is possible to estimate the correlation of substitution effects across two populations using three elements: magnitudes of additive, dominance, and additive by additive variances; relationships (Nei's minimum distances or Fst indexes); and assumed heterozygosities. Similarly, the theory applies as well to distinct generations in a population, in which case the distance across generations is a function of increase of inbreeding. Simulation results confirmed our derivations. Slight biases were observed, depending on the nonadditive mechanism and the reference allele. Our derivations are useful to understand and forecast the possibility of prediction across populations and the similarity of GWAS effects.

Keywords: QTL; substitution effects; epistasis; dominance; genetic distance

Introduction

One of the aims of quantitative genetics is to provide methods for prediction, for instance genomic prediction (prediction of livestock breeding values or of crop performance) or polygenic risk score (risk of a disease in humans). These predictions would ideally work across a range of populations (different breeds and future generations). Ideally, the prediction goes through a process of identifying causal genes, estimating their effects in some population, and transposing these effects to newly genotyped individuals (Lande and Thompson 1990; Meuwissen *et al.* 2001). These "gene effects" are substitution effects—the regression of the own phenotype (for polygenic risk scores) or expected progeny phenotypes (for estimated breeding values) on gene content at the locus. Being able to use substitution effects at causal genes across populations and generations is a goal of genomic prediction, QTL detection, and also of causal mutation finding (Grisart *et al.* 2002).

There are several obstacles for these aims. Finding and validating causal genes and understanding their functional mechanism is extremely difficult (Grobet *et al.* 1997; Bonifati *et al.* 2003; Rupp *et al.* 2015). In practice, predictions are done using SNP markers using statistical genetics techniques. In livestock, use of markers results in very good predictions within populations, but mediocre (at best) predictions across populations, even with very sophisticated techniques (Hayes *et al.* 2009; Karoui *et al.* 2012; Porto-Neto *et al.* 2015; MacLeod *et al.* 2016). Indeed, livestock and human genetics empirical results show decreasing predictive ability with increasing genetic distance across distinct populations or generations (Liu *et al.* 2016; Martin *et al.* 2019). In humans, there is a strong correlation between GWAS effect estimates across human populations, with typical values around 0.82 (Marigorta and Navarro 2013; Shi *et al.* 2021). The lack of perfect linkage disequilibrium (LD) across markers and genes has been claimed to be a reason for this decrease in accuracy. Adding extra information (more dense maps and biological prior information) should result in a better choice of markers close to causal genes, and therefore in a boost in predictive abilities across populations (Roos *et al.* 2009; MacLeod *et al.* 2016). However, in practice, the increase in predictive ability across populations is small at best (MacLeod *et al.* 2016; Moghaddar *et al.* 2019). This is against results from simulations (Roos *et al.* 2009; MacLeod *et al.* 2016)—but a problem in simulations is that typically gene effects are assumed

Received: October 30, 2020. **Accepted:** August 19, 2021

to be biologically additive and therefore constant across populations.

We argue that, although imperfect LD is a likely cause for not being able to predict across populations, it is not the only one. In fact, substitution (statistical) additive gene action is not homogeneous across populations, even for exactly the same causal mutation. Examples in livestock genetics include myostatin gene (Aiello *et al.* 2018) or DGAT1 (Gautier *et al.* 2007). For instance, in the latter, the "K" allele had rather different substitution effects across breeds: −611, −142, and −351 kg of milk for the respective breeds Montbéliarde, Normande, and Holstein, for a trait with a genetic standard deviation of ∼600 kg. Although part of these differences may be due to genotype-by-environment interactions, it is also plausible that this is due to epistasis or dominance; for instance this is the case in DGAT1 (Streit *et al.* 2011). For instance, in the 5-loci epistatic network in Carlborg *et al.* (2006) some substitution effects at genes switch signs depending on genetic background. In Drosophila, estimated substitution effects of P-element insertions switched signs depending on the genetic background (Magwire *et al.* 2010; Mackay 2015).

There is indeed widespread evidence of biological epistasis (Mackay 2014). Whereas biological epistasis does not impede (on the contrary) large additive variation (Hill *et al.* 2008; Mackay 2014; Mäki-Tanila and Hill 2014), it does imply that substitution effects do vary across genetic backgrounds. Thus, in the presence of functional dominance and epistasis, there is no stability of substitution effects across different genetic backgrounds. Even if in all of these populations, additive variation accounts for most genetic variation, and additive substitution effects are sizable (Hill *et al.* 2008), substitution effects may differ across populations.

Recent simulations (Dai *et al.* 2020; Duenk *et al.* 2020) showed that the difference in substitution effects across populations may be quite large under nonadditive biological gene action and increases with divergence of populations. It is relatively easy to derive algebraic expressions for substitution effects α assuming a specific hypothesis of biological gene action. For instance, assuming biological additive and dominance effects only (Falconer and Mackay 1996) results in $\alpha = a + (q - p)d$, whereas assuming additive, dominance and additive by additive biological gene effects results, assuming linkage equilibrium (LE), in $\alpha_1 = a_1 + (q_1 - p_1)d_1 + (p_2 - q_2)[aa]_{12}$ (Fuerst *et al.* 1997). Both of the approaches (simulation or analytical) are limited because the hypotheses of specific biological gene actions (*e.g.*, 2-loci interaction but not 3-loci interaction) are too restrictive. Classically, these hypotheses are bypassed in quantitative genetics by working on statistical effects (Falconer and Mackay 1996; Lynch and Walsh 1998; Mäki-Tanila and Hill 2014).

The key parameter to describe the resemblance of statistical effects across populations is the correlation of substitution effects across populations. Under certain assumptions (independence of allele frequencies and substitution effects, appropriate coding; Wientjes *et al.* 2017), this correlation can be estimated from SNP markers and data of two populations (Karoui *et al.* 2012; Wientjes *et al.* 2017), and can easily be accommodated into genomic prediction models (Karoui *et al.* 2012; Xiang *et al.* 2017). In human studies, the correlation is estimated through the meta-analysis of GWAS statistics (Marigorta and Navarro 2013; Shi *et al.* 2021). However, in addition to empirical results, some theories to describe the resemblance of substitution effects across different populations would be helpful to (1) better understand and quantify the change in *true* substitution effects (*i.e.*, if true genes instead of markers were being used), (2) give upper bounds for genomic prediction across populations/generations, and (3) allow *a priori* planning of genomic predictions *i.e.*, to include or not different subpopulations.

This study aims to develop a theory to understand and predict, under a neutral scenario, the extent of change of substitution effects across space (breeds and lines) and across time (generations), without invoking or assuming specific modes of biological gene action. As a result, we obtain explicit estimators that are functions of additive, dominance, and additive by additive variances, genetic distances across populations, and distributions of allele frequencies. We check and illustrate our theory using published results and simulations considering dominance and epistasis (additive by additive and complementary) from 5-loci interactions. The main factors affecting the correlation of substitution effects across populations are their genetic distance and the extent of additive-by-additive variation, which is rarely large.

Theory

Analytical results

General theory

Here, we model difference of substitution effects across two populations as Taylor expansions around one of them, the "focal" population. Using Kojima's method (Kojima 1959, 1961), we put additive substitution effects as a function of differences in allele frequencies across populations, "focal" additive substitution effects, and second-order (dominance, additive by additive epistasis) statistical effects. From here, we show that the correlation of substitution effects across populations is (approximately) a function of their differentiation (or genetic distance), the additive, dominant and additive by additive genetic variances, the average heterozygosities, and average squared heterozygosities. All these parameters can be estimated in real populations. In the following, we try to stick to Mäki-Tanila and Hill (2014) notation. Many details are given as Supplementary File S1. Main notation is presented in Table 1.

Note that our procedure is general—it does not invoke any particular mechanism for epistasis or dominance, nor knowledge of individual QTL effects and locations. We assume that the population mean is a (possibly complex) function of QTL allele frequencies $\boldsymbol{p}$ and QTL functional or biological effects. The latter, albeit unknown, are assumed to be constant across populations—we, therefore, do not consider genotype by environment interactions.

Consider the correlation $r\left(\alpha_i^b, \alpha_i^{b'}\right)$ of substitution effects α^b and $\alpha^{b'}$ across respective populations (breeds, heterotic groups, lines, or generations) b and b'. For simplicity, the allele to which α refers is random—it can be either the wild or the mutant allele. In this manner, the average value of α is 0, even in presence of deleterious mutations. It is known that, in presence of dominant and epistatic biological interactions, the value of α depends on the allele frequencies and biological effects; for instance, $\alpha_1 = a_1 + (q_1 - p_1)d_1 + (p_2 - q_2)[aa]_{12}$ for 2-loci epistasis (Fuerst *et al.* 1997), for $p_1 = 1 - q_1$ and $p_2 = 1 - q_2$ frequencies at loci 1 and 2, respectively. Inspired by this, we want to write the correlation $r\left(\alpha_i^b, \alpha_i^{b'}\right)$ for a locus i as a function of vectors of respective allele frequencies, $\boldsymbol{p}^b$ and $\boldsymbol{p}^{b'}$, but in a general manner, without defining a particular functional or biological gene action. We use a first order Taylor series expansion to approximate additive substitution effects in a population b', as a function of effects in another "focal" population b and their distance. By doing this, it can be shown that the difference of substitution effects between two

Table 1 Notation

α_i^b, $\alpha_i^{b'}$	Substitution effect of locus i in population b, in population b'	$\sigma^2_{\alpha,b}$	Variance of substitution effects in population b
d_i^{*b}	Dominance deviation at locus i in population b	$\sigma^2_{d,b}$,$\mu_{d,b}$	Variance and mean of dominance deviations in population b
$(\alpha\alpha)_{ij}^b$	Additive by additive effect of loci i,j at population b	$\sigma^2_{(\alpha\alpha,b)}$	Variance of additive-by-additive effects in population b
$(\boldsymbol{\alpha\alpha})_i^b$	Vector of additive-by-additive effects of locus i with all other loci in population b		
p_i^b, $p_i^{b'}$	Allele frequency of locus i in population b, in population b'		
$\boldsymbol{p}^b, \boldsymbol{p}^{b'}$	Vectors of allele frequencies in population b, in population b'		
$\overline{H_b}$	Average heterozygosity at population b, $\overline{H_b} = E\left(2p_i^b\left(1-p_i^b\right)\right)$		
$\overline{H_b^2}$	Average squared heterozygosity at population b, $\overline{H_b^2} = E\left(\left(2p_i^b\left(1-p_i^b\right)\right)^2\right)$		
$\overline{HH_b}$	Average product of heterozygosities at population b, $\overline{HH_b} = E_{i>j}\left(4p_i^b\left(1-p_i^b\right)p_j^b\left(1-p_j^b\right)\right)$		
ϵ_i	Difference in allele frequencies at locus i, $\epsilon_i = p_i^{b'} - p_i^b$	σ^2_ϵ	Variance of the difference in allele frequencies across all loci
$D_{b,b'}$	Nei's minimum genetic distance		
$\sigma^2_A, \sigma^2_D, \sigma^2_{AA}$	Additive, Dominance and Additive by Additive variances (in population b)		

populations is (approximately) a function of the genetic distance of the two populations, and the magnitude of dominance and second order epistatic variances in the focal population b.

To derive additive substitution effects α as function of allelic frequencies $\boldsymbol{p}$, we use Kojima's definition of statistical effects as first, second... derivatives of the mean of the population (μ) as a function of p. We do not invoke any explicit function—we just presume that there is one, in other words, change in allele frequencies of the population implies change in the total average genotypic value. Using Kojima's method, the additive substitution effect at the i-th locus is the first derivative (Kojima 1959, 1961):

$$\alpha_i = \frac{1}{2}\frac{\partial \mu}{\partial p_i}$$

Higher order statistical effects implying locus i (*i.e.*, dominance deviations and epistatic interactions) can be represented by higher order partial derivatives of μ or equivalently as derivatives of α_i. The dominance deviation at the i-th locus (that we denote as d^* to distinguish from the biological or functional effect d, Falconer and Mackay 1996) is:

$$d_i^* = -\frac{1}{4}\frac{\partial^2 \mu}{\partial p_i^2} = -\frac{1}{2}\frac{\partial \alpha_i}{\partial p_i}$$

The negative sign comes because the dominance deviation is usually understood as a feature of *heterozygosity*, in other words, it is of opposite sign than the increase of homozygosity in ∂p_i^2. Last, the epistatic pairwise deviation of locus i with j is

$$(\alpha\alpha)_{ij} = \frac{1}{4}\frac{\partial^2 \mu}{\partial p_i \partial p_j} = \frac{1}{2}\frac{\partial \alpha_i}{\partial p_j}$$

This is positive because it is the effect of increasing both p_i and p_j. Note that interaction of order k implies k-th order derivative with scaling factor $1/2^k$.

Kojima's method shows, therefore, that higher order effects of one locus are derivatives of lower order effects. With these elements, we can make a Taylor order expansion of α_i around frequencies in the "focal" population, $\boldsymbol{p}^b$, so that $\boldsymbol{p}b' = \boldsymbol{p}^b + \boldsymbol{\epsilon}$, such that from values of α in b and changes in allele frequencies $\boldsymbol{\epsilon} = \boldsymbol{p}^{b'} - \boldsymbol{p}^b$ we create a function $\alpha_i^{b'} \approx f\left(\alpha_i^b, \epsilon\right)$. In the Supplementary File S1 (Section 1.1), we show that the Taylor linear approximation of the substitution effects of population b': $\alpha_i^{b'}$, from effects from populations b, α_i^b, d_i^{*b}, $(\boldsymbol{\alpha\alpha})_i^b$ is:

$$\alpha_i^{b'} \approx \alpha_i^b + 2\epsilon_i(-d_i^{*b}) + 2\boldsymbol{\epsilon}'(\boldsymbol{\alpha\alpha})_i^b \quad (1)$$

where we use differences in allele frequencies ϵ, d_i^{*b} is the statistical dominance deviation at the locus i and $(\boldsymbol{\alpha\alpha})_i^b$ is a vector containing epistatic substitution effects of locus i with the rest of loci. By convention we assign $(\alpha\alpha)_{ii}^b = 0$.

From Equation (1), the covariance across two populations b and b' of the two substitution effects α_i^b and $\alpha_i^{b'}$ of the locus i is:

$$Cov(\alpha_i^b, \alpha_i^{b'}) \approx Cov\left(\alpha_i^b, \alpha_i^b + 2\epsilon_i(-d_i^*) + 2\boldsymbol{\epsilon}'(\boldsymbol{\alpha\alpha})_i^b\right) = Var(\alpha_i^b) \quad (2)$$

The equality holds because terms $Cov\left(\alpha_i^b, 2\epsilon_i\left(-d_i^{*b}\right)\right)$ and $Cov\left(\alpha_i^b, 2\boldsymbol{\epsilon}'(\boldsymbol{\alpha\alpha})_i^b\right)$ are null, given that the different statistical effects (α_i^b, d_i^{*b} and $(\boldsymbol{\alpha\alpha})_i^b$) are mutually orthogonal by construction.

Thus $r\left(\alpha_i^b, \alpha_i^{b'}\right) = \sqrt{\frac{Var\left(\alpha_i^b\right)}{Var\left(\alpha_i^b\right)Var\left(\alpha_i^{b'}\right)}}$, and now we need the variance of $\alpha_i^{b'}$ as a function of effects in population b, this is (see the Supplementary File S1, Section 1.2 for details):

$$Var(\alpha_i^{b'}) \approx Var\left(\alpha_i^b + 2\epsilon_i(-d_i^{*b}) + 2\boldsymbol{\epsilon}'(\boldsymbol{\alpha\alpha})_i^b\right) = Var(\alpha_i^b) + 4Var(\epsilon_i)\left(Var(d_i^{*b}) + E^2(d_i^{*b})\right) + 4tr\left(Var(\boldsymbol{\epsilon}')\mathrm{Var}((\boldsymbol{\alpha\alpha})_i^b)\right)$$

This expression is unsymmetric and seems to imply that by construction $Var\left(\alpha_i^{b'}\right) > Var\left(\alpha_i^b\right)$; the reason for this is that for the "focal" population b the variance (or at least its estimators) is better known than for the "approximated" population b'. In statistical terms, population b has $Var\left(\alpha_i^b|\sigma^2_A, \sigma^2_D, \sigma^2_{AA}, \mu_{d,b}, \overline{H}_b, \overline{H_b^2}\right)$ all of them known for population b, and population b' has $Var\left(\alpha_i^{b'}|\sigma^2_A, \sigma^2_D, \sigma^2_{AA}, \mu_{d,b}, \overline{H}_b, \overline{H_b^2}\right)$ where variances and

heterozygosities are *still* from population b (and not from b' itself). So, [3] is $Var(\alpha_i^b|\text{information on } b)$ versus $Var(\alpha_i^{b'}|\text{information on } b)$. The higher variance of α_i^b than of $\alpha_i^{b'}$ is due to this extra incertitude.

An alternative derivation in the Supplementary File S1 (Section 1.4) shows that, if the "focal" population is a third one (f) a set of expressions analogous to Equations (2) and (3) is:

$$Var(\alpha_i^{b'}) \approx Var(\alpha_i^f) + 4Var(\epsilon_i^{(b')} d_i^{*f}) + 4Var(\boldsymbol{\epsilon}^{(b')'}(\boldsymbol{\alpha\alpha})_i^f)$$

$$Var(\alpha_i^b) \approx Var(\alpha_i^f) + 4Var(\epsilon_i^{(b)} d_i^{*f}) + 4Var(\boldsymbol{\epsilon}^{(b)'}(\boldsymbol{\alpha\alpha})_i^f)$$

$$Cov(\alpha_i^{b'}, \alpha_i^b) \approx Var(\alpha_i^f) + Cov(\epsilon_i^{(b')}, \epsilon_i^{(b)})\left(Var(d_i^{*b}) + E^2(d_i^{*b})\right) + Cov(\boldsymbol{\epsilon}^{(b')'}(\boldsymbol{\alpha\alpha})_i^f, \boldsymbol{\epsilon}^{(b)'}(\boldsymbol{\alpha\alpha})_i^f)$$

Where $\epsilon_i^{(b')} = p_i^{b'} - p_i^f$ and $\epsilon_i^{(b)} = p_i^b - p_i^f$. From these expressions, we obtain Equations (2) and (3) as a particular case if the focal population is b.

In the expression (3), $Var(\boldsymbol{\epsilon}')$ and $Var\left((\boldsymbol{\alpha\alpha})_i^b\right)$ are matrices. The first one describes the variability of differences in allele frequencies:

$$Var(\boldsymbol{\epsilon}') = \begin{pmatrix} Var(\epsilon_1) & Cov(\epsilon_1,\epsilon_2) & \cdots & Cov(\epsilon_1,\epsilon_n) \\ Cov(\epsilon_2,\epsilon_1) & Var(\epsilon_2) & & Cov(\epsilon_2,\epsilon_n) \\ \cdots & & \cdots & \\ Cov(\epsilon_n,\epsilon_1) & & & Var(\epsilon_n) \end{pmatrix}$$

A locus that may diverge a lot (for instance because it is highly polymorphic) has high $Var(\epsilon)$; two loci in strong linkage will show nonzero $Cov(\epsilon_1,\epsilon_2)$. The second matrix contains (co)variances of the epistatic effects across all pairs marker i—other markers:

$$Var\left((\boldsymbol{\alpha\alpha})_i^b\right) = \begin{pmatrix} Var\left((\alpha\alpha)_1^b\right) & Cov\left((\alpha\alpha)_1^b,(\alpha\alpha)_2^b\right) & \cdots & Cov\left((\alpha\alpha)_1^b,(\alpha\alpha)_n^b\right) \\ Cov\left((\alpha\alpha)_2^b,(\alpha\alpha)_1^b\right) & Var\left((\alpha\alpha)_2^b\right) & & Cov\left((\alpha\alpha)_2^b,(\alpha\alpha)_n^b\right) \\ \cdots & & \cdots & \\ Cov\left((\alpha\alpha)_n^b,(\alpha\alpha)_1^b\right) & & & Var\left((\alpha\alpha)_n^b\right) \end{pmatrix}$$

For instance, $Var\left((\alpha\alpha)_1^b\right)$ contains the variance of the epistatic effect of marker i with marker 1, and so on. We assume $Var\left((\alpha\alpha)_i^b\right) = \boldsymbol{I}\,\sigma^2_{(\alpha\alpha,b)}$, *i.e.*, epistatic terms $(\boldsymbol{\alpha\alpha})_i^b$ can be either positive or negative, with a variance $\sigma^2_{(\alpha\alpha,b)}$ and are a priori uncorrelated to each other. Assuming null off-diagonals for $Var\left((\alpha\alpha)_i^b\right)$ results that, in the previous product $tr\left(Var(\boldsymbol{\epsilon})Var\left((\boldsymbol{\alpha\alpha})_i^b\right)\right)$, off-diagonal elements of $Var(\boldsymbol{\epsilon}')$ disappear from the result (even if they are not null). Thus, assuming that all diagonal elements of $Var(\boldsymbol{\epsilon}')$ have the same common variance σ^2_ϵ, this results in $4tr\left(Var(\boldsymbol{\epsilon})\text{Var}\left((\boldsymbol{\alpha\alpha})_i^b\right)\right) = 4n\sigma^2_\epsilon\sigma^2_{(\alpha\alpha,b)}$. Note that assuming that all diagonal elements of $Var(\boldsymbol{\epsilon})$ are equal is an approximation—for instance more polymorphic loci vary more.

The expression $Var\left(\epsilon_i d_i^{*b}\right) = Var(\epsilon_i)\left(Var\left(d_i^{*b}\right) + E^2\left(d_i^{*b}\right)\right)$ is detailed in the Supplementary File S1 (Section 1.2), and it shows that both the variability of dominance deviations $Var\left(d_i^{*b}\right) = \sigma^2_{d,b}$ and its mean $E\left(d_i^{*b}\right) = \mu_{d,b}$ which, if not zero, can be understood as the basis of inbreeding depression, enter into the expression. Also, we assume ϵ (change in allele frequencies, but not the allele frequencies themselves) and the different nonadditive effects d_i^{*b} and $(\boldsymbol{\alpha\alpha})_i^b$ to be independent (uncorrelated). This makes sense as loci may be selected for additive effects but not for nonadditive effects.

Our next goal is to relate these results in Equation (3) to quantities that are measured empirically. In particular, we need the variances of the different genetic effects and the variance of changes in allele frequencies. We address these two terms in turn.

We factorize the variance of statistical additive, dominant and additive by additive effects as follows (Mäki-Tanila and Hill 2014; Vitezica *et al.* 2017):

$$Var(\alpha_i^b) = \sigma^2_{\alpha,b} = \frac{\sigma^2_A}{n\overline{H_b}}$$

$$Var(d_i^{*b}) + E^2(d_i^{*b}) = \sigma^2_{d,b} + \mu^2_{d,b} = \frac{\sigma^2_D}{n\overline{H_b^2}}$$

(the latter is shown in the Supplementary File S1, Section 1.5) and

$$Var((\alpha\alpha)^b_{i,j,j>i}) = \sigma^2_{(\alpha\alpha,b)} = \frac{\sigma^2_{AA}}{n(n-1)\overline{H_b}\,\overline{H_b}} \approx 2\frac{\sigma^2_{AA}}{n^2\overline{H_b}\;\overline{H_b}}$$

for n the number of QTL loci and using functions of heterozygosities (more details in the Supplementary File S1, Section 1.6):

$$\overline{H_b} = E\left(2p_i^b(1-p_i^b)\right)$$
$$\overline{H_b^2} = E\left(4p_i^b(1-p_i^b)p_i^b(1-p_i^b)\right)$$
$$\overline{HH_b} = E_{i>j}\left(2p_i^b(1-p_i^b)2p_j^b(1-p_j^b)\right) \approx \frac{1}{2}\overline{H_b}\,\overline{H_b}$$

Here, we have assumed independence of QTL allele frequencies and QTL effects. All variances and effects, as well as heterozygosities H_b, refer to the focal population b with allele frequencies p^b and effects α^b. Note that we assume HWE and LE within both b and b'. Of course, we do not know allele frequencies or even numbers of true causal genes, but the first can be prudently guessed and the latter cancel out in the following.

Consider the scalar σ^2_ϵ, the variance of differences of allele frequency. In fact, $\sigma^2_\epsilon = Var(p^{b'} - p^b) = E\left(\left(p^{b'} - p^b\right)^2\right) - E\left(p^{b'} - p^b\right)^2 = E\left(\left(p^{b'} - p^b\right)^2\right)$ because $E(p^{b'} - p^b) = 0$ when averaged across loci. Therefore, $\sigma^2_\epsilon = E\left(\left(p^{b'} - p^b\right)^2\right)$ which corresponds to Nei's "minimum genetic distance" (Nei 1987; Caballero and Toro 2002), and, accordingly, will be called $D_{b,b'} = \sigma^2_\epsilon$. The value of $D_{b,b'}$ can be estimated from marker data as $\hat{D}_{b,b'} = \frac{1}{n}\sum\left(p_i^{b'} - p_i^b\right)^2$, although it is sensitive to the spectra of polymorphisms (*e.g.*, SNP chips *vs* sequencing). In addition, $D_{b,b'}$ is also the numerator of the F_{ST} fixation index, *e.g.*, Hudson's (1992) $F_{ST} = \frac{E\left(\left(p_i^{b'} - p_i^b\right)^2\right)}{E\left(p_i^{b'}\left(1-p_i^b\right)+p_i^b\left(1-p_i^{b'}\right)\right)}$ with an estimator $\hat{F}_{ST} = \frac{\frac{1}{n}\sum_i\left(\left(p_i^{b'} - p_i^b\right)^2\right)}{\frac{1}{n}\sum_i\left(p_i^{b'}\left(1-p_i^b\right)+p_i^b\left(1-p_i^{b'}\right)\right)}$ (Hudson *et al.* 1992; Bhatia *et al.* 2013). In principle, $D_{b,b'}$ and F_{ST} can be estimated from markers, but also from evolutionary distances and effective population size, or both (Weir and Hill 2002; Bonhomme *et al.* 2010). Now we can rewrite the last term of Equation (3) as

$$4tr\left(Var(\epsilon')Var((\boldsymbol{\alpha\alpha})_i^b)\right) = 4n(\sigma^2_\epsilon\sigma^2_{(\alpha\alpha,b)}) = 4nD_{b,b'}\frac{2\sigma^2_{AA}}{n^2\;\overline{H_b}\;\overline{H_b}}$$

Combining expressions above, Equation (3) becomes

$$Var(\alpha_i^{b'}) = \frac{\sigma^2_A}{n\overline{H}_b} + 4D_{b,b'}\frac{\sigma^2_D}{n\overline{H^2_b}} + 4nD_{b,b'}\,2\frac{\sigma^2_{AA}}{n^2\overline{H}_b\,\overline{H}_b}$$
$$= \frac{1}{n}\left(\frac{\sigma^2_A}{\overline{H}_b} + 4D_{b,b'}\frac{\sigma^2_D}{\overline{H^2_b}} + 8D_{b,b'}\frac{\sigma^2_{AA}}{\overline{H}_b\,\overline{H}_b}\right) \quad (4)$$

From here the correlation of α across populations is (factorizing out $\frac{1}{n}$)

$$r(\alpha_i^b, \alpha_i^{b'}) \approx \frac{Cov(\alpha_i^b, \alpha_i^{b'})}{\sqrt{\left(Var(\alpha_i^b)\right)\left(Var(\alpha_i^{b'})\right)}}$$
$$= \frac{Var(\alpha_i^b)}{\sqrt{\left(Var(\alpha_i^b)\right)\left(Var(\alpha_i^{b'})\right)}} = \frac{\sqrt{Var(\alpha_i^b)}}{\sqrt{Var(\alpha_i^{b'})}} \quad (5)$$
$$= \frac{\sqrt{\frac{\sigma^2_A}{\overline{H}_b}}}{\sqrt{\frac{\sigma^2_A}{\overline{H}_b} + 4D_{b,b'}\frac{\sigma^2_D}{\overline{H^2_b}} + 8D_{b,b'}\frac{\sigma^2_{AA}}{\overline{H}_b\,\overline{H}_b}}}$$

Expression (5) can also be obtained if we start the Taylor expansion from a third focal population that is neither b nor b', as shown in the Supplementary File S1, Section 1.4. Factorizing out $\overline{H_b}$ we arrive to the slightly clearer expression

$$r(\alpha_i^b, \alpha_i^{b'}) \approx \frac{\sqrt{\sigma^2_A}}{\sqrt{\sigma^2_A + D_{b,b'}\left(4\frac{\overline{H}_b}{\overline{H^2_b}}\,\sigma^2_D + 8\frac{1}{\overline{H}_b}\sigma^2_{AA}\right)}} \quad (6)$$

which shows well that the correlation is a function of distance across populations ($D_{b,b'}$) and weights of additive *vs* nonadditive variances. Note that $D_{b,b'}$ is always positive, which implies that $0 < r\left(\alpha_i^b, \alpha_i^{b'}\right) < 1$ as expected.

The quantities involved in Equation (6) are (1) Nei's "minimum genetic distance" $D_{b,b'}$, which describes the similarity of populations b and b', (2) the variance of statistical additive, dominant and additive by additive effects at the individual level σ^2_A, σ^2_D, σ^2_{AA} in population b (3) first and second moments of heterozygosities $\overline{H_b}$, $\overline{H^2_b}$. All these values can be, in principle, estimated from data or "prudently guessed." For the particular case of heterozygosities, SNP markers are a poor choice and it is likely better to use a guess based on sequence or evolutionary processes (coalescence).

From the definition of F_{ST} and assuming $\overline{H_b} \approx \overline{H_{b'}}$, $D_{b,b'} \approx \frac{F_{ST}}{1-F_{ST}}\overline{H_b}$ (detailed in the Supplementary File S1, Section 1.7), leading to

$$r(\alpha_i^b, \alpha_i^{b'}) \approx \frac{\sqrt{\sigma^2_A}}{\sqrt{\sigma^2_A + \frac{F_{ST}}{1-F_{ST}}\left(4\frac{(\overline{H}_b)^2}{\overline{H^2_b}}\,\sigma^2_D + 8\sigma^2_{AA}\right)}} \quad (7)$$

The advantage of the F_{ST} is that it is more robust to the spectra of allele frequencies used to estimate it (Bhatia *et al.* 2013). If we further assume that dominance variance σ^2_D is negligible, we can write a neat expression of the correlation in terms of F_{ST}:

$$r(\alpha_i^b, \alpha_i^{b'}) \approx \frac{\sqrt{\sigma^2_A}}{\sqrt{\sigma^2_A + \frac{8F_{ST}}{1-F_{ST}}\sigma^2_{AA}}} \approx \sqrt{1 - \frac{8F_{ST}}{1-F_{ST}}\frac{\sigma^2_{AA}}{\sigma^2_A}} \quad (8)$$

The second approximation involving $\frac{1}{1+x} \approx 1 - x$. This expression (8) plainly tells that the squared correlation of gene substitution effects across two populations is (to a few degrees of approximation) a linear function of the similarity of populations and the additive-by-additive variance. To our knowledge, these results showing the importance of the different factors on the difference between substitution effects had been shown before only through simulations.

Thus, the algorithm to estimate *a priori* the correlation of α across populations b and b', $r\left(\alpha_i^b, \alpha_i^{b'}\right)$, is:

1) Estimate in population b
   a) additive, dominance, and additive by additive variances
   b) average heterozygosity $\overline{H_b}$ and average squared heterozygosity $\overline{H^2_b}$
2) Estimate Nei's distance $D_{b,b'}$ and/or F_{ST} of the two populations
3) Apply equation (6) or (7).

Consideration of directionality of substitution effects

In Plant and Animal Breeding the origin of the allele is often overlooked, as a mutation may be evolutionary harmful but of interest for farming, and also because many traits selected for do not have a close relationship to fitness in the wild. However, in Evolutionary Genetics it is reasonable to think that most mutations are deleterious, thus with a negative effect of the mutant allele. In Medical Genetics reports of estimated substitution effects are also often done in terms of "susceptible" alleles. In both cases $E(\alpha) \neq 0$ or even $\alpha > 0$ for all loci. In both cases α is "oriented" and has no zero mean. In the theory, above we have considered that α is the effect of a randomly drawn allele, which leads to $E(\alpha) = 0$ and enormously simplifies the algebra. In order to consider oriented α, we propose to transform the estimate of $r\left(\alpha_i^b, \alpha_i^{b'}\right)$ into $r\left(|\alpha_i^b|, |\alpha_i^{b'}|\right)$, the correlation of the absolute values. Assuming that α follows a normal distribution, $|\alpha|$ follows a so-called folded normal distribution. From here, $r\left(|\alpha_i^b|, |\alpha_i^{b'}|\right)$ is obtained from $r\left(\alpha_i^b, \alpha_i^{b'}\right)$ using expressions (not detailed here) in Kan and Robotti (2017), conveniently programmed in the R package MomTrunc (Galarza *et al.* 2020). The specific R function is in the Supplementary File S1, Section 1.8, and we will call it `r2rabs()`.

Covariance across generations within one population

The definition above of two populations is general enough that we can consider any two populations, *e.g.*, two breeds (Angus and Hereford), two strains (New Zealand Holstein and US Holstein) or two generations or time frames (*e.g.*, animals born in 2000 *vs* animals born in 2005, or animals born in 2000 *vs* their descendants). There is evidence that across-generations genetic correlation decreases with (many) generations to values as low as 0.6 (Tsuruta *et al.* 2004; Haile-Mariam and Pryce 2015). Part of this is likely due to genotype by environment interactions. Anyway, part of the across-generation genetic correlation could be due to changes in the allele frequency due to drift, and therefore it can be accounted for by our model, based on the evolution of average coancestry in the breed. We develop this next.

We will talk about "generations" but ideas apply to pedigrees with overlapping generations as well. Consider that what we previously called populations b and b' are animals born at time t_1 and t_2, with $t_2 > t_1$. Equation (7) becomes $r\left(\alpha_i^{t_1}, \alpha_i^{t_2}\right) \approx \frac{\sqrt{\sigma^2_A}}{\sqrt{\sigma^2_A + \frac{F_{ST}}{1-F_{ST}}\left(4\frac{\left(\overline{H}_{t_1}\right)^2}{\overline{H^2_{t_1}}}\,\sigma^2_D + 8\sigma^2_{AA}\right)}}$. In pedigree-based context, the F_{ST} is simply half the increase in average relationship from t_1 to t_2 (Powell *et al.* 2010) which is approximately equal to the increase in inbreeding from $t_1 + 1$ to $t_2 + 1$, which in turn is

approximately $(t_2 - t_1)\Delta F$ for small values of ΔF and steady increase of inbreeding. Thus, $F_{ST} \approx \frac{(t_2-t_1)\Delta F}{2}$ to obtain

$$r(\alpha_i^{t_1}, \alpha_i^{t_2}) \approx \frac{\sqrt{\sigma_A^2}}{\sqrt{\sigma_A^2 + \frac{\frac{(t_2-t_1)\Delta F}{2}}{1-\frac{(t_2-t_1)\Delta F}{2}}\left(4\frac{(\overline{H_{t_1}})^2}{\overline{H_{t_1}^2}}\sigma_D^2 + 8\sigma_{AA}^2\right)}} \quad (9)$$

Assuming, like in Equation (8), small values of dominance variance and of F_{ST}, we obtain

$$r(\alpha_i^{t_1}, \alpha_i^{t_2}) \approx \sqrt{1 - 4(t_2 - t_1)\Delta F \frac{\sigma_{AA}^2}{\sigma_A^2}}$$

Thus, the correlation of substitution effects decreases when there is large drift, reflected in high values of ΔF. This is the case for instance if parents of the next generation are very highly selected without restrictions in future inbreeding, resulting in a considerable change in allele frequencies over the generations.

This agrees with classical theory (Falconer and Mackay 1996): the variance in change of allele frequencies from one generation to the next is simply ΔF. Typical values of ΔF in livestock (rabbits, pigs, cattle or sheep) are at most 0.01 per generation (Welsh *et al.* 2010; García-Ruiz *et al.* 2016; Fernández *et al.* 2017; Rodríguez-Ramilo *et al.* 2019), although this is potentially changing with genomic selection.

Simulations

Description of the simulations

The objective of the simulation is to verify our results considering different kinds of nonadditive biological gene action, and not to infer the values of the correlation of substitution effects across populations as this requires realistic presumptions of the forms and magnitudes of epistatic interactions, something that is largely unknown.

We used *macs* (Chen *et al.* 2009) to simulate a "cattle" scenario of two domestic cattle breeds which diverged from a common ancestral population, in the lines of (Pérez-Enciso 2014; Pérez-Enciso *et al.* 2015), where a large population had a 10-fold reduction bottleneck (domestication) 2000 generations ago and a split into two populations t generations ago, where we made t oscillate between 0 and 100 by steps of 10. Parameters in the simulation were tailored (Pérez-Enciso 2014; Pérez-Enciso *et al.* 2015) to mimic observed levels of diversity in cattle (Gibbs *et al.* 2009) and lead to F_{ST} values between 0 and 0.15. We considered 100 DNA stretches of 300 Kb each. We simulated 200 individuals per population, from which we obtained allele frequencies per population. Details are provided in the Supplementary File S1, Section 1.9. This provided ~400,000 segregating loci with realistic allele frequencies (L-shaped distribution of allele frequencies).

To simulate nonadditive biological gene action, we considered three scenarios. All of them involved 5000 single marker polymorphisms drawn at random from the simulated ones with no particular restriction. First scenario was Complete Dominance (within locus); second one was 1000 networks of 5 loci in Complementary Epistasis, and the third one 1000 networks of 5 loci with Multiplicative Epistasis (additive by additive). We also simulated networks of 2 and of 10 loci with similar results (not shown). These scenarios are similar to Hill *et al.* (2008) but instead of considering 2 loci we consider 5. We use these forms of epistasis, as the first two ones may be interpreted as biologically meaningful gene actions, and the third one is an extreme case for the change in substitution effect across populations. In addition, all three scenarios are analytically tractable. Complete dominance has the genotypic value of the heterozygote equal to one of the homozygotes, *e.g.*, the presence of a single copy of the "good" allele is enough to, say, avoid the disease. Complementary epistasis can be seen as a multi-loci dominance, *e.g.*, disease happens when there is a recessive deleterious genotype at *any* of the loci. Additive by additive epistasis can be understood as a pure multiplication of gene contents and has no good biological interpretation. Networks contribute additively to the total phenotype. Note that, by definition, functional gene action is the same across all populations. From the description of the nonadditive gene action and some algebra in the Supplementary File S1, we are able to derive analytically the values of α, d^* and $(\alpha\alpha)$ in each population and variances σ_A^2, σ_D^2, and σ_{AA}^2. Details are given in the Supplementary File S1, Section 1.10. For each value of t, we made 10 replicates and averaged the results.

From the true substitution effects α^b and $\alpha^{b\prime}$derived above, we obtained the true value $r(\alpha^b, \alpha^{b\prime})$. Note that because of the coalescent simulation, all reference alleles (*i.e.*, with frequency p) are "mutant" ones, and due to assumed dominance and epistatic actions, α are negative by construction (*i.e.*, the mutant allele is deleterious). To conciliate this fact with our derivations, that assume that α refers to a random allele and has null means, we did two things: (1) compute a "random allele" version of $\alpha^b, \alpha^{b\prime}$ in which α was changed sign for "odd" loci, and (2) estimate $r(\alpha^b, \alpha^{b\prime})$ using the transformation of normal distribution into folded normal (Kan and Robotti 2017), *i.e.*, the `r2rabs()` function mentioned above. Thus, we have two estimands, the correlation for the "mutant allele" effect $r_{\text{mutant}}(\alpha^b, \alpha^{b\prime})$ (which corresponds *e.g.*, to typical use in Evolutionary and Medical Genetics) and the correlation for the "random allele" effect $r_{\text{random}}(\alpha^b, \alpha^{b\prime})$ (which corresponds, *e.g.*, to genomic selection and some GWAS). We observed that $r_{\text{mutant}}(\alpha^b, \alpha^{b\prime}) < r_{\text{random}}(\alpha^b, \alpha^{b\prime})$. For instance, in one of the scenarios $r_{\text{mutant}}(\alpha^b, \alpha^{b\prime}) = 0.67$ and $r_{\text{random}}(\alpha^b, \alpha^{b\prime}) = 0.85$.

Now we describe the estimators. We considered either the use of "all" polymorphisms, or of a "SNP" selection in which MAF > 0.01 across both populations simultaneously (roughly 40,000 polymorphisms), as this corresponds to typical SNP panels in genetic improvement. Then we used the Equation (6) either for "all" or for "SNP," using their frequencies to obtain $D_{b,b\prime}$, $\overline{H}_b$, and $\overline{H_b^2}$. In "all," the spectra of allele frequencies of polymorphisms and of QTL is the same, but not in "SNP." As "SNP" tends to be biased, we also considered the Equation (7) using F_{ST} estimated from SNPs and $\overline{H_b} = 0.107$, and $\overline{H_b^2} = 0.018$ from a U-shaped distribution (Hill et al. 2008) detailed later. Note that F_{ST} is more robust to the spectra of allele frequencies used to estimate it. Thus, we obtained three estimators for "random allele" $r_{\text{random}}(\alpha^b, \alpha^{b\prime})$: $\hat{r}_{all}$, $\hat{r}_{SNP}$, and $\hat{r}_{SNPFst}$, and also the three corresponding estimators for "mutant allele" $r_{\text{mutant}}(\alpha^b, \alpha^{b\prime})$ applying function r2rabs() to $\hat{r}_{all}$, $\hat{r}_{SNP}$, and $\hat{r}_{SNPFst}$.

Results of the simulations

Simulated variance components are shown in Table 2. All scenarios yield high additive genetic variances, as expected.

True simulated values of correlation across substitution effects $r_{\text{random}}(\alpha^b, \alpha^{b\prime})$ and their estimates $\hat{r}_{all}$, $\hat{r}_{SNP}$, and $\hat{r}_{SNPFst}$, are presented in Figure 1 (for α defined for random alleles) and Figure 2 (for α defined for mutant alleles). Generally speaking, Figure 1 applies to nonfitness related traits (for "random") and Figure 2 to fitness-related traits (for "mutant"). As predicted by our derivations, there is a clear and almost linear decrease of $r(\alpha^b, \alpha^{b\prime})$ with increasing values of F_{ST}.

In Figure 1 (for α defined for random alleles) it can be observed that our expressions (equations 6 and 7) tend to slightly overestimate $r_{\text{random}}(\alpha^b, \alpha^{b\prime})$ for Complete Dominance, and Complementary Epistasis. The estimates are quite good for Multiplicative Epistasis. Estimators using SNP markers are slightly less accurate than using the complete polymorphism spectra, and using F_{ST} from SNP markers, coupled with a guess of heterozygosities, is similar to using SNP alone to infer both distances and heterozygosities.

Table 2 Additive, dominance and additive by additive variances in the simulated population 1

	Additive	Dominance	Additive by additive
Complete dominance	252	113	0
Complementary epistasis	130	53	12
Multiplicative	219	0	122

Figure 2 (for α defined for mutant alleles) presents values of $r_{\text{mutant}}(\alpha^b, \alpha^{b\prime})$ and estimates $\hat{r}_{all}$, $\hat{r}_{SNP}$, and $\hat{r}_{SNPFst}$. Estimating $r_{\text{mutant}}(\alpha^b, \alpha^{b\prime})$ is more difficult than estimating $r_{\text{random}}(\alpha^b, \alpha^{b\prime})$: depending on the scenario there is more over- or under-estimation that for $r_{\text{random}}(\alpha^b, \alpha^{b\prime})$. In this case the effect of using all polymorphisms or SNP is more marked. However, overall, we find that our estimators explain well the decay in $r(\alpha^b, \alpha^{b\prime})$. The imperfect disagreement (both in Figures 1 and 2) is probably due to several wrong assumptions; we comment some of them in the Discussion section.

Empirical examples

Estimates of correlation across populations using literature values

Literature values used and assumptions

From values of the literature, the previous expressions, Equations (6) and (7) were applied to obtain guesses of the correlation of substitution effects across different populations and generations within population. Estimates of statistical

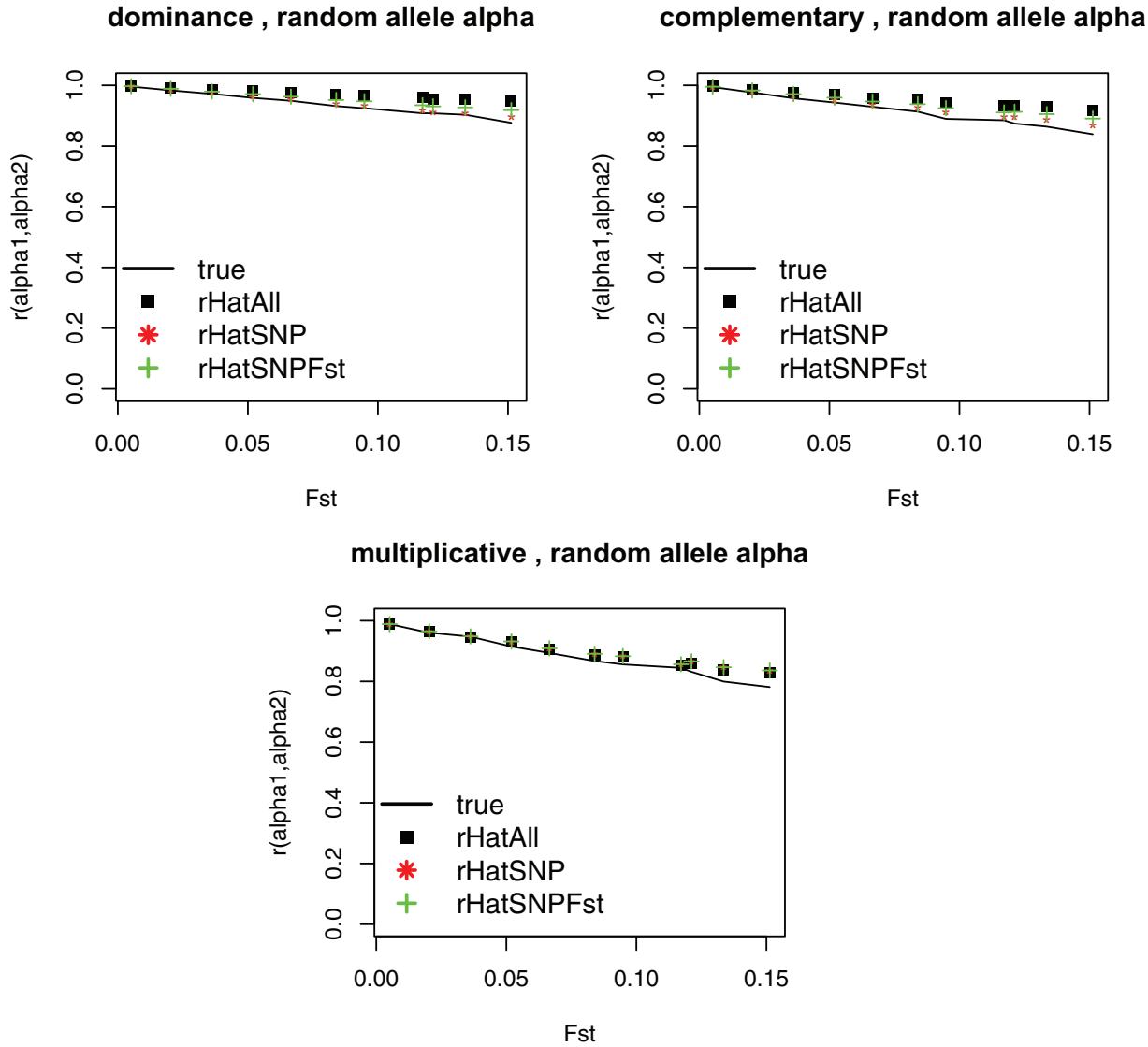

Figure 1 Simulated (straight black line) and estimated (points) correlation across QTL substitution effects of random alleles across two populations as a function of their F_{ST} differentiation coefficient. rHatAll: estimates with all polymorphism. rHatSNP: estimates using SNP-like loci. rHatSNPFst: estimates using SNP-like loci with a correction for heterozygosity. Results of 10 replicates per point with s.e. <0.01.

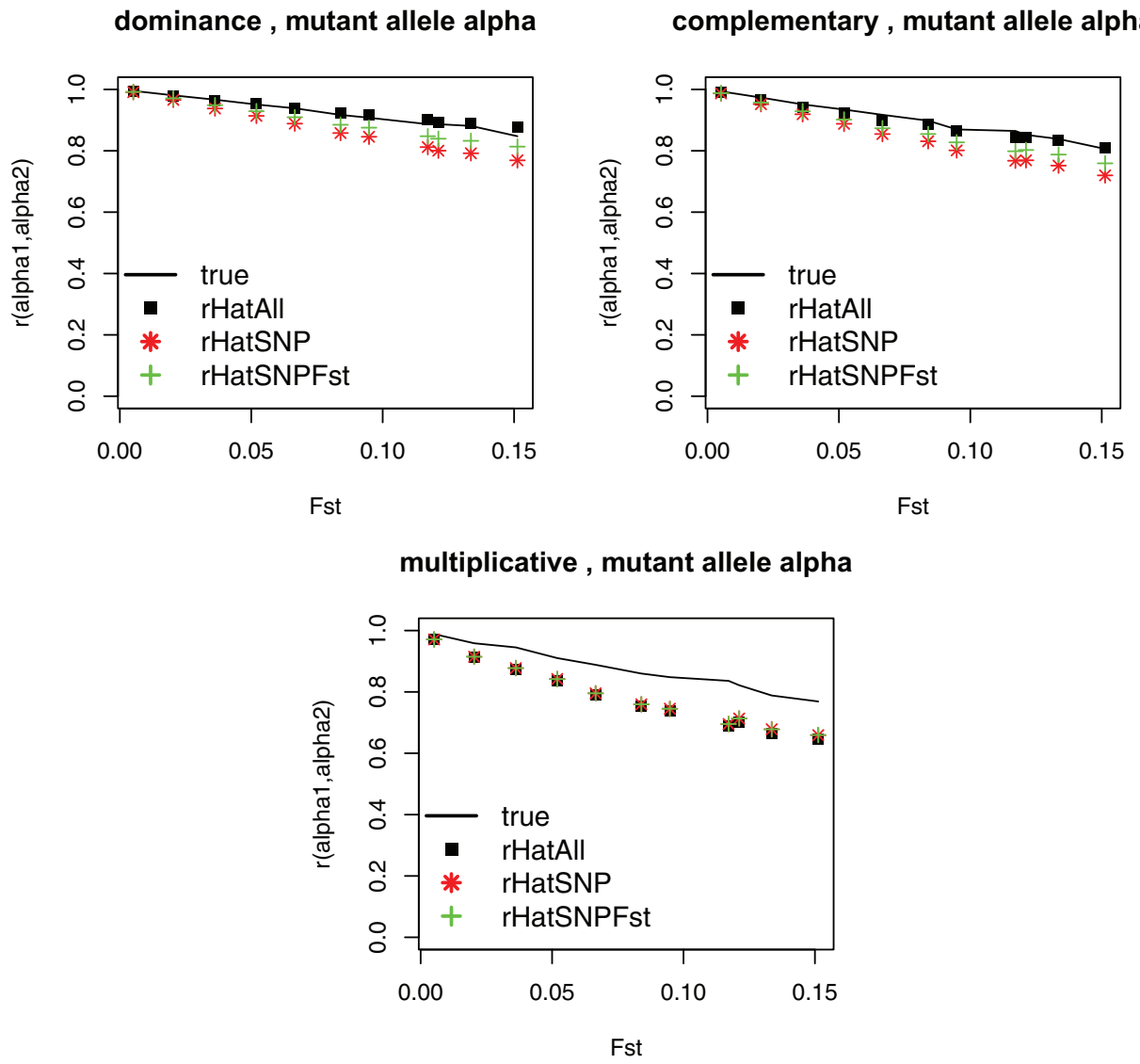

Figure 2 Simulated (straight black line) and estimated (points) correlation across QTL substitution effects of mutant alleles across two populations as a function of their F_{ST} differentiation coefficient. rHatAll: estimates with all polymorphism. rHatSNP: estimates using SNP-like loci. rHatSNPFst: estimates using SNP-like loci with a correction for heterozygosity. Results of 10 replicates per point with s.e. <0.01.

nonadditive variation are scarce, as are estimates of distances across populations in terms of covariance of allele frequencies. The differentiation F_{ST} of Jersey and Holstein breeds of cattle obtained from VanRaden *et al.* (2011) is 0.06, and among Landrace and Yorkshire is 0.16 (Xiang *et al.* 2017).

Table 3 includes literature estimates for additive, dominance and epistatic variation. We used estimates for milk yield in Simmental cattle (Fuerst and Sölkner 1994) as if it was Holstein, and for litter size in a commercial pig line (Vitezica *et al.* 2018) as if it was Landrace, because we could not find estimates in the same populations. These estimates (as most estimates of nonadditive variances) are inaccurate and are presented here only as examples. In the particular case of cattle, the estimated value of σ^2_{AA} is higher than usually expected *a priori* (Hill *et al.* 2008), so we take it as an example of an extreme, but not impossible, case.

Then, we assumed three distributions for QTL allele frequencies, following Hill *et al.* (2008): a uniform distribution ("Uniform"), and a U-shaped distributions with $f(p) \propto p^{-1}(1-p)^{-1}$ with effective population size of $Ne = 50$ ("Hill"), bounded at $\left[\frac{1}{2Ne}, 1-\frac{1}{2Ne}\right]$. We also assumed a Beta(0.04, 0.04) distribution, which is an extreme U-shaped distribution ("Extreme") in which roughly 80% of the QTLs have minor allele frequency lower than 0.01. This results in respective values of $\overline{H_b} = (0.333, 0.107, 0.037)$ and $\overline{H_b^2} = (0.133, 0.018, 0.013)$. Using the values described above, we estimated the correlation of QTL substitution effects across breeds, $r\left(\alpha_i^b, \alpha_i^{b'}\right)$, using Equation (7).

Table 3 Variance component estimates, as ratio from phenotypic variance, from literature

Species	σ^2_A	σ^2_D	σ^2_{AA}
Cattle	0.20	0.09	0.15
Pigs	0.092	0.020	0.016

Estimates for milk yield in cattle (Fuerst and Sölkner 1994) and litter size in pigs (Vitezica *et al.* 2018).

Similarly, the correlation of QTL substitution effects a certain number of generations away, $r\left(\alpha_i^{t_2}, \alpha_i^{t_1}\right)$, was obtained applying Equation (9). An increase $\Delta F = 0.01$ per generation was assumed, for the same assumed variances and allele frequency distributions.

Estimates of correlation across populations using literature values

The results for reasonable guesses of the distribution of allele frequencies at QTL are in Table 4. The correlation of QTL substitution effects across breeds was roughly 0.85 across both species (pigs and cattle) and all assumed distributions of gene frequencies. These numbers are higher than scarce estimates of genetic correlations in the literature: 0.4–0.8 for production traits (Karoui *et al.* 2012), 0.3–0.5 for milk yield (Legarra *et al.* 2014) and 0.6 for body condition (Porto-Neto *et al.* 2015), and we consider that this is because across-breeds genetic correlations also include genotype by environment interactions. It may be argued that these studies used SNPs instead of true QTLs; however, Wientjes *et al.* (2017) in a simulation study obtained unbiased estimates of correlation across QTLs using SNP markers for a true correlation of 0.5, and (Cuyabano *et al.* 2018) showed analytically that when SNP markers yield correct estimates of true relationships at QTL, estimates of genetic parameters are unbiased.

Concerning decrease of correlations across generations, results (Table 5) show a slow decrease, and this decrease is, again, not sensitive to the assumed distribution of allelic frequencies. Values in Table 5, quite close to 1, somehow disagree with few existing estimates of across-generation genetic correlations, clearly lower than 1 (Tsuruta *et al.* 2004; Haile-Mariam and Pryce 2015). Again, part of the literature estimates of correlation much lower than 1 is probably due to nonadditive gene action, and part due to genotype by environment interactions.

Discussion

The fact that most genetic variation is statistically additive is well known. However, this does not imply equality of statistical additive effects across populations, something that is possibly reflected in the difficulty of predicting breeding values or disease risks across (distinct) populations, either in animal or human genetics. Only recently has nonadditive biological gene action been identified as playing a role in determining the degree to which prediction across populations is successful (Dai *et al.* 2020; Duenk *et al.* 2020). These authors presented result of simulations, where the problem is that the results are scenario specific and do not allow a good appraisal of the different factors. Having a theory allows to derive more general expressions and understanding factors influencing across-population differences in substitution effects.

In this study, we present a general, formal framework that does not depend on specific hypothesis regarding gene action or order of the epistatic interactions. In our derivation, we approached *high-order functional* epistasis by Taylor expansions, leading to expressions that involve only *low-order statistical* additive epistasis (actually pairwise). Including one extra term in our Taylor expansion would include *three-loci statistical* epistasis and so on, but extra terms would lead to more difficult expressions (covariance of allele frequencies across loci and populations), and it is expected that the magnitude of the statistical effects is smaller and smaller with higher orders of interactions (Mäki-Tanila and Hill 2014). We find reasonable agreement with our simulation and also with values from literature. However, our simulations may not be particularly realistic, something that would require considerable thinking on how to simulate biologically meaningful epistasis mechanisms for a variety of traits. We see them as building blocks of nonadditive architecture. At any rate, the three scenarios generated sizeable nonadditive variation which is a challenging case for our expressions.

Three main factors influence the correlation of substitution effects between populations $r(\alpha^b, \alpha^{b'})$: the genetic similarity of the two populations, the magnitudes of additive, dominance and additive by additive variances, and the distribution of allele frequencies at QTL. We consider that showing explicitly these three factors is an achievement, as their role is implicit, yet not explicitly shown, in previous works in simulated and real data (Martin *et al.* 2019; Dai *et al.* 2020; Duenk *et al.* 2020). Now we discuss these three factors.

The distance across populations is summarized by Nei's minimum genetic distances $D_{b,b'}$ or F_{st} indexes. Under pure drift scenarios, these depend on divergence times and effective population sizes (Weir and Hill 2002; Bonhomme *et al.* 2010; Walsh and Lynch 2018).

The factors $\frac{\overline{H_b}}{H_b^2}$ and $\frac{1}{H_b}$ are weighting factors on dominance and additive by additive variances. If the allele frequencies are modelled using symmetric $Beta(a, a)$ distributions (see Supplementary File S1, Section 1.11), these become $\frac{\overline{H_b}}{H_b^2} = \frac{1}{2^{\frac{a+1}{a}}\left(1+\frac{a+2}{2a+2}\left(\frac{a+3}{2a+3}-2\right)\right)}$ and $\frac{1}{H_b} = 2 + \frac{1}{a}$. The first is bounded between 3 (for U-shaped distributions) and 2 (for peaked distributions), so that dominance variation does not play a big role in the difference between substitution effects across populations unless dominance variation is much larger that additive variation, something that seems unlikely based on theory and estimates. However, the second weight, due to epistasis, is not bounded and is large for small values of a (*e.g.*, 27 for Beta[0.04,0.04]), and in this case functional epistasis plays a strong role in additive variation for U-shaped distributions (Hill *et al.* 2008). The spectra of allele frequencies of causal mutations is subject to large debate but there seems to be

Table 4 Estimates of correlations of QTL effects across breeds based on values from Table 3 and **Equation 7**, for different distributions of QTL frequencies

Species	Uniform	Hill[a]	Extreme[b]
Cattle	0.83	0.82	0.84
Pigs	0.87	0.85	0.88

[a]U-shaped distribution with effective population size of 50.
[b]Beta(0.04, 0.04) distribution.

Table 5 Correlation of QTL effects within breed across time, based on values from Table 3 and Equation 7, for different distributions of QTL frequencies

Species	Distance in generations	Uniform	Hill[a]	Extreme[b]
Cattle	1	0.98	0.97	0.98
	2	0.97	0.94	0.97
	5	0.93	0.87	0.93
	10	0.86	0.78	0.87
Pigs	1	1.00	0.99	1.00
	2	0.99	0.98	0.99
	5	0.98	0.96	0.97
	10	0.96	0.93	0.93

[a]U-shaped distribution with effective population size of 50.
[b]Beta (0.04, 0.04) distribution.

a consensus that low frequency mutations make non-negligible contributions to genetic variance (Eyre-Walker 2010; Gibson 2012). It is unknown if this leads to extreme values of $\frac{1}{H_b}$.

The particular case of purifying selection deserves some attention. For a mutant recessive deleterious allele with fitness $1-s$, the substitution effect is $\alpha = -ps$ so, in principle, α changes across populations—this is similar to our simulated scenario of Complete Dominance. However, if the locus is truly at equilibrium in all populations, then $p \approx \sqrt{\frac{u}{s}}$ (Crow and Kimura 1970, Chapter 6). Provided that s and u are identical across populations, the allele frequency p *and* the substitution effect α should be identical across populations. However, s is not necessarily homogeneous across populations, likely depending on the environment.

The previous result ($\frac{1}{H_b}$ needs to be large for biological dominance to play a role) clarifies the findings of Duenk *et al.* (2020) and Dai *et al.* (2020) that it is mainly biological epistasis that generates changes in additive substitution effects across populations. For instance, when one population drifts, the additive by additive statistical variation enters into additive variation (Hill *et al.* 2006; Walsh and Lynch 2018).

Thus, $r\left(\alpha_i^b, \alpha_i^{b'}\right)$ will be very low only if (1) populations are distinct, (2) there is *statistical* additive by additive variation and (3) QTL allele frequency distributions are U- or L-shaped.

As for the magnitudes of nonadditive variance, in most conceivable cases σ^2_{AA} and σ^2_D are of magnitudes at most like additive variances, but more often less than half (Fuerst and Sölkner 1994; Palucci *et al.* 2007; Hill *et al.* 2008; Mäki-Tanila and Hill 2014; Vitezica *et al.* 2018), so in practice the most limiting factor is distance across breeds. Indeed, there are very few accurate estimates of nonadditive variation although more are becoming available. We argue that then, and based also in our results, it will be possible to make theory and data-based choices on the possibilities of using predictions and GWAS results across populations.

Then, for reasonable assumptions about allele frequency distribution of QTLs, we have shown that the correlation of substitution effects across populations is typically around 0.8 or higher, which is higher than scarce estimates of genetic correlations across populations available in the literature, which range from 0.3 to 0.8 (Karoui *et al.* 2012; Legarra *et al.* 2014; Porto-Neto *et al.* 2015; Xiang *et al.* 2017) the difference being due, probably, to genotype by environment interaction. Our results seem rather robust to different distributions of allele frequencies. These values are high but not 1, which raises the question of how to conceive optimal strategies for across populations predictions (*e.g.*, more data within breed or finer locations of causal variants across breed). This is of practical relevance, *e.g.*, for genomic predictions in livestock improvement, but also in human genetics *e.g.*, for the use of European-based Polygenic Risk Scores in individuals from other ancestries (Martin *et al.* 2019).

Similar results apply to the same populations across generations, in which case the correlation of substitution effects across generations goes from 0.99 (next 1–2 generations) to 0.80 (10 generations). This illustrates that, even if genomic predictions would use QTLs or markers in very tight LD with QTLs, there would still be, in the long run, a need for a continuous system of data collecting and re-estimation of effects.

In the simulations, we confirmed that our estimates are reasonably good although not perfect. They are, depending on the scenario and target (random allele or mutant allele), almost unbiased, slightly biased upwards, or biased downward. There are several reasons for the disagreement. The most obvious one is the inherent approximation of the Taylor series expansion. Second, splitting variances such as $Var\left(\epsilon_i d_i^{*b}\right)$ or $Var\left(\boldsymbol{\epsilon}'\left(\boldsymbol{\alpha}\boldsymbol{\alpha}\right)_i^b\right)$ into basic components implies either a strong assumption of multivariate normality and independence or a less strong one of "expectation and variance-independence" (Bohrnstedt and Goldberger 1969). For instance, it is assumed that the variance of d_i^{*b} is not related to the magnitude of the difference of allele frequencies ϵ_i, but this is not necessarily true. Third, it is further assumed, in the factorization of genetic variances, that QTL effects and allele frequencies are independent. We have also ignored that the change ϵ is proportional to heterozygosity and therefore small at extremes values of allele frequencies.

Another factor that we did not consider is the empirical evidence of an inverse relationship between heterozygosity and absolute effect at the locus (Park *et al.* 2011). It is unclear how this would affect our findings. A (rather extreme) functional gene action that generates larger α at extreme frequencies is overdominance, which is similar to our "dominance" scenario.

As for our estimators of the correlation across "negative" alleles $r_{\text{mutant}}\left(\alpha^b, \alpha^{b'}\right)$, they are less robust that the estimators for a "random" allele $r_{\text{random}}\left(\alpha^b, \alpha^{b'}\right)$. The reason for this is that obtaining $r_{\text{mutant}}\left(\alpha^b, \alpha^{b'}\right)$ from $r_{\text{random}}\left(\alpha^b, \alpha^{b'}\right)$ involves a further approximation, the normality of α.

Conclusions

We presented a coherent, approximate theory, that does not invoke any particular mechanism of gene action, to explain and appraise the change in magnitude of (additive) QTL substitution effects across populations and generations. The theory gives good approximate estimates of this correlation, that needs to be otherwise explicitly estimated. More importantly, the theory shows that the main sources for the change of effects are relationships across populations, magnitudes of additive and first-order nonadditive variances (dominance and additive by additive), and spectra of allele frequencies. These findings provide better understanding of the properties of genomic prediction methods and of quantitative genetics in general.

Data availability

Supplementary File S1 associated with this manuscript is in https://doi.org/10.25386/genetics.13168952. Code and files for simulations can also be found in https://figshare.com/articles/software/scripts_zip/14509956. All other results can be reproduced using equations and figures from the text.

Acknowledgments

The authors thank editor and reviewers for very detailed feedback, and M.A. Toro, A. Caballero, S. Boitard, and M. Bonhomme for advice. A.L., Z.G.V., and C.A.G.B. want to dedicate this study to their friend Eduardo Fernández, who could discuss substitution effects while driving through traffic jams, and who left us too soon.

Funding

Part of this study was done while Y. Wientjes was visiting the GenPhySE unit at INRAE, Toulouse, financed by the Netherlands Organisation of Scientific Research (NWO). Authors thank INRA SelGen metaprograms EpiSel and EpiFun. This project has

received funding from the European Unions' Horizon 2020 Research & Innovation programme under grant agreement N°772787 -SMARTER. We are grateful to the Genotoul Bioinformatics Platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing computing and storage resources.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

Aiello D, Patel K, Lasagna E. 2018. The myostatin gene: an overview of mechanisms of action and its relevance to livestock animals. Anim Genet. 49:505–519. doi:10.1111/age.12696.

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the impact of rare variants. Genome Res. 23:1514–1521. doi:10.1101/gr.154831.113.

Bohrnstedt GW, Goldberger AS. 1969. On the exact covariance of products of random variables. J Am Stat Assoc. 64:1439–1442. doi:10.1080/01621459.1969.10501069.

Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, *et al.* 2010. Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics. 186:241–262. doi:10.1534/genetics.110.117275.

Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, *et al.* 2003. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science. 299:256–259. doi:10.1126/science.1077209.

Caballero A, Toro MA. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv Genet. 3:289–299.

Carlborg Ö, Jacobsson L, Åhgren P, Siegel P, Andersson L. 2006. Epistasis and the release of genetic variation during long-term selection. Nat Genet. 38:418–420. doi:10.1038/ng1761.

Chen GK, Marjoram P, Wall JD. 2009. Fast and flexible simulation of DNA sequence data. Genome Res. 19:136–142. doi:10.1101/gr.083634.108.

Crow J, Kimura M. 1970. An Introduction to Population Genetics Theory. New York, NY: Harper and Row.

Cuyabano BCD, Sørensen AC, Sørensen P. 2018. Understanding the potential bias of variance components estimators when using genomic models. Genet Sel Evol. 50:41.doi:10.1186/s12711-018-0411-0.

Dai Z, Long N, Huang W. 2020. Influence of genetic interactions on polygenic prediction. G3 (Bethesda). 10:109–115. doi:10.1534/g3.119.400812.

Duenk P, Bijma P, Calus MPL, Wientjes YCJ, van der Werf JHJ. 2020. The impact of non-additive effects on the genetic correlation between populations. G3 (Bethesda). 10:783–795. doi:10.1534/g3.119.400663.

Eyre-Walker A. 2010. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc Natl Acad Sci USA. 107(Suppl. 1):1752–1756. doi:10.1073/pnas.0906182107.

Falconer DS, Mackay TFC. 1996. Introduction to Quantitative Genetics. New York, NY: Longman.

Fernández E, Sánchez J, Martínez R, Legarra A, Baselga M. 2017. Role of inbreeding depression, non-inbred dominance deviations and random year-season effect in genetic trends for prolificacy in closed rabbit lines. J Anim Breed Genet. 134:441–452.

Fuerst C, James JW, Sölkner J, Essl A. 1997. Impact of dominance and epistasis on the genetic make-up of simulated populations under selection: a model development. J Anim Breed Genet. 114:163–175. doi:10.1111/j.1439-0388.1997.tb00502.x.

Fuerst C, Sölkner J. 1994. Additive and nonadditive genetic variances for milk yield, fertility, and lifetime performance traits of dairy cattle. J Dairy Sci. 77:1114–1125. doi:10.3168/jds.S0022-0302(94)77047-8.

Galarza CE, Matos LA, Dey DK, Lachos VH. 2020. On moments of folded and doubly truncated multivariate extended skew-normal distributions. Math Stat. ArXiv200913488.

García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, *et al.* 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci USA. 113:E3995–E4004.

Gautier M, Capitan A, Fritz S, Eggen A, Boichard D, *et al.* 2007. Characterization of the DGAT1 K232A and variable number of tandem repeat polymorphisms in French Dairy Cattle. J Dairy Sci. 90:2980–2988. doi:10.3168/jds.2006-707.

Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, *et al.* 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Sci NY. 324:528–532.

Gibson G. 2012. Rare and common variants: twenty arguments. Nat Rev Genet. 13:135–145. doi:10.1038/nrg3118.

Grisart B, Coppieters W, Farnir F, Karim L, Ford C, *et al.* 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222–231.

Grobet L, Martin LJ, Poncelet D, Pirottin D, Brouwers B, *et al.* 1997. A deletion in the bovine myostatin gene causes the double–-muscled phenotype in cattle. Nat Genet. 17:71–74. doi:10.1038/ng0997-71.

Haile-Mariam M, Pryce JE. 2015. Variances and correlations of milk production, fertility, longevity, and type traits over time in Australian Holstein cattle. J Dairy Sci. 98:7364–7379. doi:10.3168/jds.2015-9537.

Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol. 41:51.

Hill WG, Barton NH, Turelli M. 2006. Prediction of effects of genetic drift on variance components under a general model of epistasis. Theor Popul Biol. 70:56–62. doi:10.1016/j.tpb.2005.10.001.

Hill WG, Goddard ME, Visscher PM. 2008. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4:e1000008.doi:10.1371/journal.pgen.1000008.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics. 132:583–589.

Kan R, Robotti C. 2017. On moments of folded and truncated multivariate normal distributions. J Comput Graph Stat. 26:930–934. doi:10.1080/10618600.2017.1322092.

Karoui S, Carabaño MJ, Díaz C, Legarra A. 2012. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. Genet Sel Evol. 44:39.

Kojima K. 1959. Role of epistasis and overdominance in stability of equilibria with selection. Proc Natl Acad Sci USA. 45:984–989.

Kojima K-I. 1961. Effects of dominance and size of population on response to mass selection. Genet Res. 2:177–188. doi: 10.1017/S0016672300000689.

Lande R, Thompson R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics. 124:743–756.

Legarra A, Baloche G, Barillet F, Astruc JM, Soulas C, *et al.* 2014. Within-and across-breed genomic predictions and genomic relationships for Western Pyrenees dairy sheep breeds Latxa, Manech, and Basco-Béarnaise. J Dairy Sci. 97:3200–3212.

Liu Z, Alkhoder H, Reinhardt F, Reents R. 2016. Accuracy and bias of genomic prediction for second-generation candidates. Interbull Bull. 50:17–23.

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

Mackay TFC. 2014. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet. 15: 22–33. doi:10.1038/nrg3627.

Mackay TFC. 2015. Epistasis for Quantitative Traits in Drosophila. In: JH Moore, SM Williams, editors. Epistasis: Methods and Protocols, Methods in Molecular Biology. New York, NY: Springer. p. 47–70.

MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, *et al.* 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genomics. 17:144. doi:10.1186/s12864-016-2443-6.

Magwire MM, Yamamoto A, Carbone MA, Roshina NV, Symonenko AV, *et al.* 2010. Quantitative and molecular genetic analyses of mutations increasing Drosophila life span. PLoS Genet. 6: e1001037. doi:10.1371/journal.pgen.1001037.

Mäki-Tanila A, Hill WG. 2014. Influence of gene interaction on complex trait variation with multilocus models. Genetics. 198: 355–367. doi:10.1534/genetics.114.165282.

Marigorta UM, Navarro A. 2013. High trans-ethnic replicability of GWAS results implies common causal variants. PLoS Genet. 9: e1003566.doi:10.1371/journal.pgen.1003566.

Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, *et al.* 2019. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 51:584–591. doi: 10.1038/s41588-019-0379-x.

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819–1829.

Moghaddar N, Khansefid M, van der Werf JHJ, Bolormaa S, Duijvesteijn N, *et al.* 2019. Genomic prediction based on selected variants from imputed whole-genome sequence data in Australian sheep populations. Genet Sel Evol. 51:72. doi: 10.1186/s12711-019-0514-2.

Nei M. 1987. Molecular Evolutionary Genetics. New York, NY: Columbia University Press.

Palucci V, Schaeffer LR, Miglior F, Osborne V. 2007. Non-additive genetic effects for fertility traits in Canadian Holstein cattle (Open Access publication). Genet Sel Evol. 39:181–193. doi: 10.1186/1297-9686-39-2-181.

Park J-H, Gail MH, Weinberg CR, Carroll RJ, Chung CC, *et al.* 2011. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proc Natl Acad Sci USA. 108:18026–18031. doi: 10.1073/pnas.1114759108.

Pérez-Enciso M. 2014. Genomic relationships computed from either next-generation sequence or array SNP data. J Anim Breed Genet. 131:85–96. doi:10.1111/jbg.12074.

Pérez-Enciso M, Rincón JC, Legarra A. 2015. Sequence- vs. chip-assisted genomic selection: accurate biological information is advised. Genet Sel Evol. 47:43.

Porto-Neto LR, Barendse W, Henshall JM, McWilliam SM, Lehnert SA, *et al.* 2015. Genomic correlation: harnessing the benefit of combining two unrelated populations for genomic selection. Genet Sel Evol. 47:84. doi:10.1186/s12711-015-0162-0.

Powell JE, Visscher PM, Goddard ME. 2010. Reconciling the analysis of IBD and IBS in complex trait studies. Nat Rev Genet. 11:800–805.

Rodríguez-Ramilo S, Elsen JM, Legarra A. 2019. Inbreeding and effective population size in French dairy sheep: comparison between genomic and pedigree estimates. J Dairy Sci. 102:4227–4237.

Roos A, Hayes BJ, Goddard ME. 2009. Reliability of genomic predictions across multiple populations. Genetics. 183:1545–1553. doi: 10.1534/genetics.109.104935.

Rupp R, Senin P, Sarry J, Allain C, Tasca C, *et al.* 2015. A point mutation in suppressor of Cytokine signalling 2 (Socs2) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. PLoS Genet. 11:e1005629. doi: 10.1371/journal.pgen.1005629.

Shi H, Gazal S, Kanai M, Koch EM, Schoech AP, *et al.* 2021. Population-specific causal disease effect sizes in functionally important regions impacted by selection. Nat Commun. 12:1098. doi:10.1038/s41467-021-21286-1.

Streit M, Neugebauer N, Meuwissen THE, Bennewitz J. 2011. Short communication: evidence for a major gene by polygene interaction for milk production traits in German Holstein dairy cattle. J Dairy Sci. 94:1597–1600. doi:10.3168/jds.2010-3834.

Tsuruta S, Misztal I, Lawlor TJ, Klei L. 2004. Modeling final scores in US Holsteins as a function of year of classification using a random regression model. Livest Prod Sci. 91:199–207. doi: 10.1016/j.livprodsci.2003.09.016.

VanRaden P, Olson K, Wiggans G, Cole J, Tooker M. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 94:5673–5682.

Vitezica ZG, Legarra A, Toro MA, Varona L. 2017. Orthogonal estimates of variances for additive, dominance and epistatic effects in populations. Genetics. 206:1297–1307. doi:10.1534/genetics.116.199406.

Vitezica ZG, Reverter A, Herring W, Legarra A. 2018. Dominance and epistatic genetic variances for litter size in pigs using genomic models. Genet Sel Evol. 50:1–8.

Walsh B, Lynch M. 2018. Evolution and Selection of Quantitative Traits. Oxford University Press.

Weir BS, Hill WG. 2002. Estimating F-statistics. Annu Rev Genet. 36: 721–750. doi:10.1146/annurev.genet.36.050802.093940.

Welsh CS, Stewart TS, Schwab C, Blackburn HD. 2010. Pedigree analysis of 5 swine breeds in the United States and the implications for genetic conservation. J Anim Sci. 88:1610–1618. doi: 10.2527/jas.2009-2537.

Wientjes YCJ, Bijma P, Vandenplas J, Calus MPL. 2017. Multi-population genomic relationships for estimating current genetic variances within and genetic correlations between populations. Genetics. 207:503–515. doi:10.1534/genetics.117.300152.

Xiang T, Christensen OF, Legarra A. 2017. Technical note: genomic evaluation for crossbred performance in a single-step approach with metafounders. J Anim Sci. 95:1472–1480.

Communicating editor: A. Agrawal