



**HAL**  
open science

## Non-additive effects in genomic selection

Luis Varona, Andres Legarra, Miguel A. Toro, Zulma G. Vitezica

► **To cite this version:**

Luis Varona, Andres Legarra, Miguel A. Toro, Zulma G. Vitezica. Non-additive effects in genomic selection. *Frontiers in Genetics*, 2018, 9, Non paginé. 10.3389/fgene.2018.00078 . hal-02626411

**HAL Id: hal-02626411**

**<https://hal.inrae.fr/hal-02626411>**

Submitted on 26 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



# Non-additive Effects in Genomic Selection

Luis Varona<sup>1,2\*</sup>, Andres Legarra<sup>3</sup>, Miguel A. Toro<sup>4</sup> and Zulma G. Vitezica<sup>5</sup>

<sup>1</sup> Departamento de Anatomía, Embriología y Genética Animal, Universidad de Zaragoza, Zaragoza, Spain, <sup>2</sup> Instituto Agroalimentario de Aragón (IA2), Zaragoza, Spain, <sup>3</sup> Génétique Physiologie et Systèmes d'Elevage (GenPhySE), Institut National de la Recherche Agronomique de Toulouse, Castanet-Tolosan, France, <sup>4</sup> Departamento Producción Agraria, ETS Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Madrid, Spain, <sup>5</sup> Génétique Physiologie et Systèmes d'Elevage (GenPhySE), Université de Toulouse, Castanet-Tolosan, France

In the last decade, genomic selection has become a standard in the genetic evaluation of livestock populations. However, most procedures for the implementation of genomic selection only consider the additive effects associated with SNP (Single Nucleotide Polymorphism) markers used to calculate the prediction of the breeding values of candidates for selection. Nevertheless, the availability of estimates of non-additive effects is of interest because: (i) they contribute to an increase in the accuracy of the prediction of breeding values and the genetic response; (ii) they allow the definition of mate allocation procedures between candidates for selection; and (iii) they can be used to enhance non-additive genetic variation through the definition of appropriate crossbreeding or purebred breeding schemes. This study presents a review of methods for the incorporation of non-additive genetic effects into genomic selection procedures and their potential applications in the prediction of future performance, mate allocation, crossbreeding, and purebred selection. The work concludes with a brief outline of some ideas for future lines of that may help the standard inclusion of non-additive effects in genomic selection.

**Keywords:** genomic selection, dominance, epistasis, crossbreeding, genetic evaluation

## INTRODUCTION

Through his experiments on pea plants, Gregor Mendel (1866) realized that some traits are dominant over others (for example “round peas” were dominant over “wrinkled peas”). In Mendel’s own words: “As a rule, hybrids do not represent the form exactly intermediate between the parental strains... Those traits that pass into hybrid association entirely or almost entirely unchanged, thus themselves representing the traits of the hybrid, are termed “dominating,” and those that become latent in the association, “recessive””. Shortly after the rediscovery of Mendel’s rules, it was observed that, in some cases, the addition of the individual action of genes could not explain the mode of inheritance, and Bateson (1909) coined the term “epistasis” to describe the cases in which the actions of two or more genes interact. A distinction must be drawn between biological (functional) genetic effects that correspond to the Mendelian definition (i.e., dominance means that the heterozygote value is higher or lower than the mean of homozygous genotypes) and statistical (population or weighted) effects which depend on allelic frequencies. In the latter, the relevant issue is the contribution of non-additive effects to genetic variance. Some authors argue that non-additive genetic effects may be a general phenomenon whose understanding is important for gaining more knowledge on the nature of quantitative traits, but whose contribution to variance is negligible (Crow, 2010).

## OPEN ACCESS

### Edited by:

Rohan Luigi Fernando,  
Iowa State University, United States

### Reviewed by:

Romdhane Rekaya,  
University of Georgia, United States  
Piter Bijma,  
Wageningen University & Research,  
Netherlands

### \*Correspondence:

Luis Varona  
lvarona@unizar.es

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 15 December 2017

**Accepted:** 19 February 2018

**Published:** 06 March 2018

### Citation:

Varona L, Legarra A, Toro MA and  
Vitezica ZG (2018) Non-additive  
Effects in Genomic Selection.  
Front. Genet. 9:78.  
doi: 10.3389/fgene.2018.00078

From the perspective of quantitative genetics, Fisher (1918) conceived the infinitesimal model which postulates that a very large number of unlinked genes control the genetic variation of quantitative traits. He described the resemblance between relatives in a pure additive model which was quickly extended to incorporate dominance (Fisher, 1918; Wright, 1921). Resemblance between relatives including epistatic effects of second and higher order was also described (Cockerham, 1954; Kempthorne, 1954). However, whilst the formulation of the infinitesimal model in the additive context is evident, its interpretation is not clear when non-additive effects are included (Barton et al., 2017).

The main goal of animal or plant breeding is to identify, select and mate the best individuals of a breeding stock in order to maximize performance in future generations (Falconer and McKay, 1996; Bernardo, 2010). The procedure for computing the breeding values (genetic evaluation) of candidates for selection plays a crucial role. Traditionally, these methods use phenotypic and genealogical information, such as the selection index (Hazel, 1943) or the Best Linear Unbiased Predictor (Henderson, 1973) and rely on the foundations of the infinitesimal model (Fisher, 1918).

Nevertheless, non-additive genetic effects have been ignored in the genetic evaluation of livestock for several reasons: (i) the lack of informative pedigrees, such as large full-sib families; (ii) the calculations involved are more complex; (iii) the fact that statistical additive variance captures biological dominance or higher order interaction effects (Hill, 2010); and, (iv) the difficulty in using dominant values in practice (mate allocation). As a consequence, estimates of non-additive genetic variances are scarce in livestock populations (Misztal et al., 1998; Nguyen and Nagyné-Kiszlinger, 2016).

## GENOMIC SELECTION

Since the late 80s and 90s, developments in molecular genetics resulted in a set of neutral molecular markers, such as microsatellites, that were commonly used to detect QTL (Quantitative Trait Loci) in almost all livestock populations. The objective of those studies was to identify polymorphic markers or genes associated with phenotypic variation of traits of interest ([www.animalgenome.org/QTL](http://www.animalgenome.org/QTL)), with the ultimate goal of using them in Marker or Gene Assisted Selection (Dekkers, 2004). However, these strategies became obsolete with the advent of dense genotyping devices (Gunderson et al., 2005) that provided a very large amount of SNP (Single Nucleotide Polymorphism) and that allowed the development of genomic selection (GS) models (Meuwissen et al., 2001).

Genomic selection has become a very successful strategy for the prediction of the breeding values of candidates for selection and has revolutionized the field of animal breeding over the past decade. The basic idea of GS is to develop the following linear model:

$$y_i = \mu + \sum_{j=1}^n t_{ij}a_j + e_i$$

The model explains the phenotypic data of  $m$  individuals ( $y_i$ ) with  $i = 1 \dots m$  (or transformations of data, such as daughter yield deviations) by the effects associated with a very large number ( $n$ ) of SNP ( $a_j$ ) with  $j = 1 \dots n$ . Moreover,  $t_{ij}$  is the genotypic configuration (coded additively, e.g., Falconer and McKay, 1996) of the  $i$ th individual and for the  $j$ th SNP (0, 1, and 2 for  $A_1A_1$ ,  $A_1A_2$ , and  $A_2A_2$  genotypes, respectively), and  $e_i$  is the residual. Furthermore, the prediction of individual breeding values ( $\hat{u}_i$ ) of the candidates for selection can be calculated *a posteriori* from marker effect estimates as  $\hat{u}_i = \sum_{j=1}^n t_{ij}\hat{a}_j$ .

A significant limitation for implementation is that most genomic evaluation models suffer the statistical problem of a larger number of parameters ( $n$ ) that must be estimated from a smaller number of data ( $m$ ). The most common method employed for resolving this problem is the use of some type of regularization of SNP marker effects (Gianola, 2013). Several approaches have been suggested, ranging from a simple Gaussian regularization (Meuwissen et al., 2001) to more complex models that involve  $t$  shaped (Meuwissen et al., 2001), double exponential (De los Campos et al., 2009b), mixtures of distributions (Meuwissen et al., 2001; Habier et al., 2011; Erbe et al., 2012), or non-parametric or semi-parametric approaches (Gonzalez-Recio et al., 2014). The predictive ability of all these approaches depends on the genetic architecture of the traits being analyzed (Daetwyler et al., 2010), although for polygenic traits, all approaches offer similar results (Wang et al., 2015).

An interesting property of the assumption of a Gaussian prior distribution for marker effects (Random Regression BLUP—RR-BLUP) is that the GS model can be reformulated in terms of individual (animal) effects, using the equations of the Henderson's classic Mixed Model that provide breeding values for all individuals, including candidates for selection (Genomic BLUP or GBLUP). The only difference with standard mixed model equations is that the numerator relationship matrix ( $\mathbf{A}$ ) is replaced by the genomic relationship matrix ( $\mathbf{G}$ ), as defined by VanRaden (2008). In addition, this approach can be extended for the genetic evaluation of non-genotyped individuals in the Single-Step approach (Aguilar et al., 2010), facilitating the integration of GS procedures in the genetic evaluation of candidates for selection in most livestock breeding programmes. More recently, Fernando et al. (2014) described a Bayesian procedure that can also simultaneously evaluate genotyped and non-genotyped individuals and allows the use of alternative regularization procedures. Nevertheless, computational costs are markedly higher with the Bayesian model than with the Single-Step approach.

Despite the regularization procedure, the genomic evaluation methods are based on the evaluation of marker substitution effects through the construction of the covariates ( $t_{ij}$ ) or the  $\mathbf{G}$  matrix (above). The additive (or breeding) values capture a large part of dominant and higher-order interaction effects (Hill et al., 2008; Crow, 2010; Hill, 2010). Substitution effects that capture dominance and epistatic functional effects are not necessarily stable across generations or populations due to changes in allelic frequencies. In any case, only additive values (substitution effects) contribute to breeding values and are therefore expressed in the next generation. However, estimates of non-additive genetic

effects may be of relevance because: (i) they may contribute to increasing the accuracy of prediction of breeding values and the response to selection (Toro and Varona, 2010; Aliloo et al., 2016; Duenk et al., 2017); (ii) they allow the definition of mate allocation procedures between candidates for selection (Maki-Tanila, 2007; Toro and Varona, 2010; Aliloo et al., 2017); and (iii) they can be used to benefit from non-additive genetic variation through the definition of appropriate crossbreeding or purebred breeding schemes (Maki-Tanila, 2007; Zeng et al., 2013).

## GENOMIC SELECTION MODELS WITH DOMINANCE

The simplest approach for the inclusion of dominance in genomic selection models is to extend the basic model with the inclusion of a dominance effect (Toro and Varona, 2010; Su et al., 2012) associated to each SNP marker:

$$y_i = \mu + \sum_{j=1}^n t_{ij}a_j + \sum_{j=1}^n c_{ij}d_j + e_i$$

where  $y_i$  is the phenotypic value of the  $i$ th individual and  $\mu$  is the population mean. For each of the  $n$  SNP markers,  $a_j$  and  $d_j$  are the additive and dominance effects for the  $j$ th marker, respectively. The covariates  $t_{ij}$  and  $c_{ij}$  are 2, 1, and 0 (coded additively) and 0, 1, and 0 (coded in a “biological dominant” manner) for the genotypes  $A_1A_1$ ,  $A_1A_2$ ,  $A_2A_2$  of each marker, respectively. In some ways, pedigree-based models for dominance were based on “expected” dominant relationships. Thus, genomic models are based on “observed” heterozygotes. However, when using this model it should be noted that that  $a_j$  is no longer the marker substitution effect, but the “biological” additive genotypic effect and individual breeding values are not predicted. In fact, the partition of variance in statistical components due to additivity, dominance, and epistasis does not reflect the “biological” (or “functional”) effect of the genes although it is useful for prediction and selection (Huang and Mackay, 2016). The model was reformulated in terms of breeding values and dominance deviations (Falconer and Mackay, 1996) by Vitezica et al. (2013) after the assumption of a Hardy-Weinberg equilibrium within each:

$$y_i = \mu + \sum_{j=1}^n w_{ij}\alpha_j + \sum_{j=1}^n g_{ij}d_j + e_i$$

where

$$w_{ij} = \begin{cases} (2 - 2p_j) & A_1A_1 \\ (1 - 2p_j) & A_1A_2 \\ -2p_j & A_2A_2 \end{cases} \quad g_{ij} = \begin{cases} -2q_j^2 & A_1A_1 \\ 2p_jq_j & A_1A_2 \\ -2p_j^2 & A_2A_2 \end{cases}$$

and  $\alpha_j = a_j + d_j(q_j - p_j)$  is now the allelic substitution effect and  $p_j$  and  $q_j$  are the allelic frequencies for  $A_1$  and  $A_2$  for the  $j$ th SNP marker. The genetic variance due to a single locus is:

$$\sigma_{G_j}^2 = 2p_jq_j[a_j + d_j(q_j - p_j)]^2 + (2p_jq_jd_j)^2$$

where the additive variance is  $\sigma_{A_j}^2 = 2p_jq_j[a_j + d_j(q_j - p_j)]^2 = 2p_jq_j\alpha_j^2$  and the dominance variance is  $\sigma_{D_j}^2 = (2p_jq_jd_j)^2$  and the multilocus variances, under linkage equilibrium (LE), are  $\sigma_G^2 = \sum_{j=1}^n \sigma_{G_j}^2$ ,  $\sigma_A^2 = \sum_{j=1}^n \sigma_{A_j}^2$  and  $\sigma_D^2 = \sum_{j=1}^n \sigma_{D_j}^2$ . In fact, “biological” (in terms of genotypic additive and dominant values) and “statistical” (in terms of breeding values and dominance deviations) models are equivalent parameterisations of the same model (Vitezica et al., 2013), and the following expressions:

$$\sigma_A^2 = \sum_{j=1}^n (2p_jq_j) \sigma_a^2 + \sum_{j=1}^n (2p_jq_j(q_j - p_j)^2) \sigma_d^2$$

$$\sigma_{A^*}^2 = \sum_{j=1}^n (2p_jq_j) \sigma_a^2$$

$$\sigma_D^2 = \sum_{j=1}^n (4p_j^2q_j^2) \sigma_d^2$$

$$\sigma_{D^*}^2 = \sum_{j=1}^n (2p_jq_j(1 - 2p_jq_j)) \sigma_d^2$$

that can be used to switch variance components estimates between “biological” ( $\sigma_{A^*}^2$  and  $\sigma_{D^*}^2$ ) and “statistical” ( $\sigma_A^2$  and  $\sigma_D^2$ ) models. It can be verified that  $\sigma_A^2 + \sigma_D^2 = \sigma_{A^*}^2 + \sigma_{D^*}^2$ . In addition, if  $p = q = 0.5$ , all variances are identical and if  $d = 0$ ,  $\sigma_A^2 = \sigma_{A^*}^2$ . A further generalization can be also achieved to avoid the requirements of the Hardy-Weinberg equilibrium (Vitezica et al., 2017), by following the NOIA model (Alvarez-Castro and Carlborg, 2007) by replacing  $w_{ij}$  and  $g_{ij}$  with:

$$w_{ij} = \begin{cases} -(-p_{12j} - 2p_{22j}) & A_1A_1 \\ -(1 - p_{12j} - 2p_{22j}) & A_1A_2 \\ -(2 - p_{12j} - 2p_{22j}) & A_2A_2 \end{cases}$$

$$g_{ij} = \begin{cases} -\frac{2p_{12j}p_{22j}}{p_{11j} + p_{22j} - (p_{11j} - p_{22j})^2} & A_1A_1 \\ \frac{4p_{11j}p_{22j}}{p_{11j} + p_{22j} - (p_{11j} - p_{22j})^2} & A_1A_2 \\ -\frac{2p_{11j}p_{12j}}{p_{11j} + p_{22j} - (p_{11j} - p_{22j})^2} & A_2A_2 \end{cases}$$

where,  $p_{11j}$ ,  $p_{12j}$ , and  $p_{22j}$  are the genotypic frequencies for  $A_1A_1$ ,  $A_1A_2$ , and  $A_2A_2$  at the  $j$ th SNP marker, respectively.

Note that all these models require a regularization process for additive and dominance effects. The simplest approach is to expand the RR-BLUP by the assumption of a prior Gaussian distribution for the additive and dominance effects. It is feasible to assume any other kind of prior distribution for the dominance (as described above) and the additive effects (Acevedo et al., 2015). However, a major advantage of using a Gaussian prior distribution is that the model can be easily transformed into Henderson’s Mixed Model equations by using the definition of additive (**G**) and dominance covariance matrices (**D**), as suggested by Vitezica et al. (2013).

Genomic selection models with dominance have been tested in several populations, including dairy cattle (Ertl et al., 2014;

Aliloo et al., 2016; Jiang et al., 2017), pigs (Esfandyari et al., 2016; Xiang et al., 2016), sheep (Moghaddar and van der Werf, 2017), and layers (Heidaritabar et al., 2016) with ambiguous results. Jiang et al. (2017) found a negligible percentage of variation explained by dominance effects for productive life in a Holstein cattle population, although Ertl et al. (2014) suggested that dominance may suppose up to 39% of the total genetic variation for Somatic Cell Score in a population of Fleckvieh cattle. In general, the increase in the accuracy of additive breeding values by including dominance was scarce, with the exception of Aliloo et al. (2016).

## DOMINANCE AND INBREEDING DEPRESSION (OR HETEROSIS)

The classical theory of quantitative genetics (Falconer and Mackay, 1996) postulates that inbreeding depression (or heterosis) occurs due to directional dominance. However, the presence of directional dominance (i.e., a higher percentage of positive than negative dominant effects) is in sharp contrast to the assumptions of the procedures described above that use symmetric prior distributions. This drawback can be overcome by the assumption of a mean of dominant effects that is different from zero, e.g.,  $E(d) = \mu_d$ , as proposed by Xiang et al. (2016). The standard model can be reformulated as:

$$y_i = \mu + \sum_{j=1}^n t_{ij}a_j + \sum_{j=1}^n c_{ij} [d_j^* + \mu_d] + e_i$$

$$= \mu + \sum_{j=1}^n t_{ij}a_j + \sum_{j=1}^n c_{ij}d_j^* + \sum_{j=1}^n c_{ij}\mu_d + e_i$$

where  $d_j^* = d_j - \mu_d$ , then  $E(d^*) = 0$ . It should be noted the term  $\sum_{j=1}^n c_{ij}\mu_d$  is an average of dominance effects for the  $i$ th individual, because  $c_{ij}$  has a value of 1 for heterozygous loci and 0 for homozygous. Inbreeding (or full homozygosity) coefficients  $f_i$  can be calculated as:

$$f_i = 1 - \frac{\sum_{j=1}^n c_{ij}}{n}$$

So,  $\sum_{j=1}^n c_{ij}\mu_d = (1 - f_i) n\mu_d = n\mu_d - f_i n\mu_d$ . The first term  $n\mu_d$  is absorbed in the overall mean of the model ( $\mu$ ), and the second ( $-f_i n\mu_d$ ) corresponds to a covariate  $b = -n\mu_d$  associated with inbreeding ( $f_i$ ). This covariate can be seen as inbreeding depression (if it has a detrimental effect) caused by genomic inbreeding. In addition, it can be also implemented in the GBLUP models described above with the introduction of a covariate within the mixed model equations.

Nonetheless, it assumes that the expected mean of the dominance effects is the same for all markers. In the literature, there are signs that the decrease in performance is associated heterogeneously within the genomic regions (Pryce et al., 2014; Howard et al., 2015; Saura et al., 2015). Models that consider alternative means of dominance effects within genomic regions

may be useful to model inbreeding depression in a more appropriate way.

An alternative approach to explain the phenomenon of inbreeding depression (or heterosis) is the consideration of a possible relationship between additive and dominance biological effects (Wellmann and Bennewitz, 2011). There is theoretical proofs (Caballero and Keightley, 1994) and empirical evidence (Bennewitz and Meuwissen, 2010) that supports this argument. Wellmann and Bennewitz (2012) expanded the “biological” model described above with regularization procedures that allows for this dependence. They defined up to four models (Bayes D0 to D3) based on the Bayes C approach (Verbyla et al., 2009). The last two models (Bayes D2 and D3) included dependencies between genotypic additive and dominance effects. In the first (D2), the dependence was modeled through the prior variance of the dominance effects ( $Var(d|a)$ ) and in the second (D3), they further expanded it to the prior mean ( $E(d|a)$ ), where  $|a|$  is the absolute value of the additive effect. Implementation of these models is extremely complex and they have not been thoroughly tested (Bennewitz et al., 2017).

## IMPRINTING

Another source of non-additive genetic variation is genomic imprinting (Reik and Walter, 2001). This involves total or partial inactivation of paternal and maternal alleles. Following the quantitative model established by Spencer (2002), Nishio and Satoh (2015) put forward two alternative genomic selection models to include imprinting effects. The first extends the “statistical” model with dominance (in terms of breeding values and dominance deviations) as:

$$y_i = \mu + \sum_{j=1}^n w_{ij}\alpha_j + \sum_{j=1}^n g_{ij}d_j + \sum_{j=1}^n r_{ij}i_j + e_i$$

where

$$w_{ij} = \begin{cases} 2 - 2p_j & A_1A_1 \\ 1 - 2p_j & A_1A_2 \\ 1 - 2p_j & A_2A_1 \\ -2p_j & A_2A_2 \end{cases}$$

$$g_{ij} = \begin{cases} -2q_j^2 & A_1A_1 \\ 2p_jq_j & A_1A_2 \\ 2p_jq_j & A_2A_1 \\ -2p_j^2 & A_2A_2 \end{cases} \text{ and } r_{ij} = \begin{cases} 0 & A_1A_1 \\ 1 & A_1A_2 \\ -1 & A_2A_1 \\ 0 & A_2A_2 \end{cases}$$

and  $i_j$  is the imprinting effects associated with  $j$ th marker. The second alternative proposed the distribution of the genetic effects into paternal ( $p_j$ ) and maternal ( $m_j$ ) gametic effects and a dominance deviation.

$$y_i = \mu + \sum_{j=1}^n l_{ij}p_j + \sum_{j=1}^n j_{ij}m_j + \sum_{j=1}^n g_{ij}d_j + e_i$$

where

$$l_{ij} = j_{ij} = \begin{cases} q_j & A_1 \\ -(1 - q_j) & A_2 \end{cases}$$



These models have been implemented in some studies with livestock data: (Hu et al., 2016) did not find an increase in predictive ability when imprinting effects were included in the model. In addition, estimates of the percentage of phenotypic variation caused by imprinting were small and ranged between 1.3 and 1.4% in pigs (Guo et al., 2016) and from 0.2 to 2.1% in dairy cattle (Jiang et al., 2017). However, this latter study reported that imprinting effects supposed more than 20% of the total genetic variance in some reproductive traits, like pregnancy or conception rate.

## EPISTASIS

The last and most complex source of non-additive genetic variation is the epistatic interactions between two or more genes. An immediate approach for genomic evaluation including epistatic interactions is to define an explicit model by including pairwise or higher order epistatic effects:

$$\begin{aligned}
 y_i = & \mu + \sum_{j=1}^n t_{ij}a_j + \sum_{j=1}^n c_{ij}d_j + \sum_{j=1}^n \sum_{k=1}^n t_{ij}t_{ik}aa_{jk} \\
 & + \sum_{j=1}^n \sum_{k=1}^n t_{ij}g_{ik}ad_{jk} + \sum_{j=1}^n \sum_{k=1}^n g_{ij}g_{ik}dd_{jk} \\
 & + \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n t_{ij}t_{ik}t_{il}aaa_{jkl} + \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n t_{ij}t_{ik}g_{il}aad_{jkl} \\
 & + \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n t_{ij}g_{ik}g_{il}add_{jkl} + \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n g_{ij}g_{ik}g_{il}ddd_{jkl} \\
 & + \dots + e_i
 \end{aligned}$$

where  $aa_{jk}$ ,  $ad_{jk}$ , and  $dd_{jk}$  are second order additive x additive, additive x dominant and dominant x dominant epistatic effects between the  $j$ th and  $k$ th SNP effects and  $aaa_{jkl}$ ,  $aad_{jkl}$ ,  $add_{jkl}$  and  $ddd_{jkl}$  are third order additive x additive x additive, additive x additive x dominant, additive x dominant x dominant and dominant x dominant x dominant epistatic effects. Despite the method of regularization used, the number of parameters to estimate is extremely large. Consequently, the computational requirements are enormous and the amount of information available, in the statistical sense, for the estimation of each epistatic effect is very small. Therefore, the most efficient (at least from a computational point of view) method for including epistatic interactions in genomic selection models is to define appropriate covariance matrices between individual effects, in the same way that the standard GBLUP model uses the genomic relationship matrix, but, in this case, taking into account the interactive nature of the genetic effects. There are two main approaches in the published literature: (1) the definition of genomic relationship matrices that consider epistatic interactions (Varona et al., 2014; Martini et al., 2016; Vitezica et al., 2017), and (2) the application of Kernel-based statistical methods (Gianola et al., 2006; de los Campos et al., 2009a; Morota and Gianola, 2014).

This simplest method for defining genomic relationship matrices is the *extended GBLUP model (EGBLUP)*, described by Jiang and Reif (2015) and Martini et al. (2016). These authors start from a reduced version of the “biological” model:

$$y_i = \mu + \sum_{j=1}^n t_{ij}a_j + \sum_{j=1}^n \sum_{k=1}^n t_{ij}t_{ik}aa_{jk} + e_i$$

and they define an equivalent model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{g}_1 + \mathbf{g}_2 + \mathbf{e}$$

where  $\mu$  is the general mean,  $\mathbf{y}$  is the vector of phenotypic data and  $\mathbf{e}$  is the vector of the residuals. In addition, the model includes one “biologically” additive ( $\mathbf{g}_1$ ) and one epistatic ( $\mathbf{g}_2$ ) multivariate Gaussian term with the following distributions:

$$\mathbf{g}_1 \sim N(0, \mathbf{G}_1\sigma_{g1}^2) \quad \mathbf{g}_2 \sim N(0, \mathbf{G}_2\sigma_{g2}^2)$$

Where  $\mathbf{G}_1 = \mathbf{T}\mathbf{T}'$  and  $\mathbf{G}_2 = \mathbf{G}_1 \circ \mathbf{G}_1$  being:

$$\mathbf{T} = \begin{bmatrix} t_{11} & \dots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{k1} & \dots & t_{kn} \end{bmatrix}$$

and the Hadamard product. Moreover,  $n$  is the number of SNP markers and  $k$  the number of individuals. However, with this model the additive and epistatic effects are not orthogonal and dominant effects are not included. Therefore, it can only be used for prediction of the phenotypes and not for the estimation of variance components (Martini et al., 2016). To avoid this inconvenience, Varona et al. (2014) and Vitezica et al. (2017) developed a full orthogonal model. They start with the expansion of the individual genotypic effect into additive, dominance and epistatic effects:

$$\begin{aligned}
 \mathbf{y} = \mathbf{1}\mu + \mathbf{g} + \mathbf{e} = & \mathbf{1}\mu + \mathbf{g}_A + \mathbf{g}_D + \sum_{i=A,D} \sum_{j=A,D} \mathbf{g}_{ij} \\
 & + \sum_{i=A,D} \sum_{j=A,D} \sum_{k=A,D} \mathbf{g}_{ijk} + \dots + \mathbf{e}
 \end{aligned}$$

Where  $\mathbf{g}$  is the vector of the individual genotypic effects,  $\mathbf{g}_A$  is the vector of additive effects,  $\mathbf{g}_D$  the vector of individual dominance effects,  $\mathbf{g}_{ij}$  is the second order epistatic effects,  $\mathbf{g}_{ijk}$  the third order epistatic effects and so on. For simplicity, each individual effect is defined by the sum of SNP (or combination of SNP) effects  $\mathbf{h}$  with equal prior Gaussian variability and weighted by an incidence matrix ( $\mathbf{H}$ ). So, for the additive and dominant effects,  $\mathbf{g}_A = \mathbf{H}_A\mathbf{a}$  and  $\mathbf{g}_D = \mathbf{H}_D\mathbf{d}$ :

$$\mathbf{H}_A = \begin{pmatrix} \mathbf{h}_{A1} \\ \dots \\ \mathbf{h}_{Ak} \end{pmatrix} \quad \text{and} \quad \mathbf{H}_D = \begin{pmatrix} \mathbf{h}_{D1} \\ \dots \\ \mathbf{h}_{Dk} \end{pmatrix}$$

Where each  $\mathbf{h}$  vector is composed by  $n$  (number of SNP markers) elements ( $\mathbf{h}_{Ai} = \{h_{Ai1}, h_{Ai2}, \dots, h_{Ain}\}$ ) and

$\mathbf{h}_{Di} = \{h_{Di1}, h_{Di2}, \dots, h_{Din}\}$  and  $\mathbf{a}$  and  $\mathbf{d}$  are the vectors of the SNP additive and dominant effects. These  $\mathbf{h}_{Ai}$  and  $\mathbf{h}_{Di}$  vectors can be defined in several ways, depending of the reference point or the assumption of the Hardy-Weinberg equilibrium, among others. However, orthogonal partitioning of variances must follow the NOIA approach (Alvarez-Castro and Carlborg, 2007):

$$h_{Aij} = \begin{cases} -(-p_{12j} - 2p_{22j}) A_1 A_1 \\ -(1 - p_{12j} - 2p_{22j}) A_1 A_2 \\ -(2 - p_{12j} - 2p_{22j}) A_2 A_2 \end{cases}$$

$$h_{Dij} = \begin{cases} -\frac{2p_{12j}p_{22j}}{p_{11j} + p_{22j} - (p_{11j} - p_{22j})^2} A_1 A_1 \\ \frac{4p_{11j}p_{22j}}{p_{11j} + p_{22j} - (p_{11j} - p_{22j})^2} A_1 A_2 \\ -\frac{2p_{11j}p_{12j}}{p_{11j} + p_{22j} - (p_{11j} - p_{22j})^2} A_2 A_2 \end{cases}$$

Therefore, and under the assumption that SNP additive or dominant effects follow a Gaussian distribution, the additive and dominant “genomic” (co) variance relationship matrices can be computed as:

$$Cov(\mathbf{g}_A) = \frac{\mathbf{H}_A \mathbf{H}'_A}{tr(\mathbf{H}_A \mathbf{H}'_A)/n} \sigma_A^2 \quad Cov(\mathbf{g}_D) = \frac{\mathbf{H}_D \mathbf{H}'_D}{tr(\mathbf{H}_D \mathbf{H}'_D)/n} \sigma_D^2$$

where the division by traces standardizes the variance components to an ideal infinite “unrelated” population. For second order epistatic effects ( $\mathbf{g}_{AA}$ ,  $\mathbf{g}_{AD}$ , and  $\mathbf{g}_{DD}$ ), Alvarez-Castro and Carlborg (2007) proved that:

$$\mathbf{h}_{AAij} = \mathbf{h}_{Ai} \otimes \mathbf{h}_{Aj} \quad \mathbf{h}_{ADij} = \mathbf{h}_{Ai} \otimes \mathbf{h}_{Dj} \quad \mathbf{h}_{DDij} = \mathbf{h}_{Di} \otimes \mathbf{h}_{Dj}$$

and, as a consequence, the matrices  $\mathbf{H}_{AA}$ ,  $\mathbf{H}_{AD}$  and  $\mathbf{H}_{DD}$  can be written as:

$$\mathbf{H}_{AA} = \begin{pmatrix} \mathbf{h}_{A1} \otimes \mathbf{h}_{A1} \\ \mathbf{h}_{A2} \otimes \mathbf{h}_{A2} \\ \vdots \\ \mathbf{h}_{An} \otimes \mathbf{h}_{An} \end{pmatrix} \quad \mathbf{H}_{AD} = \begin{pmatrix} \mathbf{h}_{A1} \otimes \mathbf{h}_{D1} \\ \mathbf{h}_{A2} \otimes \mathbf{h}_{D2} \\ \vdots \\ \mathbf{h}_{An} \otimes \mathbf{h}_{Dn} \end{pmatrix}$$

$$\mathbf{H}_{DD} = \begin{pmatrix} \mathbf{h}_{D1} \otimes \mathbf{h}_{D1} \\ \mathbf{h}_{D2} \otimes \mathbf{h}_{D2} \\ \vdots \\ \mathbf{h}_{Dn} \otimes \mathbf{h}_{Dn} \end{pmatrix}$$

and, as before, under the assumption of Gaussian distribution of second-order epistatic effects, the covariance between them can be calculated as:

$$Cov(\mathbf{g}_{AA}) = \frac{\mathbf{H}_{AA} \mathbf{H}'_{AA}}{tr(\mathbf{H}_{AA} \mathbf{H}'_{AA})/n} \sigma_{AA}^2 = \mathbf{G}_{AA} \sigma_{AA}^2$$

$$Cov(\mathbf{g}_{AD}) = \frac{\mathbf{H}_{AD} \mathbf{H}'_{AD}}{tr(\mathbf{H}_{AD} \mathbf{H}'_{AD})/n} \sigma_{AD}^2 = \mathbf{G}_{AD} \sigma_{AD}^2$$

$$Cov(\mathbf{g}_{DD}) = \frac{\mathbf{H}_{DD} \mathbf{H}'_{DD}}{tr(\mathbf{H}_{DD} \mathbf{H}'_{DD})/n} \sigma_{DD}^2 = \mathbf{G}_{DD} \sigma_{DD}^2$$

and the covariance between any higher order epistatic effects must be:

$$Cov(\mathbf{g}_{ijk}) = \frac{\mathbf{H}_{ijk} \mathbf{H}'_{ijk}}{tr(\mathbf{H}_{ijk} \mathbf{H}'_{ijk})/n} \sigma_{ijk}^2 = \mathbf{G}_{ijk} \sigma_{ijk}^2$$

However,  $\mathbf{H}$  matrices are extremely large and calculation of  $\mathbf{H}\mathbf{H}'$  cross-products is computationally expensive; each  $\mathbf{H}$  matrix has as many columns as marker interactions and as many rows as individuals. Nevertheless, Vitezica et al. (2017) provided an algebraic shortcut that allows calculation from the additive and dominance matrices, described above, as:

$$Cov(\mathbf{g}_{AA}) = \frac{\mathbf{G}_A \circ \mathbf{G}_A}{tr(\mathbf{G}_A \circ \mathbf{G}_A)/n} \sigma_{AA}^2 = \mathbf{G}_{AA} \sigma_{AA}^2$$

$$Cov(\mathbf{g}_{AD}) = \frac{\mathbf{G}_A \circ \mathbf{G}_D}{tr(\mathbf{G}_A \circ \mathbf{G}_D)/n} \sigma_{AD}^2 = \mathbf{G}_{AD} \sigma_{AD}^2$$

$$Cov(\mathbf{g}_{DD}) = \frac{\mathbf{G}_D \circ \mathbf{G}_D}{tr(\mathbf{G}_D \circ \mathbf{G}_D)/n} \sigma_{DD}^2 = \mathbf{G}_{DD} \sigma_{DD}^2$$

For higher order interactions the results are equivalent. As an example, the covariance matrix for the AAD epistatic interaction can be calculated as:

$$Cov(\mathbf{g}_{AAD}) = \frac{\mathbf{G}_A \circ \mathbf{G}_D \circ \mathbf{G}_D}{tr(\mathbf{G}_A \circ \mathbf{G}_D \circ \mathbf{G}_D)/n} \sigma_{AAD}^2 = \mathbf{G}_{AAD} \sigma_{AAD}^2$$

It should be noted that  $\mathbf{G} \circ \mathbf{G} \dots$  products tend to  $\mathbf{I}$  and higher order epistatic effects tend to be confused with residuals. Nevertheless, this orthogonal approach assumes linkage equilibrium between SNP molecular markers. Linkage disequilibrium (LD) modifies the distribution of the variance into additive, dominance and epistatic components, and orthogonal partition is not possible (Hill and Maki-Tanila, 2015). In outbred populations, substantial LD is present only between polymorphisms in tight linkage (Hill and Maki-Tanila, 2015). However, whilst the distribution of epistatic effects is still unclear (Wei et al., 2015, there is evidence of epistatic interactions between linked loci (Lynch, 1991). Alternative approaches, such as those of Akdemir and Jannick (2015) and Akdemir et al. (2017) have been developed to define locally epistatic relationship matrices. These studies used a RKHS (Reproducing Kernel Hilbert Space) to define these matrices and average them.

The RKHS approach to model epistatic interactions relies on the idea that the relationship between phenotypes and genotypes may not be linear (Gianola et al., 2006; de los Campos et al., 2009a). The main objective is to predict the performance of each individual given its marker genotype through a function that maps the genotypes into phenotypic responses. One of the simplest methods is to consider that this function is linear and, consequently, the results are equivalent to the GBLUP approach. Nevertheless, the power of the Kernel concept relies on the possibility of using alternative functions of marker genotypes. In short, RKHS procedures result in some non-parametric functions  $g(\cdot)$  of a SNP markers set ( $\mathbf{X}$ ):

$$\mathbf{y} = \mu + g(\mathbf{X}) + \mathbf{e}$$

and define a cost function to minimize

$$J = (\mathbf{y} - g(\mathbf{X}))'(\mathbf{y} - g(\mathbf{X})) + \lambda \|g(\mathbf{X})\|_H^2$$

where the term  $\|g(\mathbf{X})\|_H^2$  is a norm under a Hilbert space. Kimeldorf and Wahba (1971) found that  $g(\mathbf{X})$  can be reformulated as:

$$g(\mathbf{X}) = \alpha_0 + \sum_{i=1}^n \alpha_i \mathbf{K}(\mathbf{x} - \mathbf{x}_i)$$

where  $\mathbf{K}$  is a positive semi-definite matrix that meets the requisites of a Kernel Matrix. It defines the similarity between individuals and meets the distance requirements in a Hilbert space (Wootters, 1981). The performance of the method depends on an adequate choice of  $\mathbf{K}$  that can be chosen from among a very large number of options. The easiest RKHS option is to use the genealogical ( $\mathbf{A}$ ) or genomic ( $\mathbf{G}$ ) relationship matrices as kernel matrices (Rodríguez-Ramilo et al., 2014), this leads to the standard BLUP or the GBLUP as particular cases of RKHS. However, they only are able to capture the additive genetic variation and if the model tries to accommodate dominance or epistatic interactions, an alternative Kernel matrix has to be implemented for a pair of SNP vectors of two individuals ( $\mathbf{x}$  and  $\mathbf{x}'$ ). Most kernels proposed so far (Gianola et al., 2006; Piepho, 2009; Morota et al., 2013; Tusell et al., 2014) consider the similarity across individuals within loci (i.e., similarities within loci are summed). Using Taylor series expansions, it can be shown that kernels of this type are a weighted sum of the additive ( $\mathbf{G}$ ) and dominance covariance matrices ( $\mathbf{D}$ ), and therefore implicitly account for dominance (Piepho, 2009). However, these kernels do not consider joint similarity across loci. A kernel that includes epistasis should measure similarities simultaneously between pairs, triplets etc., of loci across individuals, as described in Jiang and Reif (2015) and Martini et al. (2016).

## APPLICATIONS OF GENOMIC SELECTION WITH NON-ADDITIVE GENETIC EFFECTS

### Predictive Performance

The most direct application of the genomic prediction models is to predict the performance of an individual for continuous or categorical phenotypes. Here the introduction of non-additive genetic effects in the procedures of prediction becomes relevant, as the main objective is to predict performance conditioned on the genotype of the individual, despite the additive, dominant or epistatic gene action. In fact, simulation studies show up to 17% more accurate predictions based on the sum of additive and dominance effects compared to prediction based on only additive effects (Wellmann and Bennewitz, 2012; Da et al., 2014). However, the performance of semi-parametric or non-parametric approaches such as RKHS methods seems to be appropriate because they are designed to maximize predicting ability over a given individual and not to predict the future performance of the progeny; they are also designed to capture complex and non-explicit interactions. Moreover, some new research fields have merged with genomic evaluation for

predicting future performance, examples include: microbiomics (Ramayo-Caldas et al., 2016; Yang et al., 2017), metabolomics (Fontanesi, 2016) and precision farming (Banhazi et al., 2012). Over time they will provide a global picture of the genetic and environmental circumstances that affect the future performance of individuals and they will contribute to the development of more accurate prediction models.

### Mate Allocation

In the past, there was a strong belief in “nicking”: pairs of individuals that, wisely selected, would give rise to very efficient offspring (Lush, 1943). In terms of quantitative genetics, the existence of “nicking” would imply that there is large variance of dominant deviations (or epistasis) compared to the variance of breeding values, something that finally turned out to be generally false. Even so, there is room for mate allocation within a population (Toro and Varona, 2010). Under models that include dominance effects, the output of the genomic selection procedure can be used to calculate the prediction of performance of future mating ( $G_{ij}$ ) between the  $i$ th and  $j$ th individual as:

$$E(G_{ij}) = \sum_{k=1}^n \left[ pr_{ijk}(A_1A_1) \hat{a}_j + pr_{ijk}(A_1A_2) \hat{d}_j - pr_{ijk}(A_2A_2) \hat{a}_j \right]$$

where  $pr_{ijk}(A_1A_1)$ ,  $pr_{ijk}(A_1A_2)$ , and  $pr_{ijk}(A_2A_2)$  are the probabilities of the genotypes  $A_1A_1$ ,  $A_1A_2$ , and  $A_2A_2$  for the combination of the  $i$ th and  $j$ th individual and the  $k$ th marker,  $\hat{a}_k$  and  $\hat{d}_k$  are the estimates of the additive and dominance effects for the same marker and  $n$  is the number of markers. Later, optimisation procedures like linear programming (Jansen and Wilton, 1985) or heuristic approximations (simulated annealing, Kirkpatrick et al., 1983) can be used to define a set of mates that maximize performance in the future generation. In a simulated example, Toro and Varona (2010) compared random mating vs. mate selection with a model including dominance and found advantages that ranged between 6 and 22% of the expected response. Sun et al. (2013), Ertl et al. (2014), and Aliloo et al. (2017) have confirmed these improvements with dairy cattle data. However, its implementation in livestock populations is limited because it must be taken into account that the accuracy of the prediction of a potential mate will be low and the advantage will be only relevant when traits have a large amount of non-additive genetic variance. In addition, it requires the genotyping of male and females in the population that is not always available. Moreover, the use of models that include more complex interactions, such as models with epistatic effects or non-parametric approaches, is not so immediate. In fact, the predicted performance of a mate should be calculated after integrating the predictive performance over all possible future genotypic configurations of the expected progeny. For epistasis (but not for dominance) these genotypic configurations also depend on recombination fractions across the genome.

### Selection for Crossbreeding

There is consensus that profit from non-additive genetic effects in a selection program can be obtained when commercial animals



are the product of mating with those that do not participate in the maintenance of a breeding population. The typical way to proceed is to produce two-way or three-way crosses between populations maintained and selected separately (i.e., in pigs). Selection is carried out within lines to benefit from additivity and, in addition, the value of the cross may increase due to the heterosis. Some of the most popular livestock production systems, including pig, poultry, and rabbit production, involve regular crossbreeding schemes, with the aim of capturing the complementarity between the performance of the purebred populations and heterosis. The breeding goal within pure lines is to select individuals to maximize the response in the crossbred population. The traditional approach for this objective was Reciprocal Recurrent Selection—RRS—(Comstock et al., 1949). RRS postulates the selection of individuals in purebred populations based on the performance of their crossbred progeny. If the source of information is the performance of these crossbred progeny, the main drawback of the practical application of RRS is the increase of generation intervals that reduce overall genetic response. In practical terms, the performance of the pure lines is used, and a high genetic purebred/crossbred correlation is sought in order to warrant correct genetic progress (Wei and van der Werf, 1994), however, this may not be the case because of non-additive effects or genotype x environment (G x E) interactions.

The use of genomic information can provide a very useful tool to improve the ability of prediction of breeding values in purebred populations based on crossbred performance without the need to wait for recording crossbred progeny. Ibáñez-Escriche et al. (2009) designed a first approach of the use of GS for crossbred performance under a purely additive model. This study defined a breed specific genomic selection model as:

$$y_i = \mu + \sum_{j=1}^n \left( t_{ijk}^S \alpha_{jk}^S + t_{ijl}^D \alpha_{jl}^D \right) + e_i$$

where  $t_{ijk}^S$  is the SNP allele at the  $j$ th locus from breed  $k$  and received from the sire of the  $i$ th individual that can take values 0 or 1, and  $\alpha_{jk}^S$  is the breed-specific substitution effect for the  $j$ th locus and the  $k$ th breed. Similarly,  $t_{ijl}^D$  and  $\alpha_{jl}^D$  were defined for the alleles received from the dam of the  $i$ th breed. The objective of this approach was to estimate allele substitution effects within breed. Even under the assumption of absence of G x E interactions, SNP allele substitution effects may differ between populations due to: (1) Specific population patterns of linkage disequilibrium with the QTL, or (2) The presence of genotypic dominance effects. The allelic substitution effects of the A (or B) population ( $\alpha_A$  or  $\alpha_B$ ) on performance of A x B depends on the biological additive ( $a$ ) and dominance ( $d$ ) effects, and the allelic frequencies of B-p<sub>B</sub>- (or A-p<sub>A</sub>-) as  $\alpha_A = a + (1 - 2p_B) d$  or  $\alpha_B = a + (1 - 2p_A) d$ . Under dominance, Kinghorn et al. (2010) demonstrated a clear advantage of this approach, assuming the estimation of SNP effects was perfect. This model has been expanded by Sevillano et al. (2017) to a three-way crossbreeding scheme, after the evaluation of a procedure to trace the breed-of-origin of alleles in three-way crossbred animals (Sevillano

et al., 2016). This is an example of the “partial genetic” approach (substitution effects defined within populations). Stuber and Cockerham (1966) showed that gene substitution effects can be defined within populations or across populations, and, if all the (non-additive) effects are accounted for, both approaches are equivalent. Christensen et al. (2015) proposed an alternative model called the “common genetic” approach. Both models were compared by Xiang et al. (2016, 2017) in the same data set with very similar results, but more research is still needed.

Crossbreeding implies mating between individuals of parental populations and a formal description of the additive and dominance variance in the crossbred population is required to evaluate the relevance of mate allocation when the crossbreds are generated. Toosi et al. (2010) and Zeng et al. (2013) extended the aforementioned model to include additive and dominance effects and proved (in both cases with simulated data) its superiority over the strictly additive model if dominance variance is present. These results were confirmed by Esfandyari et al. (2015), who proved that the response to selection for crossbreeding performance is increased by training on crossbred genotypes and phenotypes, and by tracking the allele line origin when pure lines are not closely related. Later, Vitezica et al. (2016) described the substitution effects and dominance deviations within the scope of an F1 population and showed that the additive and dominant variance in a crossbred population is:

$$\begin{aligned} \sigma_{A(A)}^2 &= 2p_A q_A \alpha_A^2 = 2 \left[ p_A q_A a^2 + 2p_A q_A (q_B - p_B) ad \right. \\ &\quad \left. + p_A q_A (q_B - p_B)^2 d^2 \right] \\ \sigma_{A(B)}^2 &= 2p_B q_B \alpha_B^2 = 2 \left[ p_B q_B a^2 + 2p_B q_B (q_A - p_A) ad \right. \\ &\quad \left. + p_B q_B (q_A - p_A)^2 d^2 \right] \\ \sigma_{A(B)}^2 &= 2p_B q_B \left[ a + (q_A - p_A) d \right]^2 \\ \sigma_D^2 &= 4p_A q_A p_B q_B d^2 \end{aligned}$$

where  $\sigma_{A(A)}^2$  and  $\sigma_{A(B)}^2$  are the additive variance generated by the purebred populations A and B, respectively,  $\sigma_D^2$  is the dominance variance,  $p_A$ ,  $q_A$ ,  $p_B$  and  $q_B$  are the allelic frequencies in purebred populations, and  $a$  and  $d$  are the additive and dominance effects.

However, all these approaches assume that the additive and dominance effects have the same magnitude in pure and crossbred populations and this implies an absence of G x E interaction. To avoid this restriction, Vitezica et al. (2016) and Xiang et al. (2016) proposed a multivariate genomic BLUP that is capable of considering different additive and dominance effects and their correlations between pure and crossbred populations.

## Selection in Purebred Populations

The response to selection in purebred populations depends on the magnitude of the additive variance and on the prediction of the additive breeding values for the candidates for reproduction. It is usually assumed that it is not worth selecting individuals with the highest dominance values because they will go back to zero as a result of random mating. However, Toro (1993, 1998)

proposed two mating strategies that can be used to take advantage of dominance in a closed population. The first (Toro, 1993), was a method that basically consists of performing two types of mating: (a) minimum coancestry mating in order to obtain the progenies that will constitute the commercial population and will also be utilized for testing, and (b) maximum coancestry mating from which the breeding population will be maintained. Toro's second strategy (Toro, 1998) advocates the use of the selection of grandparental combinations. Both strategies are analogous with reciprocal-recurrent selection (Comstock et al., 1949) in that they rely on the crucial distinction between commercial and breeding populations. Nevertheless, they have been exclusively tested by simulation and with a reduced set of genes with known additive and dominance effect. Their efficiency has yet to be verified using a large number of SNP markers.

## FINAL REMARKS

Despite huge efforts in the development of statistical models for the implementation of genomic selection with non-additive effects, there are still some issues that have to be dealt with before the use of these models in genomic evaluation becomes standard. A major obstacle is the lack of serious testing as this requires extensive data sets with genotypes and phenotypes, and these data sets are rare. In fact, non-additive genetic variance is expected to be low for most traits (Crow, 2010; Hill et al., 2010), with the exception of fitness related traits. Therefore, the inclusion of non-additive effects in genomic selection models will provide very low (or negligible) improvement in the genetic response or the ability of prediction.

Non-additive effects are easily incorporated into GBLUP procedures (Vitezica et al., 2013, 2017) but efforts must be made to define a single-step approach (Aguilar et al., 2010) that is able to use phenotypic data from non-genotyped individuals and the complete genealogical information of breeding schemes. The major limitation of the GBLUP or single-step approaches is the calculation of the inverse of the genomic relationship matrices ( $\mathbf{G}$ ), the introduction of non-additive effects will involve the calculation of the inverse of additional matrices related with dominance or epistatic effects. Nevertheless, this is really a constraint in populations with a large number of genotyped individual (i.e., Holstein), while most of the livestock populations do not suffer for any limitations. In fact, the computational cost for inverting additive and non-additive genomic relationship matrices is equivalent. On the other hand, using current pedigree-based BLUP models based on dominance (de Boer and Hoeschele, 1993) seems futile because the models are computationally complicated.

Recent studies (Xiang et al., 2016) have shown that inbreeding depression can be modeled and included in GS approaches through a covariate with the average individual heterozygosity.

## REFERENCES

- Acevedo, C. F., de Resende, M. D. V., Silva, F. F., Viana, J. M. S., Valente, M. S. F., Resende, M. F. R., et al. (2015). Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC Genetics* 16:105. doi: 10.1186/s12863-015-0264-2
- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi: 10.3168/jds.2009-2730
- Nevertheless, this approach only considers the effects of the dominance in inbreeding depression and the role of epistatic interactions in inbreeding depression (Minvielle, 1987) has not yet been studied. However, directional dominance is not necessary requisite for having a substantial dominance variance. In fact it would be interesting to know if there are traits with substantial dominance variance and without inbreeding depression, because they would be good candidates for successful strategies of using dominance. In addition, it should be mentioned that the genetic architecture of non-additive genetic effects and its relationship with inbreeding depression and heterosis is a relevant subject of future research.
- The presence of dominance with inbreeding implies the existence of up to five variance components in pedigree-based analysis (Smith and Maki-Tanila, 1990; de Boer and Hoeschele, 1993): additive; dominance between non-inbred; dominance between inbred; covariance between additive; and, inbred dominance values and inbreeding depression. As far as we know, this model has only been used twice with real data in animal breeding (Shaw and Woolliams, 1999; Fernández et al., 2017); their equivalence with the variance components captured by SNP marker effects has to be clarified.
- Finally, the parametric approach for the estimation of epistatic effects (Vitezica et al., 2017) fails when linkage disequilibrium is present. A full description of the effect of the genes and their interactions in populations under linkage disequilibrium and the definition of predictive effects has not been reformulated within the scope of genomic selection. It is unclear what we mean by genetic variances when there is linkage disequilibrium, particularly because linkage disequilibrium is population specific and unstable across generations or subpopulations. Nevertheless, Mäki-Tanila and Hill (2014) showed that when the number of loci increases, epistatic variance disappears. At the same time, the proportion of dominance variance stays the same. Thus, dominance variance is the main non-additive component even with linkage disequilibrium (Hill and Maki-Tanila, 2015).

## AUTHOR CONTRIBUTIONS

LV prepare the initial draft of the review and it was corrected and improved by AL, MT, and ZV. The final manuscript was read and approved by all the authors.

## ACKNOWLEDGMENTS

This work was financed by the INRA SELGEN metaprogram—project EpiSel (ZV, AL), CGL2016-80155 (LV) and CGL2016-75904-C2-2-P (MT) of Ministerio de Economía y Competitividad of Spain.

- Akdemir, D., and Jannick, J. (2015). Locally epistatic genomic relationships matrices for genomic association and prediction. *Genetics* 199, 857–871. doi: 10.1534/genetics.114.173658
- Akdemir, D., Jannick, J., and Isidro-Sanchez, J. (2017). Locally epistatic models for genome-wide prediction and association by importance sampling. *Genet. Sel. Evol.* 49:74. doi: 10.1186/s12711-017-0348-8
- Aliloo, H., Pryce, J. E., González-Recio, O., Cocks, B. G., and Hayes, B. J. (2016). Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. *Genet. Sel. Evol.* 48:186. doi: 10.1186/s12711-016-0186-0
- Aliloo, H., Pryce, J. E., González-Recio, O., Cocks, B. G., Goddard, M. E., and Hayes, B. J. (2017). Including non-additive genetic effects in mating programs to maximize dairy farm profitability. *J. Dairy Sci.* 100, 1203–1222. doi: 10.3168/jds.2016-11261
- Alvarez-Castro, J. M., and Carlborg, O. (2007). A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics* 176, 1151–1167. doi: 10.1534/genetics.106.067348
- Banhazi, T. M., Lehr, H., Black, J. L., Crabtree, H., Schofield, P., Tschärke, M., et al. (2012). Precision livestock farming: an international review of scientific and commercial aspects. *Int. J. Agric. Biol. Eng.* 5, 1–9. doi: 10.3965/ijabe.20120503.001
- Barton, N. H., Etheridge, A. M., and Véber, A. (2017). The infinitesimal model: definition, derivation and implications. *Theor. Pop. Biol.* 118, 50–73. doi: 10.1016/j.tpb.2017.06.001
- Bateson, W. (1909). *Mendel's Principles of Heredity*. Cambridge, UK: Cambridge University Press Warehouse. doi: 10.5962/bhl.title.44575
- Bennewitz, J., and Meuwissen, T. H. E. (2010). The distribution of QTL additive and dominance effects in porcine F2 crosses. *J. Anim. Breed. Genet.* 127, 171–179. doi: 10.1111/j.1439-0388.2009.00847.x
- Bennewitz, J., Edel, C., Fries, R., Meuwissen, T. H. E., and Wellmann, R. (2017). Application of a Bayesian dominance model improves power in quantitative trait genome-wide association analysis. *Genet. Sel. Evol.* 49:7. doi: 10.1186/s12711-017-0284-7
- Bernardo, R. (2010). *Breeding for Quantitative Traits in Plants, 2nd Edn.* Woodbury, MN: Stemma Press.
- Caballero, A., and Keightley, P. D. (1994). A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* 138, 883–900.
- Christensen, O. F., Legarra, A., Lund, M. S., and Su, G. (2015). Genetic evaluation for three-way crossbreeding. *Genet. Sel. Evol.* 47:177. doi: 10.1186/s12711-015-0177-6
- Cockerham, C. C. (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39, 859–882.
- Comstock, R. E., Robinson, H. F., and Harvey, P. H. (1949). A breeding procedure designed to make maximum use of both general and specific combining ability. *Agron. J.* 41, 360–367. doi: 10.2134/agronj1949.00021962004100080006x
- Crow, J. F. (2010). On epistasis: why it is unimportant in polygenic directional selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1241–1244. doi: 10.1098/rstb.2009.0275
- Da, Y., Wang, C., Wang, S., and Hu, G. (2014). Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. *PLoS ONE* 9:e87666. doi: 10.1371/journal.pone.0087666
- Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185, 1021–1031. doi: 10.1534/genetics.110.116855
- de Boer, I., and Hoeschele, I. (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theor. Appl. Genet.* 86, 245–258. doi: 10.1007/BF00222086
- de los Campos, G., Gianola, D., and Rosa, G. J. (2009a). Reproducing kernel Hilbert spaces regression: a general framework for genetic evaluation. *J. Anim. Sci.* 87, 1883–1887. doi: 10.2527/jas.2008-1259
- De los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. (2009b). Prediction quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182, 375–385. doi: 10.1534/genetics.109.101501
- Dekkers, J. C. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82, E313–E328. doi: 10.2527/2004.8213\_supplE313x
- Duenk, P., Calus, M. P. L., Wientjes, Y. C. J., and Bijma, P. (2017). Benefits of dominance over additive models for the estimation of average effects in the presence of dominance. *G3 Genes Genomes Genetics* 7, 3405–3414. doi: 10.1534/g3.117.300113
- Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., et al. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* 95, 4114–4129. doi: 10.3168/jds.2011-5019
- Ertl, J., Legarra, A., Vitezica, Z. G., Varona, L., Edel, C., Emmerling, R., et al. (2014). Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. *Genet. Sel. Evol.* 46:40. doi: 10.1186/1297-9686-46-40
- Esfandyari, H., Bijma, P., Henryon, M., Christensen, O. F., and Sorensen, A. C. (2016). Genomic prediction of crossbred performance based on purebred Landrace and Yorkshire data using a dominance model. *Genet. Sel. Evol.* 48:40. doi: 10.1186/s12711-016-0220-2
- Esfandyari, H., Sorensen, A. C., and Bijma, P. (2015). A crossbred reference population can improve the response to genomic selection for crossbred performance. *Genet. Sel. Evol.* 47:76. doi: 10.1186/s12711-015-0155-z
- Falconer, D. S., and McKay, T. (1996). *Introduction to Quantitative Genetics*. Harlow: Pearson Education Limited.
- Fernández, E. N., Legarra, A., Martínez, R., Sánchez, J. P., and Baselga, M. (2017). Pedigree-based estimation of covariance between dominance deviations and additive genetic effects in closed rabbit lines considering inbreeding and using a computationally simpler equivalent model. *J. Anim. Breed. Genet.* 134, 184–195. doi: 10.1111/jbg.12267
- Fernando, R. L., Dekkers, J. C., and Garrick, D. J. (2014). A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analysis. *Genet. Sel. Evol.* 46:50. doi: 10.1186/1297-9686-46-50
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh* 52, 399–433. doi: 10.1017/S0080456800012163
- Fontanesi, L. (2016). Metabolomics and livestock genomics: insights into a phenotyping frontier and its application in animal breeding. *Anim. Front.* 6, 73–79. doi: 10.2527/af.2016-0011
- Gianola, D. (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 194, 573–596. doi: 10.1534/genetics.113.151753
- Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173, 1761–1776. doi: 10.1534/genetics.105.049510
- Gonzalez-Recio, O., Rosa, G. J. M., and Gianola, D. (2014). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest. Sci.* 166, 217–231. doi: 10.1016/j.livsci.2014.05.036
- Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G., and Chee, M. S. (2005). A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genet.* 37, 549–554. doi: 10.1038/ng1547
- Guo, X., Christensen, O. F., Ostensen, T., Wang, Y., Lund, M. S., and Su, G. (2016). Genomic prediction using models with dominance and imprinting effects for backfat thickness and average daily gain in Danish Duroc pigs. *Genet. Sel. Evol.* 48:67. doi: 10.1186/s12711-016-0245-6
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186
- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics* 28, 476–490.
- Heidaritabar, M., Wolc, A., Arango, J., Zeng, J., Settar, P., Fulton, J. E., et al. (2016). Impact of fitting dominance and additive effects on accuracy of genomic prediction of breeding values in layers. *J. Anim. Breed. Genet.* 133, 334–346. doi: 10.1111/jbg.12225
- Henderson, C. R. (1973). “Sire evaluation and genetic trends,” in *Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush 10-41* (Champaign, IL: ASAS and ADSA).
- Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 73–85. doi: 10.1098/rstb.2009.0203



- Hill, W. G., and Maki-Tanila, A. (2015). Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. *J. Anim. Breed. Genet.* 132, 176–186. doi: 10.1111/jbg.12140
- Hill, W. G., Goddard, M. E., and Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 4:e1000008. doi: 10.1371/journal.pgen.1000008
- Howard, J. T., Haile-Mariam, M., Pryce, J. E., and Maltecca, C. (2015). Investigation of regions impacting inbreeding depression and their association with the additive genetic effect for United States and Australia Jersey dairy cattle. *BMC Genomics* 16:813. doi: 10.1186/s12864-015-2001-7
- Hu, Y., Rosa, G. J. M., and Gianola, D. (2016). Incorporating parent-of-origin effects in whole-genome prediction of complex traits. *Genet. Sel. Evol.* 48:34. doi: 10.1186/s12711-016-0213-1
- Huang, W., and Mackay, T. F. C. (2016). The genetic architecture of quantitative traits cannot be inferred from variance component analysis. *PLoS Genet.* 10:e1006421. doi: 10.1371/journal.pgen.1006421
- Ibáñez-Escriche, N., Fernando, R. L., Toosi, A., and Dekkers, J. C. (2009). Genomic selection of purebreds for crossbred performance. *Genet. Sel. Evol.* 41:12. doi: 10.1186/1297-9686-41-12
- Jansen, G. B., and Wilton, J. W. (1985). Selecting mating pairs with linear programming techniques. *J. Dairy Sci.* 68, 1302–1305. doi: 10.3168/jds.S0022-0302(85)80961-9
- Jiang, J., Shen, B., O'Connell, J. R., VanRaden, P. M., Cole, J. B., and Ma, L. (2017). Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. *BMC Genomics* 18:425. doi: 10.1186/s12864-017-3821-4
- Jiang, Y., and Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics* 201, 759–768. doi: 10.1534/genetics.115.177907
- Kempthorne, O. (1954). The correlation between relatives in a random mating population. *Proc. R. Soc. Lond. B Biol. Sci.* 143, 102–113. doi: 10.1098/rspb.1954.0056
- Kimeldorf, G., and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* 33, 82–95. doi: 10.1016/0022-247X(71)90184-3
- Kinghorn, B. P., Hickey, J. M., and van der Werf, J. H. J. (2010). “Reciprocal recurrent genomic selection for total genetic merit in crossbred individuals,” in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production (Leipzig)*, 36.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* 220, 671–680. doi: 10.1126/science.220.4598.671
- Lush, J. L. (1943). *Animal Breeding Plans, 2nd Edn.* Ames, IA: The Collegiate Press Inc., Iowa.
- Lynch, M. (1991). The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution* 45, 622–629. doi: 10.1111/j.1558-5646.1991.tb04333.x
- Maki-Tanila, A. (2007). An overview on quantitative and genomic tools for utilising dominance genetic variation in improving animal production. *Agric. Food Sci.* 16, 188–198. doi: 10.2137/145960607782219337
- Mäki-Tanila, A., and Hill, W. G. (2014). Influence of gene interaction on complex trait variation with multilocus models. *Genetics* 198, 355–367. doi: 10.1534/genetics.114.165282
- Martini, J. W., Wimmer, V., Erbe, M., and Simianer, H. (2016). Epistasis and covariance: how gene interaction translates into genomic relationship. *Theor. Appl. Genet.* 129, 963–976. doi: 10.1007/s00122-016-2675-5
- Mendel, G. (1866). Versuche über Pflanzen-Hybriden. – Verhandlungen des Naturforschenden Vereines, Abhandlungen, Brünn, 4, 3–47. Editions in different languages published by Matlová (1973). doi: 10.5962/bhl.title.61004
- Meuwissen, T. H., Hayes, B., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Minvielle, F. (1987). Dominance is not necessary for heterosis: a two-locus model. *Genet. Res.* 49, 245–247. doi: 10.1017/S0016672300027142
- Misztal, I., Varona, L., Culbertson, M., Gengler, N., Bertrand, J. K., Mabry, J., et al. (1998). Studies on the value of incorporating the effect of dominance in genetic evaluations of dairy cattle, beef cattle and swine. *Biotechnol. Agron. Soc. Environ.* 2, 227–233.
- Moghaddar, N., and van der Werf, J. H. J. (2017). Genomic estimation of additive and dominance effects and impact of accounting for dominance on accuracy of genomic evaluation in sheep populations. *J. Anim. Breed. Genet.* 134, 453–462. doi: 10.1111/jbg.12287
- Morota, G., and Gianola, D. (2014). Kernel-based whole-genome prediction of complex traits: a review. *Front. Genet.* 5:363. doi: 10.3389/fgene.2014.00363
- Morota, G., Koyama, M., Rosa, G. J. M., Weigel, K. A., and Gianola, D. (2013). Predicting complex traits using a diffusion kernel on genetic markers with an application to dairy cattle and wheat data. *Genet. Sel. Evol.* 45:17. doi: 10.1186/1297-9686-45-17
- Nguyen, T. N., and Nagyné-Kiszlinger, H. (2016). Dominance effects in domestic populations. *Acta Agraria Kaposvariensis* 20, 1–20.
- Nishio, M., and Satoh, M. (2015). Genomic best linear unbiased prediction method including imprinting effects for genomic evaluation. *Genet. Sel. Evol.* 47:32. doi: 10.1186/s12711-015-0091-y
- Piepho, H. P. (2009). Ridge regression and extensions for genomewide selection in maize. *Crop Sci.* 49, 1165–1176. doi: 10.2135/cropsci2008.10.0595
- Pryce, J. E., Haile-Mariam, M., Goddard, M. E., and Hayes, B. J. (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet. Sel. Evol.* 46:71. doi: 10.1186/s12711-014-0071-7
- Ramayo-Caldas, Y., Mach, N., Lepage, P., Levenez, F., Denis, C., Lemonnier, G., et al. (2016). Phylogenetic network analysis applied to pig gut microbiota identifies an ecosystem structure linked with growth traits. *ISME J.* 10, 2973–2977. doi: 10.1038/ismej.2016.77
- Reik, W., and Walter, J. (2001). Genomic imprinting, parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32. doi: 10.1038/35047554
- Rodríguez-Ramilo, S. T., García-Cortés, L. A., and González-Recio, O. (2014). Combining genomic and genealogical information in a reproducing kernel Hilbert spaces regression model for genome-enabled predictions in dairy cattle. *PLoS ONE* 9:e93424. doi: 10.1371/journal.pone.0093424
- Saura, M., Fernández, A., Varona, L., Fernández, A. I., De Cara, M. A. R., Barragán, C., et al. (2015). Detecting inbreeding depression for reproductive traits in Iberian pigs using genome wide data. *Genet. Sel. Evol.* 47:12. doi: 10.1186/s12711-014-0081-5
- Sevillano, C. A., Vandenplas, J., Bastiaansen, J. W. M., and Calus, M. P. L. (2016). Empirical determination of breed-of-origin of alleles in three-way crossbred pigs. *Genet. Sel. Evol.* 48:55. doi: 10.1186/s12711-016-0234-9
- Sevillano, C. A., Vandenplas, J., Bastiaansen, J. W. M., Bergsma, R., and Calus, M. P. L. (2017). Genomic evaluation for a three-way crossbreeding system considering breed-of-origin of alleles. *Genet. Sel. Evol.* 49:75. doi: 10.1186/s12711-017-0350-1
- Shaw, F. H., and Woolliams, J. A. (1999). Variance component analysis of skin and weight data for sheep subjected to rapid inbreeding. *Genet. Sel. Evol.* 31, 43–59. doi: 10.1186/1297-9686-31-1-43
- Smith, S. P., and Maki-Tanila, A. (1990). Genotypic covariance matrices and their inverses for models allowing dominance and inbreeding. *Genet. Sel. Evol.* 22, 65–91. doi: 10.1186/1297-9686-22-1-65
- Spencer, H. G. (2002). The correlation between relatives on the supposition of genomic imprinting. *Genetics* 161, 411–417.
- Stuber, C. W., and Cockerham, C. C. (1966). Gene effects and variances in hybrid populations. *Genetics* 64, 1279–1286
- Su, G., Christensen, O. F., Ostensen, T., Henryon, M., and Lund, M. S. (2012). Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS ONE* 7:e45293. doi: 10.1371/journal.pone.0045293
- Sun, C., VanRaden, P. M., O'Connell, J. R., Weigel, K. A., and Gianola, D. (2013). Mating programs including genomic relationships and dominance effects. *J. Dairy Sci.* 96, 8014–8023. doi: 10.3168/jds.2013-6969
- Toosi, A., Fernando, R. L., and Dekkers, J. C. (2010). Genomic selection in admixed and crossbred populations. *J. Anim. Sci.* 88, 32–46. doi: 10.2527/jas.2009-1975
- Toro, M. A. (1993). A new method aimed at using the dominance variance in closed breeding populations. *Genet. Sel. Evol.* 25, 63–74. doi: 10.1186/1297-9686-25-1-63
- Toro, M. A. (1998). Selection of grandparental combinations as a procedure designed to make use of dominance genetic effects. *Genet. Sel. Evol.* 30, 339–349. doi: 10.1186/1297-9686-30-4-339
- Toro, M. A., and Varona, L. (2010). A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol.* 42:33. doi: 10.1186/1297-9686-42-33
- Tusell, L., Pérez-Rodríguez, P., Forni, S., and Gianola, D. (2014). Model averaging for genome-enabled prediction with reproducing kernel Hilbert spaces: a case



- study with pig litter size and wheat yield. *J. Anim. Breed Genet.* 131, 105–115. doi: 10.1111/jbg.12070
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4123. doi: 10.3168/jds.2007-0980
- Varona, L., Vitezica, Z. G., Munilla, S., and Legarra, A. (2014). “A general approach for calculation of genomic relationship matrices for Epistatic effects,” in *Proceedings from the 10th World Congress on Genetics Applied to Livestock Production* (Vancouver, BC), 11–22.
- Verbyla, K. L., Hayes, B. J., Bowman, P. J., and Goddard, M. E. (2009). Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genet. Res.* 91, 307–311. doi: 10.1017/S0016672309990243
- Vitezica, Z. G., Legarra, A., Toro, M. A., and 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, Z. G., Varona, L., and Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* 195, 1223–1230. doi: 10.1534/genetics.113.155176
- Vitezica, Z. G., Varona, L., Elsen, J. M., Misztal, I., Herring, W., and Legarra, A. (2016). Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genet. Sel. Evol.* 48:6. doi: 10.1186/s12711-016-0185-1
- Wang, X., Yang, Z., and Xu, C. (2015). A comparison of genomic selection methods for breeding value prediction. *Sci. Bull.* 60, 925–935. doi: 10.1007/s11434-015-0791-2
- Wei, M., and van der Werf, J. H. J. (1994). Maximizing genetic response in crossbreds using both purebred and crossbred information. *Anim. Prod.* 59, 401–413. doi: 10.1017/S0003356100007923
- Wei, W. H., Hemani, G., and Haley, C. S. (2015). Detecting epistasis in human complex traits. *Nat. Rev. Genet.* 15, 722–733. doi: 10.1038/nrg3747
- Wellmann, R., and Bennewitz, J. (2011). The contribution of dominance to the understanding of quantitative genetic variation. *Genet. Res.* 92, 139–154. doi: 10.1017/S0016672310000649
- Wellmann, R., and Bennewitz, J. (2012). Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genet. Res.* 94, 21–37. doi: 10.1017/S0016672312000018
- Wootters, W. K. (1981). Statistical distance and Hilbert space. *Phys. Rev. D* 23, 357–363. doi: 10.1103/PhysRevD.23.357
- Wright, S. (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics* 6, 111–123.
- Xiang, T., Christensen, O. F., and Legarra, A. (2017). Technical note: genomic evaluation for crossbred performance in a single-step approach with metafounders. *J. Anim. Sci.* 95, 1472–1480. doi: 10.2527/jas2016.1155
- Xiang, T., Christensen, O. F., Vitezica, Z. G., and Legarra, A. (2016). Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genet. Sel. Evol.* 48:92. doi: 10.1186/s12711-016-0271-4
- Yang, H., Huang, X., Fang, S., He, M., Zhao, Y., Wu, Z., et al. (2017). Unraveling the fecal microbiota and metagenomic functional capacity associated with feed efficiency in pigs. *Front. Microbiol.* 8:1555. doi: 10.3389/fmicb.2017.01555
- Zeng, J., Toosi, A., Fernando, R. L., Dekkers, J. C. M., and Garrick, D. J. (2013). Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genet. Sel. Evol.* 45:11. doi: 10.1186/1297-9686-45-11

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Varona, Legarra, Toro and Vitezica. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.