



A Review of Genomic Models for the Analysis of Livestock Crossbred Data

Joana Stock¹, Jörn Bennewitz¹, Dirk Hinrichs² and Robin Wellmann^{1*}

¹ Department of Animal Breeding and Genetics, Institute of Animal Science, University of Hohenheim, Stuttgart, Germany,

² Department of Animal Breeding, University of Kassel, Witzenhausen, Germany

OPEN ACCESS

Edited by:

Xiao-Lin Wu,
University of Wisconsin-Madison,
United States

Reviewed by:

Fabyano Fonseca Silva,
Universidade Federal de Viçosa, Brazil
Luiz Brito,
Purdue University, United States
Xiangdong Ding,
China Agricultural University, China

*Correspondence:

Robin Wellmann
r.wellmann@uni-hohenheim.de

Specialty section:

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

Received: 19 February 2020

Accepted: 11 May 2020

Published: 26 June 2020

Citation:

Stock J, Bennewitz J, Hinrichs D and
Wellmann R (2020) A Review of
Genomic Models for the Analysis of
Livestock Crossbred Data.
Front. Genet. 11:568.
doi: 10.3389/fgene.2020.00568

Livestock breeding has shifted during the past decade toward genomic selection. For the estimation of breeding values in purebred breeding schemes, genomic best linear unbiased prediction has become the method of choice. Systematic crossbreeding with the aim to utilize heterosis and breed complementary effects is widely used in livestock breeding, especially in pig and poultry breeding. The goal is to improve the performance of the crossbred animals. Due to genotype-by-environment interactions, imperfect linkage disequilibrium, and the existence of dominance and imprinting, purebred and crossbred performances are not perfectly correlated. Hence, more complex genomic models are required for crossbred populations. This study reviews and compares such models. Compared to purebred genomic models, the reviewed models were of much higher complexity due to the inclusion of dominance effects, breed-specific effects, imprinting effects, and the joint evaluation of purebred and crossbred performance data. With the model assessment work conducted until now, it is not possible to come to a clear recommendation as to which existing method is most suitable for a specific breeding program and a specific genetic trait architecture. Since it is expected that a superior method includes all the different genetic effects in a single model, a dominance model with imprinting and breed-specific SNP effects is proposed. Further progress could be made by assuming realistic covariance structures between the genetic effects of the different breeding lines, and by using larger marker panels and mixture distributions for the SNP effects.

Keywords: genomic selection, genomic models, livestock, crossbreeding, heterosis

INTRODUCTION

The crossing of different lines or breeds is widely used in animal breeding with the main aim to produce superior offspring. This superiority results from heterosis and from breed complementary effects. Continuous and discontinuous crossbreeding schemes have been designed and are implemented in various livestock species (Lopez-Villalobos et al., 2000; Samorè and Fontanesi, 2016). In discontinuous schemes, crossbred animals are used solely for production and are not selected as parents of the next generation. Breeding takes place in the parental breeds and the breeding goal is usually to improve crossbred performance. The level of organization in such a system is high and it is sometimes difficult to utilize by-products, such as male offspring of mother lines. These schemes can be predominantly found in livestock species with a high female reproduction rate such as pigs and poultry. In continuous breeding schemes, the female crossbreds are used as parents to breed the next generation. These systems are sometimes implemented in

livestock species with a low female reproduction rate such as cattle. Since there are substantial non-additive effects for reproduction traits in dairy cattle (Jiang et al., 2017), the aims of crossbreeding in dairy cattle are to improve reproduction traits and other functional traits by exploiting heterosis and imprinting and by removing inbreeding depression (Sørensen et al., 2008; Buckley et al., 2014).

A further form of crossbreeding is the upgrading of low-performance breeds with high-yielding breeds. This introgression of genes from high-yielding breeds increases the production level in subsequent generations and reduces inbreeding depression by increasing the genetic diversity of the low-performance breed. This breeding system has frequently been applied to local breeds, such as the German Vorderwald cattle (Hartwig et al., 2014, 2015). However, if upgrading is repeated over several generations, then the breed eventually goes extinct because the native alleles are removed from its gene pool. The formation of a synthetic breed can also be seen as a special form of crossbreeding. A well-known example is the establishment of the so-called Schwarzbuntes Milchrind in the former East Germany (Freyer et al., 2008).

Livestock breeding has shifted toward genomic selection, which is now frequently implemented in large pure breeds. The core of the system that has been implemented in pure breeds is a reference population that consists of genotyped and phenotyped animals. The phenotypes are either the animal's own performance records, or deregressed conventional breeding values. The reference population is needed for the prediction of marker effects. The marker effects are then used for predicting genomic breeding values of the genotyped selection candidates. The reliability of genomic breeding values depends on the size of the reference population, on the effective number of chromosome segments, and on the method used for the prediction of marker effects (Goddard, 2009).

Extensive research has been dedicated to develop statistical models for the prediction of marker effects. These statistical models include the SNP-BLUP model that assumes normally distributed SNP effects, various Bayesian models that assume more heavy-tailed distributions, as well as non-parametric and semi-parametric models (Meuwissen et al., 2001; Gianola, 2013). More complex models assume different SNP variances, depending on the type of control region the SNP belongs to (MacLeod et al., 2016). Some models avoid the prediction of marker effects by building a genomic relationship matrix based on SNP genotypes. The most prominent method based on genomic relationships is GBLUP, which is an equivalent model to SNP-BLUP (VanRaden, 2008; Goddard, 2009). The genotyped selection candidates are included in the model, and their genomic breeding values are calculated by utilizing their genomic relationships with the reference population. GBLUP assumes that all animals are genotyped, which is in general not the case. Therefore, the genomic breeding values are blended in a second step with pedigree-based breeding values to obtain genomically enhanced breeding values on which selection decisions are based. This two-step procedure can be avoided with so-called single-step GBLUP models (ssGBLUP). They were developed as extensions of GBLUP. Single-step models

include genotyped and non-genotyped animals simultaneously (Legarra et al., 2009, 2014; Aguilar et al., 2010; Christensen and Lund, 2010) and assume a particular covariance structure for the breeding values that is computed from genomic and pedigree-based relationships. Fernando et al. (2014) extended the single step model toward non-normally distributed marker effects. In purebred routine application mostly additive effects are considered, with dominance being an integral part of the estimated breeding values. Some genomic models were extended toward accounting for dominance explicitly, but this increased the reliabilities of the breeding values only slightly (Su et al., 2012; Wellmann and Bennewitz, 2012; Azevedo et al., 2015).

To summarize, it seems that in practical purebred genomic evaluations, GBLUP and ssGBLUP have and will become the models of choice, and non-additive gene effects are usually not an issue. The picture is however somewhat different if data from crossbred animals in combination with the parental purebred data is analyzed. The potential applications of genomic models with non-additive genetic effects have been reviewed by Varona et al. (2018). The main breeding goal is in this case to improve the performance of the crossbred animals. Due to genotype-by-environment interaction, imperfect LD, and the existence of dominance, epistasis and imprinting, purebred and crossbred performances (PP and CP, respectively) are not perfectly correlated (e.g., Wei and van der Werf, 1995; Dekkers, 2007; Zumbach et al., 2007; Duenk et al., 2019). Wientjes and Calus (2017) reviewed existing literature about purebred-crossbred correlations in pigs. The average from 201 reported correlation coefficients was 0.63 with 50% of the reported coefficients being between 0.45 and 0.87. The purebred-crossbred correlation affects the optimal design of the reference population (van Grevenhof and van der Werf, 2015) and the choice of an appropriate genomic model.

While genomic models are well-established for pure breeds, much research has been conducted in the recent years to develop genomic models for the analysis of crossbred data. The aim of this study is to review genomic models for the prediction of crossbred performance that were recently developed and were evaluated either using simulated or real crossbred data.

GENOMIC MODELS

Genomic models for crossbred data are extensions of purebred models. The extensions were made in several directions. Most genomic models for the analysis of crossbred data are developed for two-way crosses. A two-way cross \mathcal{X} is created from a sire line \mathcal{A} and a dam line \mathcal{B} , which are usually not inbred. The pure lines have breeding values $a_{\mathcal{A}}$ and $a_{\mathcal{B}}$ for PP, and breeding values $c_{\mathcal{A}}$ and $c_{\mathcal{B}}$ for CP. Typically, some animals are genotyped, whereas others are not. The goal is to obtain accurate predictions of the breeding values for CP by utilizing phenotypic information from genotyped and ungenotyped purebred and crossbred animals. An overview over the considered models is given in **Table 1**.

The SNP alleles are usually assumed to be biallelic, so they may be coded as alleles 1 and 2. Most authors use centered allele content matrices as proposed by VanRaden (2008). The centering

TABLE 1 | Additive and dominance models for the prediction of crossbred performance.

		Data requirements		
		Phenotyped crossbreds	Genotyped crossbreds	
Additive Models	Parental Model	x		
	BSAM/ASGM	x	x	
	Single step	x	(x)	
Dominance Models	Line-independent	(x)	(x)	Provide more accurate breeding values for CP than additive models
	Line-dependent	(x)	(x)	
	Dominance + Imprinting	x	x	

(x) not necessarily needed but can be utilized.

does not affect the predictions, but affects the model-based reliabilities (Strandén and Christensen, 2011). We denote with

$$Z_G^A = G^A - 2P^A$$

the centered allele content matrix for the genotyped animals from line \mathcal{A} , whereby the allele content $G_{im}^A \in \{0, 1, 2\}$ of animal i from line \mathcal{A} is the number of copies of allele 2, animal i has at SNP m , and P_{im}^A is the frequency of allele 2 of SNP m in line \mathcal{A} . Moreover, we denote with

$$Z_{\mathcal{X}}^A = G_{\mathcal{X}}^A - P^A$$

the centered allele origin matrix for alleles from cross \mathcal{X} that originate from line \mathcal{A} . That is, $G_{im}^A \in \{0, 1\}$ is the number of copies of allele 2, crossbred animal i has obtained from sire line \mathcal{A} at SNP m . These matrices are needed to define genetic values of purebred and crossbred animals. The vector with breeding values for CP for animals from line \mathcal{A} has the representation

$$c_{\mathcal{A}} = Z_G^A \alpha^A, \tag{1}$$

where α^A is the vector with allele substitution effects for CP. The vector with breeding values for PP has the representation

$$a_{\mathcal{A}} = Z_G^A \tilde{\alpha}^A, \tag{2}$$

where $\tilde{\alpha}^A$ is the vector with allele substitution effects for PP. The equations for $a_{\mathcal{B}}$ and $c_{\mathcal{B}}$ are similarly.

Most genomic models for two-way crosses utilize, that the vector $a_{\mathcal{X}}$ with additive genetic values of the crossbred animals can be decomposed into a contribution $c_{\mathcal{X}}^A$ that comes from sire line \mathcal{A} , and a contribution $c_{\mathcal{X}}^B$ that comes from dam line \mathcal{B} . That is,

$$a_{\mathcal{X}} = c_{\mathcal{X}}^A + c_{\mathcal{X}}^B, \tag{3}$$

where

$$c_{\mathcal{X}}^A = Z_{\mathcal{X}}^A \alpha^A, \text{ and } c_{\mathcal{X}}^B = Z_{\mathcal{X}}^B \alpha^B. \tag{4}$$

The contribution $c_{\mathcal{X}}^A$ from line \mathcal{A} can be further decomposed into a contribution that comes from the breeding values $c_{\mathcal{A}}$ for CP, and into a vector $m_{\mathcal{X}}^A$ that contains the Mendelian sampling terms of the transmitted gametes (Wei and van der Werf, 1994). That is,

$$c_{\mathcal{X}}^A = 0.5 Z_{\mathcal{X}\mathcal{A}} c_{\mathcal{A}} + m_{\mathcal{X}}^A, \tag{5}$$

where matrix $Z_{\mathcal{X}\mathcal{A}}$ assigns animals from line \mathcal{A} to their crossbred offspring.

Different models have been developed for predicting CP, which can broadly be classified into additive models and dominance models. While some models predict the breeding values for CP directly with Equation (5), others predict the vector α^A with allele substitution effects for CP. In the latter case, the estimated breeding values $\hat{c}_{\mathcal{A}}$ for CP in line \mathcal{A} are obtained by substituting α^A with the prediction $\hat{\alpha}^A$ in Equation (1).

Additive Models

Different additive models have been proposed in the literature. Some models assume that the crossbred animals are genotyped, whereas others do not. The general additive model for a two-way cross is a trivariate model that has two equations for the parental lines, and one equation for the cross. It has the general representation

$$\begin{aligned} y_{\mathcal{A}} &= X_{\mathcal{A}} b_{\mathcal{A}} + Z_{\mathcal{A}} u_{\mathcal{A}} + a_{\mathcal{A}} + E_{\mathcal{A}} \\ y_{\mathcal{B}} &= X_{\mathcal{B}} b_{\mathcal{B}} + Z_{\mathcal{B}} u_{\mathcal{B}} + a_{\mathcal{B}} + E_{\mathcal{B}} \\ y_{\mathcal{X}} &= X_{\mathcal{X}} b_{\mathcal{X}} + Z_{\mathcal{X}} u_{\mathcal{X}} + \dots + E_{\mathcal{X}}, \end{aligned}$$

where $y_{\mathcal{A}}, y_{\mathcal{B}}, y_{\mathcal{X}}$ are vectors with phenotypic records of the respective subpopulation, $b_{\mathcal{A}}, b_{\mathcal{B}}, b_{\mathcal{X}}$ are vectors of fixed effects with design matrices $X_{\mathcal{A}}, X_{\mathcal{B}}, X_{\mathcal{X}}$, and $u_{\mathcal{A}}, u_{\mathcal{B}}, u_{\mathcal{X}}$ are vectors of non-genetic random effects with design matrices $Z_{\mathcal{A}}, Z_{\mathcal{B}}, Z_{\mathcal{X}}$. Finally, $a_{\mathcal{A}}, a_{\mathcal{B}}$ are the breeding values for PP, and $E_{\mathcal{A}}, E_{\mathcal{B}}, E_{\mathcal{X}}$ are the residual terms. The term “...” in the third equation depends on the respective model.

The first two model equations are needed because PP and CP are genetically correlated (Wientjes and Calus, 2017), so phenotypic records of purebred animals increase the reliabilities of the breeding values for CP.

The Parental Additive Model

The parental additive model is based on Equations (2), (3), and (5), and is suitable when the crossbred animals are not genotyped. The model assumes that the Mendelian sampling terms are part of the residuals, so the model equations become

$$\begin{aligned} y_{\mathcal{A}} &= X_{\mathcal{A}} b_{\mathcal{A}} + Z_{\mathcal{A}} u_{\mathcal{A}} + a_{\mathcal{A}} + E_{\mathcal{A}} \\ y_{\mathcal{B}} &= X_{\mathcal{B}} b_{\mathcal{B}} + Z_{\mathcal{B}} u_{\mathcal{B}} + a_{\mathcal{B}} + E_{\mathcal{B}} \\ y_{\mathcal{X}} &= X_{\mathcal{X}} b_{\mathcal{X}} + Z_{\mathcal{X}} u_{\mathcal{X}} + 0.5 Z_{\mathcal{X}\mathcal{A}} c_{\mathcal{A}} + 0.5 Z_{\mathcal{X}\mathcal{B}} c_{\mathcal{B}} + E_{\mathcal{X}}, \end{aligned}$$

where $a_{\mathcal{A}} = Z_G^A \tilde{\alpha}^A$, $a_{\mathcal{B}} = Z_G^B \tilde{\alpha}^B$, $c_{\mathcal{A}} = Z_G^A \alpha^A$, and $c_{\mathcal{B}} = Z_G^B \alpha^B$.

The BSAM and ASGM Model

The model with breed-specific allele effects (BSAM) and the model with breed-independent allele effects, which is also called the across-breed SNP genotype model (ASGM) are based on Equations (2–4), and require that the crossbred animals are genotyped. While the ASGM model predicts one effect per SNP, the BSAM model predicts one effect for the paternal allele, and one for the maternal allele of the crossbred animals. Origin-specific allele effects may occur e.g., due to a different LD pattern between the marker and the QTL, different gene frequencies at the QTL, imprinting effects, or the epistatic effects may be different in the pure breeds. This results in different effects of the marker alleles and thus affects the estimated breeding values.

The first two equations of the BSAM and ASGM model are as above, whereas the third model equation becomes for the BSAM

$$y_{\mathcal{X}} = X_{\mathcal{X}}b_{\mathcal{X}} + Z_{\mathcal{X}}u_{\mathcal{X}} + Z_{\mathcal{X}}^A\alpha^A + Z_{\mathcal{X}}^B\alpha^B + E_{\mathcal{X}}.$$

An equivalent representation for the ASGM model is

$$y_{\mathcal{X}} = X_{\mathcal{X}}b_{\mathcal{X}} + Z_{\mathcal{X}}u_{\mathcal{X}} + a_{\mathcal{X}} + E_{\mathcal{X}}.$$

Ibáñez-Escriche et al. (2009) predicted CP of the parental lines from genotyped crossbred animals with BSAM and ASGM, whereby the breed-specific allele substitution effects of the BSAM model were a priori independent. The allele substitution effects were estimated with BayesB, which is a method that assumes that many of them are actually zero. An oligogene trait was simulated with breed-independent QTL effects. Although the SNP effects are expected to be breed-specific due to differences in LD between markers and QTL, the authors found that the BSAM model outperformed ASGM only if the number of markers was low, the number of records for training was high, and if the parental breeds were distantly related.

Lopes et al. (2017) used the BSAM model with normally distributed SNP effects to predict breeding values for CP from crossbred data, and compared the results with conventional GBLUP. The model provided similar prediction accuracies as conventional GBLUP for the traits litter size and gestation length in pigs. It may be not superior to GBLUP because the allele substitution effects of the different breeds were implicitly assumed to be uncorrelated, which is an assumption that is not likely to be fulfilled.

Sevillano et al. (2019) extended the BSAM and ASGM model toward a three-way cross and distinguished SNP that showed a strong trait association from all remaining SNP. For the trait associated SNP breed-specific effects were estimated, whereas for the remaining SNP one effect was estimated, regardless of the allele origin. This model was compared with the BSAM model and with the ASGM model for the trait daily gain by assuming normally distributed SNP effects. Purebred as well as crossbred data was used for training. The results showed a superiority of their method only if the estimated genetic correlations between PP and CP for the trait associated SNPs and the remaining SNPs were unequal.

Vandenplas et al. (2017) derived equations for predicting the reliability of genomic breeding values for CP for BSAM and ASGM models and assumed normally distributed SNP effects.

The authors found that BSAM outperformed ASGM for a specific parental line, if the effective number of chromosome segments in the crossbred reference animals that originate from the parental line is less than half the effective number of all chromosome segments that are independently segregating.

Additive Single Step Model

While BSAM has the disadvantage that all crossbred animals have to be genotyped, the parental additive model has the disadvantage, that the information provided by the Mendelian sampling terms cannot be utilized for prediction. These problems could be resolved by using a trivariate model of the form

$$\begin{aligned} y_{\mathcal{A}} &= X_{\mathcal{A}}b_{\mathcal{A}} + Z_{\mathcal{A}}u_{\mathcal{A}} + a_{\mathcal{A}} + E_{\mathcal{A}} \\ y_{\mathcal{B}} &= X_{\mathcal{B}}b_{\mathcal{B}} + Z_{\mathcal{B}}u_{\mathcal{B}} + a_{\mathcal{B}} + E_{\mathcal{B}} \\ y_{\mathcal{X}} &= X_{\mathcal{X}}b_{\mathcal{X}} + Z_{\mathcal{X}}u_{\mathcal{X}} + c_{\mathcal{X}}^A + c_{\mathcal{X}}^B + E_{\mathcal{X}} \end{aligned}$$

that includes both, genotyped and phenotyped animals. Christensen et al. (2014) derived the joint covariance matrix $A_{\mathcal{A}}$ of $c_{\mathcal{X}}^A$, $c_{\mathcal{A}}$, and $a_{\mathcal{A}}$ by using the pedigree-based model of Wei and van der Werf (1994) as a starting point. The authors derived the covariance matrix $A_{\mathcal{A}}$ from pedigree relationships, and replaced it in a subsequent step by a covariance matrix $H_{\mathcal{A}}$ that combines pedigree and genotype information.

Xiang et al. (2016a) validated the model of Christensen et al. (2014) in a two-way pig cross for the trait number of piglets born. The authors found that the inclusion of crossbred genomic information improved the model-based reliabilities for CP and reduced to some extent the bias of prediction.

Tusell et al. (2016) used a single-step model for two-way crossbred pigs and the sire line \mathcal{A} , so the model reduced to a bivariate model. The purebred animals were partly genotyped. Since the crossbred animals were not genotyped, the third equation in the model of Christensen et al. (2014) was replaced by a parental additive model equation, i.e., the Mendelian sampling terms were part of the residual. This resulted in a model equation of the form

$$\begin{aligned} y_{\mathcal{A}} &= X_{\mathcal{A}}b_{\mathcal{A}} + Z_{\mathcal{A}}u_{\mathcal{A}} + a_{\mathcal{A}} + E_{\mathcal{A}} \\ y_{\mathcal{X}} &= X_{\mathcal{X}}b_{\mathcal{X}} + Z_{\mathcal{X}}u_{\mathcal{X}} + 0.5 Z_{\mathcal{X}\mathcal{A}}c_{\mathcal{A}} + 0.5 Z_{\mathcal{X}\mathcal{B}}c_{\mathcal{B}} + E_{\mathcal{X}}. \end{aligned}$$

The authors evaluated six growth and meat traits and found that the genetic correlations between purebred and CP were larger than 0.69 for all traits. The accuracies of the genomic breeding values were higher than those obtained from univariate single-step models that took either purebred or CP into account, and also higher than those obtained with pedigree-based models.

Dominance Models

Crossbreeding utilizes heterosis and breed complementarity. A widely accepted hypothesis is that heterosis arises predominantly from dominance effects. An animal carries a dominance effect only if it is heterozygous at a particular QTL. We denote with

$$Z_{im}^{\mathcal{X}} = H^{\mathcal{X}} - \bar{H}^{\mathcal{X}}$$

the centered indicator matrix for heterozygosity. That is, $H_{im}^{\mathcal{X}} \in \{0, 1\}$ equals one, if animal i is heterozygous at SNP m , and $\bar{H}_{im}^{\mathcal{X}}$

is the heterozygosity of SNP m in line \mathcal{X} . The dominance model assumes that the vector $g_{\mathcal{X}}$ with genotypic values of the crossbred animals has the representation

$$g_{\mathcal{X}} = \mu_{\mathcal{X}}1 + Z_G^{\mathcal{X}}a^{\mathcal{X}} + Z_H^{\mathcal{X}}d^{\mathcal{X}}, \tag{6}$$

where $\mu_{\mathcal{X}}$ is the population mean, $a^{\mathcal{X}}$ is the vector with population-dependent additive effects, and $d^{\mathcal{X}}$ is vector with population-dependent dominance effects. The genotypic values of purebred animals are defined accordingly. The trivariate dominance model for a two-way cross and the parental lines has therefore the representation

$$\begin{aligned} y_{\mathcal{A}} &= X_{\mathcal{A}}b_{\mathcal{A}} + Z_{\mathcal{A}}u_{\mathcal{A}} + Z_G^{\mathcal{A}}a^{\mathcal{A}} + Z_H^{\mathcal{A}}d^{\mathcal{A}} + E_{\mathcal{A}} \\ y_{\mathcal{B}} &= X_{\mathcal{B}}b_{\mathcal{B}} + Z_{\mathcal{B}}u_{\mathcal{B}} + Z_G^{\mathcal{B}}a^{\mathcal{B}} + Z_H^{\mathcal{B}}d^{\mathcal{B}} + E_{\mathcal{B}} \\ y_{\mathcal{X}} &= X_{\mathcal{X}}b_{\mathcal{X}} + Z_{\mathcal{X}}u_{\mathcal{X}} + Z_G^{\mathcal{X}}a^{\mathcal{X}} + Z_H^{\mathcal{X}}d^{\mathcal{X}} + E_{\mathcal{X}}, \end{aligned} \tag{7}$$

which we call the dominance model with line-dependent effects. The vector $c_{\mathcal{A}}$ with breeding values for CP from breed \mathcal{A} has the representation given in Equation (1), but the vector with allele substitution effects for CP is

$$\alpha^{\mathcal{A}} = a^{\mathcal{X}} + (1 - 2p^{\mathcal{B}}) \circ d^{\mathcal{X}},$$

where $p^{\mathcal{B}}$ is vector with allele frequencies in the opposite line, and the Hadamard product “ \circ ” is the component-wise product. The breeding values and allele substitution effects for line \mathcal{B} are defined accordingly. Predictions $\hat{a}^{\mathcal{X}}$ and $\hat{d}^{\mathcal{X}}$ of $a^{\mathcal{X}}$ and $d^{\mathcal{X}}$ are needed to get predictions of the allele substitution effects for CP in line \mathcal{A} with equation

$$\hat{\alpha}^{\mathcal{A}} = \hat{a}^{\mathcal{X}} + (1 - 2p^{\mathcal{B}}) \circ \hat{d}^{\mathcal{X}}.$$

Some solvers are unable to account for the fact that $E(d^{\mathcal{X}}) = \mu_d^{\mathcal{X}}1 \neq 0$ for most traits. As shown by Xiang et al. (2016b), one may write $d^{\mathcal{X}} = d_*^{\mathcal{X}} + \mu_d^{\mathcal{X}}1$ such that $E(d_*^{\mathcal{X}}) = 0$. Then, the term $Z_H^{\mathcal{X}}d^{\mathcal{X}}$ in Equation (7) equals $Z_H^{\mathcal{X}}1\mu_d^{\mathcal{X}} + Z_H^{\mathcal{X}}d_*^{\mathcal{X}}$, where $\mu_d^{\mathcal{X}}$ is treated as an additional fixed effect. The same needs to be done for the parental lines. We can write $Z_H^{\mathcal{X}}1\mu_d^{\mathcal{X}} = \mu_d^{\mathcal{X}}M(\hat{h}_{\mathcal{X}} - \bar{h}_{\mathcal{X}}1)$, where M is the number of SNPs, $\hat{h}_{\mathcal{X}}$ is the vector with heterozygosities of the crossbred animals, and $\bar{h}_{\mathcal{X}}$ is the average heterozygosity of the crossbred animals. Hence, the value $-\mu_d^{\mathcal{X}}M$ quantifies the inbreeding depression per unit of genomic inbreeding.

Vitezica et al. (2016) demonstrated how dominance models with normally distributed SNP effects can be transformed into equivalent dominance models with animal effects, whereby different covariance matrices are needed for the additive component and the dominance component of the animal effects. That is, if all SNP effects are normally distributed, then the

SNP effects model can be replaced by the equivalent animal effects model

$$\begin{aligned} y_{\mathcal{A}} &= X_{\mathcal{A}}b_{\mathcal{A}} + Z_{\mathcal{A}}u_{\mathcal{A}} + Z_H^{\mathcal{A}}1\mu_d^{\mathcal{A}} + \tilde{a}_{\mathcal{A}} + \tilde{d}_{\mathcal{A}}^* + E_{\mathcal{A}} \\ y_{\mathcal{B}} &= X_{\mathcal{B}}b_{\mathcal{B}} + Z_{\mathcal{B}}u_{\mathcal{B}} + Z_H^{\mathcal{B}}1\mu_d^{\mathcal{B}} + \tilde{a}_{\mathcal{B}} + \tilde{d}_{\mathcal{B}}^* + E_{\mathcal{B}} \\ y_{\mathcal{X}} &= X_{\mathcal{X}}b_{\mathcal{X}} + Z_{\mathcal{X}}u_{\mathcal{X}} + Z_H^{\mathcal{X}}1\mu_d^{\mathcal{X}} + \tilde{a}_{\mathcal{X}} + \tilde{d}_{\mathcal{X}}^* + E_{\mathcal{X}} \end{aligned}$$

from which the SNP effects can be backsolved. Thereby, the animal effects satisfy $\tilde{a}_{\mathcal{X}} = Z_G^{\mathcal{X}}a^{\mathcal{X}}$, and $\tilde{d}_{\mathcal{X}}^* = Z_H^{\mathcal{X}}d_*^{\mathcal{X}}$, and so on. The joint covariance matrices of the animal effects are given in Christensen et al. (2019).

The SNP effects in Equation (7) were assumed to be line-dependent, which may be the case because the LD between SNP and QTL differs between lines. This may be neglected if the marker panel is sufficiently large. In this case, the SNP effects can assumed to be line-independent, and we obtain the simplified model

$$\begin{aligned} y_{\mathcal{A}} &= X_{\mathcal{A}}b_{\mathcal{A}} + Z_{\mathcal{A}}u_{\mathcal{A}} + Z_G^{\mathcal{A}}a + Z_H^{\mathcal{A}}d + E_{\mathcal{A}} \\ y_{\mathcal{B}} &= X_{\mathcal{B}}b_{\mathcal{B}} + Z_{\mathcal{B}}u_{\mathcal{B}} + Z_G^{\mathcal{B}}a + Z_H^{\mathcal{B}}d + E_{\mathcal{B}} \\ y_{\mathcal{X}} &= X_{\mathcal{X}}b_{\mathcal{X}} + Z_{\mathcal{X}}u_{\mathcal{X}} + Z_G^{\mathcal{X}}a + Z_H^{\mathcal{X}}d + E_{\mathcal{X}}, \end{aligned}$$

which we call the dominance model with line-independent effects.

Vitezica et al. (2013) emphasized that two different parameterizations of the dominance model exist. The first parameterization, which is given by Equation (6), is suitable for two-way crosses, and includes the additive and dominant SNP effects. In contrast, the second parameterization includes the allele substitution effects and the dominance deviations of the SNP. Both parameterizations are equivalent, but their interpretation is different.

Model Evaluation

Zeng et al. (2013) compared a Bayesian dominance model with the corresponding BSAM model and the corresponding ASGM model. A BayesC π type method was used to estimate the marker effects, so the prior assumption was that the SNP effects are either zero, or come from a normal distribution. The comparison was done for a simulated two-way crossbreeding program. A number of 20 generations of selection was simulated with the aim to improve CP in both parental lines. The marker effects were estimated only once in generation one from crossbred animals and used in all subsequent generations. The simulated traits showed a different degree of dominance variance, ranging from “large” to “realistic,” or null. The dominance model was superior to the BSAM model and to the ASGM model. This superiority depended on the fraction of dominance and thus heterosis in the data, but even for situations where no dominance was simulated, the accuracy of the dominance model was similar to the additive model, indicating the robustness of the model. It can tentatively be concluded, that the use of a dominance model is in general advisable, even if dominance is not an important source of trait variability.

Xiang et al. (2016b) used a dominance model with line-dependent effects for a two-way cross and the parental breeds.

The SNP effects were normally distributed, and the additive and dominance effects of the three different populations were correlated. The authors found that the increased predictive ability of the dominance model arose solely from capturing inbreeding depression. This suggests that dominance effects of individual QTL have not been captured. The reason may be that a 60K SNP panel is not sufficient for achieving high LD between markers and QTL, and that the normality assumption is unlikely to be fulfilled.

Esfandyari et al. (2016) compared a Bayesian dominance model with the corresponding Bayesian ASGM model at the example of litter size in a two-way pig cross, whereby BayesC of Habier et al. (2011) was used for prediction. Training was on the parental lines. The prediction accuracies for PP and CP obtained with the dominance model were both higher than those for PP obtained with the ASGM model.

Implications for Breeding Programs

All additive models for predicting CP rely on phenotypic data collected from crossbred animals. This can be problematic in situations where the crossbred animals are not individually identified and thus such data collection pipeline is not implemented. This is likely the case on many farms housing crossbred animals. While additive models require phenotypes from crossbred animals, this is not the case for dominance models because the breeding values for CP can be derived from additive and dominance effects that are predicted in the pure breed, and from the allele frequencies in the opposite breed. Esfandyari et al. (2015a) proposed therefore to use dominance models for selecting purebred animals for CP based on purebred phenotypic and genotypic information only. They did a simulation study and estimated the marker effects with Bayesian LASSO (Park and Casella, 2008; los Campos et al., 2009). The results showed that the gain in CP was higher when the purebreds were selected for CPs, which demonstrated the feasibility of the method even when no crossbred data is available. Moreover, combining several related lines into a single reference population increased the prediction accuracy. However, as shown by Esfandyari et al. (2015b), training on crossbred animals leads to a higher selection response than training on purebred animals. A likely explanation is, that the level of heterozygosity was higher than in the purebred data.

Although genomic selection for CP is a promising strategy to increase selection response for CP in the short and medium term, Esfandyari et al. (2018) found that genomic selection for CP leads eventually to lower CP in the long term than genomic selection on PP. This hold regardless of whether training was on purebred or crossbred animals.

Dominance Model With Imprinting

Dominance effects, as well as additive effects may depend on the breed of origin, which may be due to imprinting or breed complementarity. It could therefore be advantageous to account for imprinting explicitly. A dominance model with imprinting needs to distinguish between the paternal and the maternal allele. If an animal has received allele A_1 from line \mathcal{A} and allele A_2 from line \mathcal{B} , then we denote its genotype as A_1A_2 . The centered

indicator matrix for genotype A_1A_2 is given by

$$W_{\mathcal{X}}^{A_1A_2} = H_{\mathcal{X}}^{A_1A_2} - \bar{H}_{\mathcal{X}}^{A_1A_2},$$

where $H_{\mathcal{X}im}^{A_1A_2} \in \{0, 1\}$ equals one, if animal i from cross \mathcal{X} has genotype A_1A_2 at SNP m , and $\bar{H}_{\mathcal{X}im}^{A_1A_2}$ is the proportion of animals from cross \mathcal{X} that have this genotype at SNP m .

The dominance model with imprinting assumes that the vector $g_{\mathcal{X}}$ with genotypic values of the crossbred animals has the representation

$$g_{\mathcal{X}} = \mu_{\mathcal{X}}1 + (W_{\mathcal{X}}^{21} + W_{\mathcal{X}}^{22})a_{\mathcal{A}}^{\mathcal{X}} + (W_{\mathcal{X}}^{12} + W_{\mathcal{X}}^{22})a_{\mathcal{B}}^{\mathcal{X}} + W_{\mathcal{X}}^{21}d_{\mathcal{A}}^{\mathcal{X}} + W_{\mathcal{X}}^{12}d_{\mathcal{B}}^{\mathcal{X}}, \tag{8}$$

where $\mu_{\mathcal{X}}$ is the population mean, vectors $a_{\mathcal{A}}^{\mathcal{X}}$ and $a_{\mathcal{B}}^{\mathcal{X}}$ contain breed-of-origin dependent additive effects, and vectors $d_{\mathcal{A}}^{\mathcal{X}}$ and $d_{\mathcal{B}}^{\mathcal{X}}$ contain breed-of-origin dependent dominance effects. The model equation for the crossbred animals becomes

$$y_{\mathcal{X}} = X_{\mathcal{X}}b_{\mathcal{X}} + Z_{\mathcal{X}}u_{\mathcal{X}} + (W_{\mathcal{X}}^{21} + W_{\mathcal{X}}^{22})a_{\mathcal{A}}^{\mathcal{X}} + (W_{\mathcal{X}}^{12} + W_{\mathcal{X}}^{22})a_{\mathcal{B}}^{\mathcal{X}} + W_{\mathcal{X}}^{21}d_{\mathcal{A}}^{\mathcal{X}} + W_{\mathcal{X}}^{12}d_{\mathcal{B}}^{\mathcal{X}} + E_{\mathcal{X}}. \tag{9}$$

If imprinting in the parental lines is neglected, then the model equations for the parental lines remain as in Equation (7). The vector with allele substitution effects for CP of line \mathcal{A} is in this case

$$\alpha^{\mathcal{A}} = a_{\mathcal{A}}^{\mathcal{X}} + (1 - p^{\mathcal{B}}) \circ d_{\mathcal{A}}^{\mathcal{X}} - p^{\mathcal{B}} \circ d_{\mathcal{B}}^{\mathcal{X}}, \tag{10}$$

where $p^{\mathcal{B}}$ is the vector with allele frequencies in the opposite line. The proof is given in the **Supplementary Material**. When the SNP effects in the cross do not depend on the breed of origin, then the model simplifies, and becomes identical to the dominance model with line-dependent effects.

Nishio and Satoh (2015) proposed two alternative parameterizations for models with dominance and imprinting and fitted them by assuming normally distributed SNP effects. Their first model includes an additive effect, a dominance effect, and an imprinting effect for the heterozygous genotype, while their second model includes a paternal and a maternal gametic effect, and a dominance effect. The models provided in a simulation study more accurate estimates of genotypic values than GBLUP. While the models of Nishio and Satoh (2015) have the advantage that only 3 effects are needed in the equivalent SNP model for modeling the contribution of each SNP to the genotypic value of an animal, the model in Equation (9) has the advantage that more rigorous prior assumptions can be made for the joint distribution of the effects. That is, if the paternal lines are closely related, then the additive effects $a_{\mathcal{A}}^{\mathcal{X}}$ and $a_{\mathcal{B}}^{\mathcal{X}}$ could assumed to be *a priori* highly correlated, as well as the dominance effects $d_{\mathcal{A}}^{\mathcal{X}}$ and $d_{\mathcal{B}}^{\mathcal{X}}$. However, the parameterization does not allow to predict the vectors $a_{\mathcal{A}}^{\mathcal{X}}$, $a_{\mathcal{B}}^{\mathcal{X}}$, $d_{\mathcal{A}}^{\mathcal{X}}$ and $d_{\mathcal{B}}^{\mathcal{X}}$ individually.

Esfandyari et al. (2015b) compared in a simulation study a Bayesian dominance model with imprinting with the corresponding dominance model with line-independent effects, but used a different parameterization. The model considered

imprinting because it included a separate effect for each phased genotype. Compared to the model proposed above, it has the disadvantage that the effects have no direct interpretation as additive and dominance effects. The genetic effects of the parental breeds were a priori independent. Even though the authors did not simulate imprinting, they found that the dominance model with imprinting was superior, if the reference population was sufficiently large, and if both lines were not closely related. The reason may be that the LD between markers and QTL was different in the cross and in the parental lines, so the additive effects and dominance effects were population-dependent.

DISCUSSION

In this paper, genomic models for the analysis of discontinuous crossbred data were reviewed. Compared to purebred genomic models, the reviewed models were of much higher complexity due to the inclusion of dominance effects, breed-specific effects, imprinting effects, and the use of PP and/or CP data. In the following some additional aspects regarding the distribution of the SNP effects and the model choice are considered.

Distribution of SNP Effects

The normal distribution is the most common assumption about the distribution of SNP effects. Such models have the advantage, that they have an equivalent representation as animal models with genomic covariance matrices for which fast solvers exist, such as DMU (Madsen et al., 2010), WOMBAT (Meyer, 2007), ASReml (Gilmour et al., 2009), blupf90 (Misztal, 1999), or MiX99 (Vuori et al., 2006). Although the assumption of a normal distribution is not likely to be fulfilled when large marker panels are used, the experience with purebred data suggest that the reliabilities of the breeding values are only slightly worse than those obtained with non-normally distributed marker effects. However, the situation in crossbreeding is different because the parental lines are commonly distantly related, and it may be envisaged to evaluate all lines simultaneously in order to increase the reliabilities of the breeding values. This requires that all QTL are in high LD with at least one marker, which implies the necessity to use a large marker panel. However, if the marker panel is large, then only few markers are needed to capture the effect of any QTL. Consequently, the true effects of most markers are actually zero. The model for genomic selection should account for this and assume as a prior distribution for the SNP effects a mixture of two distributions. One component provides the distribution for markers that are in strong LD with a QTL, and the other one is actually zero. In this case, a random-variable γ_m is commonly introduced, which indicates whether the effects of an SNP m are different from zero. Well-known examples are BayesB (Meuwissen et al., 2001), BayesC (Habier et al., 2011), and BayesR (Erbe et al., 2012). Such algorithms are usually implemented with MCMC algorithms, which results in long computation times. However, alternative and faster implementations are available for some models (e.g., Meuwissen, 2009; Shepherd et al., 2010).

For models with additive and dominance effects, an important aspect is, whether these effects are a priori independent or not. It may be advantageous to assume that all effects of a particular SNP m are of the same order of magnitude. This is possible if all effects of a particular SNP m have conditionally on the common covariance matrix $\gamma_m \sigma_m^2 \Sigma$ a normal distribution, where $\sigma_m^2 \sim \text{Inv-}\chi^2(\nu, s)$ and Σ is an appropriately chosen covariance matrix. For the dominance model with line-dependent effects, this means that

$$(a_m^A, a_m^B, a_m^X, d_m^A, d_m^B, d_m^X) | \sigma_m^2, \gamma_m \sim N(0, \gamma_m \sigma_m^2 \Sigma).$$

It can be shown that in this case, all effects of SNP m would have for $\gamma_m = 1$ a t -distribution with ν degrees of freedom, and are for $\gamma_m = 0$ equal to zero. Moreover, the magnitude of the effect size would be similar for all effects of a given SNP m , which reduces the proportion of overdominant SNP. Developing a fast algorithm for such a model is an area for future research.

Model Choice

The most suitable model for a breeding program depends on the achievable accuracies for the breeding values of the selection candidates, and on the available data. Among the additive models, the parental model provided the least accurate predictions for CP, which is because the Mendelian sampling terms are part of the residual and can therefore not be utilized for prediction. It has, however, the advantage that the crossbred animals do not need to be genotyped and may therefore be suitable for animals with low economic value.

The BSAM and ASGM models provided similar results in most cases. The BSAM model, however, needs the trace of the alleles from the purebred parent breed to the crossbred end product, which is a source of potential errors. This might even be more a problem when more complex crossbred structures are involved, e.g., three- or four-way crossbred data. Vandenplas et al. (2016) and Sevillano et al. (2016) developed a statistical pipeline for this purpose and applied it to a three-way crossbred pig data set.

The reviewed papers suggest that the dominance models provide more accurate genomic breeding values for CP than the additive models. Although Xiang et al. (2016b) showed that this gain in accuracy results in the case of normally distributed SNP effects almost solely from capturing inbreeding depression, this may be not the case when large marker panels and appropriate Bayesian models are used for evaluation. Dominance models have the additional advantage that breeding values for crossbred performance can be obtained from purebred animals, so phenotyping and genotyping crossbred individuals may not be necessary. However, as shown by Esfandyari et al. (2015b), the accuracy of the breeding values can be increased when phenotyped and genotyped crossbred individuals are included in the reference population.

Three different dominance models have been applied to crossbred data, which are the dominance model with line-independent effects, the dominance model with line-dependent effects, and the dominance model with imprinting. The

dominance model with line-dependent effects is likely to be inferior to the model with line-independent effects if the SNP effects of the different lines are falsely assumed to be statistically independent, the reference population is small, and the lines are closely related. This could be avoided by specifying a covariance between the SNP effects of the different lines.

When imprinting is relevant, then a dominance model with imprinting is of interest. For example, Jiang et al. (2017) found that there is substantial imprinting for reproduction traits in dairy cattle. The application of imprinting models requires that the crossbred animals are genotyped and that the alleles are traced from the parental lines to the crossbred animals. Unfortunately, to the best of our knowledge, these models are not well-analyzed yet. More research should be done in this area, which includes to analyze all models with common data sets.

CONCLUSION

Genomic models for crossbred data are of much higher complexity than models for purebred data, which results from the inclusion of dominance effects, breed-specific effects, imprinting effects, and from the joint evaluation of PP and CP. Although much research has already been done to develop genomic models for crossbred data, it can be expected that further progress can be made by developing statistical models that include all the different genetic effects in a single model, assume realistic covariance structures between the genetic effects of different breeding lines, use large marker panels, and assume realistic distributions for the SNP effects. The comparisons made in the reviewed papers are not sufficiently comprehensive to come to a clear recommendation as to which existing method is most suitable for a specific breeding program and a specific genetic trait architecture. Some papers suggested

a superiority of dominance models. In the reviewed papers, the focus was on discontinuous crossbreeding schemes. This was because, to our best knowledge, no genomic models have been published that are specifically designed for continuous crossbreeding schemes.

AUTHOR CONTRIBUTIONS

RW did the formal model comparison. JS, RW, and JB wrote the paper. DH contributed to the writing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the German Federal Ministry of Food and Agriculture (BMEL) through the Federal Office for Agriculture and Food (BLE), grant number 2817ERA10D. The project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 696231 - ReDiverse (Biodiversity within and between European Red Dairy Breeds). JS was partly supported by the H. Wilhelm Schaumann Foundation, Hamburg, Germany, which is gratefully acknowledged.

ACKNOWLEDGMENTS

The manuscript has benefitted from the critical and helpful comments of the reviewers.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00568/full#supplementary-material>

REFERENCES

- 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
- Azevedo, C. F., de Resende, M. D. V., E 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 Genet.* 16:105. doi: 10.1186/s12863-015-0264-2
- Buckley, F., Lopez-Villalobos, N., and Heins, B. J. (2014). Crossbreeding: implications for dairy cow fertility and survival. *Anim. Int. J. Anim. Biosci.* 8(Suppl. 1), 122–133. doi: 10.1017/S1751731114000901
- Christensen, O. F., and Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2. doi: 10.1186/1297-9686-42-2
- Christensen, O. F., Madsen, P., Nielsen, B., and Su, G. (2014). Genomic evaluation of both purebred and crossbred performances. *Genet. Sel. Evol.* 46:23. doi: 10.1186/1297-9686-46-23
- Christensen, O. F., Nielsen, B., Su, G., Xiang, T., Madsen, P., Ostensen, T., et al. (2019). A bivariate genomic model with additive, dominance and inbreeding depression effects for sire line and three-way crossbred pigs. *Genet. Sel. Evol.* 51:45. doi: 10.1186/s12711-019-0486-2
- Dekkers, J. C. M. (2007). Marker-assisted selection for commercial crossbred performance. *J. Anim. Sci.* 85, 2104–2114. doi: 10.2527/jas.2006-683
- Duenk, P., Calus, M. P. L., Wientjes, Y. C. J., Breen, V. P., Henshall, J. M., Hawken, R., et al. (2019). Estimating the purebred-crossbred genetic correlation of body weight in broiler chickens with pedigree or genomic relationships. *Genet. Sel. Evol.* 51:6. doi: 10.1186/s12711-019-0447-9
- 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
- Esfandyari, H., Berg, P., and Sørensen, A. C. (2018). Balanced selection on purebred and crossbred performance increases gain in crossbreds. *Genet. Sel. Evol.* 50:8. doi: 10.1186/s12711-018-0379-9
- Esfandyari, H., Bijma, P., Henryon, M., Christensen, O. F., and Sørensen, 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., Sørensen, A. C., and Bijma, P. (2015a). Maximizing crossbred performance through purebred genomic selection. *Genet. Sel. Evol.* 47:16. doi: 10.1186/s12711-015-0099-3
- Esfandyari, H., Sørensen, A. C., and Bijma, P. (2015b). 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

- 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 analyses. *Genet. Sel. Evol.* 46:50. doi: 10.1186/1297-9686-46-50
- Freyer, G., König, S., Fischer, B., Bergfeld, U., and Cassell, B. G. (2008). Invited review: crossbreeding in dairy cattle from a German perspective of the past and today. *J. Dairy Sci.* 91, 3725–3743. doi: 10.3168/jds.2008-1287
- Gianola, D. (2013). Priors in whole-genome regression: the bayesian alphabet returns. *Genetics* 194, 573–596. doi: 10.1534/genetics.113.151753
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., and Thompson, R. (2009). *AS-REML User Guide. Release 3.0. VSN International*. Hemel Hempstead.
- Goddard, M. E. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi: 10.1007/s10709-008-9308-0
- 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
- Hartwig, S., Wellmann, R., Emmerling, R., Hamann, H., and Bennewitz, J. (2015). Short communication: importance of introgression for milk traits in the German vorderwald and hinterwald cattle. *J. Dairy Sci.* 98, 2033–2038. doi: 10.3168/jds.2014-8571
- Hartwig, S., Wellmann, R., Hamann, H., and Bennewitz, J. (2014). The contribution of migrant breeds to the genetic gain of beef traits of German vorderwald and hinterwald cattle. *J. Anim. Breed. Genet.* 131, 496–503. doi: 10.1111/jbg.12099
- Ibáñez-Escriche, N., Fernando, R. L., Toosi, A., and Dekkers, J. C. M. (2009). Genomic selection of purebreds for crossbred performance. *Genet. Sel. Evol.* 41:12. doi: 10.1186/1297-9686-41-12
- 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
- Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi: 10.3168/jds.2009-2061
- Legarra, A., Christensen, O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livest. Sci.* 166, 54–65. doi: 10.1016/j.livsci.2014.04.029
- Lopes, M. S., Bovenhuis, H., Hidalgo, A. M., van Arendonk, J. A. M., Knol, E. F., and Bastiaansen, J. W. M. (2017). Genomic selection for crossbred performance accounting for breed-specific effects. *Genet. Sel. Evol.* 49:51. doi: 10.1186/s12711-017-0328-z
- Lopez-Villalobos, N., Garrick, D. J., Holmes, C. W., Blair, H. T., and Spelman, R. J. (2000). Profitabilities of some mating systems for dairy herds in New Zealand. *J. Dairy Sci.* 83, 144–153. doi: 10.3168/jds.S0022-0302(00)74865-X
- los Campos, G., de Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182, 375–385. doi: 10.1534/genetics.109.101501
- MacLeod, I. M., Bowman, P. J., Vander Jagt, C. J., Haile-Mariam, M., Kemper, K. E., Chamberlain, A. J., et al. (2016). Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genom.* 17:144. doi: 10.1186/s12864-016-2443-6
- Madsen, P., Su, G., Labouriau, R., and Christensen, F. (2010). “DMU—a package for analyzing multivariate mixed models,” in *Proceedings of 9th World Congress on Genetics Applied to Livestock Production* (Leipzig: Gesellschaft für Tierzuchtwissenschaft e.V.), 137
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Meuwissen, T. H. E. (2009). Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41:35. doi: 10.1186/1297-9686-41-35
- Meyer, K. (2007). WOMBAT-A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* 8, 815–821. doi: 10.1631/jzus.2007.B0815
- Misztal, I. (1999). “Complex models, more data: simpler programming?,” in *Interbull Proceedings of the Interbull Workshop Computers and Cattle Breeds* (Tuusula: Interbull Bulletin 20), 33–42.
- 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
- Park, T., and Casella, G. (2008). The bayesian lasso. *J. Am. Stat. Assoc.* 103, 681–686. doi: 10.1198/01621450800000337
- Samorè, A. B., and Fontanesi, L. (2016). Genomic selection in pigs: state of the art and perspectives. *Ital. J. Anim. Sci.* 15, 211–232. doi: 10.1080/1828051X.2016.1172034
- Sevillano, C. A., Bovenhuis, H., and Calus, M. P. L. (2019). Genomic evaluation for a crossbreeding system implementing breed-of-origin for targeted markers. *Front. Genet.* 10:418. doi: 10.3389/fgene.2019.00418
- 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-breed cross pigs. *Genet. Sel. Evol.* 48:55. doi: 10.1186/s12711-016-0234-9
- Shepherd, R. K., Meuwissen, T. H. E., and Woolliams, J. A. (2010). Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. *BMC Bioinformatics* 11:529. doi: 10.1186/1471-2105-11-529
- Sørensen, M. K., Norberg, E., Pedersen, J., and Christensen, L. G. (2008). Invited review: crossbreeding in dairy cattle: a danish perspective. *J. Dairy Sci.* 91, 4116–4128. doi: 10.3168/jds.2008-1273
- Strandén, I., and Christensen, O. F. (2011). Allele coding in genomic evaluation. *Genet. Sel. Evol.* 43:25. doi: 10.1186/1297-9686-43-25
- 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
- Tusell, L., Gilbert, H., Riquet, J., Mercat, M.-J., Legarra, A., and Larzul, C. (2016). Pedigree and genomic evaluation of pigs using a terminal-cross model. *Genet. Sel. Evol.* 48:32. doi: 10.1186/s12711-016-0211-3
- van Grevenhof, I. E. M., and van der Werf, J. H. J. (2015). Design of reference populations for genomic selection in crossbreeding programs. *Genet. Sel. Evol.* 47:14. doi: 10.1186/s12711-015-0104-x
- Vandenplas, J., Calus, M. P. L., Sevillano, C. A., Windig, J. J., and Bastiaansen, J. W. M. (2016). Assigning breed origin to alleles in crossbred animals. *Genet. Sel. Evol.* 48:61. doi: 10.1186/s12711-016-0240-y
- Vandenplas, J., Windig, J. J., and Calus, M. P. L. (2017). Prediction of the reliability of genomic breeding values for crossbred performance. *Genet. Sel. Evol.* 49:43. doi: 10.1186/s12711-017-0318-1
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Varona, L., Legarra, A., Toro, M. A., and Vitezica, Z. G. (2018). Non-additive effects in genomic selection. *Front. Genet.* 9:78. doi: 10.3389/fgene.2018.00078
- 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
- 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
- Vuori, K., Strandén, I., Lidauer, M., and Mäntysaari, E. A. (2006). “MiX99 - effectivesolver for large and complex linear mixed models,” in *Proceedings of 8th World Congress on Genetics Applied to Livestock Production* (Belo Horizonte: Instituto Prociência), 27–33.
- Wei, M., and van der Werf, J. H. (1995). Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. *J. Anim. Sci.* 73, 2220–2226. doi: 10.2527/1995.7382220x
- Wei, M., and van der Werf, J. H. J. (1994). Maximizing genetic response in crossbreds using both purebred and crossbred information. *Anim. Sci.* 59, 401–413. doi: 10.1017/S0003356100007923
- 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/S001667231200018

- Wientjes, Y. C. J., and Calus, M. P. L. (2017). Board invited review: the purebred-crossbred correlation in pigs: a review of theory, estimates, and implications. *J. Anim. Sci.* 95, 3467–3478. doi: 10.2527/jas2017.1669
- Xiang, T., Christensen, O. F., Vitezica, Z. G., and Legarra, A. (2016b). 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
- Xiang, T., Nielsen, B., Su, G., Legarra, A., and Christensen, O. F. (2016a). Application of single-step genomic evaluation for crossbred performance in pig. *J. Anim. Sci.* 94, 936–948. doi: 10.2527/jas.2015-9930
- 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
- Zumbach, B., Misztal, I., Tsuruta, S., Holl, J., Herring, W., and Long, T. (2007). Genetic correlations between two strains of durocs and crossbreds from differing production environments for slaughter traits. *J. Anim. Sci.* 85, 901–908. doi: 10.2527/jas.2006-499

Conflict of Interest: 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 © 2020 Stock, Bennowitz, Hinrichs and Wellmann. 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(s) 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.