Invited Review

Genomic selection in multi-breed dairy cattle populations

John Bruce Cole[1], Marcos Vinicius Gualberto Barbosa da Silva[2]

[1] United States Department of Agriculture, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, Maryland, USA.
[2] Embrapa Gado de Leite, Juiz de Fora, MG, Brazil.

ABSTRACT - Genomic selection has been a valuable tool for increasing the rate of genetic improvement in purebred dairy cattle populations. However, there also are many large populations of crossbred dairy cattle in the world, and multi-breed genomic evaluations may be a valuable tool for improving rates of genetic gain in those populations. Multi-breed models are an extension of single-breed genomic models in which a genomic relationship matrix is used to account for the breed origin of alleles in the population, as well as allele frequency differences between breeds. Most studies have found little benefit from multi-breed evaluations for pure breeds that have large reference populations. However, breeds with small reference populations may benefit from inclusion in a multi-breed evaluation without adversely affecting evaluations for purebred performance. Most research has been conducted in taurine breeds, so additional research is needed to determine the value of multi-breed reference populations for composite and synthetic breeds that include both indicine and taurine cattle adapted to tropical climates.

Key Words: crossbred cattle, crossbreeding, genomic evaluation, multi-breed populations

Introduction

Genomic selection was rapidly adopted in purebred cattle populations (Hayes et al., 2009b; Wiggans et al., 2011), but there are many populations that include crossbred animals which can contribute to genetic progress, including crossbred bulls used in New Zealand (Harris and Johnson, 2010; Olson et al., 2012). There also are many breeds with small reference populations that could potentially be combined with similar breeds from other countries (e.g., Pryce et al., 2011). When the number of individuals in a reference population is limited, genomic breeding values are not accurately estimated, and predictions made in one breed do not perform well when applied to other breeds (Hayes et al., 2009a; Olson et al., 2012). Lund et al. (2014) recently provided a comprehensive review of genomic selection in multiple breed populations. Research conducted to date has focused on several problems, including use of information from one breed to improve predictions in one or more other breeds (Boerner et al., 2014; Porto-Neto et al., 2015), use of data from crossbreds to improve purebred predictions

(Grevenhof and Werf, 2015), use of data from purebreds to improve crossbred predictions (Esfandyari et al., 2015a), use of data from crossbreds to improve crossbred predictions (Esfandyari et al., 2015b), and use of genomic information to estimate genomic predicted transmitting ability (PTA) for both purebred and crossbred animals (Christensen et al., 2014; Hozé et al., 2014). The objective of this review is to describe how genomic selection can be applied to dairy cattle populations that include purebred and crossbred animals.

Principles underlying multi-breed genomic selection

The fundamental principle underlying genomic selection is that linkage disequilibrium (LD) can be exploited to track causal variants using a large number of DNA markers (Nejati-Javaremi et al., 1997; Meuwissen et al., 2001; Dekkers, 2007), and phenotypes measured in one population may be used to accurately compute breeding values in other populations (de los Campos et al., 2013). Genomic relationships (e.g., Hayes et al., 2009b) and Mendelian sampling also (Cole and VanRaden, 2011) can be predicted using DNA marker data. However, most causal variants are unknown, and prediction accuracy in genomic selection programs is driven largely by the size of the reference population (Goddard, 2008) and the linkage disequilibrium (LD) between markers and quantitative trait loci (QTL) (Daetwyler et al., 2012). The

size of reference populations can be limited for smaller breeds because there may be few animals with traditional evaluations that have high reliabilities, as well as due to financial constraints. The use of reference populations that include animals from multiple breeds help improve accuracies by using information from large populations to improve predictions for small ones. Veroneze (2015) also noted that genomic selection of purebreds for crossbred performance is expected to be successful because crossbred animals are typically closely related to their pure line ancestors.

The consistency of the LD phase is very important for genomic selection, as well as the amount of LD in the population. The accuracy of across- and multi-breed genomic prediction is influenced by the consistency of the LD phase between populations. Characterizing LD patterns in different pig populations, Veroneze (2015) verified that an inconsistent LD phase can explain why a marker associated with an important effect in one population may not be effective for selection in another.

Linkage disequilibrium between markers and causal variants may arise in several ways, including *de novo* mutation, genetic drift, strong selection pressure, and differential use of animals as parents. When breeds are closely related, composite reference populations are useful because the LD relationships are similar. For example, Zhou et al. (2013) found that the use of a combined Nordic and Chinese reference population resulted in substantial gains in reliability. However, when breeds are distantly related or unrelated, the linkage relationships differ, and the ability of a predictor population of one breed to predict genetic merit of animals in another is very limited (e.g., Olson et al., 2012).This is due to the decay in LD due to within-breed selection following breed divergence, as well as crossbreeding. In a study using pure-breed and multi-breed beef cattle data, Kizilkaya et al. (2010) found that LD was retained over greater genomic distance in pure-breed than in multi-breed populations, and that some markers were less informative when used in another breed. They also showed that QTL can account for a substantial amount of within-breed variation when the number of QTL is small, but predictive ability decreases as the number of QTL increases. This is consistent with expectations under an infinitesimal model for quantitative traits, which Cole et al. (2009) validated using dairy traits, and differing LD between breeds.

The use of high-density single nucleotide polymorphism (SNP) platforms with hundreds of thousands of markers has been proposed as a possible solution for problems due to breed-specific LD and ascertainment bias

on SNP arrays (de Roos et al., 2008). Erbe et al. (2012) reported small gains in accuracy when using a high-density panel with a combined Holstein and Jersey reference panel, but they found that a panel constructed using only SNP from transcribed genes was more promising. In breeding programs, causal mutations might be important and could aid genomic selection and genetic gain (Snelling et al., 2012). According to Fortes et al. (2014), causal mutations have an advantage in comparison to random SNP: they are not dependent on LD, and so they can be used for selection over generations, across breeds, and in breeds that were not in the reference population.

Theory of multi-breed genomic selection

The principal difference between single-breed and multiple-breed genomic prediction is in the relationship matrix used to relate SNP effects to phenotypes. The following discussion is based on the derivation of a genomic best linear unbiased prediction (GBLUP) model presented in Harris and Johnson (2010), who used three breed groupings, New Zealand Holsteins, foreign Holsteins, and Jerseys, and some intermediate steps have been omitted for brevity. The model does not include a polygenic effect to account for additive genetic variance not explained by the markers. This material can be extended to an arbitrary number of breeds without loss of generality. While the discussion focuses on GBLUP, there are several other approaches that can be used to estimate genomic breeding values, including the single-step approach (Misztal et al., 2009; Liu et al., 2014) and various Bayesian methods (Gianola et al., 2006; Moser et al., 2009; Habier et al., 2011).

The linear mixed model:

$$y = Xb + Zu + e \qquad [1]$$

relates phenotype y to fixed effects b and random SNP effects u through design matrices X and Z, respectively. The residual variance e has a diagonal variance matrix R. The *m*-th column of Z corresponds to genotypes for the *m*-th SNP, which are coded −1, 0, and 1 for homozygotes, heterozygotes, and other homozygotes, respectively. The vector of breeding values, a = Zu, is assumed to equal the sum over all SNP effects. If we assume that fixed effects (Xb) are known, then a may be estimated as:

$$\hat{a} = (R^{-1} + \sigma_u^{-2}(ZZ')^{-1})^{-1} R^{-1}(y - Xb) \qquad [2]$$

and all SNP are assumed to have equal variances.

The matrix ZZ' has an expectation of

$$E(ZZ') = 11' \sum_m (p_m - q_m)^2 + A \sum_m 2p_m q_m \qquad [3]$$

in which p$_m$ is the frequency of the *m*-th allele, q$_m$ = 1 − p$_m$; A is the average relationship matrix; and 1 are *m*-by-*1* vectors of 1s (Habier et al., 2007). The diagonal elements of ZZ′ are counts of the number of homozygous loci for each individual, and the off-diagonals measure the number of alleles shared by two individuals. A genomic relationship matrix can be estimated from [3] using the regression method of VanRaden (2008):

$$ZZ' = b_1 11' + b_2 A + E \qquad [4]$$

in which E includes differences between the true and expected fractions of DNA shared by individuals, as well as measurement error due to the use of a subset of SNP in place of the full DNA sequence. While E(E) = 0, E would probably be non-zero in practice even if full sequence data were used because relatives do not share exactly the same fraction of DNA as expected. (Note that [3] was rewritten to emphasize its similarity to [4].) The regression model in [4] does not require estimates of founder population allele frequencies. The expectations of the regression coefficients, *b₁* and *b₂*, are:

$$E(b_1) = \sum_m (p_m - q_m)^2$$

$$E(b_2) = \sum_m 2 p_m q_m \qquad [5]$$

The matrix A in [4] is the expected value of the genomic relationship matrix, G, and can be estimated by:

$$G = \frac{ZZ' - \widehat{b_1} 11'}{\widehat{b_2}} \qquad [6]$$

See VanRaden (2008) and VanRaden et al. (2011) for additional discussion related to estimation of G. When ZZ' from [6] is substituted into [2] and the regression coefficients are replaced with their expectations from [5] we obtain the estimated breeding value based on the SNP effects:

$$\hat{a} = (R^{-1} + \sigma_a^{-2} G^{-1})^{-1} R^{-1}(y - Xb) \qquad [7]$$

in which

$$\sigma_a^2 = 2 \sum_m p_m q_m \sigma_u^2 \qquad [8]$$

In practice, the markers do not account for all of the additive genetic variance associated with a trait, and a random polygenic term commonly is added to models to account for this (Calus and Veerkamp, 2007).

The genomic relationship matrix estimated in [6] can now be generalized to the case of a multi-breed population. In order to properly account for breed-specific allele frequencies, equation [6] must be replaced with a multiple regression equation that accounts for different expected means and variances. If we let λ$_{ik}$

denote the fraction of breed *k* in individual *i*, then, in the two-breed case, we can sum over breeds *k* and *l*. This results in the generalization of equation [6] to:

$$ZZ' = \sum_{k \le l} b_{1(kl)} J_{kl} + \sum_{k \le l} b_{1(kl)} K_{kl} + E \qquad [9]$$

in which J$_{(kl)}$ and K$_{(kl)}$ are the covariates for the intercepts and regressions associated with the (co)variances among relatives, respectively. The diagonals in K differ from those in the relationship matrix of a purebred population because they are now partitioned into breed fractions to account for different variances and allele frequencies among breeds. The generalization of equation [4] shown above in equation [9] accounts for breed-specific founder population allele frequencies. The expectations of *b₁* and *b₂* in equation [9] are analogous to those in equation [5]. If *p$_{km}$* is the allele frequency of marker *m* in breed *k*:

$$E(b_{1(kl)}) = \sum_m (p_{km} - q_{km})(p_{lm} - q_{lm})$$

$$E(b_{2(kl)}) = \begin{cases} \sum_m 0.5(p_{km} - p_{lm})^2, & k < l \\ \sum_m 2 p_{km} q_{km}, & k = l \end{cases} \qquad [10]$$

When *k* is equal to *l*, the expectations are identical to the purebred case. This emphasizes that the new information being added in the multi-breed case is information about (co)variances among the breeds. If two breeds are closely related in a phylogenetic sense and have not been separated for very long, then the (co)variances will be close to 0 and the expectations of *b₁* and *b₂* in the multi-breed case will be very similar to those in the single-breed case. When breeds have a low relationship to one another, then the expectations in the multiple-breed case may differ substantially from those in the single-breed cases, and information from one breed may contribute little or no information about the other.

Finally, omitting several intermediate steps discussed in Harris and Johnson (2010), the multi-breed genomic relationship matrix can be written as:

$$G = L_1 \hat{F}_1^{-1} \left(ZZ' - \sum_{k \le l} b_{1(kl)} J_{(kl)} \right)_1 \hat{F}_1'^{-1} L_1' \qquad [11]$$

The L and F matrices are derived from Cholesky factorization of A$_1$, the submatrix of the pedigree relationship matrix that includes the genotyped animals in the population. The resulting G has the same general form as in equation [6], with ZZ' being adjusted to account for the intercepts and regressions associated with the (co)variances among relatives.

As mentioned above, methodologies other than GBLUP can be used to predict genomic estimated breeding

values (GEBV). Of particular interest is the single-step GBLUP method (ssGBLUP; Misztal et al., 2009; Aguilar et al., 2010), which permits the simultaneous estimation of SNP effects and GEBV. In the single-step model, the relationship matrix, A, is replaced with a matrix, H, which includes both pedigree-based relationships and differences between pedigree-based and genomic-based relationships, and many different forms of the relationship matrix can be accommodated (Forni et al., 2011). However, the extension of two-step GBLUP models to ssGBLUP has not been straightforward. Poultry populations that include two distinct lines have been successfully evaluated using ssGBLUP and appropriately scaled G (Simeone et al., 2012), and Carillier et al. (2014) applied a similar strategy to French dairy goats, but neither analysis included crossbred individuals. Misztal et al. (2013) demonstrated that unknown-parent groups can be used in ssGBLUP to accommodate crossbred animals because unknown parents of different breeds are assigned to different groups. Christensen et al. (2014) extended the model of Wei and van der Werf (1994), which models two purebred lines and their F1 crosses to ssGBLUP using partial relationship matrices.

Applications of multi-breed genomic selection

A number of authors have reported one aspect or another of genomic selection in multi-breed populations. Hayes et al. (2009a) used single-breed and multiple-breed reference populations to predict breeding values for purebred Holsteins and Jerseys and found that the agreement of realized with expected reliabilities was lower with crossbred than purebred predictor sets under GBLUP; predictions of opposite-breed genomic predicted transmitting ability (GPTA) had accuracies near 0; the G matrix for multi-breed populations must be scaled to achieve appropriate expected accuracies; and Bayesian approaches produce higher accuracies for some traits, particularly when a large QTL is segregating (e.g., *DGAT1*). These results were generally consistent with those of Ibáñez-Escriche et al. (2009), who simulated 6,000 SNP and 30 QTL in four breeds and reported that SNP effects depend on their line of origin when purebred lines are not closely related, as is the case with Gir and Holstein, for instance. They concluded that breed-specific models rarely out-perform across-breed models, and models with breed-specific allele effects may not be necessary when many SNP are used because high-density panels are likely to have markers in tight LD with the causal variants.

Kizilkaya et al. (2010) performed a study that combined real genotypes with simulated phenotypes, reporting results which support the earlier findings that multiple-breed training sets produce higher correlations of predicted genetic merit with phenotype, but also reported that purebred training sets produce higher correlations of predicted with actual genetic merit. A simulation study by Toosi et al. (2010) suggested that the use of multi-breed reference populations could produce better evaluations for crossbred cattle without adversely affecting the purebred cattle in the evaluation. Harris and Johnson (2010) showed that use of a multi-breed genomic evaluation rather than a single-breed produced similar reliabilities for proven bulls, and slightly higher reliabilities for young bulls, and Harris et al. (2011) subsequently reported that increasing SNP density in a reference population including purebred and crossbred animals improved prediction accuracy of one pure breed from another, but not crossbred animals.

Pryce et al. (2011) compared single- and multi-breed reference populations for Fleckvieh, Holstein, and Jersey cattle, concluding that gains were minimal when using a multi-breed reference population. However, a multi-breed reference population produces better results than a single-breed reference population when predicting breeding values for a breed with no genotyped individuals in the reference population. This is in agreement with the results of Olson et al. (2012), who compared three methods of multi-breed evaluation against single-breed results. They found that application of single-breed SNP effects estimates in other breeds performed quite poorly; estimates from a multi-breed population performed better than parent average but more poorly than single-breed estimates; and a multiple trait model in which breeds were treated as traits performed similarly to single-breed estimates. Weber et al. (2012) also reported that accuracies were lower when using single-breed beef cattle reference populations, but that gains in accuracy were highest for the breeds best represented in the multi-breed training population.

More recently, Makgahlela et al. (2013) concluded that models which include breed-specific effects were not notably better than models that ignored breed effects. In a follow-up study, there was no increase in prediction accuracy when the breed origin of alleles in an admixed population of Nordic cattle was accounted for in the model (Makgahlela et al., 2014). Hozé et al. (2014) recently showed that multi-breed genomic evaluation can be very helpful for breeds with small (<500 animals) reference populations, but that the gains decrease as the size of the reference population increases.

VanRaden and Cooper (2015) showed that genomic evaluations for crossbreds can be computed by taking a weighted sum of purebred marker effects based on the breed composition of the admixed animals. Recent work by Wientjes et al. (2015) has found that some of the discrepancy between results predicted in simulation and those achieved in real populations may be attributable to trait architecture. While Cole et al. (2009) showed that most traits of interest in dairy cattle improvement programs are well-described by an infinitesimal model in which all markers have small effects, Hayes et al. (2010) reported that prediction accuracies are higher for traits with some large-effect loci if the model accounts for those effects. Tiezzi and Maltecca (2015) also showed that rescaling G to account for realized marker variance increases prediction accuracy. If models are not capable of reflecting the architecture of traits, for example, by constraining all marker effects to have equal variance, then there will be discrepancies between expectations based on simulation and actual performance.

Recently, Christensen et al. (2014) extended the two-breed model of Wei and van der Werf (1994) to include genomic information, using partial relationship matrices to combine pedigree and marker information. Their results showed promising results, and could be readily applied to a population such as the Girolando breed because the method allows information from crossbred animals to be incorporated in a coherent manner for such a crossbreeding system, but additional validation of their approach is desirable.

Strandén and Mäntysaari (2013) used a random regression model to include breed composition information and genetic variances of origin breeds in multi-breed analyses as a computationally tractable approximation for use in admixed populations. Using the same data, they reported a correlation of 0.987 with results of García-Cortés and Toro (2006). This approach may be more computationally appealing in large genotyped populations, but their model focused only on prediction of purebred performance in admixed populations, not prediction of crossbred performance. The model would need refinement for use in populations composed almost entirely of crossbred individuals.

While early results suggested that higher SNP density might be a solution to challenges with across-breed evaluation because they depend on similar LD among SNP and QTL in training and test data, high-density data have not improved predictions in multi-breed populations (Ibáñez-Escriche et al., 2009; Olson et al., 2012; Makgahlela et al., 2013). Whole genome sequencing is rapidly dropping in price and will support the discovery of many causal variants, and replacing markers with causal variants should increase the accuracy of genomic prediction. This will eliminate many issues related to different patterns of LD in different populations, but variants may differ across breeds.

Discussions and Conclusions

Ibáñez-Escriche et al. (2009) discussed four reasons to favor genomic selection for crossbred performance over more conventional breeding programs that incorporate purebred and crossbred animals:

1) Genomic selection does not require pedigree information on crossbreds;

2) Once SNP effects have been estimated, prediction can continue for several generations without the need to collect additional phenotypes (Meuwissen et al., 2001);

3) Rates of inbreeding can be better managed under genomic selection (Daetwyler et al., 2007); and

4) It is easier to accommodate non-additive gene action in a genomic selection program (Dekkers, 2007).

Each of these points is relevant to dairy production in Brazil, particularly with respect to the Girolando breed. The dairy sector plays a significant part in Brazilian agribusiness, with approximately 80% of the milk produced by crossbred cows, mostly Girolando cows. While many farmers do record pedigree information, there are many who do not, and with genomics the performance data for animals with no pedigree still can be used because relationships can be estimated using genotypes.

With some breeds that have only a small number of animals on test, as in the Girolando breed, there are not enough new data to support frequent computation of genetic evaluations and updating of SNP effects, so the ability to potentially use allele substitution effects for several years allows the breeder to maximize the value of genotyping. However, it is not possible, as thought early in the development of genomic selection, to use marker effects predicted in a population that is distant in time from the population where they are being applied because prediction accuracy decreases as the LD between the true QTL and the markers decays (e.g., Goddard, 2008). The use of the method of VanRaden and Cooper (2015) for computing genomic breed composition might be useful in such populations as well, perhaps with genomic breed composition included as a fixed effect in the genomic prediction model.

Daetwyler et al. (2007) showed that inbreeding should be reduced under genomic selection, but this has not been observed in practice (e.g., Cole and VanRaden, 2011), although new tools for managing inbreeding under genomic selection are now available (Pryce et al., 2012). Some work

has been done recently on the use of non-additive effects in genomic selection (Sun et al., 2013; Da et al., 2014), but it is still a challenge to obtain good estimates of non-additive genetic (co)variances.

While comparisons of single- to multi-breed genomic evaluations in dairy cattle have consistently shown that there were no substantial gains in prediction accuracy for most populations, there are gains for animals with small reference populations when combined with large reference populations from other breeds. However, most studies in dairy cattle have focused on the use of multi-breed animals to predict the performance of pure-breed animals, rather than using all available information to predict the performance of composite animals, and additional research is needed to extend the utility of genomic selection to those populations. Key points that need to be better understood include the value of extensive pedigree data on multi-breed as well as pure-breed animals, and the effectiveness of imputation in multi-breed populations. Most of this research has been conducted in taurine breeds, not in indicine or taurine-indicine crossbred populations, so additional research is needed to determine the value of multi-breed reference populations for use in mixed-breed cattle adapted to tropical climates.

Large multi-breed populations can implement genomic predictions with their own data; smaller multi-breed populations may include information from larger purebred populations; and haplotype-based models may be more helpful than SNP-based models because the effective number of chromosome segments is estimated to be fairly small (e.g., Misztal, 2015). Bastiaansen et al. (2014) recently used haplotypes to determine the genetic background of SNP based on the parent (breed) of origin. In so doing, they were able to identify interactions among SNP and the environment, as well as among SNP and their genetic background. The use of haplotypes, rather than individual SNP, appears likely to provide more computationally tractable, and accurate predictions with genetic merit that will remain stable over time.

Acknowledgments

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 score1. Journal of Dairy Science 93:743-752. doi:10.3168/jds.2009-2730.

Bastiaansen, J. W. M.; Bovenhuis, H.; Lopes, M. S.; Silva, F. F.; Megens, H. J. and Calus, M. P. L. 2014. SNP Effects depend on genetic and environmental context. p.17-22. In: Proceedings of the 10th World Congress on Genetics Applied to Livestock Production, Vancouver.

Boerner, V.; Johnston, D. J. and Tier, B. 2014. Accuracies of genomically estimated breeding values from pure-breed and across-breed predictions in Australian beef cattle. Genetics, Selection, Evolution 46:61. doi:10.1186/s12711-014-0061-9.

Calus, M. P. and Veerkamp, R. F. 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. Journal of Animal Breeding and Genetics 124:362-368.

Carillier, C.; Larroque, H. and Robert-Granié, C. 2014. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. Genetics, Selection, Evolution 46:67. doi:10.1186/s12711-014-0067-3.

Christensen, O. F.; Madsen, P.; Nielsen, B. and Su, G. 2014. Genomic evaluation of both purebred and crossbred performances. Genetics, Selection, Evolution 46:23. doi:10.1186/1297-9686-46-23.

Cole, J. B. and VanRaden, P. M. 2011. Use of haplotypes to estimate Mendelian sampling effects and selection limits. Journal of Animal Breeding and Genetics 128:446-455. doi:10.1111/j.1439-0388.2011.00922.x.

Cole, J. B.; VanRaden, P. M.; O'Connell, J. R.; Van Tassell, C. P.; Sonstegard, T. S.; Schnabel, R. D.; Taylor, J. F. and Wiggans, G. R. 2009. Distribution and location of genetic effects for dairy traits. Journal of Dairy Science 92:2931-2946. doi:10.3168/jds.2008-1762.

Daetwyler, H. D.; Kemper, K. E.; van der Werf, J. H. J. and Hayes, B. J. 2012. Components of the accuracy of genomic prediction in a multi-breed sheep population. Journal of Animal Science 90:3375-3384. doi:10.2527/jas.2011-4557.

Daetwyler, H. D.; Villanueva, B.; Bijma, P. and Woolliams, J. A. 2007. Inbreeding in genome-wide selection. Journal of Animal Breeding and Genetics 124:369-376.

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.

de los Campos, G.; Vazquez, A. I.; Fernando, R.; Klimentidis, Y. C. and Sorensen, D. 2013. Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS Genetics 9:e1003608. doi:10.1371/journal.pgen.1003608.

de Roos, A. P. W.; Hayes, B. J.; Spelman, R. J. and Goddard, M. E. 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus Cattle. Genetics 179:1503-1512. doi:10.1534/genetics.107.084301.

Dekkers, J. C. M. 2007. Marker-assisted selection for commercial crossbred performance. Journal of Animal Science 85:2104-2114. doi:10.2527/jas.2006-683.

Erbe, M.; Hayes, B. J.; Matukumalli, L. K.; Goswami, S.; Bowman, P. J.; Reich, C. M.; Mason, B. A. and Goddard, M. E. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. Journal of Dairy Science 95:4114-4129. doi:10.3168/jds.2011-5019.

Esfandyari, H.; Sørensen, A. C. and Bijma, P. 2015a. Maximizing crossbred performance through purebred genomic selection. Genetics, Selection, Evolution 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. Genetics, Selection, Evolution 47:76. doi:10.1186/s12711-015-0155-z.

Forni, S.; Aguilar, I. and Misztal, I. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genetics, Selection, Evolution 43:1. doi:10.1186/1297-9686-43-1.

Fortes, M. R. S.; Suhaimi, A. H. M. S.; Porto-Neto, L. R.; McWilliam, S. M.; Flatscher-Bader, T.; Moore, S. S.; D'Occhio, M. J.; Meira, C. T.; Thomas, M. G.; Snelling, W. M.; Reverter, A. and Lehnert, S. A. 2014. Post-partum anoestrus in tropical beef cattle: A systems approach combining gene expression and genome-wide association results. Livestock Science 166:158-166. doi:10.1016/j.livsci.2014.06.017.

García-Cortés, L. A. and Toro, M. A. 2006. Multibreed analysis by splitting the breeding values. Genetics, Selection, Evolution 38:601-615.

Gianola, D.; Fernando, R. L. and Stella, A. 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173:1761-1776.

Goddard, M. 2008. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245-257. doi:10.1007/s10709-008-9308-0.

Grevenhof, I.E. van, and van der Werf, J. H. 2015. Design of reference populations for genomic selection in crossbreeding programs. Genetics, Selection, Evolution 47:14. doi:10.1186/s12711-015-0104-x.

Habier, D.; Fernando, R. L. and Dekkers, J. C. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389-2397.

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.

Harris, B. L.; Creagh, F. E.; Winkelman, A. M. and Johnson, D. L. 2011. Experiences with the Illumina high density bovine BeadChip. Interbull Bulletin 44:3-7.

Harris, B. L. and Johnson, D. L. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. Journal of Dairy Science 93:1243-1252. doi:10.3168/jds.2009-2619.

Hayes, B. J.; Bowman, P. J.; Chamberlain, A. C.; Verbyla, K. and Goddard, M. E. 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics, Selection, Evolution 41:51. doi:10.1186/1297-9686-41-51.

Hayes, B. J.; Bowman, P. J.; Chamberlain, A. J. and Goddard, M. E. 2009b. Invited review: Genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science 92:433-443.

Hayes, B. J.; Pryce, J.; Chamberlain, A. J.; Bowman, P. J. and Goddard, M. E. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genetics 6:e1001139. doi:10.1371/journal.pgen.1001139.

Hozé, C.; Fritz, S.; Phocas, F.; Boichard, D.; Ducrocq, V. and Croiseau, P. 2014. Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. Journal of Dairy Science 97:3918-3929. doi:10.3168/jds.2013-7761.

Ibáñez-Escriche, N.; Fernando, R. L.; Toosi, A. and Dekkers, J. C. M. 2009. Genomic selection of purebreds for crossbred performance. Genetics, Selection, Evolution 41:1-10.

Kizilkaya, K.; Fernando, R. L and Garrick, D. J. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. Journal of Animal Science 88:544-551. doi:10.2527/jas.2009-2064.

Liu, Z.; Goddard, M. E.; Reinhardt, F. and Reents, R. 2014. A single-step genomic model with direct estimation of marker effects. Journal of Dairy Science doi:10.3168/jds.2014-7924.

Lund, M. S.; Su, G.; Janss, L.; Guldbrandtsen, B. and Brøndum, R. F. 2014. Genomic evaluation of cattle in a multi-breed context. Livestock Science 166:101-110. doi:10.1016/j.livsci.2014.05.008.

Makgahlela, M. L.; Mäntysaari, E. A.; Strandén, I.; Koivula, M.; Nielsen, U. S.; Sillanpää, M. J. and Juga, J. 2013. Across breed multi-trait random regression genomic predictions in the Nordic Red dairy cattle. Journal of Animal Breeding and Genetics 130:10-19. doi:10.1111/j.1439-0388.2012.01017.x.

Makgahlela, M. L.; Strandén, I.; Nielsen, U. S.; Sillanpää, M. J. and Mäntysaari, E. A. 2014. Using the unified relationship matrix adjusted by breed-wise allele frequencies in genomic evaluation of a multibreed population. Journal of Dairy Science 97:1117-1127. doi:10.3168/jds.2013-7167.

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.

Misztal, I. 2015. Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics. doi:10.1534/genetics.115.182089.

Misztal, I.; Legarra, A. and Aguilar, I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. Journal of Dairy Science 92:4648-4655.

Misztal, I.; Vitezica, Z. G.; Legarra, A.; Aguilar, I. and Swan, A. A. 2013. Unknown-parent groups in single-step genomic evaluation. Journal of Animal Breeding and Genetics 130:252-258. doi:10.1111/jbg.12025.

Moser, G.; Tier, B.; Crump, R. E.; Khatkar, M. S. and Raadsma, H. W. 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genetics, Selection, Evolution 41:56.

Nejati-Javaremi, A.; Smith, C. and Gibson, J. P. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. Journal of Animal Science 75:1738-1745.

Olson, K. M.; VanRaden, P. M. and Tooker, M. E. 2012. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. Journal of Dairy Science 95:5378-5383. doi:10.3168/jds.2011-5006.

Porto-Neto, L. R.; Barendse, W.; Henshall, J. M.; McWilliam, S. M.; Lehnert, S. A. and Reverter, A. 2015. Genomic correlation: harnessing the benefit of combining two unrelated populations for genomic selection. Genetics, Selection, Evolution 47:84. doi:10.1186/s12711-015-0162-0.

Pryce, J. E.; Gredler, B.; Bolormaa, S.; Bowman, P. J.; Egger-Danner, C.; Fuerst, C.; Emmerling, R.; Sölkner, J.; Goddard, M. E. and Hayes, B. J. 2011. Short communication: Genomic selection using a multi-breed, across-country reference population. Journal of Dairy Science 94:2625-2630. doi:10.3168/jds.2010-3719.

Pryce, J. E.; Hayes, B. J. and Goddard, M. E. 2012. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. Journal of Dairy Science 95:377-388. doi:10.3168/jds.2011-4254.

Simeone, R.; Misztal, I.; Aguilar, I. and Vitezica, Z. G. 2012. Evaluation of a multi-line broiler chicken population using a single-step genomic evaluation procedure. Journal of Animal Breeding and Genetics 129:3-10. doi:10.1111/j.1439-0388.2011.00939.x.

Snelling, W. M.; Cushman, R. A.; Fortes, M. R. S.; Reverter, A.; Bennett, G. L.; Keele, J. W.; Kuehn, L. A.; McDaneld, T. G.; Thallman, R. M. and Thomas, M. G. 2012. Physiology and Endocrinology Symposium: How single nucleotide polymorphism chips will advance our knowledge of factors controlling puberty and aid in selecting replacement beef females. Journal of Animal Science 90:1152-1165. doi:10.2527/jas.2011-4581.

Strandén, I. and Mäntysaari, E. A. 2013. Use of random regression model as an alternative for multibreed relationship matrix. Journal of Animal Breeding and Genetics 130:4-9. doi:10.1111/jbg.12014.

Sun, C.; VanRaden, P. M.; O'Connell, J. R.; Weigel, K. A. and Gianola, D. 2013. Mating programs including genomic

relationships and dominance effects. Journal of Dairy Science 96:8014-8023. doi:10.3168/jds.2013-6969.

Tiezzi, F. and Maltecca, C. 2015. Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. Genetics, Selection, Evolution 47:24. doi:10.1186/s12711-015-0100-1.

Toosi, A.; Fernando, R. L. and Dekkers, J. C. M. 2010. Genomic selection in admixed and crossbred populations. Journal of Animal Science 88:32-46.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science 91:4414-4423.

VanRaden, P. M. and Cooper, T. A. 2015. Genomic evaluations and breed composition for crossbred U.S. dairy cattle. Interbull Bulletin 49:19-23.

VanRaden, P. M.; Olson, K. M.; Wiggans, G. R.; Cole, J. B. and Tooker, M. E. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. Journal of Dairy Science 94:5673-5682. doi:10.3168/jds.2011-4500.

Veroneze, R. 2015. Linkage disequilibrium and genomic selection in pigs. Thesis (Ph.D.). Wageningen University, Wageningen, The Netherlands.

Weber, K. L.; Thallman, R. M.; Keele, J. W.; Snelling, W. M.; Bennett, G. L.; Smith, T. P. L.; McDaneld, T. G.; Allan, M. F.; Van Eenennaam, A. L. and Kuehn, L. A. 2012. Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. Journal of Animal Science 90:4177-4190. doi:10.2527/jas.2011-4586.

Wei, M. and van der Werf, J. H. J. 1994. Maximizing genetic response in crossbreds using both purebred and crossbred information. Animal Production 59:401-413.

Wientjes, Y. C.; Calus, M. P.; Goddard, M. E. and Hayes, B. J. 2015. Impact of QTL properties on the accuracy of multi-breed genomic prediction. Genetics, Selection, Evolution 47:42. doi:10.1186/s12711-015-0124-6.

Wiggans, G. R.; VanRaden, P. M. and Cooper, T. A. 2011. The genomic evaluation system in the United States: Past, present, future. Journal of Dairy Science 94:3202-3211. doi:10.3168/jds.2010-3866.

Zhou, L.; Ding, X.; Zhang, Q.; Wang, Y.; Lund, M. S. and Su, G. 2013. Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. Genetics, Selection, Evolution 45:7. doi:10.1186/1297-9686-45-7.