Including crossbred pigs in the genomic relationship matrix through utilization of both linkage disequilibrium and linkage analysis¹

M. W. Iversen,^{*†2} Ø. Nordbø,^{*‡} E. Gjerlaug-Enger,^{* E. Grindflek,^{* M.S. Lopes,^{§#} and T. H. E. Meuwissen[†]}}

^{*}Topigs Norsvin, Storhamargata 44, 2317 Hamar, Norway; [†]Norwegian University of Life

Sciences, Postboks 5003 NMBU, 1432 Ås, Norway; [‡]GENO SA, Storhamargata 44, 2317 Hamar, Norway;

[§]Topigs Norsvin Research Center, Beuningen 6641 SZ, the Netherlands; [#]Topigs Norsvin, Curitiba 80420-210, Brazil

ABSTRACT: In pig breeding, the final product is a crossbred (CB) animal, while selection is performed at the purebred (PB) level using mainly PB data. However, incorporating CB data in genetic evaluations is expected to result in greater genetic progress at the CB level. Currently, there is no optimal way to include CB genotypes into the genomic relationship matrix. This is because, in single-step genomic BLUP, which is the most commonly used method, genomic and pedigree relationships must refer to the same base. This may not be the case when several breeds and CB are included. An alternative to overcome this issue may be to use a genomic relationship matrix (G matrix) that accounts for both linkage disequilibrium (LD) and linkage analysis (LA), called G_{LDLA} . The objectives of this study were to further develop the G_{LDLA} matrix approach to utilize both PB and CB genotypes simultaneously, to investigate its performance, and the general added value of including CB genotypes in genomic evaluations. Data were available on Dutch Landrace, Large White, and the F1 cross of those breeds. In total, 7 different G matrix compositions (PB

alone, PB together, each PB with the CB, all genotypes across breeds, and G_{LDLA}) were tested on 3 maternal traits: total number born (TNB), live born (LB), and gestation length (GL). Results show that G_{LDLA} gave the greatest prediction accuracy of all the relationship matrices tested for PB prediction, but not for CB prediction. Including CB genotypes in general increased prediction accuracy for all breeds. However, in some cases, these increases in prediction accuracy were not significant (at $P < 0.05$). To conclude, CB genotypes increased prediction accuracy for some of the traits and breeds, but not for all. The G_{LDLA} matrix had significantly greater prediction accuracy in PB than the other G matrix with both PB and CB genotypes, except in one case. While for CB, the G matrix with genotypes across all breeds gave the greatest accuracy, though this was not significantly different from G_{LDLA} . Computation time was high for G_{LDLA} , and research will be needed to reduce its computational costs to make it feasible for use in routine evaluations. The main conclusion is that inclusion of CB genotypes is beneficial for both PB and CB animals.

Key words: crossbreds, genomic relationship matrix, linkage analysis, linkage disequilibrium, maternal traits, pigs

© 2017 American Society of Animal Science. All rights reserved. J. Anim. Sci. 2017.95:5197–5207
doi:10.2527/jas2017.1705

INTRODUCTION

In pig breeding, the final product is the crossbred (CB) animal. However, selection is performed at the

purebred (PB) level using mainly PB data, which may result in a reduced genetic progress at the CB level (Dekkers, 2007; Toosi et al., 2010; Bloemhof et al., 2012; Bijma and Bastiaansen, 2014; Esfandyari et al., 2015). Hence, the inclusion of CB data in the breeding value estimation of PB is expected to improve genetic progress of CB animals. Currently, there is no optimal way to include CB genotypes in the genomic relationship matrix (Misztal et al., 2014) or in single-step GBLUP (ssGBLUP; Christensen et al., 2014). In ssGBLUP, an **H** matrix (combination of pedigree-based

¹This study was partly financed by Topigs Norsvin and the Research Council of Norway through project no. 244434/I10. There are no known conflicts of interests.

²Corresponding author: maja.iversen@norsvin.no

Received May 8, 2017.

Accepted September 22, 2017.

and genomic-based relationships) is used, assuming that both relationships refer to the same base (Legarra et al., 2015). However, this assumption does not hold when several breeds and CB are included. Using breed-specific allele frequencies from genomic information may alleviate this problem, but this is not possible in ssGBLUP (Loureiro et al., 2016). An alternative option is to use the $\mathbf{G}_{\text{LDLA}}$ relationship matrix, which utilizes both linkage disequilibrium (**LD**) and linkage analysis (**LA**), making use of genotypes, genotype probabilities, and pedigree relationships (Meuwissen et al., 2015). Genetic groups can be accounted for so that the base animals of different breeds can be entered as alternative genetic groups, and thus use allele frequencies according to breed rather than across all animals. This would accommodate CB because they are linked to the PB through the pedigree. The $\mathbf{G}_{\text{LDLA}}$ matrix has shown promising results analyzing PB data (Meuwissen et al., 2015), but has not yet been applied to CB data. Therefore, the aim of this study was to further develop the $\mathbf{G}_{\text{LDLA}}$ matrix approach to combine PB and CB genotypes simultaneously, to investigate its performance, and the general added value of including CB genotypes in genomic evaluations.

MATERIALS AND METHODS

Care and Use of Animals

Data recording and sample collection were conducted strictly in line with the laws given by Dutch animal research authorities on the protection of animals (Gezondheids-en welzijnswet voor dieren). The data were obtained as part of routine data recording in commercial breeding programs. Samples collected for DNA extraction were only used for the routine diagnostic purpose of the breeding program.

Animals and Data

Data were available on 2 PB populations (Dutch Landrace and Large White) and their F1 cross, hereafter referred to as **A**, **B**, and **X**, respectively. The traits evaluated in this study were total number born (**TNB**, sum of alive and dead piglets), live born (**LB**, number of piglets born alive), and gestation length (**GL**, number of days between insemination and farrowing). Phenotypic observations were available on 11,491 sows, and genotypes were available on 8,350 animals. Both males and females were genotyped in the PB, but in the F1 population, only females were genotyped. All animals were genotyped using the Illumina Porcine SNP60 Beadchip (Illumina Inc., San Diego, CA). Quality control consisted of excluding SNP with GenCall < 0.15, call rate <

0.95, minor allele frequency < 0.01, and strong deviations from Hardy-Weinberg equilibrium ($\chi^2 > 600$). The SNP located on sex chromosomes and unmapped SNP were also excluded. Positions of the SNP were based on the Sscrofa10.2 assembly of the reference genome (Groenen et al., 2012). All genotyped animals had a frequency of missing genotypes above the threshold of 0.05 for excluding poorly genotyped animals. After quality control, SNP not segregating in all breeds were excluded, leaving 36,778 SNP common to all breeds for further analysis. An overview of phenotypic and genotypic data can be found in Tables 1 and 2.

Statistical Analysis

The Linkage Disequilibrium Multilocus Iterative Peeling (**LDMIP**) program (Meuwissen and Goddard, 2010) was used to get genotype probabilities for the genotyped animals and their ungenotyped ancestors. It is a method for imputation of phase and missing genotypes, and sets up the linkage analysis part of the identity by descent (**IBD**) matrix (Meuwissen and Goddard, 2010). There is an option in LDMIP to run both with (genetic) groups and without groups. Here, the base animals of the 2 PB were entered into different genetic groups according to breed. With this, LDMIP accounts for differences in allele frequencies according to which breed(s) the animals originate from (Meuwissen et al., 2015). This will also apply to CB because they are linked to the PB through the pedigree. The option to run without genetic groups was also used to determine the importance of including genetic groups when having a multibreed data set. Information from neighboring loci was not used when running LDMIP because Meuwissen et al. (2015) found better accuracies of genomic selection when not using information from neighboring loci.

This estimation of genotype probabilities was followed by setting up the $\mathbf{G}_{\text{LDLA}}$ matrix. The genotype probabilities from LDMIP were used to set up the gametic relationship matrix:

$$\dot{\mathbf{G}} = \mathbf{W}\mathbf{W}'/\Sigma_j p_j(1 - p_j),$$

where $\dot{\mathbf{G}}$ was a $(2n \times 2n)$ matrix of gametic relationships (n = number of animals); and $\mathbf{W}$ was a $(2n \times m)$ matrix of standardized genotype probabilities (m = number of markers), with two rows per animal (one row each for the paternal and the maternal allele). Element W_{ij} is obtained by taking the probability of a '1' allele of gamete i at marker j and subtracting the corresponding marker allele frequency (across all animals), p_j (Meuwissen et al., 2015). The expectation is that the diagonal of $\dot{\mathbf{G}}$ is 1, because the relationship of a gamete with itself is 1. Off-diagonals represent inbreeding because

Table 1. Number of genotyped animals, individuals in pedigree and, phenotypic observations

Breed	No. of genotyped animals	No. of animals in pedigree
Dutch Landrace (A)	3,238	6,620
Large White (B)	3,735	7,240
F1 crosses (X)	1,377	1,377
Total	8,350	15,932 ¹
Trait	No. of observations	No. of phenotyped animals
Total number born (TNB)	67,063	11,491
Live born (LB)	66,958	11,491
Gestation length (GL)	61,015	10,768
Total		11,491

¹Includes pseudoparents for animals with one missing parent.

non-zero off-diagonals indicate that the maternal and paternal gamete is related. However, the diagonal of $\hat{G}$ may deviate from 1, either due to sampling or because genotype probabilities have lower variance than actual genotypes (Meuwissen et al., 2015). This may lead to underestimated relationships, and $\hat{G}$ was adjusted for this through the following formula:

$$\mathbf{G}_{\text{LDLA}} = [\mathbf{S}(\mathbf{D}\hat{\mathbf{G}}\mathbf{D} + \Delta\tilde{\mathbf{A}})\mathbf{S}']/2,$$

where $\mathbf{D}$ was a diagonal matrix with elements $\sqrt{(1/\hat{G}_{ii})}$ when $\hat{G}_{ii}$ is greater than 1, or 1 elsewhere, Δ was a diagonal matrix with elements $\sqrt{(1-\hat{G}_{ii})}$ when $\hat{G}_{ii}$ was less than 1, or 0 elsewhere, $\tilde{\mathbf{A}}$ was the pedigree-based gametic relationship matrix, and $\mathbf{S}$ was the design matrix that indicated which gametes belong to which animals, which reduces the size of the gametic relationship matrix to number of animals squared. This equation will also make the $\mathbf{G}_{\text{LDLA}}$ matrix positive semidefinite (Meuwissen et al., 2015). For further details, and an example, see Meuwissen et al. (2015). Ungenotyped descendants of the genotyped animals were added to the relationship matrix according to Henderson's rules.

In addition to the $\mathbf{G}_{\text{LDLA}}$ matrix, 7 other relationship matrices were built for comparison. These were the pedigree-based A matrix ($\mathbf{PED}$), G matrix for breed A ($\mathbf{G}_A$, i.e., a G matrix with breed A genotypes), breed B ($\mathbf{G}_B$), breed A and B together (including marker-based relationships between breeds, $\mathbf{G}_{AB}$), each of the PB with the CB ($\mathbf{G}_{AX}$ and $\mathbf{G}_{BX}$), and a G matrix including all of the genotypes across breeds ($\mathbf{G}_{ABX}$). The G matrices (except $\mathbf{G}_{\text{LDLA}}$) were built with the Gmatrix program (Su and Madsen, 2014) that is part of the DMU package. After building the different G matrices, ssGBLUP was used as implemented in DMU (Madsen and Jensen, 2008) for analyzing the full data set for all of the matrices using a multitrait model. The G-ADJUST option in DMU (adjusts genomic relationships so that

Table 2. Number of animals with phenotypes per breed, and number of animals for each breed with both genotype and phenotype information

Breed	Phenotype ¹		Genotype and phenotype ²
	TNB and LB	GL	
Dutch Landrace (A)	4,802	4,792	2,377
Large White (B)	5,377	5,217	2,859
F1 crosses (X)	1,312	759	1,312

¹Phenotypes: TNB = total number born, LB = live born, GL = gestation length.

²Number of genotyped animals with at least one phenotype.

they correspond to average relationships in the A matrix (Gao et al., 2012)) was used for all G matrices except $\mathbf{G}_{\text{LDLA}}$ when building the $\mathbf{H}$ matrix used in ssGBLUP. The full pedigree (including all breeds) was used for building the A matrix in all analyses.

The focus was on maternal traits, and the majority of the sows had more than one observation per trait. Consequently, a repeatability model was used to account for the effect of permanent maternal environment. The model was

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Mm} + \mathbf{Vv} + \mathbf{e},$$

where $\mathbf{y}$ was a vector of observations (TNB, LB, and GL) $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{M}$, and $\mathbf{V}$, known incidence matrices; $\mathbf{b}$ a vector of fixed effects (see below); $\mathbf{u}$ a vector of random additive genetic effects, with $\mathbf{u} \sim N(0, \mathbf{A}\sigma_u^2 \text{ or } \mathbf{G}\sigma_u^2)$, where σ_u^2 was the additive genetic variance; $\mathbf{m}$ a vector of permanent maternal environmental effects, with $\mathbf{m} \sim N(0, \mathbf{I}_m\sigma_m^2)$, where σ_m^2 was the non-genetic maternal environmental variance; $\mathbf{v}$ a vector of herd-year-season effects, with $\mathbf{v} \sim N(0, \mathbf{I}_v\sigma_v^2)$, where σ_v^2 was the variance of herd-year-season effects; and $\mathbf{e}$ a vector of residuals with $\mathbf{e} \sim N(0, \mathbf{I}_e\sigma_e^2)$, where σ_e^2 was the residual variance. $\mathbf{I}_m$, $\mathbf{I}_v$, and $\mathbf{I}_e$ were identity matrices of the appropriate dimensions, $\mathbf{A}$ was a matrix of pedigree-based, additive genetic relationships ($\mathbf{PED}$), and $\mathbf{G}$ a matrix of genomic relationships between all individuals. Here, $\mathbf{G}$ represents the aforementioned G matrices ($\mathbf{G}_A$, $\mathbf{G}_B$, $\mathbf{G}_{AB}$, $\mathbf{G}_{AX}$, $\mathbf{G}_{BX}$, $\mathbf{G}_{ABX}$, or $\mathbf{G}_{\text{LDLA}}$). Variance components (σ_u^2 , σ_m^2 , σ_v^2 , σ_e^2) were estimated using REML by DMU from the data. Fixed effects were breed, parity, farm, and farrowing quarter for TNB and LB, and breed, farm, and farrowing quarter for GL. Random effects were genetic effects of animal, permanent environmental effect (nongenetic maternal effects), and herd-year-season effects. The same model was used with all of the relationship matrices. The traits were not considered different in PB and CB animals because Lopes et al. (2017) found a purebred-crossbred correlation (r_{pc}) of ~0.9 for this data set.

The analysis was performed using the entire data set, and the solutions for fixed and random effects were extracted from this analysis for each of the relationship matrices. The data set was then modified to mask phenotypes for the validation animals (1,000 animals) and their offspring (both PB and CB offspring if applicable), for either breed A, breed B, or CB (X). Thus, validation was either in the A, B, or X animals, not in all at the same time. Two validation sets were used for each of the PB ($2 \times 1,000$ animals), where validation animals were different between the validation sets. For CB, only 1 validation set was used due to the availability of data. The training set was the remaining animals after validation animals had been removed. Note that, when validating in CB, few CB observations were available in the training set. To create the validation sets, the first (youngest) 1,000 animals in the list of genotyped animals in each breed were chosen, and then the next 1,000 animals in the next validation set for that breed (for PB). Thus, there were 2 validation sets with 1,000 animals each for each PB (A or B; ~4,790 observations per validation data set). Only animals that had both genotypes and phenotypes were included in the validation set. Matrices $\mathbf{G}_A$ and $\mathbf{G}_{AX}$ were only used in the validation of breed A and X animals, and likewise matrices $\mathbf{G}_B$ and $\mathbf{G}_{BX}$ were only used in the validation of breed B and X animals. After analyzing with the reduced data set, fixed and nongenetic random effects from the full analysis (for each matrix) were included before predicting phenotypes of validation animals to avoid altering the precision of estimates of fixed and nongenetic random effects by using a smaller data set. The prediction accuracy of the cross-validation was estimated by the following formula: $r = \frac{\text{corr}(\text{EBV}, \text{AdjPheno})}{\sqrt{h^2}}$, where EBV was

estimated breeding value, **AdjPheno** was phenotype (of validation animal) adjusted for fixed and (nongenetic) random effects, and h^2 was the heritability of the trait. Regression coefficients between EBV and adjusted phenotypes were estimated by fitting a linear model with adjusted phenotype as response variable and EBV as the explanatory variable. Standard deviations of estimated breeding values were also estimated. Except when the pedigree-based relationship matrix (**PED**) was used, EBV were genomic EBV (GEBV).

In addition, accuracy was also estimated when using fixed and random effects from the analysis using the full data set with the **PED** matrix. Thus, **AdjPheno** was adjusted for fixed and random effects from **PED**, and not from each matrix. This reduces the accuracy for each method, but makes methods comparable when testing for significant differences. This does not, however, change the EBV from the methods.

Table 3. Descriptive statistics

Trait ¹	Mean	SD	Minimum	Maximum	h^2
TNB	15.34	3.46	1	32	0.10
LB	14.02	3.23	1	28	0.07
GL, d	115.48	1.65	105	124	0.34

¹TNB = Total number born, LB = Live born, GL = Gestation length in days.

Bootstrapping was used to test whether accuracies of the different genomic prediction methods were significantly different from each other. This was done on the results from using the fixed and random effects from **PED**. The EBV from 2 methods at a time were compared against each other (pairwise comparison), to see which was best to predict the adjusted phenotypes (of the validation animals) from the **PED** evaluation. The size of the input data for the bootstrapping procedure was the same as for the validation sets. The Bootstrap procedure randomly samples with replacement data point triplets: the adjusted phenotype and their predictions (EBV) using 2 methods. It estimates which of the methods yields a greater correlation with the adjusted phenotype in each Bootstrap sample. A total of 10,000 Bootstrap samples were constructed for each pairwise comparison. By counting the number of times 1 method had a greater correlation than the other, the 2 methods (matrices) were considered significantly different if 1 of the methods had a greater correlation in at least 97.5% of Bootstrap samples (at a P -value of 5% due to the 2-sided nature of the test). This is similar to the method used by Vazquez et al. (2016).

The relationships that were common to both $\mathbf{G}_{LDLA}$ and $\mathbf{G}_{ABX}$ were plotted against each other to see if these would differ between the 2 matrices. The average relationship within and across breeds were estimated for $\mathbf{G}_{LDLA}$ and $\mathbf{G}_{ABX}$.

RESULTS

Trait means, standard deviations, number of observations, and other trait statistics are presented in Table 3. Heritability estimates are across all breeds. The mean number of parities for sows with observations was 5.85.

In terms of prediction accuracy (Table 4), the $\mathbf{G}_{LDLA}$ matrix had the greatest accuracy for all traits for PB, followed by $\mathbf{G}_{ABX}$, although this difference was small. For CB, the $\mathbf{G}_{ABX}$ matrix had the greatest accuracy for all traits, followed by $\mathbf{G}_{LDLA}$ (Table 5). Mean gain (in accuracy) from using $\mathbf{G}_{LDLA}$ over $\mathbf{G}_{ABX}$ was 1.0 and 1.1 percentage points across traits for breeds A and B, respectively. Including CB in the genomic relationship matrix ($\mathbf{G}_{AX}$, $\mathbf{G}_{BX}$, $\mathbf{G}_{ABX}$, $\mathbf{G}_{LDLA}$) always gave a greater accuracy than not including CB genotypes ($\mathbf{G}_A$, $\mathbf{G}_B$, $\mathbf{G}_{AB}$) for all breeds. Including both PB in the same

Table 4. Accuracy estimates of the relationship matrices for PB (mean accuracy between 2 validation sets of 1,000 animals each per breed)¹

Matrix ²	Total number born		Live born		Gestation length	
	Breed A ³	Breed B ³	Breed A	Breed B	Breed A	Breed B
PED	0.446	0.434	0.411	0.425	0.562	0.587
G _A	0.521	–	0.508	–	0.706	–
G _B	–	0.575	–	0.569	–	0.713
G _{AB}	0.567	0.593	0.552	0.593	0.712	0.716
G _{AX}	0.596	–	0.559	–	0.715	–
G _{BX}	–	0.615	–	0.619	–	0.750
G _{ABX}	0.632	0.626	0.599	0.633	0.726	0.755
G _{LDLA}	0.637	0.640	0.608	0.647	0.743	0.760

¹Bold numbers indicate the largest accuracy within each column.

²PED = pedigree-based relationship matrix, G_A = genomic relationship matrix for breed A, G_B = genomic relationship matrix for breed B, G_{AB} = genomic relationship matrix for breed A and B combined, G_{AX} = genomic relationship matrix for breed A and crossbreds (X), G_{BX} = genomic relationship matrix for breed B and crossbreds (X), G_{ABX} = genomic relationship matrix for both purebreds and crossbreds combined, G_{LDLA} = genomic relationship matrix for both purebreds and crossbreds utilizing linkage disequilibrium and linkage analysis.

³A = Dutch Landrace, B = Large White.

genomic relationship matrix (G_{AB}) increased accuracy compared to not including both PB in the same matrix for PB animals. For CB, this was only true for trait GL, while for TNB and LB, only when moving from G_A to G_{AB}. The increase in accuracy when moving from a single-breed matrix to G_{AB} was larger for Dutch Landrace (A) than for Large White (B). Overall, breed A benefitted more from including more animals in the genomic relationship matrix than breed B. In terms of individual traits, GL had the lowest increase in accuracy by including more animals in the genomic relationship matrix, but had the largest initial and overall accuracy. Accuracies were greater for matrices including CB genotypes and PED when validating in CB for traits TNB and GL, compared to validating in PB. This was also true for LB, except for matrix G_{LDLA} and PED.

When using fixed and random effects from PED for all matrices, the accuracies reduced with 3.8 to 15.4 percentage points for PB, and 0.0 to 9.8 percentage points for CB. Very small increases in accuracy (< 0.3 percentage points) were seen for G_A for trait TNB and GL in CB when using fixed and random effects from PED. In terms of differences in accuracies between matrices, not all of these were significant (Tables 6 and 7). For breed A, G_{LDLA} had a significantly greater accuracy than all other matrices for traits TNB and GL. For TNB, also G_{ABX} had a significantly greater accuracy than the other matrices, except G_{LDLA}. For LB, G_{LDLA} was not significantly more accurate than G_{ABX} and G_{AB}. Nor was G_{ABX}'s accuracy significantly different from that of G_A. For breed B, G_{LDLA} had a significantly greater

Table 5. Accuracy estimates of the relationship matrices for CB (1 validation set of 1,000 animals)¹

Matrix ²	Total number born	Live born	Gestation length
PED	0.450	0.403	0.619
G _A	0.432	0.374	0.635
G _B	0.508	0.459	0.630
G _{AB}	0.496	0.440	0.646
G _{AX}	0.615	0.578	0.745
G _{BX}	0.659	0.624	0.783
G _{ABX}	0.683	0.661	0.835
G _{LDLA}	0.670	0.631	0.826

¹Bold numbers indicate the largest accuracy within each column.

²PED = pedigree-based relationship matrix, G_A = genomic relationship matrix for breed A, G_B = genomic relationship matrix for breed B, G_{AB} = genomic relationship matrix for breed A and B combined, G_{AX} = genomic relationship matrix for breed A and crossbreds (X), G_{BX} = genomic relationship matrix for breed B and crossbreds (X), G_{ABX} = genomic relationship matrix for both purebreds and crossbreds combined, G_{LDLA} = genomic relationship matrix for both purebreds and crossbreds utilizing linkage disequilibrium and linkage analysis.

accuracy than all other matrices except G_{BX} for TNB and GL. For LB, G_{LDLA} only had a significantly greater accuracy than G_{AB} and G_{ABX}. For CB, there was no significant difference between G_{LDLA} and G_{ABX}. For trait TNB and LB, G_{LDLA} was not significantly different from either of the other matrices where CB were included, while for GL, G_{ABX} and G_{LDLA} had significantly greater accuracy than all other matrices. All pairwise comparisons are found in Tables 6 and 7.

Regression coefficients between EBV and adjusted phenotypes were close to 1 and similar across methods and no relationship matrix was clearly better than the others (Table 8).

Within trait, G_{LDLA} always had the greatest standard deviation of estimated breeding values (for validation animals; Tables 9 and 10). For PB, this was followed by including genotypes from all breeds (G_{ABX}), and for TNB and LB, by including CB with the PB (G_{AX} and G_{BX}). For GL, including CB had the third greatest standard deviation for B (G_{BX}), but for A, G_{AB} produced the third greatest standard deviation. For CB, G_{ABX} resulted in the second greatest standard deviation for LB and GL, while for TNB, this was G_{BX}.

There was very little difference in terms of accuracy (< 0.01 percentage points) and regression coefficients (0.01 increase for G_{LDLA} for TNB), and no difference for standard deviations of EBV in analyzing without genetic groups compared to with groups, and therefore these results are not presented here.

The correlation between allele frequencies of A and B was 0.25. The correlations between allele frequencies of PB and CB were 0.77 and 0.76 for A and B, respectively.

When plotting relationships common to G_{LDLA} and G_{ABX} against each other, there was some discordance

Table 6. Accuracies for PB animals (mean accuracy between 2 validation sets of 1,000 animals each per breed) of the different relationship matrices when using fixed and random effects from PED to create adjusted phenotype

Matrix ¹	Total number born		Live born		Gestation length	
	Breed A ²	Breed B ²	Breed A	Breed B	Breed A	Breed B
PED	0.446 ^{a,b}	0.434 ^a	0.411 ^a	0.425 ^{ab}	0.562 ^a	0.587 ^a
G _A	0.455 ^a	—	0.451 ^{a,b,c}	—	0.668 ^b	—
G _B	—	0.466 ^a	—	0.460 ^{a,b}	—	0.645 ^b
G _{AB}	0.480 ^{a,b}	0.472 ^a	0.473 ^{b,d}	0.460 ^a	0.656 ^b	0.639 ^b
G _{AX}	0.486 ^b	—	0.462 ^{a,b}	—	0.668 ^b	—
G _{BX}	—	0.482 ^{a,b}	—	0.474 ^{a,b}	—	0.672 ^{c,d}
G _{ABX}	0.511 ^c	0.488 ^a	0.489 ^{c,d}	0.479 ^a	0.663 ^b	0.671 ^c
G _{LDLA}	0.524 ^d	0.506 ^b	0.496 ^d	0.498 ^b	0.684 ^c	0.684 ^d

¹PED = pedigree-based relationship matrix, G_A = genomic relationship matrix for breed A, G_B = genomic relationship matrix for breed B, G_{AB} = genomic relationship matrix for breed A and B combined, G_{AX} = genomic relationship matrix for breed A and crossbreds (X), G_{BX} = genomic relationship matrix for breed B and crossbreds (X), G_{ABX} = genomic relationship matrix for both purebreds and crossbreds combined, G_{LDLA} = genomic relationship matrix for both purebreds and crossbreds utilizing linkage disequilibrium and linkage analysis.

²A = Dutch Landrace, B = Large White.

^{a-d}Accuracies within column with different superscript letters are significantly different ($P < 0.05$).

between the matrices (Fig. 1). This was especially true for low relationships (< 0.4). Some animals were seemingly unrelated in one of the matrices, but had relationships as strong as 0.6 in the other. This type of discordance went both ways. This led to the discovery of some pedigree errors, although not all could be corrected because not all of the animals in the pedigree had genotypes. Self-relationships were generally larger in G_{LDLA} than in G_{ABX}. Note that this was before G_{ABX} was adjusted for pedigree relationships with G-ADJUST, while G_{LDLA} is already adjusted for pedigree relationships (as this is part of the method).

Average relationships within and across breeds for G_{LDLA} and G_{ABX} are shown in Tables 11 and 12, respectively. In general, mean relationships were greater in G_{LDLA} than in G_{ABX}, except between A and B. The greatest relationship between individuals within A was between two inbred full-sibs, whose parents were also full-sibs. The greatest relationship within B is a sire-offspring relationship. The greatest relationship between A and B seems unreasonable and is likely due to some animals being assigned the wrong breed and therefore seeming unrelated in the pedigree even though that is not the case.

The computation time for building the G matrices was not the same between matrices. Computation time for LDMIP (pre-program for G_{LDLA}) was from 11.5 to 14.0 h (18 parallel jobs on the Abel computer cluster, dual Intel E5–2570 based, 2.6 GHz per node;

Table 7. Accuracies for CB animals (1 validation set of 1,000 animals) of the different relationship matrices when using fixed and random effects from PED to create adjusted phenotype

Matrix ¹	Total number born	Live born	Gestation length
PED	0.450 ^a	0.403 ^a	0.620 ^a
G _A	0.435 ^{a,b}	0.374 ^b	0.633 ^b
G _B	0.502 ^{c,d}	0.452 ^{c,d}	0.627 ^{a,b}
G _{AB}	0.495 ^d	0.436 ^{a,d}	0.642 ^c
G _{AX}	0.537 ^{b,d,e}	0.501 ^{b,d,e}	0.697 ^d
G _{BX}	0.567 ^{d,e}	0.536 ^{c,e,f}	0.727 ^d
G _{ABX}	0.582 ^{c,e}	0.563 ^f	0.755 ^e
G _{LDLA}	0.584 ^c	0.544 ^{c,f}	0.753 ^e

¹PED = pedigree-based relationship matrix, G_A = genomic relationship matrix for breed A, G_B = genomic relationship matrix for breed B, G_{AB} = genomic relationship matrix for breed A and B combined, G_{AX} = genomic relationship matrix for breed A and crossbreds (X), G_{BX} = genomic relationship matrix for breed B and crossbreds (X), G_{ABX} = genomic relationship matrix for both purebreds and crossbreds combined, G_{LDLA} = genomic relationship matrix for both purebreds and crossbreds utilizing linkage disequilibrium and linkage analysis.

^{a-f}Accuracies within column with different superscript letters are significantly different ($P < 0.05$).

UiO, 2017) and for building G_{LDLA} from 11.5 to 12.5 h (on the Abel computer cluster). In comparison, computation time for building the other G matrices was a couple of minutes (on a 46-bit physical Intel Core i7, 3.40-GHz core processor running Linux). Computation time for DMU5 (on a 46-bit physical Intel Core i7, 3.40-GHz core processor running Linux) increased with the size of the G matrix, thus G_{LDLA} was the slowest, with a computation time of about 1 h, and G_{ABX} second slowest with approximately 40 min.

DISCUSSION

Including CB genotypes in the genomic relationship matrix, whether based on markers only (G_{AX}, G_{BX}, or G_{ABX}) or through using G_{LDLA}, increased prediction accuracy for both PB and CB. However, not all of these increases were significantly different from not including CB. Depending on trait and breed, prediction accuracy for PB increased with between 0.9 to 11.6 percentage points compared to using a PB G matrix (G_A or G_B), while for CB, prediction accuracy increased with between 10.7 to 28.7 percentage points. This also led to a larger standard deviation of estimated breeding values (increase of 0.005 to 0.244). The increase in accuracy from including CB genotypes was larger for total number born (TNB; 4.0 to 11.6 percentage points) and live born (LB; 5.1 to 10.0 percentage points) than for gestation length (GL; 0.9 to 4.7 percentage points) in PB. In CB, the increase was similar across traits, though the greatest increase was seen for LB (28.7 percentage points difference between G_{ABX}

Table 8. Regression coefficients between estimated breeding values and adjusted phenotypes, across traits and breeds (i.e., mean across traits and breeds), for the different relationship matrices

Matrix ¹	Mean	Minimum	Maximum
PED	0.97	0.88	1.02
G _A ²	0.96	0.81	1.04
G _B ²	0.96	0.91	1.03
G _{AB}	0.99	0.91	1.07
G _{AX} ²	1.01	1.00	1.05
G _{BX} ²	0.99	0.93	1.04
G _{ABX}	1.01	0.92	1.07
G _{LDLA}	0.98	0.92	1.03

¹PED = pedigree-based relationship matrix, G_A = genomic relationship matrix for breed A, G_B = genomic relationship matrix for breed B, G_{AB} = genomic relationship matrix for breed A and B combined, G_{AX} = genomic relationship matrix for breed A and crossbreeds (X), G_{BX} = genomic relationship matrix for breed B and crossbreeds (X), G_{ABX} = genomic relationship matrix for both purebreds and crossbreeds combined, G_{LDLA} = genomic relationship matrix for both purebreds and crossbreeds utilizing linkage disequilibrium and linkage analysis.

²Based on fewer validation sets due to validating in only two breeds.

and G_A). The increase in accuracy was greatest for CB for all traits, but it was also greater for Dutch Landrace (A) than for Large White (B), except for the trait GL.

Accuracy

The greater prediction accuracy with G_{LDLA} for PB could be a result of this matrix utilizing more information than the other matrices. The increase in accuracy compared to using only 1 breed or both PB is probably mainly due to the utilization of across breed information. However, when comparing G_{LDLA} to the G matrix that also utilizes all the genotypes across all breeds (G_{ABX}), G_{LDLA} still has a significantly greater accuracy. An exception exists for the trait LB for breed A, although the difference is small. However, for CB, G_{ABX} had a greater accuracy than G_{LDLA}, although this difference was not statistically significant. This might suggest that G_{LDLA} is a more appropriate relationship matrix when dealing with crosses between breeds, when improvement is desired in the PB. Often, marker and QTL linkage phases will be different between different breeds (Zhou et al., 2014), which supports this notion. For CB prediction, it seems sufficient to include CB genotypes in one form or another. In theory, it should not be possible to include CB in ssGBLUP, because pedigree and genomic relationships will not refer to the same base and due to differences in allele frequencies between breeds (Legarra et al., 2015; Lourenco et al., 2016). However, in practice, G_{ABX} yields quite accurate predictions for PB (at least in the current data set) and the greatest accuracy for CB. For LB and GL in

Table 9. Standard deviations of breeding values for the different relationship matrices for PB (mean over 2 validation sets of 1,000 animals each per breed)¹

Matrix ²	Total number born		Live born		Gestation length	
	Breed A ³	Breed B ³	Breed A	Breed B	Breed A	Breed B
PED	0.454	0.406	0.333	0.321	0.472	0.448
G _A	0.524	—	0.390	—	0.561	—
G _B	—	0.607	—	0.480	—	0.611
G _{AB}	0.544	0.600	0.412	0.484	0.571	0.613
G _{AX}	0.582	—	0.434	—	0.566	—
G _{BX}	—	0.614	—	0.499	—	0.626
G _{ABX}	0.608	0.621	0.608	0.510	0.582	0.635
G _{LDLA}	0.634	0.636	0.634	0.527	0.598	0.644

¹Bold numbers indicate the largest standard deviation within each column.

²PED = pedigree-based relationship matrix, G_A = genomic relationship matrix for breed A, G_B = genomic relationship matrix for breed B, G_{AB} = genomic relationship matrix for breed A and B combined, G_{AX} = genomic relationship matrix for breed A and crossbreeds (X), G_{BX} = genomic relationship matrix for breed B and crossbreeds (X), G_{ABX} = genomic relationship matrix for both purebreds and crossbreeds combined, G_{LDLA} = genomic relationship matrix for both purebreds and crossbreeds utilizing linkage disequilibrium and linkage analysis.

³A = Dutch Landrace, B = Large White.

CB, G_{ABX} is significantly better than only including the PB animals (G_A, G_B, or G_{AB}), but for TNB, it is not significantly different from G_B. In PB, G_{ABX} is not significantly better than only including the PB (G_A, G_B, or G_{AB}), except for GL in breed B and TNB in breed A. The small differences between G_{LDLA} and G_{ABX} could be because the genotyped animals make up 52.4% of the animals in the pedigree and 56.6% of the animals with phenotypes, i.e., the proportion of genotyped animals was rather large. It is therefore likely that if a smaller proportion of the animals were genotyped, the relative difference between G_{LDLA} and G_{ABX} might increase.

It is expected that using population-specific allele frequencies (as in G_{LDLA}, for example) can accurately predict breeding values across populations (Wientjes et al., 2015). If the allele frequencies between the 2 PB were highly correlated (correlation close to 1), there would be little difference between these matrices, but this was not the case here. However, large differences in allele frequencies between the PB do not seem to be a problem when prediction is in CB. Unlike the other matrices, G_{LDLA} also uses identity by descent (IBD) information, whereas the other G matrices use only identity by state (IBS) information. As seen here, this might be problematic if there is discordance between the pedigree and the genotypes. Pedigree errors led to some strange relationships between individuals for G_{LDLA} that were not in agreement with the genotypes. On the other hand, this approach may be a useful tool to detect pedigree errors that are not detected by comparison of genotypes of parents and their offspring. For example, it could detect

Table 10. Standard deviations of breeding values for the different relationship matrices for CB (1 validation set of 1,000 animals)¹

Matrix ²	Total number born	Live born	Gestation length
PED	0.474	0.363	0.489
G _A	0.464	0.365	0.490
G _B	0.491	0.386	0.499
G _{AB}	0.468	0.372	0.487
G _{AX}	0.573	0.453	0.604
G _{BX}	0.623	0.494	0.602
G _{ABX}	0.621	0.496	0.633
G _{LDLA}	0.633	0.508	0.649

¹Bold numbers indicate the largest standard deviation within each column.

²PED = pedigree-based relationship matrix, G_A = genomic relationship matrix for breed A, G_B = genomic relationship matrix for breed B, G_{AB} = genomic relationship matrix for breed A and B combined, G_{AX} = genomic relationship matrix for breed A and crossbreds (X), G_{BX} = genomic relationship matrix for breed B and crossbreds (X), G_{ABX} = genomic relationship matrix for both purebreds and crossbreds combined, G_{LDLA} = genomic relationship matrix for both purebreds and crossbreds utilizing linkage disequilibrium and linkage analysis.

errors in the relationships between genotyped and ungenotyped individuals that cannot be detected by comparing genotypes. Possibly, the G_{LDLA} approach is more sensitive to pedigree errors, and it may be expected that without pedigree errors, G_{LDLA} will yield greater accuracies relative to G_{ABX} than the results shown here.

Combining genotypes from the 2 PB in the G matrix (G_{AB}) seemed to increase prediction accuracy over using single-breed matrices in PB, although this increase was not statistically significant. This increase is in agreement with findings by Esfandyari et al. (2016), who did not test for its significance. In CB, this increase was statistically significant for GL, but in TNB and LB, only when going from G_A to G_{AB} (also statistically significant). However, Hidalgo et al. (2015) found that including 2 PB in the G matrix reduced accuracies for all 4 traits under study. Relationships between the breeds will contribute to this increase in accuracy, even though the breeds are assumed unrelated. The mean relationship between animals of these breeds would indicate unrelatedness, but there were some notable relationships that could have influenced the results. It is also possible that this increase would be insignificant with larger population sizes for the PB as these are relatively small in the current study.

Regression Coefficients

Regression coefficients did not really differ between the different relationship matrices and no particular matrix was clearly best. These results are in agreement with Xiang et al. (2016), who did not find differences in regression coefficients between different

Figure 1. Common matrix elements of G_{ABX} and G_{LDLA} plotted against each other. Matrix G_{ABX} is the genomic relationship matrix for both purebreds and crossbreds combined, and matrix G_{LDLA} is the genomic relationship matrix for both purebreds and crossbreds utilizing linkage disequilibrium and linkage analysis.

scenarios either. This indicates that all models are unbiased and that the heritability estimates are reasonable.

Traits

There were differences between traits in terms of accuracy and standard deviations, although the overall trend was that G_{LDLA} and G_{ABX} performed the best across traits. More was gained in TNB and LB than in GL in terms of accuracy by adding CB genotypes to the relationship matrix when predicting in PB. This could be because the proportion of CB animals with phenotypes for GL was lower than for the other 2 traits. In addition, GL had a greater initial accuracy, which could be due to greater heritability for GL than for TNB and LB. It is also possible that the purebred-crossbred correlation (r_{pc}) for GL is greater than for TNB and LB, thus adding CB data does not really add more information than just adding more animals in general (i.e., if $r_{pc} = 1$, CB data equal PB data). A r_{pc} of 0.70 to 0.78 has been reported for TNB in Landrace and 0.57 to 0.68 for Yorkshire (Xiang et al., 2016), and Lopes et al. (2017) found an r_{pc} of 0.89 and 0.88 for TNB and GL, respectively, in breed A and 0.91 and 0.92 for TNB and GL, respectively, for breed B on the same data as were used in this study. Thus, adding the CB genotypes would not be much different than adding more animals in general when PB and CB performance is considered as the same trait, and it is difficult to conclude whether the increase is due to adding CB specifically or just adding more information. If r_{pc} is low, PB

Table 11. Off-diagonal relationship coefficients within and across breeds for G_{LDLA}

Relationship ¹	Mean	Minimum	Maximum	Median	Variance
Within A	0.183	-0.227	1.229	0.178	0.012
Within B	0.164	-0.062	0.977	0.155	0.004
Within X	0.019	-0.136	0.819	0.015	0.003
Between A-B	-0.155	-0.281	0.595	-0.156	0.001
Between A-X	0.014	-0.167	0.676	0.009	0.004
Between B-X	-0.002	-0.208	0.680	-0.003	0.001

¹A = Dutch Landrace, B = Large White, X = F1 Crossbreds.

and CB performance should be considered as different traits, and it would be more important to use CB data to predict CB-GEV for PB animals. Due to the high r_{pc} , this was not done here, as not much would be gained from it. When predicting in CB, the gain in accuracy was similar across traits, but it should be noted that when prediction was in CB, few CB observations were available on the traits, and thus CB are mainly predicted on PB data. This is possible when r_{pc} is high, but would be problematic if it was not, as several factors affecting the trait would not be accounted for.

Whether to consider PB and CB performance as the same or different traits depends on several factors. However, in most cases, it is the genetic correlation between PB and CB performance (r_{pc}) that is taken into consideration, both in terms of whether to include CB data at all, but also in terms of whether to consider PB and CB performance as different, but genetically correlated traits. The exact size of r_{pc} at which CB data are beneficial is debated, but most studies recommend inclusion of CB data when r_{pc} is below 0.7 or 0.8 (Dekkers, 2007; Bloemhof et al., 2012; Hidalgo et al., 2015; Tusell et al., 2016). Differences in r_{pc} are affected by several factors such as nonadditive effects (dominance, heterosis, etc.), genotype by environment interactions, breed of origin effects, differences in allele frequencies between breeds, etc. (Dekkers and Chakraborty, 2004; Christensen et al., 2014; van Grevenhof and van der Werf, 2015; Lopes et al., 2017). If the r_{pc} is low, it would make sense to consider PB and CB performance as different traits. This makes it possible to select for animals that have a better breeding value for CB performance. The traits could also be weighed differently in the breeding goal, ensuring genetic progress in both PB and CB performance. A limitation to viewing PB and CB performance as different traits might be data availability. If there are few observations on either PB or CB, it might not be enough data available for analysis. Considering PB and CB performance as the same trait would result in more available data, but if r_{pc} is low, this would result in poor prediction, due to factors such as those mentioned above. A

Table 12. Off-diagonal relationship coefficients within and across breeds for G_{ABX}

Relationship ¹	Mean	Minimum	Maximum	Median	Variance
Within A	0.162	-0.215	1.149	0.157	0.010
Within B	0.129	-0.210	0.833	0.123	0.003
Within X	0.010	-0.133	0.660	0.006	0.002
Between A-B	-0.144	-0.259	0.603	-0.146	0.001
Between A-X	0.009	-0.188	0.760	0.004	0.003
Between B-X	-0.012	-0.226	0.544	-0.014	0.001

¹A = Dutch Landrace, B = Large White, X = F1 Crossbreds.

benefit of considering PB and CB performance as the same trait is that one can have observations on both parents and offspring, or other close relationships such as half-sib PB and CB, increasing prediction accuracy. In the current study, PB and CB performance was analyzed as the same trait. However, the methods used here could also be applied when analyzing PB and CB observations as correlated traits. The reason for analyzing PB and CB performance as the same trait is mainly based on Lopes et al. (2017) finding an r_{pc} of 0.88–0.92 for TNB and GL on the same data set as was used in the current study. Possible reasons for this high r_{pc} could be lack of nonadditive effects (Hidalgo, 2015) and genotype-by-environment (**GxE**) interactions for these traits. The traits for both PB and CB were measured on many different farms, and this may have diluted or hidden any GxE effects, if present. However, it is difficult to conclude based on the information available what the specific reason for such a high r_{pc} is.

Breeds

Crossbreds gained the most from including CB genotypes. This makes sense, because the utilization of their genomic relationships tie them to more observations (and relatives), leading to more accurate prediction. When only utilizing PB genotypes, only pedigree relationships will link CB animals to each other and to PB relatives. The exception being that the PB relatives will have more accurate relationships between themselves if they are genotyped. In terms of the PB, Dutch Landrace (breed A) gained more in accuracy by using CB genotypes, or even just by adding genotypes from the other PB, than Large White (breed B). One reason might be that breed A had lower accuracy to start with, thus more to gain in general. Esfandyari et al. (2016) also found greater prediction accuracies in one breed (Yorkshire) over another (Landrace). They suggested that this may be due to a larger variance in genomic relationships for the Yorkshire animals (Esfandyari et al., 2016), but this does not agree with the results in the current study, where larger variation in relation-

ships was found in breed A. However, the mean relationship between breed A and CB was larger than the mean relationship between breed B and CB, and could contribute to the greater gain in breed A when including CB genotypes. An increase in accuracy when including more than 1 breed could be explained by relatively close relationships between the breeds (Zhou et al., 2014), but this was not the case in the current study, although there were some high relationships between the breeds. The average relationship between individuals of A and B was -0.15 and -0.14 for G_{LDLA} and G_{ABX} , respectively. According to Lourenco et al. (2016), using breed-specific allele frequencies will pull across-breed relationships closer to zero, but no such effect was seen when using G_{LDLA} compared to G_{ABX} in the current study. When the correlation between allele frequencies of 2 breeds is low, this may lead to negative relationships between breeds (Lourenco et al., 2016), which is in agreement with the findings of the current study.

Computation Time

The computation time for G_{LDLA} was considerable longer than for the other G matrices. In part this is because computations for G_{LDLA} have not been optimized as is the case for DMU routines, but still the calculation of genotype probabilities for all animals in the pedigree and all SNP on the chip implies substantial computational costs. Possibly, in the future the calculation of genotype probabilities for ungenotyped animals may become an integral part of the genotype imputation algorithms, which are routinely used to impute missing genotypes. In any case, more research is needed to reduce the computational costs of the G_{LDLA} approach, especially when applied to larger data sets than the current one.

Grouping

The G_{LDLA} matrix only outperforms G_{ABX} by a small amount across traits (1.0 to 1.1 percentage points) for PB, but unlike G_{ABX} , it keeps track of which breed each animal comes from, or at least which genetic group it belongs to. Thus, for estimation of genetic trends or levels of genetic groups, it might be more accurate, although this was not attempted in this study. It is possible that to get a marked difference between grouping strategies, groups may need to be more detailed than simply using breeds as genetic groups. It is likely that a year-effect within-breed may more accurately reflect the genetic differences, because not all animals with unknown parents will be base animals, and will thus have different genetic levels to start with.

Breed Composition

Crossbred data make up a relatively large part of the genotypes (almost 16.5%) and phenotypes (11.4%) in this study, which may not be the case in routine evaluations where PB data dominate. For most breeding organizations, most of the available data are on PB, although this is likely to change in the future. Thus, it is possible that the gain from including CB genotypes would be less in routine situations than in the current study, but the genotyping and phenotyping of CB individuals will improve this situation. The proportion of CB phenotypes for GL (7.0%) was less than for TNB and LB, and may be one of the reasons that the gain in accuracy for PB by adding CB genotypes is less for GL (because there was less data available).

Conclusions

Including CB genotypes is beneficial for prediction accuracies in CB animals, but also in PB animals when these are parents of the CB for some traits, but not for all. Prediction accuracies increase with 10.7 to 28.7 percentage points for CB and 0.9 to 11.6 percentage points for PB by including CB genotypes. For PB, the G_{LDLA} matrix gave a significantly greater accuracy than G_{ABX} in all but one scenario (LB for breed A), although the gain in accuracy was less than 2 percentage points. However, for CB, the prediction accuracy was greater with G_{ABX} , although it was not significantly better than G_{LDLA} . Computation time for G_{LDLA} was much longer than for the other relationship matrices. Thus, research on how to reduce computational costs will be needed to make the G_{LDLA} approach feasible in large-scale routine evaluations. The results, both in terms of accuracy and in terms of computation time, do not justify the use of G_{LDLA} for PB, and for CB the accuracy is greater with G_{ABX} . Thus, the inclusion of CB genotypes is beneficial, regardless of the method used to build the genomic relationship matrix.

LITERATURE CITED

- Bijma, P., and J. W. M. Bastiaansen. 2014. The standard error of the estimated purebred-crossbred genetic correlation. In: Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada.
- Bloemhof, S., A. Kause, E. F. Knol, J. A. M. Van Arendonk, and I. Misztal. 2012. Heat stress effects on farrowing rate in sows: Genetic parameter estimation using within-line and crossbred models. *J. Anim. Sci.* 90(7):2109–2119. doi:10.2527/jas.2011-4650
- Christensen, O. F., P. Madsen, B. Nielsen, and G. Su. 2014. Genomic evaluation of both purebred and crossbred performances. *Genet. Sel. Evol.* 46(1):23–32. doi:10.1186/1297-9686-46-23
- Dekkers, J. C. M., and R. Chakraborty. 2004. Optimizing purebred selection for crossbred performance using QTL with different degrees of dominance. *Genet. Sel. Evol.* 36(3):297–324. doi:10.1186/1297-9686-36-3-297

- Dekkers, J. C. M. 2007. Marker-assisted selection for commercial crossbred performance. *J. Anim. Sci.* 85(9):2104–2114. doi:10.2527/jas.2006-683
- Esfandyari, H., A. C. Sørensen, and P. Bijma. 2015. Maximizing crossbred performance through purebred genomic selection. *Genet. Sel. Evol.* 47(1):16. doi:10.1186/s12711-015-0099-3
- Esfandyari, H., P. Bijma, M. Henryon, O. F. Christensen, and A. C. Sørensen. 2016. Genomic prediction of crossbred performance based on purebred Landrace and Yorkshire data using a dominance model. *Genet. Sel. Evol.* 48(1):40. doi:10.1186/s12711-016-0220-2
- Gao, H., O. F. Christensen, P. Madsen, U. S. Nielsen, Y. Zhang, M. S. Lund, and G. Su. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet. Sel. Evol.* 44(1):8. doi:10.1186/1297-9686-44-8
- Groenen, M. A., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard, C. Park, D. Milan, and H.-J. Megens. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491(7424):393–398. doi:10.1038/nature11622
- Hidalgo, A. M., J. W. M. Bastiaansen, M. S. Lopes, B. Harlizius, M. A. M. Groenen, and D.-J. de Koning. 2015. Accuracy of Predicted Genomic Breeding Values in Purebred and Crossbred Pigs. *G3 (Bethesda)* 5(8):1575–1583. doi:10.1534/g3.115.018119
- Hidalgo, A. M. 2015. Exploiting genomic information on purebred and crossbred pigs. PhD Diss. Swedish University of Agricultural Science, Sweden, and Wageningen University, the Netherlands. <http://edepot.wur.nl/364564>. (Accessed 19 September 2017.)
- Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. *Genetics* 200(2):455–468. doi:10.1534/genetics.115.177014
- Lopes, M. S., H. Bovenhuis, A. M. Hidalgo, J. A. van Arendonk, E. F. Knol, and J. W. Bastiaansen. 2017. Genomic selection for crossbred performance accounting for breed-specific effects. *Genet. Sel. Evol.* 49(1):51–60. doi:10.1186/s12711-017-0328-z
- Lourenco, D., S. Tsuruta, B. Fragomeni, C. Chen, W. Herring, and I. Misztal. 2016. Crossbreed evaluations in single-step genomic best linear unbiased predictor using adjusted realized relationship matrices. *J. Anim. Sci.* 94(3):909–919. doi:10.2527/jas.2015-9748
- Madsen, P., and J. Jensen. 2008. DMU: A user's guide. A package for analysing multivariate mixed models, Version 6. http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf. (Accessed 4 May 2016).
- Meuwissen, T., and M. Goddard. 2010. The use of family relationships and linkage disequilibrium to impute phase and missing genotypes in up to whole-genome sequence density genotypic data. *Genetics* 185(4):1441–1449. doi:10.1534/genetics.110.113936
- Meuwissen, T. H., M. Svendsen, T. Solberg, and J. Ødegård. 2015. Genomic predictions based on animal models using genotype imputation on a national scale in Norwegian Red cattle. *Genet. Sel. Evol.* 47(1):79. doi:10.1186/s12711-015-0159-8
- Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. *J. Dairy Sci.* 97(6):3943–3952. doi:10.3168/jds.2013-7752
- Su, G., and P. Madsen. 2014. User's guide for Gmatrix Ver2. A program for computing Genomic relationship matrix. <http://dmu.agrsci.dk/Gmatrix/Doc/Current/GmatrixV2-User-Guide.pdf>. (Accessed 26 September 2016.)
- Toosi, A., R. L. Fernando, J. C. M. Dekkers, and R. L. Quaas. 2010. Genomic selection in admixed and crossbred populations. *J. Anim. Sci.* 88(1):32–46. doi:10.2527/jas.2009-1975
- Tusell, L., H. Gilbert, J. Riquet, M.-J. Mercat, A. Legarra, and C. Larzul. 2016. Pedigree and genomic evaluation of pigs using a terminal-cross model. *Genet. Sel. Evol.* 48(1):32. doi:10.1186/s12711-016-0211-3
- UiO. 2017. The Abel computer cluster. <http://www.uio.no/english/services/it/research/hpc/abel/more/index.html>. (Accessed 27 February 2017.)
- van Grevenhof, I. E. M., and J. H. J. van der Werf. 2015. Design of reference populations for genomic selection in crossbreeding programs. *Genet. Sel. Evol.* 47(1):14. doi:10.1186/s12711-015-0104-x
- Vazquez, A. I., Y. Venturi, M. Behring, S. Shrestha, M. Kirst, M. F. R. Resende, and G. de los Campos. 2016. Increased proportion of variance explained and prediction accuracy of survival of breast cancer patients with use of whole-genome multiomic profiles. *Genetics* 203(3):1425–1438. doi:10.1534/genetics.115.185181
- Wientjes, Y. C., R. F. Veerkamp, P. Bijma, H. Bovenhuis, C. Schrooten, and M. P. Calus. 2015. Empirical and deterministic accuracies of across population genomic prediction. *Genet. Sel. Evol.* 47(5). doi:10.1186/s12711-014-0086-0
- Xiang, T., B. Nielsen, G. Su, A. Legarra, and O. F. Christensen. 2016. Application of single-step genomic evaluation for crossbred performance in pig. *J. Anim. Sci.* 94(3):936–948. doi:10.2527/jas.2015-9930
- Zhou, L., M. S. Lund, Y. Wang, and G. Su. 2014. Genomic predictions across Nordic Holstein and Nordic Red using the genomic best linear unbiased prediction model with different genomic relationship matrices. *J. Anim. Breed. Genet.* 131(4):249–257. doi:10.1111/jbg.12089