Application of single-step genomic evaluation for crossbred performance in pig

Tao Xiang, B. Nielsen, G. Su, Andres Legarra, O. F. Christensen

Published March 21, 2016

Application of single-step genomic evaluation for crossbred performance in pig[1]

T. Xiang,*†[2] B. Nielsen,‡ G. Su,* A. Legarra,† and O. F. Christensen*

*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark; †INRA, UR1388 GenPhyse, CS-52627, F-31326 Castanet-Tolosan, France; and ‡SEGES, Pig Research Centre, DK-1609 Copenhagen, Denmark

ABSTRACT: Crossbreding is predominant and intensively used in commercial meat production systems, especially in poultry and swine. Genomic evaluation has been successfully applied for breeding within purebreds but also offers opportunities of selecting purebreds for crossbred performance by combining information from purebreds with information from crossbreds. However, it generally requires that all relevant animals are genotyped, which is costly and presently does not seem to be feasible in practice. Recently, a novel single-step BLUP method for genomic evaluation of both purebred and crossbred performance has been developed that can incorporate marker genotypes into a traditional animal model. This new method has not been validated in real data sets. In this study, we applied this single-step method to analyze data for the maternal trait of total number of piglets born in Danish Landrace, Yorkshire, and two-way crossbred pigs in different scenarios. The genetic correlation between purebred and crossbred performances was investigated first, and then the impact of (crossbred) genomic information on prediction reliability for crossbred performance was explored. The results confirm the existence of a moderate genetic correlation, and it was seen that the standard errors on the estimates were reduced when including genomic information. Models with marker information, especially crossbred genomic information, improved model-based reliabilities for crossbred performance of purebred boars and also improved the predictive ability for crossbred animals and, to some extent, reduced the bias of prediction. We conclude that the new single-step BLUP method is a good tool in the genetic evaluation for crossbred performance in purebred animals.

Key words: crossbred performance, genetic correlation, genomic evaluation, pig, reliability, single-step method

 J. Anim. Sci. 2016.94:936–948
doi:10.2527/jas2015-9930

INTRODUCTION

Crossbreding is predominant and intensively used in meat production systems (Wei, 1992), especially in swine and chicken. In two-way crossbreding schemes, selection of purebreds for their crossbred performance is the ultimate goal (Wei, 1992; Bijma and Bastiaansen, 2014). Because there exist genetic differences between breeds and genotype × environment interaction effects, additive genetic effects estimated based on purebred performance cannot be used to perfectly predict the crossbred performance (Lo et al., 1997). Ideally, combined purebred and crossbred information is required to implement the genetic evaluation for crossbred performance (Wei and van der Werf, 1994). However, due to the difficulty and high cost of collection of data from crossbred animals (Dekkers, 2007), it is not common to have access to crossbred data.

Genomic selection has been successfully applied in purebreds based on data from purebred animals (Loberg and Dürr, 2009; Fulton, 2012), but it also offers opportunities for selecting purebreds for

[1]The project was funded through the Green Development and Demonstration Programme (grant number 34009-12-0540) by the Danish Ministry of Food, Agriculture and Fisheries, the Pig Research Centre, and Aarhus University. The first author benefited from a joint grant from the European Commission and Aarhus University, within the framework of the Erasmus-Mundus joint doctorate "EGS-ABG." A.L. thanks financing from INRA SelGen metaprogram projects X-Gen and SelDir. The authors also thank for the valuable suggestions given by Per Madsen from Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University.

[2]Corresponding author: Tao.Xiang@mbg.au.dk

Received October 1, 2015.
Accepted December 31, 2015.

crossbred performance by using combined information from purebreds and crossbreds (Ibáñez-Escriche et al., 2009; Zeng et al., 2013) or by using only purebred data (Esfandyari et al., 2015). However, it generally requires that all relevant animals are genotyped. Recently, a novel single-step BLUP method (Christensen et al., 2014) for genomic evaluation of both purebred and crossbred performance in a two-way crossbreding system was developed that is an extension of a single-step BLUP method (Legarra et al., 2009; Christensen and Lund, 2010) from purebred performance to combined purebred and crossbred performances.

The aim of this study is to implement the new single-step BLUP method by using both purebred and crossbred data of total number of piglets born (**TNB**) in different scenarios, estimate the genetic correlation between purebred and crossbred performance, and then explore the impact of (crossbred) genomic information on prediction reliability for crossbred performance.

MATERIALS AND METHODS

Data

For this study, all data sets were provided by Danish Pig Research Centre. Three populations were simultaneously analyzed: Danish Landrace (**LL**), Danish Yorkshire (**YY**), and two-way crossbred Danish Landrace–Yorkshire. Crossbred animals that had a Landrace sire and a Yorkshire dam were termed "Landrace_Yorkshire" (**LY**), whereas "Yorkshire_Landrace" (**YL**) represented crossbreds with Yorkshire sires and Landrace dams. The TNB data in this study comprised the records of the first parity in all three populations. Altogether, TNB was recorded in 293,339 LL, 180,112 YY, and 10,974 crossbred animals. This data set is termed the "full population" throughout the whole paper.

Among the crossbreds, 7,407 were LY and 3,567 were YL. All of the purebred animals had first farrowing dates between 2003 and 2013, and the crossbred animals first farrowed between 2010 and 2013. The pedigree for both purebred and crossbred animals was available, and all the crossbreds were traced back to their purebred ancestors until 1994 by the DMU Trace program (Madsen, 2012). Consequently, 332,929 LL, 210,554 YY, and 10,974 crossbreds were in the pedigree. Among those animals, 7,723 LL and 7,785 YY were genotyped with an Illumina PorcineSNP60 Genotyping BeadChip (Ramos et al., 2009). Two-thirds of purebred genotyped animals were boars. For the crossbreds, 5,203 animals (4,077 LY and 1,126 YL) were genotyped with a 8.5K GGP-Porcine Low Density Illumina Bead SNP Chip (GeneSeek, 2012). Single nucleotide polymorphism quality controls were applied on the same data set in a previous study (Xiang et al., 2015), where more details can be found. Finally, 41,009 SNP and 7,916 SNP in autosome chromosomes were accessible in purebreds and crossbreds, respectively. Imputation was implemented in crossbreds from 7,916 SNP to 41,009 SNP with software Beagle (Browning, 2008), which outputs phased SNP for both reference and imputed population, by using a joint reference panel of the two pure breeds (Xiang et al., 2015). As a result, 41,009 phased genotyped SNP were available for the genotyped animals in both purebreds and crossbreds for the current study.

Single-Step BLUP Model for Purebred and Crossbred Performances

The new single-step BLUP method of evaluating both purebred and crossbred performance was developed by Christensen et al. (2014). The model reformulates the "full" Wei and van der Werf (1994) A1 model and incorporates genomic information by using two breed-specific combined relationship matrices, which extend the marker-based relationship matrices to the non-genotyped animals.

The Wei and van der Werf model is a trivariate model:

$$\mathbf{y}_L = \mathbf{X}_L\beta_L + \mathbf{Z}_L\mathbf{a}_L + \mathbf{e}_L,$$

$$\mathbf{y}_Y = \mathbf{X}_Y\beta_Y + \mathbf{Z}_Y\mathbf{a}_Y + \mathbf{e}_Y, \text{ and}$$

$$\mathbf{y}_{LY} = \mathbf{X}_{LY}\beta_{LY} + \mathbf{Z}_{LY}\mathbf{c}_{LY} + \mathbf{e}_{LY},$$

in which $\mathbf{y}_L$, $\mathbf{y}_Y$, and $\mathbf{y}_{LY}$ contain phenotypes for purebred Landrace (**L**), purebred Yorkshire (**Y**), and crossbred LY animals, respectively; $\mathbf{X}_L\beta_L$, $\mathbf{X}_Y\beta_Y$, and $\mathbf{X}_{LY}\beta_{LY}$ represent fixed effects; $\mathbf{e}_L$, $\mathbf{e}_Y$, and $\mathbf{e}_{LY}$ were overall random residual effects, assumed to be independently normally distributed with mean 0 and variance $\mathbf{I}\sigma^2_{e_L}$, $\mathbf{I}\sigma^2_{e_Y}$, and $\mathbf{I}\sigma^2_{e_{LY}}$, respectively; $\mathbf{a}_L$ and $\mathbf{a}_Y$ contain breeding values for breed L and breed Y for their purebred performance (mating within each own breed); $\mathbf{c}_{LY}$ stands for the additive genetic effects of crossbred LY animals; and $\mathbf{Z}_L$, $\mathbf{Z}_Y$, and $\mathbf{Z}_{LY}$ are the respective incidence matrices. Note that the $\mathbf{c}_{LY}$ animal additive genetic effects are actually formed as the sum of two additive gametic effects, one from L and another from Y. In other words, a crossbred diploid genome decomposes into two purebred haploid genomes.

The Christensen et al. (2014) method first assumes that effects of markers across the different origins (Yorkshire and Landrace, in this case) are unrelated. Under this assumption, the additive effect of the genome of an F_1 crossbred animal can be split into

the sum of two additive gametic effects, one gamete from each breed, where the two gametic effects are uncorrelated by assumption of the model. Therefore, separate matrices of pedigree-based or genomic-based relationships can be set up within each breed and then be combined according to purebred theory for the single step (Legarra et al., 2009; Christensen and Lund, 2010). The analysis proceeds by estimating solutions to two different breed-specific random effects. The key to disentangle the breeds of origin for the genetic effect of the F_1 individuals is the ability to construct pedigree-based partial relationship matrices (García-Cortés and Toro, 2006) or separate (by origin) genomic matrices, which, in turn, requires ascertainment of breed origin of the marker genotypes. More specifically, there are three steps:

Step 1). Reformulate the Wei and van der Werf model by splitting additive genetic effects for crossbred animals (LY) into breed of origin specific genetic effects, that is, split the additive genetic value of the *i*th F1 crossbred in two additive genetic values, one from each origin (LL or YY): $c_{\mathrm{LY}i} = c^{\mathrm{L}}{}_{\mathrm{LY}i} + c^{\mathrm{Y}}{}_{\mathrm{LY}i}$. It has to be understood that neither of these is a breeding value *sensu stricto*; instead, they are additive effects in the statistical sense as "regression of value on gene dosage" as explained by Falconer et al. (1985), who clarifies the various definitions of average effect of genes in absence of random mating. Note that the new single-step model (Christensen et al., 2014) is not the animal model used by Lo et al. (1997) and Lutaaya et al. (2001). Actually, the new single-step model is a reformulation of the full model from Wei and van der Werf (1994; equation A1), whereas Lo et al. (1997) and Lutaaya et al. (2001) refer to the reduced animal model from Wei and van der Werf (1994; equation A2). In the presence of only pedigree information, the full and the reduced animal model are equivalent, but in the presence of crossbred genomic information, this is no longer the case. In the papers of Lo et al. (1997) and Lutaaya et al. (2001), the additive genetic value of the *i*th F_1 crossbred is $u_{\mathrm{LY}i} = (u^{\mathrm{L}}{}_{\mathrm{LY}p(i,\,\mathrm{L})} + \Phi_{\mathrm{L}i}) + (u^{\mathrm{Y}}{}_{\mathrm{LY}p(i,\,\mathrm{Y})} + \Phi_{\mathrm{Y}i})$. Here, $u^{\mathrm{L}}{}_{\mathrm{LY}p(i,\,\mathrm{L})}$ and $u^{\mathrm{Y}}{}_{\mathrm{LY}p(i,\,\mathrm{Y})}$ are half the additive genetic values of the purebred parents $p(i, \mathrm{Y})$ and $p(i, \mathrm{L})$, which are common to all the offspring of the same sire or dam, and $\Phi_{\mathrm{L}i}$ and $\Phi_{\mathrm{Y}i}$ are the respective Mendelian samplings, which are different for each offspring. In the reduced animal model, both Mendelian sampling terms are included in the residual effect of the crossbred animals, and only $u^{\mathrm{L}}{}_{\mathrm{LY}p(i,\,\mathrm{L})}$ and $u^{\mathrm{Y}}{}_{\mathrm{LY}p(i,\,\mathrm{Y})}$ are estimated. This is for two reasons: first, with only pedigree information, this term cannot be estimated; second, setting up matrices of additive relationships (and their inverse) for crossbred animals at the animal model is not straightforward (Lo et al., 1993; García-Cortés and Toro, 2006). Therefore, in the works of Lo et al. (1997) and Lutaaya et al. (2001), the additive genetic value of the *i*th F_1 crossbred $u_{\mathrm{LY}i}$ is replaced by $u^{\mathrm{L}}{}_{\mathrm{LY}p(i,\,\mathrm{L})} + u^{\mathrm{Y}}{}_{\mathrm{LY}p(i,\,\mathrm{Y})}$. With genomic relationships and in the model of Christensen et al. (2014), these Mendelian sampling terms are embedded into a genomic relationship matrix (relationships across animals for purebreds and gametes for crossbreds) and they are no longer uncorrelated. Therefore, the absorption of this term into the residual error term is not suitable. In the current study, $c^{\mathrm{L}}{}_{\mathrm{LY}i} = u^{\mathrm{L}}{}_{\mathrm{LY}p(i,\,\mathrm{L})} + \Phi^{\mathrm{L}}{}_{\mathrm{LY}i}$ and $c^{\mathrm{Y}}{}_{\mathrm{LY}i} = u^{\mathrm{Y}}{}_{\mathrm{LY}p(i,\,\mathrm{Y})} + \Phi^{\mathrm{L}}{}_{\mathrm{LY}i}$. The additive genetic value of the *i*th F_1 crossbred $c_{\mathrm{LY}i}$ is not identical to $u^{\mathrm{L}}{}_{\mathrm{LY}p(i,\,\mathrm{L})} + u^{\mathrm{Y}}{}_{\mathrm{LY}p(i,\,\mathrm{Y})}$ in Lo et al. (1997) and Lutaaya et al. (2001). Therefore, our model (which is a gametic model at the level of crossbreds) is not a single-step model equivalent of Lo et al. (1997) and Lutaaya et al. (2001), which, at the level of crossbreds, are reduced animal models.

Step 2). Construct breed-specific partial relationship matrices for each breed of origin genetic effects. Considering pedigree relationships, the variance and covariance between additive genetic purebred (**a**) and crossbred (**c**) effects of breed LL is described as

$$\mathrm{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{c}\end{bmatrix} = \begin{bmatrix}\sigma^2_{a_{\mathrm{L}}} & \sigma_{a_{\mathrm{L}},c_{\mathrm{L}}}\\ \sigma_{c_{\mathrm{L}},a_{\mathrm{L}}} & \sigma^2_{c_{\mathrm{L}}}\end{bmatrix} \otimes \mathbf{H}^{(\mathrm{L})}.$$

This is a two-trait representation. For better understanding, the genetic effects can be split into animal effects belonging to purebred animals ($\mathbf{a}_{\mathrm{L}}$ and $\mathbf{c}_{\mathrm{L}}$) and gametic effects belonging to crossbred animals ($\mathbf{a}^{(\mathrm{L})}_{\mathrm{LY}}$ and $\mathbf{c}^{(\mathrm{L})}_{\mathrm{LY}}$):

$$\mathrm{var}\begin{bmatrix}\mathbf{a}_{\mathrm{L}}\\ \mathbf{a}^{(\mathrm{L})}_{\mathrm{LY}}\\ \mathbf{c}_{\mathrm{L}}\\ \mathbf{c}^{(\mathrm{L})}_{\mathrm{LY}}\end{bmatrix} = \begin{bmatrix}\sigma^2_{a_{\mathrm{L}}} & \sigma_{a_{\mathrm{L}},c_{\mathrm{L}}}\\ \sigma_{c_{\mathrm{L}},a_{\mathrm{L}}} & \sigma^2_{c_{\mathrm{L}}}\end{bmatrix} \otimes \mathbf{H}^{(\mathrm{L})} = \begin{bmatrix}\sigma^2_{a_{\mathrm{L}}} & \sigma_{a_{\mathrm{L}},c_{\mathrm{L}}}\\ \sigma_{c_{\mathrm{L}},a_{\mathrm{L}}} & \sigma^2_{c_{\mathrm{L}}}\end{bmatrix} \otimes \begin{bmatrix}\mathbf{H}_{\mathrm{L,L}} & \mathbf{H}^{(\mathrm{L})}_{\mathrm{L,LY}}\\ \mathbf{H}^{(\mathrm{L})}_{\mathrm{LY,L}} & \mathbf{H}^{(\mathrm{L})}_{\mathrm{LY,LY}}\end{bmatrix},$$

in which matrix $\mathbf{H}^{(\mathrm{L})}$ is a matrix of partial relationships that contains 4 blocks, 1 for within purebred animals ($\mathbf{H}_{\mathrm{L,\,L}}$), 2 for purebred with crossbred animals ($\mathbf{H}^{(\mathrm{L})}_{\mathrm{L,LY}}$) and vice versa ($\mathbf{H}^{(\mathrm{L})}_{\mathrm{LY,L}}$), and 1 for within crossbred animals ($\mathbf{H}^{(\mathrm{L})}_{\mathrm{LY,LY}}$). If there are *n*L pure Landrace animals and *n*LY crossbred animals, the size of $\mathbf{H}^{(\mathrm{L})}$ is (*n*L + *n*LY) × (*n*L + *n*LY). The *n*L purebred animals have additive effects, which are breeding values, $\mathbf{a}_{\mathrm{L}}$ (when mated within breed) and $\mathbf{c}_{\mathrm{L}}$ (when mated to the other breed). The *n*LY purebred gametes of crossbred animals have additive effects $\mathbf{c}^{(\mathrm{L})}_{\mathrm{LY}}$ (within the cross itself). The covariance structure includes, for ease of representation, $\mathbf{a}^{(\mathrm{L})}_{\mathrm{LY}}$, which are

effects of crossbred gametes in purebred performance; these effects are merely conceptual but they simplify the representation and computation. The covariance structure for breed YY is similar:

$$\text{var}\begin{bmatrix}\mathbf{a}_{\text{Y}}\\ \mathbf{a}^{(\text{Y})}_{\text{LY}}\\ \mathbf{c}_{\text{Y}}\\ \mathbf{c}^{(\text{Y})}_{\text{LY}}\end{bmatrix}=\begin{bmatrix}\sigma^2_{a_{\text{Y}}} & \sigma_{a_{\text{Y}},c_{\text{Y}}}\\ \sigma_{c_{\text{Y}},a_{\text{Y}}} & \sigma^2_{c_{\text{Y}}}\end{bmatrix}\otimes\mathbf{H}^{(\text{Y})},$$

$$=\begin{bmatrix}\sigma^2_{a_{\text{Y}}} & \sigma_{a_{\text{Y}},c_{\text{Y}}}\\ \sigma_{c_{\text{Y}},a_{\text{Y}}} & \sigma^2_{c_{\text{Y}}}\end{bmatrix}\otimes\begin{bmatrix}\mathbf{H}_{\text{Y,Y}} & \mathbf{H}^{(\text{Y})}_{\text{Y,LY}}\\ \mathbf{H}^{(\text{Y})}_{\text{LY,Y}} & \mathbf{H}^{(\text{Y})}_{\text{LY,LY}}\end{bmatrix}$$

with the size of $\mathbf{H}^{(\text{Y})}$ equal to $(n^{\text{Y}} + n^{\text{LY}}) \times (n^{\text{Y}} + n^{\text{LY}})$, and both structures are assumed independent, that is, there is no covariance between LL effects and YY effects. As in Wei and van der Werf (1994), there are six genetic variance or covariance components, three for each breed.

Matrix $\mathbf{H}^{(\text{L})}$ can be constructed based on available information (pedigree and markers) as follows. The pedigree-based and marker-based breed LL partial relationship matrices are $\mathbf{A}^{(\text{L})}=\begin{bmatrix}\mathbf{A}_{\text{L,L}} & \mathbf{A}^{(\text{L})}_{\text{L,LY}}\\ \mathbf{A}^{(\text{L})}_{\text{LY,L}} & \mathbf{A}^{(\text{L})}_{\text{LY,LY}}\end{bmatrix}$ and $\mathbf{G}^{(\text{L})}=\begin{bmatrix}\mathbf{G}_{\text{L,L}} & \mathbf{G}^{(\text{L})}_{\text{L,LY}}\\ \mathbf{G}^{(\text{L})}_{\text{LY,L}} & \mathbf{G}^{(\text{L})}_{\text{LY,LY}}\end{bmatrix}$, respectively, in which the partition divides purebred animals from purebred gametes in crossbred animal. Because of the split into breed-specific gametes, the pedigree-based partial relationship matrices $\mathbf{A}^{(\text{L})}$ and $\mathbf{A}^{(\text{Y})}$ must be computed as in García-Cortés and Toro (2006).

Construction of the breed-specific marker-based relationship matrices assumes that the breed of origin of phased alleles in crossbred animals is known. In other words, it is known which phased allele in a crossbred animal LY is from breed LL and which one is from breed YY. Then, the marker-based partial relationship matrix contains cross products of centered genotypes:

$$\mathbf{G}_{\text{L, L}} = (\mathbf{m}^{\text{L}} - 2\mathbf{p}^{\text{L}}\mathbf{1}')(\mathbf{m}^{\text{L}} - 2\mathbf{p}^{\text{L}}\mathbf{1}')',$$

$$\mathbf{G}^{(\text{L})}_{\text{L,LY}} = (\mathbf{m}^{\text{L}} - 2\mathbf{p}^{\text{L}}\mathbf{1}')(\mathbf{q}^{\text{LY}} - \mathbf{p}^{\text{L}}\mathbf{1}')', \text{ and}$$

$$\mathbf{G}^{(\text{L})}_{\text{LY,LY}} = (\mathbf{q}^{\text{LY}} - \mathbf{p}^{\text{L}}\mathbf{1}')(\mathbf{q}^{\text{LY}} - \mathbf{p}^{\text{L}}\mathbf{1}')',$$

in which $\mathbf{m}^{\text{L}}$ and $\mathbf{q}^{\text{LY}}$ contain breed-specific allele contents of the second allele for purebred LL (coded as 0, 1, or 2) and crossbred animal LY (coded as 0 or 1), respectively, and vector $\mathbf{p}^{\text{L}}$ are breed LL specific allele frequencies based on marker genotypes for purebred and crossbred animals.

Later, matrix $\mathbf{G}^{(\text{L})}$ is adjusted to be compatible with $\mathbf{A}^{(\text{L})}$: $\mathbf{G}^{(\text{L})}_{\text{a}} = \mathbf{G}^{(\text{L})}\beta + \mathbf{K}\alpha$, in which $\mathbf{K} = \begin{bmatrix}\mathbf{J} & \mathbf{J}/2\\ \mathbf{J}/2 & \mathbf{J}/4\end{bmatrix}$ and $\mathbf{J}$ denotes a matrix of one's partitioned as $\mathbf{G}^{(\text{L})}$. Scalars α and β are estimated through solving the 2 following equations:

$$\overline{\mathbf{A}}^{(\text{L})}_{22} = \overline{\mathbf{G}}^{(\text{L})}\beta + \overline{\mathbf{K}}\alpha \text{ and}$$

$$\overline{\mathbf{dA}}^{(\text{L})}_{22} = \overline{\mathbf{dG}}^{(\text{L})}\beta + \overline{\mathbf{dK}}\alpha,$$

for example, equating the averages of the full matrices and equating the averages of the diagonals of pedigree and genomic relationships for genotyped individuals (Christensen et al., 2012). Matrix $\mathbf{A}^{(L)}_{22}$ contains pedigree relationships for genotyped LL individuals. The procedure is identical for breed YY.

Step 3). Combine the pedigree-based and adjusted marker-based partial relationship matrices to a combined partial relationship matrix $\mathbf{H}^{(\text{L})}$, which is similar to the $\mathbf{H}$ matrix used in the single-step method for purebred animals (Legarra et al., 2009; Christensen and Lund, 2010). The inverse of $\mathbf{H}^{(\text{L})}$ is

$$\left(\mathbf{H}^{(L)}\right)^{-1} = \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \left(\mathbf{G}^{(L)}_{\omega}\right)^{-1} - \left(\mathbf{A}^{(L)}_{22}\right)^{-1}\end{bmatrix} + \left(\mathbf{A}^{(L)}\right)^{-1},$$

in which $\mathbf{G}^{(L)}_{\omega} = (1 - \omega)\,\mathbf{A}^{(L)}_{22} + \omega\,\mathbf{A}^{(L)}_{22}$. Parameter ω is the relative weight on the residual polygenic effect. Many other studies have investigated the weighting factors between the pedigree-based and marker-based relationship matrices (Christensen and Lund, 2010; Christensen et al., 2012; Gao et al., 2012; Su et al., 2012; Guo et al., 2015) and, commonly, they put forward that the weighting factors should be determined by the specific trait and the data set analyzed. We investigated weighting factors from 0.1 to 0.5. Preliminary analysis (results not shown) for different weighing factors showed that $\omega = 0.4$ was appropriate, in terms of balance between predictive abilities and biases for crossbred animals. The procedure is identical for breed YY. The sparse inverse partial relationship matrices $(\mathbf{H}^{(\text{L})})^{-1}$ and $(\mathbf{H}^{(\text{Y})})^{-1}$ are used as input to solve the mixed model equations of the model.

Step 4). Therefore, the complete representation of the final model for genetic evaluation is

$$\mathbf{y}_{\text{L}} = \mathbf{X}_{\text{L}}\beta_{\text{L}} + \mathbf{Z}_{\text{L}}\mathbf{a}_{\text{L}} + \mathbf{e}_{\text{L}},$$

$$\mathbf{y}_{\text{Y}} = \mathbf{X}_{\text{Y}}\beta_{\text{Y}} + \mathbf{Z}_{\text{Y}}\mathbf{a}_{\text{Y}} + \mathbf{e}_{\text{Y}},$$

$$\mathbf{y}_{LY} = \mathbf{X}_{LY}\beta_{LY} + \mathbf{c}^{(L)}_{LY} + \mathbf{c}^{(Y)}_{LY} + \mathbf{e}_{LY},$$

$$\text{var}\begin{bmatrix} \mathbf{a}_L \\ \mathbf{a}^{(L)}_{LY} \\ \mathbf{c}_L \\ \mathbf{c}^{(L)}_{LY} \end{bmatrix} = \begin{bmatrix} \sigma^2_{a_L} & \sigma_{a_L,c_L} \\ \sigma_{c_L,a_L} & \sigma^2_{c_L} \end{bmatrix} \otimes \mathbf{H}^{(L)},$$

$$\text{var}\begin{bmatrix} \mathbf{a}_Y \\ \mathbf{a}^{(Y)}_{LY} \\ \mathbf{c}_Y \\ \mathbf{c}^{(Y)}_{LY} \end{bmatrix} = \begin{bmatrix} \sigma^2_{a_Y} & \sigma_{a_Y,c_Y} \\ \sigma_{c_Y,a_Y} & \sigma^2_{c_Y} \end{bmatrix} \otimes \mathbf{H}^{(Y)},$$

$$\text{var}(\mathbf{e}_L) = \mathbf{I}\sigma^2_{e_L}, \text{var}(\mathbf{e}_Y) = \mathbf{I}\sigma^2_{e_Y}, \text{ and var}(\mathbf{e}_{LY}) = \mathbf{I}\sigma^2_{e_{LY}}.$$

This is a three *observed* trait model (performance in LL, YY, and LY) but with two genetic effects (LL and YY), each with two *genetic* traits: purebred and crossbred performance. Estimation of genetic parameters by REML and BLUP predictions were done using the DMU software (Madsen and Jensen, 2013).

Crossbred Allele Tracing

Software Beagle, which was used to impute and phase genotypes in crossbred animals, does not give breed allele origins as an output. Therefore, to infer the allele origins in crossbred animals, we proceeded as follows. The allele tracing was processed separately on each chromosome per individual.

Among the 5,203 genotyped crossbred animals, sires of 4,520 crossbreds were genotyped, whereas neither parent of the other 683 crossbreds was genotyped. When the sire was genotyped, total differences between the two sets of phased imputed alleles of a crossed animal and two sets of phased alleles of its corresponding purebred sire were compared. Comparisons between crossbred and purebred phased alleles were made on each SNP along the chromosome. For a specific comparison, if a crossbred allele was different from the corresponding purebred allele, that SNP was counted as one difference. Along the chromosome, if the sum of differences between one set of crossbred phased alleles and one set of specific purebred phased alleles was lowest among the 4 comparisons, then this set of specific crossbred phased alleles was considered as originating from the breed of the sire. Logically, the other set of crossbred phased alleles was assigned to the other breed.

When neither parent was genotyped, one of the two sets of phased imputed crossbred alleles was studied segment by segment. Each crossbred phased chromosome was split into several small segments, which consisted of 50 consecutive SNP markers. These were compared with the corresponding collection of segments from phased chromosomes of two purebred reference populations LL and YY, which were used for imputing crossbred genotypes. Each small segment in the crossbred animals should exactly match at least one segment in the reference panel, because each crossbred segment was imputed by the purebred reference population. Copies of that specific segment being detected in the reference population of LL and YY were counted separately and were divided by the total number of segments in the same position in the reference panel of LL and YY to get proportions of the matched segment. If the proportion was higher in one breed, the crossbred segment was considered to originate from this breed. Throughout all the segments within a crossbred phased chromosome, if the vast majority of segments were considered as originating from one specific breed, then the crossbred phased chromosome was assigned to that breed. Consequently, 5,203 crossbred phased alleles were traced to either breed LL or YY.

Statistical Model

For Landrace and Yorkshire, the statistical model was as follows:

$$y_{ijklmn} = \mu + \text{hys}_i + \text{month}_j + \text{hybrid}_k + b_1 \times \text{age}_{ijklmn} + b_2 \times \text{age}^2_{ijklmn} + a_m + sb_n + e_{ijlkmn},$$

in which the dependent variable, y_{ijklmn}, represented TNB in the first parity in breed LL or YY; μ was the general mean; hys_i, month_j, and hybrid_k represented fixed effects of herd–year–season, month at farrowing, and hybrid indicator of service sire (same or different breed as sow); age_{ijklmn} and age^2_{ijklmn} were covariates for the age of farrowing and its squared value, with regression coefficient b_1 and b_2, respectively; a_m was the random additive genetic effect of sow; sb_n was a random service sire effect; and e_{ijklmn} was the random residual effect. Random effects were assumed to be independently normally distributed, $\boldsymbol{a} \sim N(\mathbf{0}, \mathbf{H}^{(L)}\sigma^2_{a_L})$ or $\boldsymbol{a} \sim N(\mathbf{0}, \mathbf{H}^{(Y)}\sigma^2_{a_Y})$, depending on pure breed; $\boldsymbol{sb} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_{sb})$ and $\boldsymbol{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_e)$, in which $\mathbf{H}^{(L)}$ and $\mathbf{H}^{(Y)}$ were previously defined; $\mathbf{I}$ was the identify matrix; $\sigma^2_{a_L}$ and $\sigma^2_{a_Y}$ were additive genetic variances for breed LL and YY for purebred performances, respectively; and σ^2_{sb} and σ^2_e were the variance of service boar effect and the variance of the residual effect, respectively.

The model for crossbred records was

$$y_{ijlm} = \mu + \text{hys}_i + \text{month}_j + b_1 \times \text{age}_{ijlm} + b_2 \times \text{age}^2_{ijlm} + c^{(L)}_m + c^{(Y)}_m + e_{ijlm},$$

Table 1. Scenarios for model-based reliability

Scenario[1]	Genotypes[2]	Phenotypes
Nogen_SC	No genotypes	Full data: 293,339 LL, 180,112 YY, and 10,974 crossbred animals
Genpure_SC	7,723 LL and 7,785 YY	
Genall_SC	7,723 LL, 7,785 YY, and 5,203 crossbreds	

[1]Nogen_SC = only pedigree information; Genpure_SC = pedigree information and purebred genotypes; Genall_SC = pedigree information and all purebred and crossbred genotypes.

[2]LL = Danish Landrace; YY = Danish Yorkshire.

in which the dependent variable, y_{ijlm}, represented TNB in the first parity in crossbred animals; μ, hys_i, $month_j$, age_{ijlm}, and e_{ilmn} represented the same effects as in the model for purebred records; and $c^{(L)}_m$ and $c^{(Y)}_m$ were breed LL and YY origin additive genetic effects, respectively. The two additive genetic effects were assumed to be independently normally distributed, $\mathbf{c}^{(L)} \sim N(\mathbf{0}, \mathbf{H}^{(L)}\sigma^2_{c_L})$ and $\mathbf{c}^{(Y)} \sim N(\mathbf{0}, \mathbf{H}^{(Y)}\sigma^2_{c_Y})$, in which $\mathbf{H}^{(L)}$ and $\mathbf{H}^{(Y)}$ were breed LL or YY specific partial additive genetic relationships and $\sigma^2_{c_L}$ and $\sigma^2_{c_Y}$ were additive genetic variances for crossbred performances of breed LL and YY, respectively.

Scenarios

Variance components, heritabilities, and genetic correlations between purebred and crossbred performances (r_{pc}) were first investigated in the full population. Heritability for purebred performance was defined as the ratio of additive genetic variances for purebred performance (σ^2_a) to phenotypic variances ($\sigma^2_p = \sigma^2_a + \sigma^2_{sb} + \sigma^2_e$), whereas heritability for crossbred animals was defined as the ratio of total additive genetic variance of crossbred performance for two breed-specific gametes ($0.5(\sigma^2_{c_L} + \sigma^2_{c_Y})$) to phenotypic variances ($0.5(\sigma^2_{c_L} + \sigma^2_{c_Y}) + \sigma^2_{e_{LY}}$). To explore the effect of different genotyping strategies on genetic evaluation for crossbred performance, the breed-specific partial relationship matrices were constructed based on three different scenarios (SC): 1) Nogen_SC, only pedigree information, which represented the traditional BLUP method; 2) Genpure_SC, pedigree information and purebred genotypes (7,723 LL and 7,785 YY), representing genotyping only purebreds; and 3) Genall_SC, pedigree information and all purebred and crossbred genotypes (7,723 LL, 7,785 YY, and 5,203 crossbreds). The purposes of studying Genall_SC were to check the necessity of including crossbred genomic information, which is normally not available, and to study the improvement of genomic prediction of purebred animals for crossbred performance. Information on each scenario is shown in Table 1. To make the results comparable across all studies, specific relationship matrices for breed LL and YY were calculated using allelic frequencies estimated from "old" purebred population (born before January 1, 2011), which were 2,210 LL and 2,161 YY, respectively. For each scenario, the variance components for purebred and crossbred performances were estimated and the genetic correlation between them was obtained.

Second, model-based reliabilities of crossbred performance for purebred boars were calculated in the 3 different scenarios mentioned. According to pedigree, 7,407 LY and 3,567 YL were offspring of 765 LL and 465 YY sires, respectively. These sires were divided into 2 subgroups of genotyped and non-genotyped animals and mean model-based reliabilities were computed in each subgroup. Mean model-based reliability was calculated as (Mrode and Thompson, 2005): $r^2 = \sum_{i=1}^{n}\left(1 - SEP_i^2 / \sigma_c^2\right)/n$, in which SEP_i was the SE of prediction for animal i, $\sigma_c{}^2$ was the variance of additive genetic effect for crossbred performance, and n was the number of purebred boars that were studied. In addition, the proportion of animals that have higher model-based reliabilities in one scenario compared with another scenario in each subgroup was also investigated.

Finally, the predictive ability for crossbred animals in the validation population (4,195 crossbreds) was investigated in different scenarios. The farrowing date of January 1, 2012, was used as the cut-off date to divide recorded sows in the full population into training and validation populations. For purebred genotyped boars, only birth dates were accessible, not days of farrowing. Therefore, for genotyped animals, the birth date of January 1, 2011, was, instead, used as the cut-off date. As a result, 240,543 LL, 139,868 YY, and 6,779 crossbreds were contained in training population, with 2,210 genotyped LL, 2,161 genotyped YY, and 2,357 genotyped crossbreds being included as well. The validation population for crossbred performance included 4,195 crossbreds, among which 2,846 were genotyped. Phenotypes of crossbred animals in the validation population were corrected for fixed and random effects other than additive genetic effect ($\mathbf{Y}_c = \mathbf{c}^{(L)}_m + \mathbf{c}^{(Y)}_m + \mathbf{e}$). $\mathbf{Y}_c$ were obtained by using full population data, with partial relationship matrices constructed in Genall_SC.

Breed-specific partial relationship matrices were constructed based on scenarios, concerning genotypes of animals in the training population: Nogen_T is the scenario in which relationship matrices $\mathbf{H}$ contained only pedigree information (i.e., $\mathbf{H}^{(L)} = \mathbf{A}^{(L)}$ and $\mathbf{H}^{(Y)} = \mathbf{A}^{(Y)}$);

Table 2. Scenarios for predictive ability

Scenario[1]	Genotypes[2]	Phenotypes
Nogen_T	No genotypes	Training: 240,543 LL, 139,868 YY, and 6,779 crossbred animals Validation: 52,796 LL, 40,244 YY, and 4,195 crossbred animals
Genpure_T	2,210 LL and 2,161 YY	
Genpc_T	2,210 LL, 2,161 YY, and 2.357 crossbreds	
Genall_T	2,210 LL, 2,161 YY, and 5,203 crossbreds	

[1]Nogen_T = the scenario in which relationship matrices **H** contained only pedigree information; Genpure_T = the scenario in which relationship matrices **H** contained pedigree information and purebred genotypes of 2,210 LL and 2,161 YY; Genpc_T = the scenario in which relationship matrices **H** contained pedigree information and genotypes of the 2,210 LL, 2,161 YY, and 2,357 crossbreds that were involved in the training data set; Genall_T = the scenario in which relationship matrices **H** comprised all information in Genpc_T plus extra genomic information (but not the phenotypic information) of the 2,846 crossbreds in the validation population.

[2]LL = Danish Landrace; YY = Danish Yorkshire.

Genpure_T is the scenario in which relationship matrices **H** contained pedigree information and purebred genotypes of 2,210 LL and 2,161 YY; Genpc_T is the scenario in which relationship matrices **H** contained pedigree information and genotypes of the 2,210 LL, 2,161 YY, and 2,357 crossbreds that were involved in the training data set; and Genall_T is the scenario in which relationship matrices **H** comprised all information in Genpc_T plus extra genomic information (but not the phenotypic information) of the 2,846 crossbreds in the validation population. Detailed information on each scenario is shown in Table 2. Variance components were estimated based on phenotypes from the training population in each scenario, being only slightly different from those based on phenotypes from the full population. The predictive ability of crossbreds was measured by validation correlations cor($\hat{\mathbf{c}}$, $\mathbf{Y}_c$) in each scenario, in which $\hat{\mathbf{c}}$ were the estimated additive genetic effects for crossbreds ($\hat{\mathbf{c}} = \mathbf{c}^{(L)}_m + \mathbf{c}^{(Y)}_m$) in the validation population from different scenarios. For Genall_T, the validation population was divided into 2 subgroups of genotyped and non-genotyped animals and the validation correlations were made in the subgroup as well as in the whole validation population. A Hotelling–Williams *t* test at a 5% confidence level was applied to evaluate the significance for the differences of validation correlations in each scenario. Moreover, to detect the possible inflation or deflation of predictions, the regression coefficients of $\mathbf{Y}_c$ on $\hat{\mathbf{c}}$ were explored to check whether they were close to 1. In addition, to measure uncertainty associated with results, bootstrap sampling (Mäntysaari and Koivula, 2012; Cuyabano et al., 2015) was used in the test population to estimate means and SE of correlations. Results were similar to the Hotelling–Williams test above and are not shown.

To check the possible impact of different genotyping scenarios on the ranking and selection of purebred animals for their crossbred performance, Spearman's rank correlations (Spearman, 1904) between breeding values of purebred sires (765 LL and 465 YY) for crossbred performance were calculated across different scenarios. In addition, the breeding values for crossbred performance were ranked from highest to lowest in different scenarios, and then the consistency of the purebred boars in the top 5% highest breeding values was checked across different scenarios. Furthermore, to investigate re-rankings in a situation closer to the way selection for crossbred performance could be implemented in practice for such a sow trait, the Spearman's rank correlation and the top 5% studies were also made on the "young" sows that were included in the validation population (52,796 LL and 40,244 YY), that is, purebred animals without their own records. Among these young sows, 1,103 LL and 1,085 YY were genotyped. These two studies were separately processed on the genotyped and non-genotyped young sows.

The new single-step BLUP method for crossbreds is complex, and therefore, we tried a simpler single-trait single-step BLUP method (Legarra et al., 2009; Christensen and Lund, 2010). This method assumed that all animals belonged to a single population, using a single relationship matrix, where the compatibility adjustment of **G** to $\mathbf{A}_{22}$ was done as in Christensen et al. (2012). Predictive abilities for crossbred animals in the validation population were also measured as cor($\hat{\mathbf{c}}$, $\mathbf{Y}_c$).

RESULTS

Variance Components, Heritabilities, and Genetic Correlations

Estimates of variance components and genetic correlations between purebred and crossbred performances for Landrace and Yorkshire in each scenario are shown in Table 3 together with calculated heritabilities. For each scenario, both pure breeds showed higher additive genetic variances for purebred performance (σ_a^2) than for crossbred performance (σ_c^2). Residual variances for purebred animals (σ_e^2) were larger than those for crossbred animals ($\sigma^2_{e_{LY}}$). For all scenarios, the estimated heritabilities for purebred performance (h^2) were always 0.11 and 0.09 for Landrace and Yorkshire, respectively. Heritabilities for crossbred animals (h^2_{LY}) were around

Table 3. Variance components,[1] heritabilities for purebred performance,[2] genetic correlation between purebred and crossbred performance for Landrace and Yorkshire,[3] and heritabilities for crossbred animals[4]

Scenario[5]	Breed	σ^2_a	$\sigma_{a,c}$	σ^2_c	σ^2_{sb}	σ^2_e	r_{pc} (SE)	h^2	$\sigma^2_{e_{LY}}$	h^2_{LY}
Nogen_SC	Landrace	1.63	0.62	0.48	0.83	12.07	0.70 (0.12)	0.11	8.36	0.08
	Yorkshire	1.23	0.61	0.92	0.73	11.47	0.57 (0.13)	0.09		
Genpure_SC	Landrace	1.65	0.78	0.68	0.88	12.16	0.73 (0.11)	0.11	8.40	0.09
	Yorkshire	1.21	0.64	0.96	0.72	11.67	0.59 (0.12)	0.09		
Genall_SC	Landrace	1.65	0.89	0.79	0.88	12.16	0.79 (0.09)	0.11	8.33	0.10
	Yorkshire	1.23	0.75	0.99	0.72	11.67	0.68 (0.10)	0.09		

[1]σ^2_a = additive genetic variance for purebred performance; $\sigma_{a,c}$ = genetic covariance between purebred and crossbred performance; σ^2_c = additive genetic variance for crossbred performance; σ^2_{sb} = variance of service-boar effect; σ^2_e = residual variance for purebred performance; $\sigma^2_{e_{LY}}$ = residual variance for crossbred animals.

[2]h^2 = heritability for purebred performance ($\sigma^2_a/(\sigma^2_a + \sigma^2_{sb} + \sigma^2_e)$).

[3]r_{pc} = genetic correlation between purebred and crossbred performance.

[4]h^2_{LY} = heritability for crossbred animals ($0.5(\sigma^2_{c_L} + \sigma^2_{c_Y})/[0.5(\sigma^2_{c_L} + \sigma^2_{c_Y}) + \sigma^2_{e_{LY}}]$).

[5]Nogen_SC = only pedigree information; Genpure_SC = pedigree information and purebred genotypes; Genall_SC = pedigree information and all purebred and crossbred genotypes.

0.09 in the different scenarios. The estimated genetic correlation between purebreds and crossbreds ranged from 0.70 in Nogen_SC to 0.78 in Genall_SC for the Landrace breed and ranged from 0.57 in Nogen_SC to 0.68 in Genall_SC for the Yorkshire breed. Standard errors were generally large but kept decreasing from around 0.12 (Nogen_SC) to 0.1 (Genall_SC) for both breeds. Slight differences of the estimated genetic correlation were observed between the two breeds. The Landrace breed showed slightly higher genetic correlation between purebred and crossbred performance than that for the Yorkshire breed.

Model-Based Reliability

Table 4 compares the mean model-based reliabilities for purebred sires for their crossbred performance in different scenarios across all boars and for genotyped and non-genotyped subgroups. The genotyped subgroup always had higher model-based reliabilities than the non-genotyped group, and for the group of all boars, model-based reliabilities were in between those of the subgroups of genotyped and non-genotyped animals in each scenario. Model-based reliabilities increased from about 0.28 to 0.39 for the Landrace breed and from about 0.22 to 0.37 for the Yorkshire breed from Nogen_SC to Genall_SC. From Nogen_SC to Genall_SC, model-based reliabilities kept increasing in all three groups. Overall, methods with marker information (Genpure_SC and Genall_SC) presented higher model-based reliabilities than the pedigree-based scenario (Nogen_SC). In addition, proportions of purebred boars that have larger model-based reliabilities between pairwise scenarios were also studied. Result shows that 100% of LL and YY boars had larger model-based reliabilities in the Genall_SC compared with the Nogen_SC and Genpure_SC (results not shown). Concerning the single-trait, single-step BLUP model, model-based reliabilities for purebred LL and YY in Genall_SC were 0.70 ± 0.12 and 0.69 ± 0.12, respectively. Although these values are much higher than results shown in Table 4, they cannot be directly compared because they represent the reliability of animals drawn from a breed that would be a mixture of YY and LL, which is not the case. In fact, this single-trait model has lower predictive abilities than Christensen's model, as will be shown next.

Predictive Abilities

Predictive abilities for crossbred pigs in the validation group for different scenarios are shown in Table 5. The Pearson correlation between the corrected phenotypes and the EBV ($\text{cor}(\hat{\mathbf{c}}, \mathbf{Y}_c)$) range from 0.084 in Nogen_T to 0.120 in Genall_T, as shown in the second row of Table 5. No statistically significant differences between Genpure_T and Nogen_T were found, but Genpc_T and Genall_T were statistically significantly more accurate than those two scenarios. For the Genall_T, the subgroup of 2,846 genotyped crossbred pigs reveals larger correlation coefficients than that in the subgroup of non-genotyped pigs. Furthermore, the subgroup of non-genotyped pigs in Genall_T shows larger correlation coefficients than those in other scenarios.

Regression coefficients of corrected phenotypes on the EBV are shown in Table 5. In general, regression coefficients were a little bit larger than 1 for all the scenarios. Regression coefficients for scenarios with marker information (Genpure_T, Genpc_T, and Genall_T) were closer to 1 than that for pedigree-based scenario (Nogen_T). Among scenarios with marker information, in terms of unbiasedness, there was no clear trend showing which scenario performed better, but none was clearly biased. For the Genall_T,

Table 4. Mean model-based reliabilities of purebred boars for their crossbred performance

	All[2]			Genotyped[3]			Non-genotyped[4]		
Breed[1]	Nogen_SC	Genpure_SC	Genall_SC	Nogen_SC	Genpure_SC	Genall_SC	Nogen_SC	Genpure_SC	Genall_SC
LL	0.303	0.332	0.385	0.307	0.341	0.391	0.280	0.279	0.346
YY	0.262	0.284	0.365	0.264	0.288	0.369	0.218	0.223	0.301

[1]LL = Danish Landrace; YY = Danish Yorkshire.

[2]All = all the sires of crossbred animals, consisting of 765 Landrace and 465 Yorkshire. Nogen_SC = only pedigree information; Genpure_SC = pedigree information and purebred genotypes; Genall_SC = pedigree information and all purebred and crossbred genotypes.

[3]Genotyped = genotyped sires of crossbred animals, consisting of 656 Landrace and 443 Yorkshire.

[4]Non-genotyped = Non-genotyped sires of crossbred animals, consisting of 109 Landrace and 22 Yorkshire.

the subgroup of genotyped animals had less bias than the subgroup of non-genotyped animals.

Single-Trait Single-Step BLUP Predictive Abilities

Predictive abilities by a single-trait single-step BLUP method for crossbred animals in the validation population are shown in last 2 rows in Table 5. They increase from 0.082 in Nogen_T to 0.106 in Genall_T. It can also be seen that the predictive abilities calculated based on the single-trait model show trends similar to those calculated from the three-trait model but are smaller than in each corresponding scenario. Regression coefficients increase slightly from 0.61 in Nogen_T to 0.71 in Genall_T but are further from 1 when compared with regression coefficients calculated based on the three-trait model. For Genall_T, the genotyped subgroup also had higher predictive abilities than that in non-genotyped subgroup.

Re-ranking of Purebred Animals across Scenarios

The Spearman's rank correlations between estimated crossbred breeding values of purebred boars (765 LL and 465 YY) in pairwise scenarios are shown in Table 6. For both breeds, it can be seen that the pairwise correlations are always smaller than 1. In terms of the "top 5%" study, from 60% to 82% of purebred boars (either LL or YY) were shared from one scenario to another in the top 5% highest breeding values (results not shown). Similar results were observed in young purebred sows (results not shown).

DISCUSSION

This study implemented the single-step BLUP method of Christensen et al. (2014) by using both purebred and crossbred data from LL and Yorkshire in several scenarios with regard to different amounts of genomic information. Results indicated that the model was applicable. The genetic correlation between purebred and crossbred performance for TNB was successfully estimated. Methods with marker information were powerful for genetic evaluation for crossbred performance with regard to the predictive ability and unbiasedness. In addition, this study demonstrated that, to implement genetic evaluation for crossbred performance, crossbred genomic information is useful in addition to purebred genotypes.

In the model, a key assumption was that breed origins of phased marker genotypes for crossbred animals were known. In this study, 60K crossbred genotypes were imputed from a 8K crossbred panel. Although Xiang et al. (2015) concluded that the imputation accuracies would be larger than 99% in terms of allele correct rates and 95% in terms of correlation coefficients between imputed genotypes and true genotypes, the uncertainty of crossbred genotypes cannot be totally eliminated. The algorithm of tracing alleles in the current study was considered to be working efficiently, because the differences between two purebred reference panels were considerably large in several sampled chromosomes. However, errors of tracing alleles probably still appeared if the similarity of two phased crossbred segments were high. All in all, a hidden risk of using incorrect alleles may still exist when building the breed-specific partial relationship matrix. This needs further research.

The additive genetic variances for purebred performance (σ^2_a) were larger than those for crossbred performance (σ^2_c), implying that the phenotypes of purebred animals could be more diverse than for the crossbred animals, which was in line with the phenotypic variances for purebred animals (15.12 and 14.14 for LL and YY, respectively) being larger than those for crossbred animals (9.49). The heritabilities for crossbred animals (h^2_{LY}) were not dramatically different from heritabilities for purebred performance (h^2), which was opposite to results in Wei and van der Werf (1995) and is due to the fact that in the current study, variances of environmental effects for crossbreds ($\sigma^2_{e_{\mathrm{LY}}}$) were only two-thirds of those for purebreds (σ^2_e), which could be a consequence of heterosis and phenotypic plasticity (better fitness) to the multiple herds for crossbreds than for purebreds (Misztal and Løvendahl, 2012) or could,

Table 5. Predictive abilities for crossbred animals in the validation population in different scenarios

Perdiction	Nogen_T[1]	Genpure_T[1]	Genpc_T[1]	Genall_T[1]		
				All	Genotyped	Non-genotyped
cor($\hat{\mathbf{c}}$, $\mathbf{Y}_c$)[2]	0.084[a]	0.088[a]	0.097[b]	0.120[c]	0.126	0.106
Regression coefficients[3]	1.179	1.049	1.081	1.067	1.048	1.105
Single-trait cor($\hat{\mathbf{c}}$, $\mathbf{Y}_c$)[4]	0.079	0.084	0.088	0.106	0.109	0.103
Single-trait regression coefficients[5]	0.588	0.644	0.644	0.698	0.875	0.647

[a–c]Different superscripts of small letters among scenarios indicate significant differences ($P < 0.05$) by a Hotelling–Williams t test.

[1]Nogen_T = the scenario in which relationship matrices **H** contained only pedigree information; Genpure_T = the scenario in which relationship matrices **H** contained pedigree information and purebred genotypes of 2,210 LL and 2,161 YY; Genpc_T = the scenario in which relationship matrices **H** contained pedigree information and genotypes of the 2,210 LL, 2,161 YY, and 2,357 crossbreds that were involved in the training data set; Genall_T = the scenario in which relationship matrices **H** comprised all information in Genpc_T plus extra genomic information (but not the phenotypic information) of the 2,846 crossbreds in the validation population.

[2]cor($\hat{\mathbf{c}}$, $\mathbf{Y}_c$) is correlation coefficients between corrected phenotypes and EBV.

[3]Regression coefficients of corrected phenotypes on EBV.

[4]Single-trait cor($\hat{\mathbf{c}}$, $\mathbf{Y}_c$) is correlation coefficients between corrected phenotypes and EBV based on the single trait single-step BLUP method.

[5]Single-trait regression coefficients is regression coefficients of corrected phenotypes on EBV based on the single trait single-step BLUP method.

alternatively, be due to fact that only three different herds were used for crossbreds. Crossbreding capitalizes on heterosis effects and complementarity between breeds and results in an increased performance of crossbreds compared with purebreds (Dekkers, 2007).

When selection is based on purebred performance, the genetic correlation between purebred and crossbred performance (r_{pc}) is a key genetic parameter in crossbreding schemes (Bell, 1982; Bijma and Bastiaansen, 2014). The genetic correlations between purebred and crossbred performance for TNB were around 0.75 and 0.63 for Landrace and Yorkshire, respectively, which confirmed the existence of a moderate correlation. The r_{pc} is smaller than 1, which is due to different environments for purebreds and crossbreds (Lutaaya et al., 2001) and the presence of dominant gene action combined with different allele frequencies in the two breeds (Lo et al., 1997; Christensen et al., 2014). This result was in line with Wong et al. (1971), who reported that the r_{pc} for litter size was 0.74. However, Wei (1992) reviewed some other studies that reported low or even negative genetic correlations between purebred and crossbred performance for litter size. A change of r_{pc} over time was reported to be caused by long-term purebred selection (Pirchner and VonKrosigk, 1973), and therefore, it needs to be estimated regularly. The SE on the estimated genetic correlations were generally large in the current study, which implies that the sample size was not large enough, especially for crossbreds. Taking the SE into account, the estimated correlations in different scenarios were not very different. Nevertheless, the slight decrease of SE with an increased amount of genomic information indicated that genotypes, especially crossbred genotypes, would reduce the uncertainty of r_{pc}. The decreasing SE demonstrated the better performance of the new single-step model incorporating crossbred marker information compared with the pedigree-based selection of purebred animals for crossbred performance. Bijma and Bastiaansen (2014) showed that when using pedigree relationships, the SE of r_{pc} was determined by number of sire families and reliabilities of EBV and suggested that the SE should not exceed 0.05. In the current study, the TNB was a low heritable trait (around 0.1) and only 1,018 sires of the 1,230 sires of 10,974 crossbred animals were genotyped, which was also low. Therefore, large SE were expected. Results in the current study showed that the r_{pc} for Landrace was slightly larger than that for Yorkshire, although the SE were large. Genetic correlations (r_{pc}) also consistently increase with number of genotypes used. One possible explanation could be that there are still some discordances between the definition of base populations in genomic and pedigree relationships. Concerning heritabilities for purebred performance, our estimates confirmed the results of Guo et al. (2015), who estimated heritabilities of 0.11 and 0.09 for TNB in Landrace and Yorkshire, respectively.

The model-based reliabilities for purebred boars for their crossbred performance were generally low in the current study. The magnitude of these reliabilities is a direct function of the prediction error variances, which, in this case, are mostly determined by the numbers of offspring per boar (Dufrasne et al., 2011). In the current study, the numbers of crossbred offspring for each boar ranged from 1 to 11, with an average of 5, which were low and led to a high uncertainty of prediction. According to Table 4, model-based reliabilities tended to increase as the amount of genomic information increased for both breeds. The scenarios with marker information presented larger model-based reliabilities than the pedigree-based scenario, which may be due to

Table 6. Spearman's rank correlations between crossbred breeding values for 765 Landrace boars (above the diagonal) and 465 Yorkshire boars (below the diagonal) of crossbred animals in pairwise scenarios

	Nogen_SC[1]	Genpure_SC[1]	Genall_SC[1]
Nogen_SC	1.00	0.92	0.90
Genpure_SC	0.93	1.00	0.98
Genall_SC	0.87	0.95	1.00

[1]Nogen_SC = only pedigree information; Genpure_SC = pedigree information and purebred genotypes; Genall_SC = pedigree information and all purebred and crossbred genotypes.

the additional marker information. Reliabilities for the subgroup of genotyped animals were larger than that for the subgroup of non-genotyped animals in each scenario, but non-genotyped animals also benefitted from the genomic information of genotyped animals, as the reliabilities for the non-genotyped subgroup kept increasing from Nogen_SC to Genall_SC. These results are in line with Lourenco et al. (2015). Reliabilities for non-genotyped animals in Genall_SC were even larger than those in Nogen_SC and Genpure_SC for genotyped animals, implying the benefit of genotyping crossbred animals. In addition, 100% of purebred boars had larger model-based reliabilities in Genall_SC than that in the other two scenarios, which also provided evidence that the model incorporating crossbred marker information was useful for genetic evaluation for crossbred performance in purebred boars. We concluded that crossbred genomic information plays a role in improving reliabilities for crossbred performance in purebred boars. Nevertheless, it has been reported that the model-based reliabilities overestimated the true reliabilities (VanRaden et al., 2009), because the markers may overfit the data set (Su et al., 2012). Therefore, further investigation on true reliabilities is needed, potentially by a simulation study.

Correlation coefficients between corrected phenotypes and EBV for TNB in crossbred animals were lower than results for daily gain and feed conversion ratio in Christensen et al. (2012). This may be related to the fact that the heritability was higher for the traits of daily gain and feed conversion ratio than for the TNB in the current study. Moreover, the additive genetic effects for crossbred animals required estimating 2 breed of origin genetic effects $\mathbf{c}^{(L)}$ and $\mathbf{c}^{(Y)}$, which may lead to more uncertainty for crossbred animals than studies for purebred animals in Christensen et al. (2012). The cor($\hat{\mathbf{c}}$, $\mathbf{Y}_c$) in different scenarios confirmed that the methods with marker information would enhance the predictive ability. The crossbred genomic information was useful to improve the prediction, because scenarios with only purebred genotypes did not show significant improvement compared with the pedigree-based scenario but significantly improved when crossbred genomic information was also involved. Results showed that genotyped animals had larger cor($\hat{\mathbf{c}}$, $\mathbf{Y}_c$) than non-genotyped animals, which was opposite to studies by Guo et al. (2015). This could be because in the current study, the validation group consisted of crossbred animals among which the genotyped subset was a random sample, without biases for prediction (Su et al., 2012), whereas in Guo et al. (2015), the validation group consisted of purebred animals among which the genotyped subset was a preselected group. Pre-selection reduces accuracies of EBV (Bijma, 2012; Lourenco et al., 2015). The non-genotyped subgroup of crossbred animals in Genall_T had larger accuracies than those in other scenarios, indicating that non-genotyped validated animals benefited from crossbred genotyped animals in the validation population. Therefore, we suggest genotyping crossbred animals as well as purebred animals when implementing genomic selection for crossbred animals.

Regression coefficients of corrected phenotypes on EBV did not show a clear preference for a specific scenario, but coefficients in all scenarios with marker information were closer to 1 than in the pedigree-based scenario. All the regression coefficients were larger than 1, suggesting the underestimation (deflation) of variation of the estimated genomic breeding values (Gao et al., 2012).

Both the values of Spearman's rank correlations lower than 1 and the top 5% study indicated that rankings of purebred animals' breeding values for crossbred performance were not consistent across different scenarios. The selected purebred candidates for crossbred performance will be different with the availability of (crossbred) genomic information.

In terms of the predictive abilities and bias, the single-trait model was less robust than the three-trait model, although it was easier to implement. With crossbred genomic information, the three-trait model showed up to 13% higher predictive abilities than the single-trait model, which seems an interesting gain for this low heritable trait.

Conclusion

The new single-step model works well for genetic evaluation for crossbred performance in pigs. A moderate, positive genetic correlation between purebred and crossbred performance (r_{pc} ranged from 0.57 to 0.78) for TNB in purebred Landrace and Yorkshire is confirmed. Crossbred genomic information reduces the SE on the estimate of this genetic correlation. Models with marker information, especially crossbred genomic information, improve model-based reliabilities for crossbred performance of purebred boars and also improve the predictive ability for validated crossbred animals and somehow reduce the bias of prediction. The single-

step model that considered the 3 populations as a single one resulted in lower predictive abilities. The model is a good tool in the genetic evaluation for crossbred performance in purebred animals.

LITERATURE CITED

Bell, A. E. 1982. Selection for heterosis – Results with laboratory and domestic animals. In: 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain, October 4–8, 1982. Vol. 6. p. 206–227.

Bijma, P. 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. J. Anim. Breed. Genet. 129(5):345–358. doi:10.1111/j.1439-0388.2012.00991.x.

Bijma, P., and J. W. Bastiaansen. 2014. Standard error of the genetic correlation: How much data do we need to estimate a purebred-crossbred genetic correlation? Genet. Sel. Evol. 46(1):79. doi:10.1186/s12711-014-0079-z.

Browning, S. R. 2008. Missing data imputation and haplotype phase inference for genome-wide association studies. Hum. Genet. 124(5):439–450. doi:10.1007/s00439-008-0568-7.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42(1):2. doi:10.1186/1297-9686-42-2.

Christensen, O. F., P. Madsen, B. Nielsen, T. Ostersen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. Animal 6(10):1565–1571. doi:10.1017/S1751731112000742.

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. doi:10.1186/1297-9686-46-23.

Cuyabano, B. C., G. Su, G. J. Rosa, M. S. Lund, and D. Gianola. 2015. Bootstrap study of genome-enabled prediction reliabilities using haplotype blocks across Nordic Red cattle breeds. J. Dairy Sci. 98(10):7351–7363. doi:10.3168/jds.2015-9360.

Dekkers, J. 2007. Marker-assisted selection for commercial crossbred performance. J. Anim. Sci. 85(9):2104–2114. doi:10.2527/jas.2006-683.

Dufrasne, M., M. Rustin, V. Jaspart, J. Wavreile, and N. Gengler. 2011. Using test station and on-farm data for the genetic evaluation of Pietrain boars used on Landrace sows for growth performance. J. Anim. Sci. 89(12):3872–3880. doi:10.2527/jas.2010-3816.

Esfandyari, H., A. Sorensen, and P. Bijma. 2015. Maximizing crossbred performance through purebred genomic selection. Genet. Sel. Evol. 47(1):16. doi:10.1186/s12711-015-0099-3.

Falconer, D. S. 1985. A note on Fisher's 'average effect' and 'average excess'. Genet. Res. 46(03):337–347. doi:10.1017/S0016672300022825.

Fulton, J. 2012. Genomic selection for poultry breeding. Anim. Front. 2(1):30–36. doi:10.2527/af.2011-0028.

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.

García-Cortés, L. A., and M. Á. Toro. 2006. Multibreed analysis by splitting the breeding values. Genet. Sel. Evol. 38:601–615.

GeneSeek. 2012. GeneSeek Genomic Profiler for Porcine LD. http://www.neogen.com/Genomics/pdf/Slicks/GGP_PorcineFlyer.pdf. (Accessed June 1 2013.)

Guo, X., O. F. Christensen, T. Ostersen, Y. Wang, M. S. Lund, and G. Su. 2015. Improving genetic evaluation of litter size and piglet mortality for both genotyped and non-genotyped individuals using a single-step method. J. Anim. Sci. 93(2):503–512. doi:10.2527/jas.2014-8331.

Ibáñez-Escriche, N., R. L. Fernando, A. Toosi, and J. C. M. Dekkers. 2009. Genomic selection of purebreds for crossbred performance. Genet. Sel. Evol. 41:12. doi:10.1186/1297-9686-41-12.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92(9):4656–4663. doi:10.3168/jds.2009-2061.

Lo, L. L., R. L. Fernando, and M. Grossman. 1993. Covariance between relatives in multibreed populations: Additive model. Theor. Appl. Genet. 87(4):423–430. doi:10.1007/BF00215087.

Lo, L. L., R. L. Fernando, and M. Grossman. 1997. Genetic evaluation by BLUP in two-breed terminal crossbreding systems under dominance. J. Anim. Sci. 75:2877–2884.

Loberg, A., and J. W. Dürr. 2009. Interbull survey on the use of genomic information. Interbull Bull. 39:3–14.

Lourenco, D. A. L., B. O. Fragomeni, S. Tsuruta, I. Aguilar, B. Zumbach, R. J. Hawken, A. Legarra, and I. Misztal. 2015. Accuracy of estimated breeding values with genomic information on males, females, or both: An example on broiler chicken. Genet. Sel. Evol. 47(1):56. doi:10.1186/s12711-015-0137-1.

Lutaaya, E., I. Misztal, J. W. Mabry, T. Short, H. H. Timm, and R. Holzbauer. 2001. Genetic parameter estimates from joint evaluation of purebreds and crossbreds in swine using the crossbred model. J. Anim. Sci. 79:3002–3007.

Madsen, P. 2012. DMU Trace, A program to trace the pedigree for a subset of animals from a large pedigree file. Version 2. Center for Quantitative Genetics and Genomics, Dep. of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark.

Madsen, P., and J. Jensen. 2013. A user's guide to DMU. Version 6, release 5.2. Center for Quantitative Genetics and Genomics, Dep. of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark.

Mäntysaari, E. A., and M. Koivula. 2012. GEBV validation test revisited. Interbull Bull. 45:11–16.

Misztal, I., and P. Løvendahl. 2012. Environmental physiology of livestock. John Wiley and Sons, West Sussex, UK.

Mrode, R. A., and R. Thompson, editors. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CAB Int. Publ., Midlothian, UK. doi:10.1079/9780851990002.0000.

Pirchner, F., and C. VonKrosigk. 1973. Genetic parameters of cross- and purebred poultry. Br. Poult. Sci. 14(2):193–202. doi:10.1080/00071667308416014.

Ramos, A. M., R. P. M. A. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald, J. E. Beever, C. Bendixen, C. Churcher, R. Clark, and P. Dehais. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PLoS One 4(8):e6524. doi:10.1371/journal.pone.0006524.

Spearman, C. 1904. The proof and measurement of association between two things. Am. J. Psychol. 15(1):72–101. doi:10.2307/1412159.

Su, G., P. Madsen, U. S. Nielsen, E. A. Mäntysaari, G. P. Aamand, O. F. Christensen, and M. S. Lund. 2012. Genomic prediction for Nordic Red cattle using one-step and selection index blending. J. Dairy Sci. 95(2):909–917. doi:10.3168/jds.2011-4804.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92(1):16–24. doi:10.3168/jds.2008-1514.

Wei, M. 1992. Combined crossbred and purebred selection in animal breeding. PhD Thesis, Wageningen University and Research Centre, Wageningen, the Netherlands.

Wei, M., and J. H. J. van der Werf. 1994. Maximizing genetic response in crossbreds using both purebred and crossbred information. Anim. Prod. 59(03):401–413. doi:10.1017/S0003356100007923.

Wei, M., and J. H. J. van der Werf. 1995. Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. J. Anim. Sci. 73:2220–2226.

Wong, W. C., W. J. Boylan, and W. E. Rempel. 1971. Purebred versus crossbred performance as a basis of selection in swine. J. Anim. Sci. 32:605–610.

Xiang, T., P. Ma, T. Ostersen, A. Legarra, and O. F. Christensen. 2015. Imputation of genotypes in Danish purebred and two-way crossbred pigs using low-density panels. Genet. Sel. Evol. 47(1):54. doi:10.1186/s12711-015-0134-4.

Zeng, J., A. Toosi, R. Fernando, J. Dekkers, and D. Garrick. 2013. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. Genet. Sel. Evol. 45(1):11. doi:10.1186/1297-9686-45-11.