Published February 2, 2017

Using markers with large effect in genetic and genomic predictions[1]

M. S. Lopes,*†[2] H. Bovenhuis,† M. van Son,‡ Ø. Nordbø,‡ E. H. Grindflek,‡ E. F. Knol,* and J. W. M. Bastiaansen†

*Topigs Norsvin Research Center, P.O. Box 43, 6640 AA Beuningen, the Netherlands; †Wageningen University, Animal Breeding and Genomics Centre, P.O. Box 338, 6708 PB Wageningen, the Netherlands; and ‡Norsvin, Storhamargata 44, 2317 Hamar, Norway

ABSTRACT: The first attempts of applying marker-assisted selection (MAS) in animal breeding were not very successful because the identification of markers closely linked to QTL using low-density microsatellite panels was difficult. More recently, the use of high-density SNP panels in genome-wide association studies (GWAS) have increased the power and precision of identifying markers linked to QTL, which offer new possibilities for MAS. However, when GWAS started to be performed, the focus of many breeders had already shifted from the use of MAS to the application of genomic selection (using all available markers without any preselection of markers linked to QTL). In this study, we aimed to evaluate the prediction accuracy of a MAS approach that accounts for GWAS findings in the prediction models by including the most significant SNP from GWAS as a fixed effect in the marker-assisted BLUP (MA-BLUP) and marker-assisted genomic BLUP (MA-GBLUP) prediction models. A second aim was to compare the prediction accuracies from the marker-assisted models with those obtained from a Bayesian variable selection (BVS) model. To compare the prediction accuracies of traditional BLUP, MA-BLUP, genomic BLUP (GBLUP), MA-GBLUP, and BVS, we applied these models to the trait "number of teats" in 4 distinct pig populations, for validation of the results. The most significant SNP in each population was located at approximately 103.50 Mb on chromosome 7. Applying MAS by accounting for the most significant SNP in the prediction models resulted in improved prediction accuracy for number of teats in all evaluated populations compared with BLUP and GBLUP. Using MA-BLUP instead of BLUP, the increase in prediction accuracy ranged from 0.021 to 0.124, whereas using MA-GBLUP instead of GBLUP, the increase in prediction accuracy ranged from 0.003 to 0.043. The BVS model resulted in similar or higher prediction accuracies than MA-GBLUP. For the trait number of teats, BLUP resulted in the lowest prediction accuracies whereas the highest were observed when applying MA-GBLUP or BVS. In the same data set, MA-BLUP can yield similar or superior accuracies compared with GBLUP. The superiority of MA-GBLUP over traditional GBLUP is more pronounced when training populations are smaller and when relationships between training and validation populations are smaller. Marker-assisted GBLUP did not outperform BVS but does have implementation advantages in large-scale evaluations.

Key words: Bayesian variable selection, genome-wide association study, genomic selection, marker-assisted selection

© 2017 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2017.95:59–71
doi:10.2527/jas2016.0754

INTRODUCTION

The first attempts of applying marker-assisted selection (**MAS**) were not very successful. One of the reasons was the difficulty in identifying markers closely linked to QTL using low-density microsatellite panels (Heffner et al., 2009). More recent-

[1]This work was financed by the Research Council of Norway and the Breed4Food Consortium. The authors thank CIGENE (NMBU, Norway) and Hanne Hamland in Norsvin for sample handling and genotyping.

[2]Corresponding author: marcos.lopes@topigsnorsvin.com

Received June 26, 2016.
Accepted November 15, 2016.

www.manaraa.com

ly, dense panels of SNP became available, enabling researchers to perform genome-wide associations studies (**GWAS**) in different species. These GWAS enabled the identification of novel QTL and the reduction of confidence intervals of previously identified QTL (Lopes et al., 2014). However, GWAS findings have not been extensively used in MAS schemes because the focus of many breeders has shifted to genomic selection (**GS**; Meuwissen et al., 2001) using all markers without preselection.

A common GS strategy is to replace the traditional BLUP with the genomic BLUP (**GBLUP**). With GBLUP, it is assumed that quantitative traits are controlled by a large number of genes and each gene explains a small amount of the variance of the trait (Goddard, 2009). However, this assumption of GBLUP leads to suboptimal prediction accuracy for quantitative traits that are controlled by a limited number of genes with moderate to large effects (Meuwissen and Goddard, 2010).

In this study, we aimed to show that with improved technologies, such as dense SNP panels, we can revive "old" strategies, such as MAS, to improve the accuracy of prediction. This approach consisted of including the most significant SNP from GWAS as a fixed effect in the prediction models: marker-assisted BLUP (**MA-BLUP**) and marker-assisted GBLUP (**MA-GBLUP**). We also compared the prediction accuracies from MA-BLUP and MA-GBLUP with those obtained by applying a Bayesian variable selection (**BVS**; George and McCulloch, 1993) model. The model trait used in this study was "number of teats," which is very relevant for pig breeding and for which an important QTL is known to be segregating (Duijvesteijn et al., 2014; Lopes et al., 2014).

METHODS

Ethics Statement

The data used for this study was 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. Data recording and sample collection were conducted strictly in line with the rules given by Dutch and Norwegian animal research authorities.

Data

Number of teats was recorded at birth in 4 pig populations: Large White, Dutch Landrace, Norwegian Landrace, and Duroc (see Table 1 for descriptive statistics). The Large White and Dutch Landrace populations were located in Dutch nucleus farms. The Norwegian Landrace and Duroc populations were located in Norwegian nucleus farms and a boar testing station. Three data sets from each population were used in this study: ALL, TRAINING, and VALIDATION.

Table 1. Descriptive statistics

Population	Data set[1]	No.[2]	Mean	SD[3]
Large White	ALL	322,887	15.05	1.05
	TRAINING	2,620	15.37	0.96
	VALIDATION	665	15.65	0.98
Dutch Landrace	ALL	439,809	15.27	1.07
	TRAINING	2,491	15.61	1.02
	VALIDATION	622	15.78	1.04
Norwegian Landrace	ALL	210,289	15.70	0.99
	TRAINING	6,090	15.92	0.95
	VALIDATION	1,522	16.06	0.97
Duroc	ALL	8,118	13.02	1.05
	TRAINING	3,798	12.98	1.04
	VALIDATION	950	13.00	1.00

[1]ALL: the whole population used in the preadjustment of the phenotypes, which includes the animals from TRAINING, VALIDATION, and their contemporaries; TRAINING: genotyped and phenotyped animals used for the genome-wide association studies and also as reference population in the genetic prediction analysis; VALIDATION: data set used to measure prediction accuracy.

[2]Number of phenotyped animals.

[3]Standard deviation of number of teats in each data set of each population.

The data set ALL consisted of all genotyped animals and their contemporaries that had phenotypes (322,887 Large White, 439,809 Dutch Landrace, 210,289 Norwegian Landrace, and 8,118 Duroc). Using ALL, the phenotypes (number of teats) were precorrected for all nongenetic effects. The precorrected phenotype was used as the response variable in further analyses. The nongenetic effects were estimated by a pedigree-based linear model in ASReml version 3.0 (Gilmour et al., 2009):

$$y_{ijkl} = \mu + \text{sex}_i + \text{hy}_j + \text{u}_k + \text{litter}_l + e_{ijkl}, \quad (1)$$

in which y_{ijkl} was the number of teats of the kth animal, μ was the overall mean, sex_i was the fixed effect of sex i, hy_j was the fixed effect of the herd–year j of birth, u_k was the random additive genetic effect of animal k, litter_l was the random effect of litter l, and e_{ijkl} was the random residual effect. The vector of additive genetic effects was assumed to be distributed as $\sim N(\mathbf{0}, \mathbf{A}\sigma_a^2)$, which accounted for the variances and covariances between animals due to relationships by formation of an **A** matrix (pedigree-based numerator relationship matrix), with σ_a^2 being the additive genetic variance. The vector of litter effects was assumed to be distributed as $\sim N(\mathbf{0}, \mathbf{I}\sigma_l^2)$, with **I** being an identity matrix and σ_l^2 the litter variance. The vector of residual effects was assumed to be distributed as $\sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, with σ_e^2 being the residual variance.

www.manaraa.com

The data set TRAINING was a subset of ALL consisting of the oldest 80% of the animals that had both phenotypes and genotypes (2,620 Large White, 2,491 Dutch Landrace, 6,090 Norwegian Landrace, and 3,798 Duroc animals). This data set was used to perform the GWAS and was also used as the reference population for prediction of breeding values.

The data set VALIDATION consisted of the remaining 20% youngest animals that had both phenotypes and genotypes (665 Large White, 622 Dutch Landrace, 1,522 Norwegian Landrace, and 950 Duroc animals). This data set was used to assess the prediction accuracy of the evaluated models as described below in the "Prediction of Breeding Values" section.

Genotypes

Genotyping was performed at the Centre for Integrative Genetics (University of Life Sciences, Ås, Norway) and at GeneSeek, Inc. (Lincoln, NE), mainly using the Illumina Porcine SNP60 BeadChip (Illumina, Inc., San Diego, CA). Part of the animals from the Large White ($n = 820$) and Dutch Landrace ($n = 873$) population were genotyped using the (Illumina, Inc.) GeneSeek Custom 80K SNP chip (GeneSeek Inc., Lincoln, NE). The number of animals genotyped per chip is shown in Table 2. Quality control consisted of excluding SNP with GenCall < 0.15 (Illumina Inc., 2005), call rate < 0.95, minor allele frequency < 0.02, strong deviation from Hardy–Weinberg equilibrium ($\chi^2 > 600$), SNP located on sex chromosomes, and unmapped SNP. The positions of the SNP were based on the Sscrofa10.2 assembly of the reference genome (Groenen et al., 2012). Animals with frequency of missing genotypes ≥ 0.05 would be removed from the data set. However, all genotyped animals had a frequency of missing genotypes < 0.05 and were therefore kept for further analyses. After quality control, the remaining missing genotypes of the animals genotyped with the SNP60 BeadChip were imputed using Beagle version 3.3.2 (Browning and Browning, 2007). At the same time, the animals genotyped with the GeneSeek Custom 80K SNP chip had their genotypes imputed to the set of SNP on the SNP60 BeadChip that passed the quality control. After quality control, 43,439 SNP for Large White, 41,077 SNP for Dutch Landrace, 38,085 SNP for Norwegian Landrace, and 36,131 SNP for Duroc were available from the SNP60 BeadChip and composed the final set of SNP used in further analyses. The number of SNP from the GeneSeek Custom 80K SNP chip that passed the quality control and were also presented in the SNP60 BeadChip (being, therefore, used in the imputation) was 34,436 SNP for Large White and 32,645 SNP for Dutch Landrace (Table 2).

Table 2. Description of genotypic data before and after quality control

SNP chip	Description	Population: Large White	Dutch Landrace	Norwegian Landrace	Duroc
Before quality control					
60K	Number of animals	2,465	2,240	7,612	4,748
	Number of SNP	64,232	64,232	64,232	64,232
80K	Number of animals	820	873		
	Number of SNP	68,528	68,528		
After quality control					
60K	Number of animals	2,465	2,240	7,612	4,748
	Number of SNP	43,439	41,077	38,085	36,131
80K	Number of animals	820	873		
	Number of SNP	34,436[1]	32,645[1]		
Final data set					
60K[2]	Number of animals	3,285	3,113	7,612	4,748
	Number of SNP	43,439	41,077	38,085	36,131

[1]Number of SNP of the GeneSeek Custom 80K SNP chip (80K) that passed quality control and is also present in the Illumina Porcine SNP60 BeadChip (60K).

[2]Illumina Porcine SNP60 BeadChip (60K) and GeneSeek Custom 80K SNP chip (80K) that was imputed to Illumina Porcine SNP60 BeadChip.

Genome-wide Association Studies

A single-SNP GWAS was performed within population using the following animal model:

$$y^*_k = \mu + X\hat{\beta} + u_k + e_k, \quad (2)$$

in which y^*_k was the precorrected phenotype of the kth animal, μ and animal$_k$ were as defined above for model [1], X was the genotype (0, 1, or 2) of the kth animal for the evaluated SNP, $\hat{\beta}$ was the unknown allele substitution effect of the evaluated SNP, and e_k was the random residual effect, which was assumed to be distributed as $\sim N(\mathbf{0}, \mathbf{I}\sigma^2_e)$. The association analyses were performed with the TRAINING data set within each population using ASReml version 3.0 (Gilmour et al., 2009).

The genetic variance explained by a SNP ($\sigma^2_{snp} = 2pq\alpha^2$) was estimated based on the allele frequencies (p and q) and the estimated allele substitution effect (α). The proportion of phenotypic variance explained by the SNP was defined as $\sigma^2_{snp}/\sigma^2_p$, in which σ^2_p is the total phenotypic variance (the sum of the additive and residual variances), which was estimated based on model [2] without a SNP effect.

Best Linear Unbiased Prediction and Genomic BLUP Models

Four models were evaluated: BLUP, GBLUP, MA-BLUP, and MA-GBLUP. All models were implemented in ASReml version 3.0 (Gilmour et al., 2009). From the single-SNP GWAS, we selected the most

www.manaraa.com

significant (smallest *P*-value derived from the ANOVA *F*-testing) SNP in each population to be included in the marker-assisted models for within-line prediction. Models fitting more than one marker were also applied when more than one QTL region in different locations (chromosomes) explaining >1% of the phenotypic variance were identified.

The models MA-BLUP and MA-GBLUP were equal to model [2], except that in MA-GBLUP, a genomic relationship matrix (**G** matrix) instead of an **A** matrix was used to account for the genomic variances and covariances between animals. The **G** matrix was built according to VanRaden (2008), using $\mathbf{G} = \mathbf{ZZ}'/2\sum pq$, in which **Z** is a matrix of centered genotypes and p and q are the allele frequencies of the SNP. The allele frequencies were estimated separately within each population and were based on the current genotyped population (recent allele frequencies). We used the allele frequencies of the current population instead of the base population because the evaluated populations, as most of the modern livestock genetic lines, were generated decades ago and genetic material from the founders is not available for genotyping. As concluded by VanRaden (2008), genomic inbreeding coefficients required accurate estimates of allele frequencies in the base population. However, predictions of genetic merit were much less sensitive to allele frequency estimates. Therefore, we do not expect a large influence of the allele frequency estimates in our results. For MA-BLUP and MA-GBLUP, the effect of the SNP (QTL) and the polygenic effect are estimated simultaneously, as already described in the late 1980s (Fernando and Grossman, 1989). The models BLUP and GBLUP were similar to model [2] but without the fixed effect $X\hat{\beta}$.

In MA-GBLUP, the SNP that was fitted as a fixed effect was also used to build the **G** matrix. To test whether using this SNP in both parts of the model has an effect on the accuracy of the MA-GBLUP, the SNP used as a fixed effect and all other SNP in high linkage disequilibrium (**LD**; $r^2 > 0.50$) with it were excluded from the set of SNP used to build the **G** matrix. The pairwise LD between the SNP used as a fixed effect in the model and all other SNP on the chromosome was estimated on the TRAINING data set using the software PLINK version 1.07 (Purcell et al., 2007).

Bayesian Variable Selection

Breeding values of the validation animals were also estimated fitting all SNP simultaneously in a BVS model (George and McCulloch, 1993):

$$\mathbf{y}^* = 1\mu + \mathbf{Z}\hat{\boldsymbol{\beta}}^* + \mathbf{e}, \quad (3)$$

in which $\mathbf{y}^*$ was a vector of precorrected phenotypes; μ was the mean number of teats; **Z** was a design matrix with SNP genotypes coded as 0, 1, or 2 copies of a given allele; $\hat{\boldsymbol{\beta}}^*$ was a vector of unknown SNP effects; and **e** was a vector of random residual effects assumed to be normally distributed $\sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, in which σ_e^2 was the residual variance and **I** was an identity matrix. A Bernoulli distribution was assumed for the SNP effects:

$$\hat{\boldsymbol{\beta}}^* \sim \begin{cases} N\left(0, \mathbf{I}\sigma_{g0}^2\right) \text{with probability } \pi_0 \\ N\left(0, \mathbf{I}\sigma_{g1}^2\right) \text{with probability } \pi_1 = 1 - \pi_0 \end{cases},$$

in which the first distribution was the null distribution, which contains SNP with small effects and explaining a small proportion of variance (σ_{g0}^2), and the second distribution contains SNP with large effects and explaining a large proportion of variance (σ_{g1}^2) of the trait. The probability to be in the null distribution (π_0) was set to 0.999 (Duijvesteijn et al., 2014; Sell-Kubiak et al., 2015; Verardo et al., 2016), meaning that only 1 in every 1,000 SNP will be in the second distribution, which is, on average, 38 SNP per cycle. The BVS model was implemented in the program Bayz (http://bayz.biz; accessed June 1, 2016). A total of 250,000 Markov chain Monte Carlo chains with a burn-in of 50,000 cycles were run and a Metropolis–Hastings sampler was applied to obtain good convergence.

Prediction Accuracy

The prediction accuracy of the models was measured using the correlation between the estimated breeding values and the corrected phenotypes of animals in the VALIDATION data set. For the models BLUP and GBLUP, breeding values were obtained directly from the analysis; for example, the polygenic breeding value of animal k ($\hat{u}_k$) was defined as the term animal$_k$ from model [2]. For MA-BLUP and MA-GBLUP, the breeding value was defined as the sum of the marker breeding value ($\hat{u}_{\text{snp}} = X\hat{\beta}$) and the polygenic breeding value ($\hat{u}_k$). The vector of breeding values from the BVS model ($\hat{u}_{\text{bvs}}$) was obtained as $\hat{u}_{\text{bvs}} = \mathbf{Z}\hat{\boldsymbol{\beta}}^*$. Finally, prediction bias was assessed by regressing the corrected phenotypes on the estimated breeding values.

RESULTS

Association Analyses

Three different SNP were identified as most significant in the different populations, but in all cases, the most significant SNP was located at approximately 103.5 Mb

www.manaraa.com

on chromosome 7 (Fig. 1; Table 3). The most significant SNP identified in the Large White population showed a $-\log_{10}$(*P*-value) equal to 16.12 and explained 3.48% of the phenotypic variance. The phenotypic variance in the Large White population was 0.89 ± 0.03 teats^2 and the corresponding heritability was 0.41 ± 0.04. The most significant SNP identified in the Dutch Landrace population showed a $-\log_{10}$(*P*-value) equal to 15.44 and explained 3.67% of the phenotypic variance. The phenotypic variance in the Dutch Landrace population was 0.98 ± 0.03 teats^2 and the corresponding heritability was 0.36 ± 0.04. The most significant SNP identified in the Norwegian Landrace population showed a $-\log_{10}$(*P*-value) equal to 34.09 and explained 3.30% of the phenotypic variance. The phenotypic variance in the Norwegian Landrace population was 0.76 ± 0.02 teats^2 and the corresponding heritability was 0.27 ± 0.03. In the Duroc population, the most significant SNP was the same as in the Large White population. In the Duroc population, this SNP showed a $-\log_{10}$(*P*-value) equal to 42.26 and explained 6.13% of the phenotypic variance, which is almost twice the variance explained by this SNP in the Large White population. The phenotypic variance in the Duroc population was 1.00 ± 0.03 teats^2 and the corresponding heritability was 0.29 ± 0.04.

Although the same QTL region on chromosome 7 was identified in all evaluated populations, the most significant SNP was not the same across populations. MARC0038565 was the most significant SNP in 2 populations, the Large White and the Duroc populations (Table 3). In the Norwegian Landrace, MARC0038565 was the second most significant SNP, being in high LD to INRA0027623, the most significant SNP in this population (r^2 = 0.99). In the Dutch Landrace, however, the most significant SNP (ASGA0035500) showed no LD with MARC0038565 (r^2 = 0). The LD between SNP located between 103 and 104 Mb on chromosome 7 of both Dutch and Norwegian Landrace is graphically represented in Fig. 2, which was built using Haploview software (Barrett et al., 2004).

Prediction of Breeding Values

In all populations, the lowest prediction accuracy was observed for BLUP and the highest for either MA-GBLUP or BVS (Table 4). In the Dutch Landrace population, we observed the lowest accuracies compared with the other populations for all models, except for BLUP, where the lowest accuracy was observed for the Duroc population. In the Norwegian Landrace population, which had the largest training data set, the highest prediction accuracies were observed compared with the other populations for all models.

In the Norwegian Landrace, highly significant peaks ($-\log_{10}$(*P*-value) > 10) were also observed on chromosomes 1, 4, and 14 (Fig. 1), with the most significant SNP in these regions explaining >1.00% of the phenotypic variance (Supplemental Table S1; see the online version of the article at http://journalofanimalscience.org). In this population, accounting only for the most significant SNP from chromosome 7, the prediction accuracies of MA-BLUP and MA-GBLUP were 0.336 and 0.477, respectively. Additionally accounting for the most significant SNP from chromosome 14, the prediction accuracies became 0.372 and 0.474, respectively (Table 5). With further additions of the most significant SNP from the peaks on chromosomes 1 and 4, accuracies became 0.399 and 0.482, respectively. Therefore, including a marker from each of these other 3 QTL regions in Norwegian Landrace increased the prediction accuracy of MA-BLUP and MA-GBLUP by 0.063 and 0.005, respectively, above the effect of the marker from chromosome 7. The prediction accuracy of MA-GBLUP using multiple QTL was, therefore, closer to the one obtained using the BVS model (0.498; Table 4), which allows multiple QTL with large effects.

Using BLUP, predictions were more biased, overestimating the genetic variance, compared with MA-BLUP in the Large White, Dutch Landrace, and Duroc population (Table 4). In the Norwegian Landrace population, the regression coefficients were 1.12 using both BLUP and BVS, 0.87 using MA-BLUP, 1.10 using GBLUP, and 1.11 using MA-GBLUP. The GBLUP, MA-GBLUP, and BVS resulted in similar bias of prediction in all populations.

DISCUSSION

In this study, we showed that accounting for the most significant SNP (identified in GWAS) in the genetic predictions resulted in improved prediction accuracy for the trait number of teats in all evaluated populations (Table 4). Replacing BLUP with MA-BLUP increased the prediction accuracy between 0.021 and 0.124, whereas replacing GBLUP with MA-GBLUP resulted in increases between 0.003 and 0.043. Meuwissen and Goddard (1996) described that the advantage of MAS over non-MAS is related to the proportion of variance explained by the QTL linked to the markers used in the prediction. Changing either from BLUP to MA-BLUP or from GBLUP to MA-GBLUP, the highest increase in prediction accuracy was observed in the Duroc population and the lowest in the Norwegian Landrace. This result is concordant with the total phenotypic variance explained by the SNP used in the predictions (6.13% in the Duroc and 3.30% in the Norwegian Landrace; Table 3).

www.manaraa.com

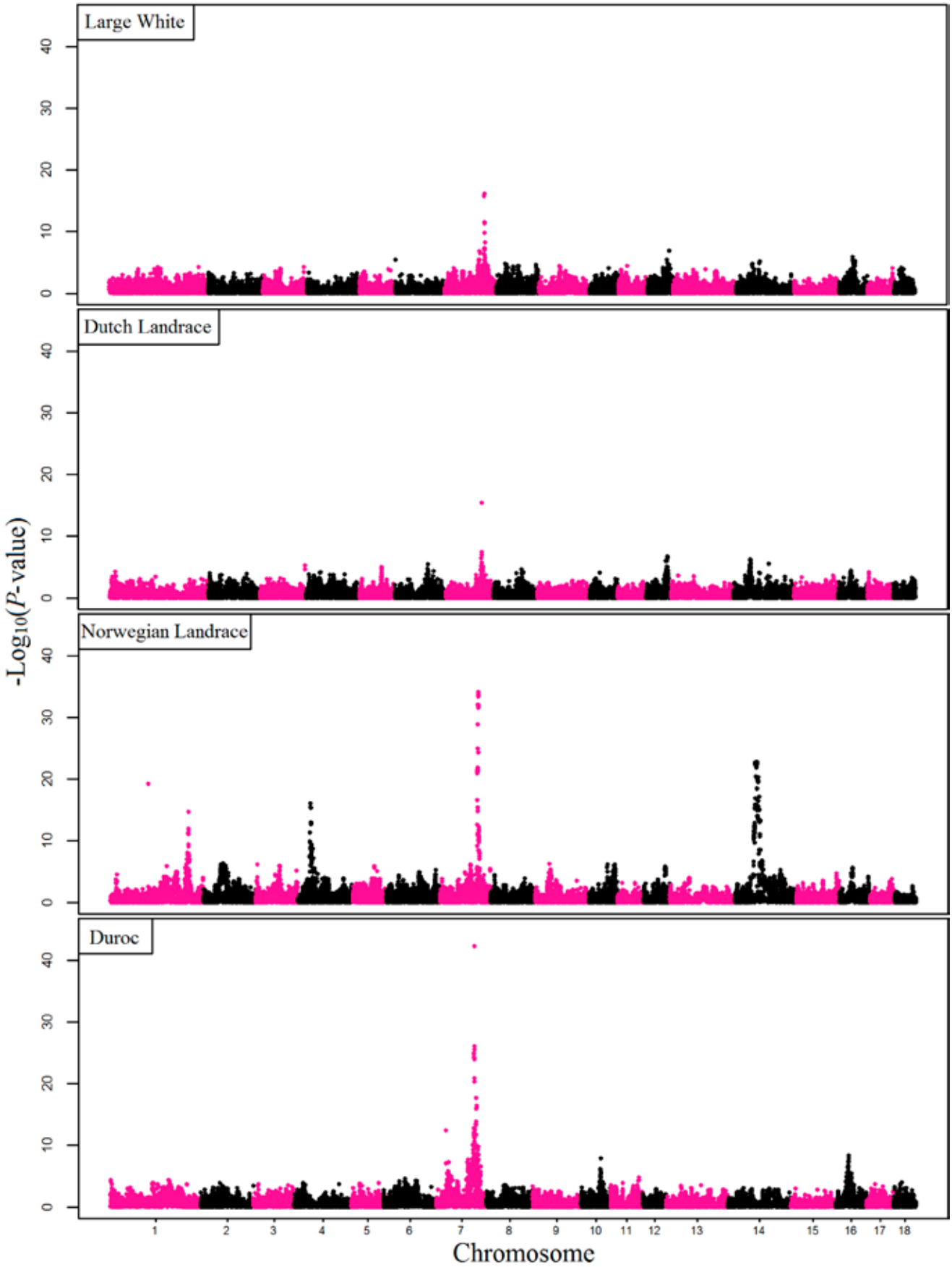

Figure 1. Genome-wide association studies on number of teats in 4 pig populations. On the *y*-axis is the $-\log_{10}$(*P*-values) of single SNP association with number of teats in pigs. On the *x*-axis is the physical position of the SNP across the 18 autosomes.

www.manaraa.com

Table 3. Description of the most significant SNP (Most Sig. SNP) in each population

Population	Most Sig. SNP	Chr[1]	Pos[2]	$-\log_{10}$(*P*-value)	Allele freq.[3]	Effect[4]	Var. explained,[5] %
Large White	MARC0038565	7	103.50	16.12	0.30	0.27	3.48
Dutch Landrace	ASGA0035500	7	103.57	15.44	0.69	0.29	3.67
Norwegian Landrace	INRA0027623	7	103.37	34.09	0.71	0.25	3.30
Duroc	MARC0038565	7	103.50	42.26	0.38	0.36	6.13

[1]Chr = chromosome.

[2]Pos = position in megabase pairs.

[3]Allele freq. = frequency of the allele related to higher number of teats.

[4]Effect represents the allele substitution effect.

[5]Var. explained = percentage of the total phenotypic variance explained by the most significant SNP.

The smaller improvement observed when replacing GBLUP with MA-GBLUP compared with replacing BLUP with MA-BLUP is most likely explained by the fact that GBLUP accounts for the Mendelian sampling, which is one of the greatest advantages of GS compared with pedigree-based selection (VanRaden, 2008; Lopes et al., 2013). Applying MA-GBLUP, however, we account for Mendelian sampling and also for some prior information on SNP with large effect, which has additional benefits for the prediction accuracy.

Applying the BVS model, we had the possibility of putting emphasis on SNP with large effect without requiring any prior knowledge on the QTL affecting the evaluated trait. We expected that the BVS model would result in higher prediction accuracies than the MA-GBLUP model because the BVS model allows multiple markers (QTL) with large effect, whereas with MA-GBLUP, we accounted for only 1 marker (QTL). However, the prediction accuracies of the BSV model were higher than those of the MA-GBLUP only in the Large White (0.370 for MA-GBLUP and 0.383 for BVS) and Norwegian Landrace (0.477 for MA-GBLUP and 0.498 for BVS) populations. In the other 2 populations, MA-GBLUP resulted in slightly higher accuracies than the BVS model (Table 4). These results indicate that for traits affected by QTL of large effect, such as number of teats, a simple model such as MA-GBLUP can yield prediction accuracies similar to more sophisticated models, such as BVS. The simplicity of MA-GBLUP is an advantage over BVS for practical application in breeding programs. With the increasing size of genomic data sets and the current computational resources, the number of markers is becoming too large for all markers to be included in the model at the same time (Brøndum et al., 2015), besides of the "large *p*, small *n*" paradigm.

Influence of Reference and Training Populations on the Prediction Accuracy

The Norwegian Landrace presented the highest prediction accuracies for all models, whereas the Dutch Landrace population presented the lowest (except for BLUP; Table 4). In both simulated and real data (Habier et al., 2007; Wu et al., 2015), it has been shown that higher relationship between training and validation populations can lead to higher prediction accuracies. The highest average pedigree-based relationship between the TRAINING and the VALIDATION data sets was observed for the Norwegian Landrace (0.06 ± 0.03) and the lowest for the Dutch Landrace (0.03 ± 0.04). For the Norwegian Landrace population, pairwise pedigree-based relationship coefficients between the animals from the TRAINING and VALIDATION data sets were all greater than 0 (Fig. 3). On the other hand, for the Dutch Landrace, a large proportion of the pairwise pedigree-based relationship coefficients were 0. Intermediate prediction accuracies and relationships between training and validation data sets were observed for both Large White and Duroc populations. These differences between populations indicate that the relationship between training and validation populations may indeed have affected the observed accuracies of prediction.

Another factor that influences the accuracies of breeding values is the size of the training population (Daetwyler et al., 2010). The size of the TRAINING data set in this study varied considerably across populations, ranging from 2,491 for the Dutch Landrace to 6,090 for the Norwegian Landrace. The population with the highest prediction accuracy (Norwegian Landrace) also had the largest training population. To evaluate the effect of the size of the TRAINING data set on the value of adding individual QTL in the model, we performed the prediction analysis in a smaller data set (n = 3,000) within the Norwegian Landrace. The 3,000 oldest animals of this population were divided into training (n = 2,400) and validation (n = 600) data sets according to their date of birth (validation animals were the 20% youngest animals from the data set). In this scenario, the prediction accuracies for BLUP, MA-BLUP, GBLUP, and MA-GBLUP were 0.287, 0.339, 0.423, and 0.446, respectively. Using the complete data (training on 6,090 animals), the prediction accuracies for BLUP, MA-

www.manaraa.com

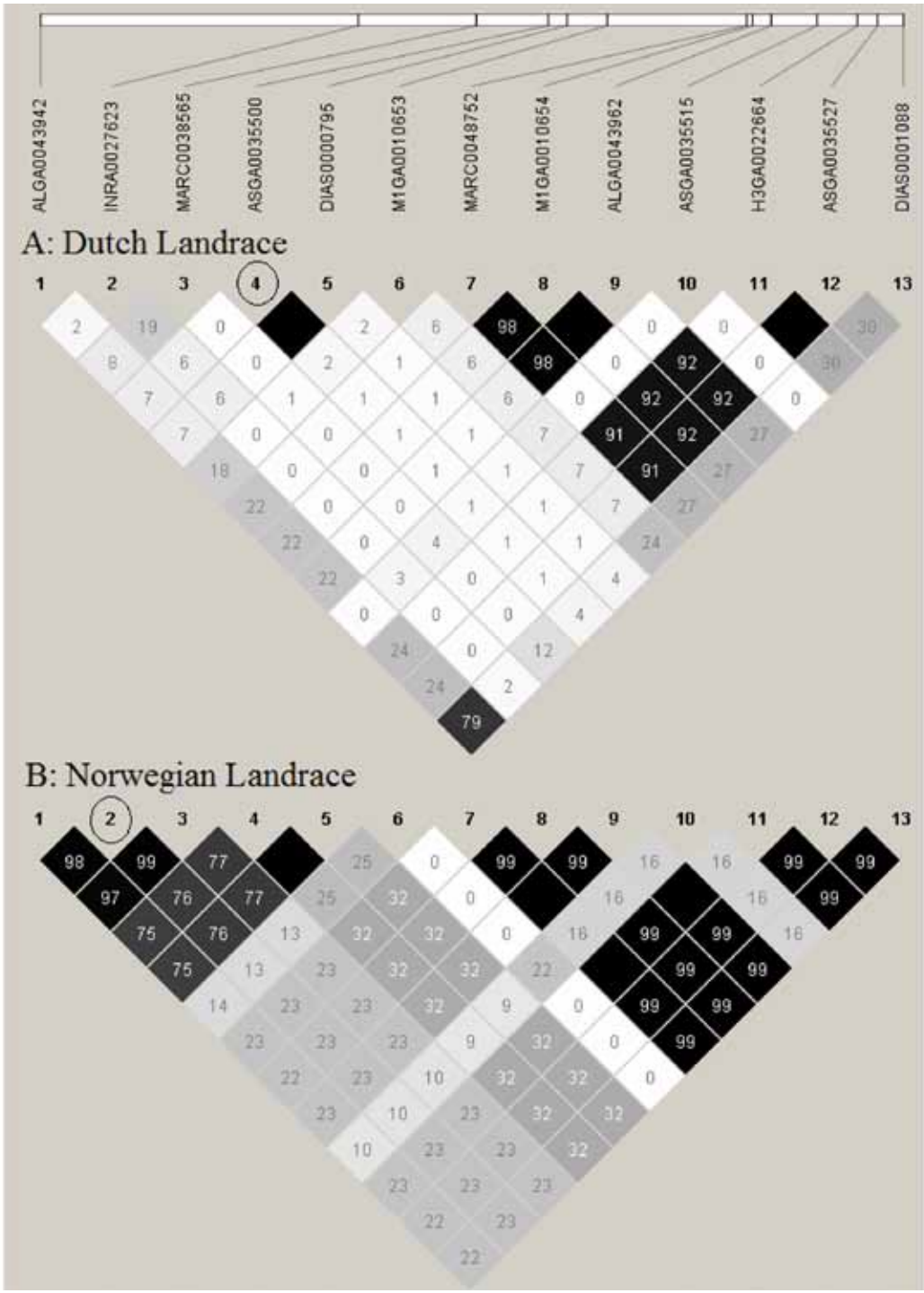

Figure 2. Linkage disequilibrium on chromosome 7. Linkage disequilibrium (r^2) between SNP located between 103 and 104 Mb in the Dutch Landrace population (A) and the Norwegian Landrace Population (B). The most significant SNP in each population is marked with a circle. The numbers inside the diamonds are the r^2 values on a scale of 0 to 100%.

BLUP, GBLUP, and MA-GBLUP were 0.315, 0.336, 0.474, and 0.477, respectively (Table 4). As expected, the prediction accuracies tended to decrease with the smaller training population. The decrease was bigger for the traditional models (BLUP and GBLUP), indicating that MAS has more added value with smaller training populations. Increases in accuracy were 0.021 (MA-BLUP) and 0.003 (MA-GBLUP) compared with BLUP and GBLUP, respectively. With the reduced data set, these increases were 0.052 and 0.023, respectively. Although these increase in prediction accuracy followed our expectations, it is also important to keep in mind that these results might also reflect the differences in the relationship between the training and validation animals from the 2 data sets.

www.manaraa.com

Table 4. Accuracy of prediction and bias

Population	BLUP	MA-BLUP[1]	GBLUP[2]	MA-GBLUP[3]	BVS[4]
Accuracy[5]					
Large White	0.238	0.266	0.361	0.370	0.383
Dutch Landrace	0.199	0.259	0.239	0.271	0.269
Norwegian Landrace	0.315	0.336	0.474	0.477	0.498
Duroc	0.192	0.316	0.319	0.362	0.359
Bias[6]					
Large White	0.84	0.87	0.97	0.96	1.00
Dutch Landrace	0.85	0.92	0.68	0.71	0.69
Norwegian Landrace	1.12	0.87	1.10	1.11	1.12
Duroc	0.82	1.01	0.80	0.88	0.86

[1]MA-BLUP = marker-assisted BLUP.

[2]GBLUP = genomic BLUP.

[3]MA-GBLUP = marker-assisted GBLUP.

[4]BVS = Bayesian variable selection.

[5]Accuracy is the correlation between the corrected phenotypes and breeding values.

[6]Bias is the regression coefficient of the corrected phenotypes on the breeding values.

As discussed above and in previous studies (Habier et al., 2007; Daetwyler et al., 2010; Wu et al., 2015), the accuracies of breeding values is influenced by the relationships between validation and training populations and the size of the reference population. However, the estimation of the SNP effect seems to be less affected by these 2 factors. The correlation between the marker breeding value (from both MA-BLUP and MA-GBLUP) and the corrected phenotype of the VALIDATION data set was 0.132 in the Large White, 0.150 in the Dutch Landrace, 0.175 in the Norwegian Landrace, and 0.260 in the Duroc population. These values seem to correlate with the proportion of phenotypic variance explained by the marker (approximately 3.48% in the dam lines and 6.13% in the Duroc) and not with the relationships between training and validation or the size of the reference population.

Table 5. Accuracy of prediction in the Norwegian Landrace population using multiple QTL regions

	MA-BLUP[2]			MA-GBLUP[3]		
QTL regions included[1]	$\hat{u}_g$	$\hat{u}_{snp}$	$\hat{u}$	$\hat{u}_g$	$\hat{u}_{snp}$	$\hat{u}$
4	0.315	–	0.315	0.474	–	0.474
7	0.296	0.175	0.336	0.453	0.173	0.477
7 and 14	0.302	0.245	0.372	0.423	0.245	0.474
7, 14, and 4	0.296	0.293	0.392	0.409	0.296	0.479
7, 14, 4, and 1	0.291	0.302	0.399	0.400	0.306	0.482

[1]Number of the chromosome from where the most significant SNP from the most pronounced peaks were selected to be included in the prediction analysis.

[2]MA-BLUP = marker-assisted BLUP; $\hat{u}_g$ = polygenic breeding value; $\hat{u}_{snp}$ = marker breeding value ($\sum_{m=1}^{M}\mathbf{X}_m\hat{\beta}_m$, with M being the number of markers included in the model, which ranged from 1 to 4); $\hat{u}$ = total breeding value ($\hat{u}_g + \hat{u}_{snp}$).

[3]MA-GBLUP = marker-assisted genomic BLUP.

[4]No SNP were used; therefore, it corresponds to traditional BLUP and genomic BLUP.

Using Multiple QTL

As previously discussed, and as expected (Sato et al., 2006; Guo et al., 2008; Ding et al., 2009; Duijvesteijn et al., 2014; Lopes et al., 2014), the QTL region on chromosome 7 was the most significant region for number of teats in all populations. However, in the Norwegian Landrace, highly significant peaks ($-\log_{10}$(P-value) > 10) were also observed on chromosomes 1, 4, and 14 (Fig. 1). As shown in Table 5, the prediction accuracy of the polygenic breeding value when no SNP was included in the model was 0.315 for BLUP and 0.474 for GBLUP. When all 4 SNP from chromosomes 1, 4, 7, and 14 were included in the model, the prediction accuracy of the polygenic breeding value was 0.291 for MA-BLUP and 0.400 for MA-GBLUP. This decrease in prediction accuracy was accompanied by a decrease in variance explained by the polygenic effect (from 0.27 to 0.22 for MA-BLUP and from 0.26 to 0.19 for MA-GBLUP). The decrease in polygenic variance indicates that the SNP included as fixed effects indeed were explaining part of the phenotypic variance of the evaluated trait. The variance explained by each SNP included in the model ranged from 0.01 to 0.03 (Supplemental Table S1; see the online version of the article at http://journalofanimalscience.org).

When using only the marker breeding value ($\hat{u}_{snp} = \sum_{m=1}^{4}\mathbf{X}_m\hat{\beta}_m$) based on the 4 markers described above, the prediction accuracy for the Norwegian Landrace was 0.302 (using SNP effects estimated in MA-BLUP; Table 5). This accuracy is similar to the polygenic breeding value accuracy from BLUP (0.315), indicating that for number of teats in this population, 4 markers have almost the same prediction ability as pedigree. With such accuracies, marker breeding values can be an important tool for selection in groups of animals that have no phenotypes or pedigree information, which is often the case in crossbred sows.

Furthermore, including all markers above our threshold of explaining >1% of the phenotypic variance lead to increased prediction accuracy. However, this threshold was arbitrary and the marker selection strategy will need to be further evaluated to apply the marker-assisted models. Performing a simulation study that evaluates all the different parameters such as number of QTL, effect sizes of QTL, heritability of the trait, and size of training

www.manaraa.com

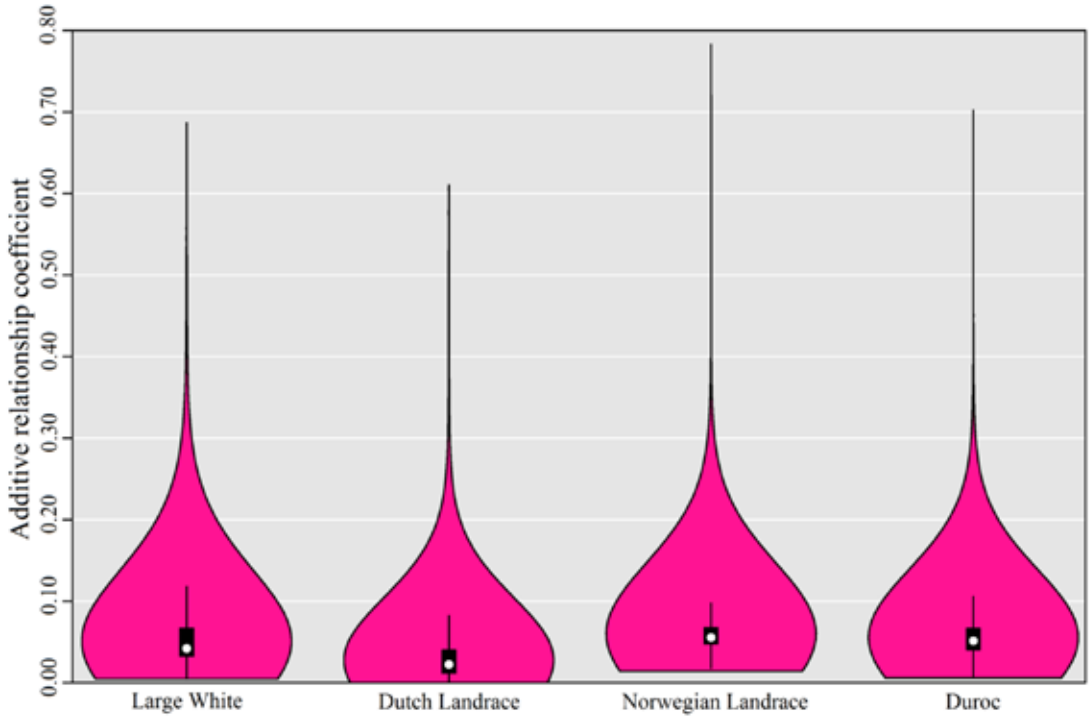

Figure 3. Additive relationship coefficient between animals from the TRAINING and VALIDATION data sets. Violin plot (box plot and probability density) of the pedigree-based relationship coefficient between the TRAINING and VALIDATION data sets of the 4 evaluated populations. The median relationship coefficient is indicated with a white dot inside the box plot.

and validation sets would be a good option to improve the marker selection strategy and also for further validation of the marker-assisted approach presented here.

Alternative Approaches

Brøndum et al. (2015) reported that the reliability of the breeding values increased up to 5 percentage points when accounting for GWAS findings compared with traditional GBLUP. In their study, these authors applied a model including 2 **G** matrices: one based on the markers from a 54,000 SNP panel and the other based on the significant markers from the whole-genome sequence data. In our study, we only used 1 **G** matrix, and the significant markers were included in the model as a fixed effect. Although the approach proposed by Brøndum et al. (2015) is interesting and showed an increase in the reliability of the breeding values, we expect that selecting only the most significant SNP per QTL region and including this SNP as a fixed effect in the model would result in higher prediction accuracies compared with using a second **G** matrix. This is because when including SNP as fixed effects in the model, a specific set of SNP can be used per trait, which gives a higher weight to each marker with large effect. On the other hand, building a second **G** matrix implies that all SNP for all traits are analyzed together under the assumption that all markers (including those not associated with the target trait) will explain the same proportion of the variance of all traits, which may limit the effect of markers associated with the target trait.

An alternative approach for using GWAS results in genetic predictions was described by Zhang et al. (2010). With this approach, the traditional **G** matrix in GBLUP is replaced with a trait-specific **G** matrix that gives different weights to each SNP. This approach favors (i.e., gives more weight to) SNP that contribute more to the genetic variance of the evaluated trait. In a traditional genomic relationship matrix (**G** matrix), all SNP are expected to contribute equally (i.e., they have the same weights). Zhang et al. (2010) showed that the breeding values from the model that applies the trait-specific **G** matrix were more accurate but also more biased than the breeding values from both BLUP and GBLUP. Recently, Veroneze et al. (2016) showed that using a trait-specific **G** matrix built using weights from a multipopulation GWAS increased the prediction accuracy of across-breed prediction compared with a traditional **G** matrix. The practical application of this approach is troublesome because the **G** matrix is trait specific and would, therefore, require single-trait genetic evaluations. However, breeding programs, in general, apply multitrait genetic evaluation to capitalize on the genetic correlations between traits. Using the MA-BLUP or MA-GBLUP, GWAS results could be incorporated into the genetic prediction using both

www.manaraa.com

single-trait and multitrait genetic evaluation, as the same **G** matrix could be used for all traits of interest.

For the marker-assisted models evaluated in this study, a modification of the **G** matrix would also be required if the double use of the SNP as both a fixed effect and a contributor to the **G** matrix would create problems. If double counting of effects would occur, the SNP used in the prediction models should not be used to build the **G** matrix. However, when we excluded the SNP used as a fixed effect and all other SNP in high LD ($r^2 > 0.50$) with it (Supplemental Table S2; see the online version of the article at http://journalofanimalscience.org) from the set of SNP used to build the **G** matrix, we obtained prediction accuracies and biases that were very similar to those obtained using all SNP (Supplemental Table S3, see the online version of the article at http://journalofanimalscience.org; Table 4), which indicates that there are limited effects due to double counting.

Another alternative for the use of GWAS results in genetic predictions was described by (Boichard et al., 2012). These authors showed that including a random effect of haplotypes in significant regions from GWAS was more accurate than traditional BLUP and GBLUP. The marker-assisted models presented in the current study are, however, easier than the haplotype approach because phasing of haplotypes is not required and a much lower number of SNP is included in the model.

Further Steps

The increasing amount of research aiming to develop models and methods for GS is showing the potential of this relatively novel breeding tool (Jonas and de Koning, 2015). With this study, we showed that with improved technologies, such as dense SNP panels, we can also revive "old" models and methods, such as MAS, to improve the accuracy of prediction. We found prediction to be improved using the marker-assisted models compared with BLUP and GBLUP models. In this study, we used the trait number of teats as a model trait because we had prior knowledge that an important QTL for this trait segregates in 2 of the 4 evaluated pig populations (Duijvesteijn et al., 2014; Lopes et al., 2014). Number of teats is a very relevant trait for breeding programs because a lower number of teats than the number of piglets increases suckling competition, which can result in lower preweaning growth and survival (Hirooka et al., 2001). However, this trait will not be the most relevant trait for GS because it can easily be recorded right after birth, is present in both sexes, and presents moderate heritability (Lopes et al., 2014). Therefore, as a further step, the marker-assisted models need to be considered for prediction of breeding values in other traits, especially those with well-defined QTL regions of large effect. In pigs, these traits would include, for example, androstenone level (Duijvesteijn et al., 2010; Hidalgo et al., 2014) and host response to porcine reproductive and respiratory syndrome virus challenge (Boddicker et al., 2012). In dairy cattle, traits affected by the *DGAT1* region on chromosome 14 (Jiang et al., 2010; Bouwman et al., 2012) would be some of the alternatives. However, QTL regions that explain a substantial proportion of the phenotypic variance will not be identified for all traits. In situations such as this (e.g., in the absence of QTL of large effect or traits affected by a large number of QTL with small effects), the application of traditional GBLUP is likely to be sufficient to obtain all or most of the advantages of genomic data for prediction.

Furthermore, the genetic evaluations in the major pig breeding companies are currently based on the so-called single-step genetic evaluations (Legarra et al., 2009; Misztal et al., 2009; Christensen and Lund, 2010), which can accommodate both genotyped and nongenotyped animals. For the incorporation of the markers-assisted models in single-step genetic evaluations, it will be required to estimate genotype probabilities (for the SNP used as a fixed effect in the model) for the nongenotyped animals. The estimation of genotype probabilities for nongenotyped animals has been described by Mulder et al. (2010) and Bouwman et al. (2014); however, its effectiveness in marker-assisted models needs to be evaluated.

Conclusions

For the model trait in this study, number of teats, BLUP resulted in the lowest prediction accuracies whereas the highest were observed when applying either MA-GBLUP or BVS. Results also show that MA-BLUP can yield similar or superior accuracies compared with GBLUP. The superiority of MA-GBLUP over traditional GBLUP is more pronounced when training populations are smaller and with more distant relationships between training and validation populations.

LITERATURE CITED

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2004. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21(2):263–265. doi:10.1093/bioinformatics/bth457

Boddicker, N., E. H. Waide, R. R. R. Rowland, J. K. Lunney, D. J. Garrick, J. M. Reecy, and J. C. M. Dekkers. 2012. Evidence for a major QTL associated with host response to porcine reproductive and respiratory syndrome virus challenge. J. Anim. Sci. 90:1733–1746. doi:10.2527/jas.2011-4464

Boichard, D., F. Guillaume , A. Baur , P. Croiseau , M. N. Rossignol , M. Y. Boscher , T. Druet , L. Genestout , J. J. Colleau , L. Journaux, V. Ducrocq, and S. Fritz. 2012. Genomic selection in French dairy cattle. Anim. Prod. Sci. 52:115–120. doi:10.1071/AN11119

www.manaraa.com

Bouwman, A. C., J. M. Hickey, M. P. Calus, and R. F. Veerkamp. 2014. Imputation of non-genotyped individuals based on genotyped relatives: Assessing the imputation accuracy of a real case scenario in dairy cattle. Genet. Sel. Evol. 46:6. doi:10.1186/1297-9686-46-6

Bouwman, A. C., M. H. Visker, J. A. van Arendonk, and H. Bovenhuis. 2012. Genomic regions associated with bovine milk fatty acids in both summer and winter milk samples. BMC Genet. 13:93. doi:10.1186/1471-2156-13-93

Brøndum, R. F., G. Su, L. Janss, G. Sahana, B. Guldbrandtsen, D. Boichard, and M. Lund. 2015. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. J. Dairy Sci. 98:4107–4116. doi:10.3168/jds.2014-9005

Browning, S. R., and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81:1084–1097. doi:10.1086/521987

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

Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021–1031. doi:10.1534/genetics.110.116855

Ding, N., Y. Guo, C. Knorr, J. Ma, H. Mao, L. Lan, S. Xiao, H. Ai, C. S. Haley, B. Brenig, and L. Huang. 2009. Genome-wide QTL mapping for three traits related to teat number in a White Duroc × Erhualian pig resource population. BMC Genet. 10:6. doi:10.1186/1471-2156-10-6

Duijvesteijn, N., E. F. Knol, J. W. M. Merks, R. P. M. A. Crooijmans, M. A. M. Groenen, H. Bovenhuis, and B. Harlizius. 2010. A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC Genet. 11:42. doi:10.1186/1471-2156-11-42

Duijvesteijn, N., J. M. Veltmaat, E. F. Knol, and B. Harlizius. 2014. High-resolution association mapping of number of teats in pigs reveals regions controlling vertebral development. BMC Genomics 15:542. doi:10.1186/1471-2164-15-542

Fernando, R., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467. doi:10.1186/1297-9686-21-4-467

George, E. I., and R. E. McCulloch. 1993. Variable selection via Gibbs sampling. J. Am. Stat. Assoc. 88:881–889. doi:10.1080/01621459.1993.10476353

Gilmour, A. R., B. Gogel, B. Cullis, and R. Thompson. 2009. ASReml user guide release 3.0. VSN International Ltd., Hemel Hempstead, UK.

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica (The Hague) 136:245–257.

Groenen, M. A. M., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard, C. Park, D. Milan, H.-J. Megens, S. Li, D. M. Larkin, H. Kim, L. A. F. Frantz, M. Caccamo, H. Ahn, B. L. Aken, A. Anselmo, C. Anthon, L. Auvil, B. Badaoui, C. W. Beattie, C. Bendixen, D. Berman, F. Blecha, J. Blomberg, L. Bolund, M. Bosse, S. Botti, Z. Bujie, M. Bystrom, B. Capitanu, D. Carvalho-Silva, P. Chardon, C. Chen, R. Cheng, S.-H. Choi, W. Chow, R. C. Clark, C. Clee, R. P. M. A. Crooijmans, H. D. Dawson, P. Dehais, F. De Sapio, B. Dibbits, N. Drou, Z.-Q. Du, K. Eversole, J. Fadista, S. Fairley, T. Faraut, G. J. Faulkner, K. E. Fowler, M. Fredholm, E. Fritz, J. G. R. Gilbert, E. Giuffra, J. Gorodkin, D. K. Griffin, J. L. Harrow, A. Hayward, K. Howe, Z.-L. Hu, S. J. Humphray, T. Hunt, H. Hornshøj, J.-T. Jeon, P. Jern, M. Jones, J. Jurka, H. Kanamori, R. Kapetanovic, J. Kim, J.-H. Kim, K.-W. Kim, T.-H. Kim, G. Larson, K. Lee, K.-T. Lee, R. Leggett, H. A. Lewin, Y. Li, W. Liu, J. E. Loveland, Y. Lu, J. K. Lunney, J. Ma, O. Madsen, K. Mann, L. Matthews, S. McLaren, T. Morozumi, M. P. Murtaugh, J. Narayan, D. T. Nguyen, P. Ni, S.-J. Oh, S. Onteru, F. Panitz, E.-W. Park, H.-S. Park, G. Pascal, Y. Paudel, M. Perez-Enciso, R. Ramirez-Gonzalez, J. M. Reecy, S. Rodriguez-Zas, G. A. Rohrer, L. Rund, Y. Sang, K. Schachtschneider, J. G. Schraiber, J. Schwartz, L. Scobie, C. Scott, S. Searle, B. Servin, B. R. Southey, G. Sperber, P. Stadler, J. V. Sweedler, H. Tafer, B. Thomsen, R. Wali, J. Wang, J. Wang, S. White, X. Xu, M. Yerle, G. Zhang, J. Zhang, J. Zhang, S. Zhao, J. Rogers, C. Churcher, and L. B. Schook. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491:393–398. doi:10.1038/nature11622

Guo, Y. M., G. Lee, A. Archibald, and C. Haley. 2008. Quantitative trait loci for production traits in pigs: A combined analysis of two Meishan × Large White populations. Anim. Genet. 39:486–495. doi:10.1111/j.1365-2052.2008.01756.x

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

Heffner, E. L., M. E. Sorrells, and J.-L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1–12. doi:10.2135/cropsci2008.08.0512

Hidalgo, A. M., J. W. M. Bastiaansen, B. Harlizius, H.-J. Megens, O. Madsen, R. P. M. A. Crooijmans, and M. A. M. Groenen. 2014. On the relationship between an Asian haplotype on chromosome 6 that reduces androstenone levels in boars and the differential expression of SULT2A1 in the testis. BMC Genet. 15:4. doi:10.1186/1471-2156-15-4

Hirooka, H., D. J. de Koning, B. Harlizius, J. A. van Arendonk, A. P. Rattink, M. A. Groenen, E. W. Brascamp, and H. Bovenhuis. 2001. A whole-genome scan for quantitative trait loci affecting teat number in pigs. J. Anim. Sci. 79(9):2320–2326. doi:10.2527/2001.7992320x

Illumina Inc. 2005. Technology spotlight. Illumina GenCall data analysis software: GenCall software algorithms for clustering, calling, and scoring genotypes. Illumina Publ. No. 370-2004-009. Illumina, Inc., San Diego, CA. Accessed October 13, 2016. http://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf.

Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5:e13661. doi:10.1371/journal.pone.0013661

Jonas, E., and D.-J. de Koning. 2015. Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Front. Genet. 6:49. doi:10.3389/fgene.2015.00049

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

Lopes, M. S., J. W. Bastiaansen, B. Harlizius, E. F. Knol, and H. Bovenhuis. 2014. A genome-wide association study reveals dominance effects on number of teats in pigs. PLoS One 9:e105867. doi:10.1371/journal.pone.0105867

Lopes, M. S., F. F. Silva, B. Harlizius, N. Duijvesteijn, P. S. Lopes, S. E. F. Guimarães, and E. F. Knol. 2013. Improved estimation of inbreeding and kinship in pigs using optimized SNP panels. BMC Genet. 14:92. doi:10.1186/1471-2156-14-92

Meuwissen, T., and M. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. 28:161–176. doi:10.1186/1297-9686-28-2-161

Meuwissen, T., and M. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185:623–631. doi:10.1534/genetics.110.116590

Meuwissen, T., B. Hayes, and M. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648–4655. doi:10.3168/jds.2009-2064

Mulder, H. A., M. P. Calus, and R. F. Veerkamp. 2010. Prediction of haplotypes for ungenotyped animals and its effect on marker-assisted breeding value estimation. Genet. Sel. Evol. 42:10. doi:10.1186/1297-9686-42-10

www.manaraa.com

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575. doi:10.1086/519795

Sato, S., K. Atsuji, N. Saito, M. Okitsu, S. Sato, A. Komatsuda, T. Mitsuhashi, K. Nirasawa, T. Hayashi, Y. Sugimoto, and E. Kobayashi. 2006. Identification of quantitative trait loci affecting corpora lutea and number of teats in a Meishan × Duroc F_2 resource population. J. Anim. Sci. 84:2895–2901. doi:10.2527/jas.2006-176

Sell-Kubiak, E., N. Duijvesteijn, M. S. Lopes, L. L. G. Janss, E. F. Knol, P. Bijma, and H. A. Mulder. 2015. Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. BMC Genomics 16:1049. doi:10.1186/s12864-015-2273-y

VanRaden, P. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980

Verardo, L. L., F. F. Silva, M. S. Lopes, O. Madsen, J. W. M. Bastiaansen, E. F. Knol, M. Kelly, L. Varona, P. S. Lopes, and S. E. F. Guimarãe. 2016. Revealing new candidate genes for reproductive traits in pigs: Combining Bayesian GWAS and functional pathways. Genet. Sel. Evol. 48(1):9. doi:10.1186/s12711-016-0189-x

Veroneze, R., P. S. Lopes, M. S. Lopes, A. M. Hidalgo, S. E. F. Guimarães, B. Harlizius, E. F. Knol, J. A. M. van Arendonk, F. F. Silva, and J. W. M. Bastiaansen. 2016. Accounting for genetic architecture in single- and multipopulation genomic prediction using weights from genomewide association studies in pigs. J. Anim. Breed. Genet. 133:187–196. doi:10.1111/jbg.12202

Wu, X., M. Lund, D. Sun, Q. Zhang, and G. Su. 2015. Impact of relationships between test and training animals and among training animals on reliability of genomic prediction. J. Anim. Breed. Genet. 132:366–375. doi:10.1111/jbg.12165

Zhang, Z., J. Liu, X. Ding, P. Bijma, D. J. de Koning, and Q. Zhang. 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. PLoS One 5:e12648. doi:10.1371/journal.pone.0012648

www.manaraa.com

Reproduced with permission of copyright owner.
Further reproduction prohibited without permission.

www.manaraa.com