Marker-assisted breeding value estimation for mastitis resistance in Finnish Ayrshire cattle

H. A. Mulder,^{*1} M. H. Lidauer,[†] J. H. Vilkki,[†] I. Strandén,[†] and R. F. Veerkamp^{*}

^{*}Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, the Netherlands

[†]MTT Agrifood Research Finland, Biotechnology and Food Research, Genetics Research, FI-31600 Jokioinen, Finland

ABSTRACT

Marker-assisted breeding value estimation is expected to increase the accuracy of estimated breeding values, especially for traits with low heritability. Several quantitative trait loci (QTL) have been found for somatic cell score and clinical mastitis. The objective of this study was to demonstrate marker-assisted breeding value estimation, combining data of genotyped and ungenotyped animals in a large pedigree population using either identical-by-descent (IBD) or identical-by-state (IBS) haplotypes for some previously identified QTL regions for somatic cell score and clinical mastitis in Finnish Ayrshire cattle. For both methods, QTL variances were estimated based on daughter yield deviations of genotyped bulls. The QTL explained only a small proportion of genetic variance, especially with IBS haplotypes. Using IBD haplotypes gave more reranking of bulls and cows than using IBS haplotypes. Cross-validation showed no increase in predictive ability when using IBS haplotypes compared with conventional breeding value estimation, whereas a decrease in predictive ability was observed with IBD haplotypes. Furthermore, computing time was lower and convergence was better with IBS haplotypes than with IBD haplotypes. In this study on mastitis resistance in Finnish Ayrshire, conventional breeding value estimation would be advocated because of the lack in improvement of accuracy and predictive ability when using marker-assisted breeding value estimation. However, in situations where IBS haplotypes would explain 10 to 20% or more of the genetic variance, marker-assisted breeding value estimation with IBS haplotypes may yield greater accuracy and predictive ability than conventional breeding value estimation.

Key words: mastitis, marker-assisted breeding value estimation, haplotype, ungenotyped animal

INTRODUCTION

Mastitis is one of the most costly diseases in dairy cattle. It has a low heritability, and recording is lacking in most countries. Therefore, in most countries, SCS, which is measured routinely in milk recording, is often used as an indicator trait for mastitis resistance. The advantage of SCS is that it is continuous and easy to measure, and it has higher heritability than mastitis resistance. Advanced use of the dynamics of SCS during lactation can further improve the accuracy of EBV for mastitis resistance (de Haas et al., 2008; Windig et al., 2010). In the Nordic countries, mastitis incidence is recorded on a national basis, making it possible to use direct information to predict EBV for mastitis resistance.

Another way to increase the accuracy of EBV for mastitis resistance is to perform marker-assisted breeding value estimation. Marker-assisted breeding value estimation increases accuracy mainly for genotyped animals, and the increase in accuracy is larger for low heritable traits compared with moderately or highly heritable traits (Lande and Thompson, 1990; Meuwissen and Goddard, 1996; Mulder et al., 2010a). Due to its low heritability, mastitis resistance is, therefore, a candidate trait for marker-assisted breeding value estimation. Considerable efforts have been made to find QTL for SCS and mastitis resistance (e.g., Schulman et al., 2004, 2009; Lund et al., 2007, 2008; Sahana et al., 2008; see review by Khatkar et al., 2004) or for SCS solely (e.g., Schrooten et al., 2000; Cole et al., 2009). In Nordic breeds, QTL have been reported on chromosomes 9, 14, and 18 (e.g., Schulman et al., 2004, 2009; Lund et al., 2007, 2008; Sahana et al., 2008). In other studies, QTL have been found on other chromosomes (Rupp and Boichard, 2003; Khatkar et al., 2004). Currently, high-density SNP chips are used for finding QTL; that is, genome-wide associations (e.g., Cole et al., 2009; Verbyla et al., 2010). Although some QTL have been found, marker-assisted breeding value estimation has not yet been implemented for mastitis resistance. One of the major challenges in this method is dealing with genotyped and ungenotyped animals in

Received December 20, 2010.

Accepted April 6, 2011.

¹Corresponding author: herman.mulder@wur.nl

Table 1. Characteristics and number of genotyped bulls for each marker for the 5 selected QTL regions

Marker	Chromosome	Position (cM)	QTL designation	Bulls (n)	Reference
BM4208	BTA9	74.0	1	739	Sahana et al. (2008)
BMS2819	BTA9	74.0	1	1,121	Sahana et al. (2008)
INRA144	BTA9	74.2	1	695	Sahana et al. (2008)
MAP4K4	BTA11	10.5	2	817	Schulman et al. (2009)
MNB40	BTA11	16.0	3	626	Schulman et al. (2009)
BM716	BTA11	17.9	3	708	Schulman et al. (2009)
DIK2653	BTA11	18.1	3	608	Schulman et al. (2009)
HELMTT44	BTA11	61.2	4	940	Schulman et al. (2009)
TGLA58	BTA11	63.1	4	1,126	Schulman et al. (2009)
TGLA227	BTA18	110.0	5	1,110	Viitala et al. (2003)

one evaluation. One approach is to use identical-by-descent (**IBD**) haplotypes with use of IBD matrices to account for phase differences between QTL and markers. A second approach is to use identical-by-state (**IBS**) haplotypes, which do not require setting up large IBD matrices. Mulder et al. (2010a) developed a simple method to predict haplotypes of ungenotyped animals based on the mixed model method as presented in Gengler et al. (2007, 2008). With simulated data, they showed that with 4-marker haplotypes spaced at 0.1 cM, 90% of the QTL variance could be captured. However, so far this method has only been tested with simulated data and not directly compared with a method using IBD haplotypes. In addition, no study is known in which both methods were implemented in a large-scale, marker-assisted genetic evaluation.

The objective of this study was to demonstrate marker-assisted breeding value estimation combining data of genotyped and ungenotyped animals in a large pedigree population using IBD or IBS haplotypes, using some previously identified QTL regions for SCS and clinical mastitis in Finnish Ayrshire cattle.

MATERIALS AND METHODS

Marker Data

In this study, we focused on Finnish Ayrshire cattle. In previous studies, several bulls were genotyped for different regions of the genome to identify QTL (Viitala et al., 2003; Schulman et al., 2004; Sahana et al., 2008). These bulls were in a granddaughter design originating from 12 sires. In total, we had 3 generations of genotyped bulls, and birth years of these bulls ranged between 1975 and 2003. In total, genotypes of 1,157 bulls were available. We selected 10 markers belonging to 5 identified QTL regions being associated with either SCS or clinical mastitis (e.g., Viitala et al., 2003; Schulman et al., 2004; Sahana et al., 2008). Table 1 shows the number of genotyped bulls for each marker.

In total, 498 bulls had genotypes for all markers and 640 bulls had at least one marker for each QTL region.

Phenotypic Data

From the national Finnish database, first-lactation records of cows were extracted from cows calving the first time between January 1, 1999, and December 31, 2008, a 10-yr period resulting in 591,596 records for lactation-average SCS, calculated as $\ln(\text{lactation-average SCC}/1,000)$; clinical mastitis between d -15 and d 50 (**CM1**); and clinical mastitis between d 51 and d 300 (**CM2**). Table 2 shows some characteristics of the phenotypic data in relation to the genotyped bulls. The Finnish Ayrshire pedigree was pruned to include only cows with observations and cows and bulls that built up informative links between the cows with observations. An animal without observation was pruned if it had no offspring or if it had no ancestors and fewer than 2 offspring. The pruned pedigree contained 1,124,763 animals.

Construction of IBS and IBD Haplotypes

For the genotyped bulls, haplotypes were constructed and phased using Haprec (Windig and Meuwissen, 2004). The phased genotypes were subsequently used to construct either IBS or IBD haplotypes. The IBD haplotypes were constructed using linkage disequilibrium and linkage analysis information (**LDLA**) following Meuwissen and Goddard (2001) assuming an effective population size of 100 and 100 generations between the first generation in the pedigree and the founder generation. All markers on a chromosome were used to calculate IBD probabilities for each QTL region. In the case of QTL region 1 (QTL1), the QTL was assumed to be in the middle between microsatellite markers 2 (BMS2819) and 3 (INRA144) (Sahana et al., 2008); for QTL region 2 (QTL2), the QTL was assumed to be at the SNP marker (MAP4K4); for QTL region 3 (QTL3),

Table 2. Summary statistics of phenotypic data used in breeding value estimation

Parameter	Overall	Trait ¹		
		SCS	CM1	CM2
Records (n)	591,596	578,626	563,523	523,663
Average		0.688	0.072	0.065
SD		0.433	0.259	0.246
Minimum		0.005	0.000	0.000
Maximum		3.197	1.000	1.000
Genotyped bulls (n)	1,157			
Genotyped bulls with daughters with records (n)	714			
Daughters with records of genotyped bulls (n)	255,219			
Average daughters/genotyped bull (n)	357			
Median daughters/genotyped bull (n)	144			

¹SCS = lactation-average SCS; CM1 = clinical mastitis between d -15 and 50; CM2 = clinical mastitis between d 51 and 300.

the QTL was assumed to be between microsatellite markers 1 (MNB40) and 2 (BM716), because of the large distance and the fact that markers 2 (BM716) and 3 (DIK2653) are very close to each other. For QTL region 4 (QTL4), the QTL was assumed to be between the 2 microsatellite markers (HELMT44 and TGLA58); and for QTL region 5 (QTL5), the QTL was assumed to be at the microsatellite marker (TGLA227), in absence of flanking markers. Distances between markers were based on Haldane's mapping function (Haldane, 1919). For animals with no genotypes, haplotypes and IBD-matrices were constructed using the Fernando and Grossman (1989) rules.

The IBS haplotypes were constructed based on the phased genotypes and on the alleles of the different microsatellites. For QTL2, the IBS haplotype was based on a single SNP allele and for QTL5 the IBS haplotypes were based on different alleles of the microsatellite. For some bulls, no genotypes were available for certain markers in a QTL region. It was assumed that those bulls had a zero allele for those markers not genotyped. Subsequently, markers were concatenated and renumbered from 1 to the maximum number of haplotypes. In this way, we constructed 143, 2, 163, 43, and 11 haplotypes for QTL 1, 2, 3, 4, and 5, respectively. The next stage was to predict the number of haplotype alleles (*n_{hc}*) for ungenotyped animals using the method described in Mulder et al. (2010a).

Variance Component Estimation for Each QTL Region

The QTL variances were estimated, because QTL effects were modeled as random effects with either IBS or IBD haplotypes. The QTL variances were estimated using daughter yield deviations (**DYD**) of the genotyped bulls with effective daughter contributions (**EDC**) as weights on the residual variance to account

for differences in accuracy of the DYD. The calculated DYD (Mrode and Swanson, 2004) and EDC (Fikse and Banos, 2001) were based on all Finnish phenotypic data of 1.3 million records used in the joint Nordic evaluation. For the variance component analysis, we used only bulls that had at least one marker for each QTL region genotyped (640 bulls).

Three univariate analyses were done in ASREML (Gilmour et al., 2006) fitting the model

$$DYD = \mu + u_{pol} + \sum_{i=1}^5 (h_{i,p} + h_{i,m}) + e, \quad [1]$$

where *DYD* is the daughter yield deviation for SCS, CM1, or CM2; μ is the overall mean; u_{pol} is the random polygenic effect $\left[N\left(\mathbf{0}, \mathbf{A}\sigma_{u_{pol}}^2 \right) \right]$, where **A** is the additive genetic or numerator relationship matrix and $\sigma_{u_{pol}}^2$ is the polygenic genetic variance; $h_{i,p}$ ($h_{i,m}$) is the random paternal (maternal) haplotype effect for QTL *i*; and *e* is the random residual $\left[N\left(\mathbf{0}, \mathbf{D}\sigma_e^2 \right) \right]$, where **D** is the diagonal matrix with the reciprocal of EDC of each bull (Mrode, 2005) and σ_e^2 is the residual variance. When using IBD haplotypes the haplotype effects were assumed to be distributed as $N\left(\mathbf{0}, \mathbf{IBD}_i\sigma_{h_i}^2 \right)$, where **IBD_i** is the IBD matrix based on LDLA for QTL *i*, and $\sigma_{h_i}^2$ is the haplotype variance of QTL *i*. When using IBS haplotypes, the haplotype effects were assumed to be independent $\left[N\left(\mathbf{0}, \mathbf{I}\sigma_{h_i}^2 \right) \right]$. The estimated variances per QTL were calculated as proportions of the total genetic variance. Multivariate analyses did not converge.

Breeding Value Estimation

Three multivariate models were applied for breeding value estimation: conventional breeding value estima-

tion (**CONBLUP**), marker-assisted breeding value estimation with IBD haplotypes (**MABLUP-IBD**), and marker-assisted breeding value estimation using IBS haplotypes (**MABLUP-IBS**). A multivariate model was used for breeding value estimation. Effects in the CONBLUP model were based on the Nordic evaluation before 2010 (Johansson et al., 2006):

$$y = H5Y + AGE + YM + hy + u_{pol} + e, \quad [2]$$

where y is SCS, CM1, and CM2; $H5Y$ is the fixed effect for herd-5-yr period; AGE is fixed age at first calving as a fixed class effect; YM is the fixed year-month effect; hy is the random herd-year effect [$MVN(\mathbf{0}, \mathbf{I} \otimes \mathbf{HY})$], where MVN = multivariate normal; u_{pol} is the random polygenic effect [$MVN(\mathbf{0}, \mathbf{A} \otimes \mathbf{G})$]; e is the random residual [$MVN(\mathbf{0}, \mathbf{I} \otimes \mathbf{R})$]; $\mathbf{G}$ is the polygenic variance-covariance matrix; $\mathbf{HY}$ is the diagonal variance-covariance matrix of herd-year effects; $\mathbf{I}$ is the identity matrix; and $\mathbf{R}$ is the residual variance-covariance matrix.

The model for MABLUP-IBD was the combination of model [2] with model [1]:

$$y = H5Y + AGE + YM + hy + u_{pol} + \sum_{i=1}^{i=5} (h_{i,p} + h_{i,m}) + e, \quad [3]$$

where $h_{i,p}$ ($h_{i,m}$) is the paternal (maternal) haplotype; [$MVN(\mathbf{0}, \mathbf{IBD}_i \otimes \mathbf{H}_i)$] and $\mathbf{H}_i$ is the variance-covariance matrix for QTL i . The haplotype EBV was $\sum_{i=1}^{i=5} (h_{i,p} + h_{i,m})$, and the total EBV was calculated as $u_{pol} + \sum_{i=1}^{i=5} (h_{i,p} + h_{i,m})$, the sum of the haplotype EBV and polygenic effect.

The model for MABLUP-IBS was model [2] extended with a regression on the number of haplotype alleles (see Mulder et al., 2010a) and therefore slightly differently parameterized than model [1], but equivalent to model [1]:

$$y = H5Y + AGE + YM + hy + u_{pol} + \sum_{i=1}^5 \sum_{j=1}^{n_i} (n\hat{h}c_{ij} \times h_{ij}) + e, \quad [4]$$

where $n\hat{h}c_{ij}$ is the number of haplotype alleles for haplotype j of QTL i ; h_{ij} is the regression coefficient for haplotype j of QTL i [$MVN(\mathbf{0}, \mathbf{I} \otimes \mathbf{H}_i)$]; and n_i is the number of haplotypes used for QTL i . Because of the large number of haplotypes for QTL 1, 3, and 4, and to limit computing time, we selected for each animal the

20 haplotypes with the largest $n\hat{h}c_{ij}$, because the others had rather small values, close to zero, which would have had a negligible effect on breeding value estimation. When using the 20 haplotypes with the largest $n\hat{h}c_{ij}$, the average sum of $n\hat{h}c_{ij}$ was 1.76, 1.42, and 1.96, respectively for QTL1, QTL3, and QTL4, indicating that most of the haplotype information was captured. The haplotype EBV was $\sum_{i=1}^{i=5} \sum_{j=1}^n (n\hat{h}c_{ij} \times h_{ij})$ and the total

EBV was calculated as $u_{pol} + \sum_{i=1}^{i=5} \sum_{j=1}^n (n\hat{h}c_{ij} \times h_{ij})$, the sum of the haplotype EBV and polygenic effect.

Variance components for herd-year effects (**HY**), polygenic effects, and residual effects (**R**) were provided by E. Negussie (MTT Agrifood Research Finland, Biotechnology and Food Research, Genetics Research, Jokioinen, Finland, personal communication). The variance components for QTL effects (**H_i**) were derived using the proportions of total genetic variance based on the variance component estimation for each QTL region. The variance due to polygenic effects (**G**) was the total additive genetic variance minus the sum of the QTL variances. The covariances for QTL effects were derived assuming that the genetic correlations for QTL effects were equal to the polygenic genetic correlations. All breeding value estimations were performed with MiXBUP (Mulder et al., 2010b; www.mixbup.eu).

Cross-Validation

We validated the methods using cross-validation in 2 ways: (1) on bulls and (2) on cows. For bulls, the validation set consisted of 104 genotyped bulls born in 2002 and 2003 and having a reliability of at least 0.80 for SCS. Reliabilities of conventional EBV were calculated using the approximation by Tier and Meyer (2004). By means of this criterion, the EBV could be considered as proxies of the true EBV. The restriction was not applied for CM1 and CM2 because too few bulls met the criterion of reliability of 0.80 for these traits. Records of daughters of these 104 bulls were set to missing, and CONBLUP, MABLUP-IBD, and CONBLUP-IBS were run on the reduced data set. The correlation between the EBV of the whole data set and EBV of the reduced data set were calculated. The correlation is an indication of the accuracy that can be obtained for young bulls before going into a progeny test.

For cows, we performed a 10-fold cross-validation where 10% of the cow phenotypes were deleted. For each cow with phenotypes, the phenotypes were set to missing in 1 of the 10 reduced data sets. To limit

computing time, we performed only CONBLUP and MABLUP-IBS on each of the 10 reduced data sets. Instead of calculating correlations between EBV and observed phenotypes for validation cows, which is not very meaningful for binomial traits, we calculated the incidence of CM1 and CM2 per 10%-quantile of EBV, either from CONBLUP or the total EBV from MABLUP-IBS. The incidence was averaged across the 10 replicates.

RESULTS

Variance Component Estimation for Each QTL Region

Estimated variance components and the proportions of genetic variance explained by QTL were sensitive to the model used and the markers used to construct IBD matrices (results not shown). For instance, excluding the polygenic effect from equation [1] resulted in extremely high proportions of genetic variance explained, up to 71.6%, due to capturing family-specific effects rather than true QTL effects of the genetic variance (e.g., Kennedy et al., 1992). In addition, the genetic variance explained by QTL2 with IBD depended largely on whether the IBD matrices were based on the single SNP or on all markers on chromosome 11. When the IBD matrices were constructed using only the single SNP (MAP4K4) on chromosome 11, the SNP explained 57.2% of the genetic variance, whereas it explained very little variance when the IBD matrices were constructed using all markers on chromosome 11.

The final estimated QTL variances (Table 3) were estimated with either IBD or IBS haplotypes and expressed as proportions of the total genetic variance (Table 4). The model with IBD haplotypes explained more genetic variance than the model with IBS haplotypes. In both models, QTL1 and QTL2 explained hardly any genetic variance, whereas QTL4 and QTL5 explained substantial proportions of genetic variance, especially for CM1 and CM2. Strikingly, QTL3 explained 13.1% of the genetic variance for CM1 when using IBD, whereas it explained hardly any variance with IBS. Standard errors on the estimated variances per QTL were large relative to the estimate, indicating low precision of the estimated variance components. The estimated residual variances as well as the total amount of genetic variance were very similar across models and similar as well to the estimates by Negussie (MTT Agrifood Research Finland, Biotechnology and Food Research, Genetics Research, Jokioinen, Finland, personal communication). Note that the estimated polygenic and QTL variances were one-quarter of the true polygenic or QTL variance because of using DYD.

Breeding Value Estimation

Table 5 shows correlations between conventional EBV and marker-assisted total, polygenic, and haplotype EBV using either IBD or IBS haplotypes. Correlations between conventional EBV and total EBV were higher when using IBS haplotypes than when using IBD haplotypes, indicating less reranking; reranking was smallest for SCS. Correlations were quite similar for bulls and cows. The correlation between haplotype EBV and conventional EBV was high for IBD haplotypes but low for IBS haplotypes. Furthermore, the correlations between haplotype EBV of both methods were between 0.12 and 0.32, indicating that both methods are different and that IBD haplotypes may capture family-specific effects rather than QTL effects. It can be concluded that IBD haplotypes caused more reranking than IBS haplotypes; IBD haplotypes also explained more genetic variance than IBS haplotypes.

Predictive Ability

Bulls. In all cases, the results (Table 6) of cross-validation based on bulls indicated that predictive ability was lower for MABLUP than for CONBLUP, except for CM2 in MABLUP-IBS, where it was slightly higher. The correlations between EBV of the full and reduced data sets were, in general, high for these traits with a low heritability, because pedigree information was still weighted heavily in both sets of EBV, inflating the correlation due to partly using the same information. Correlations between IBS haplotype EBV of the full and reduced data sets were very high (>0.92), whereas correlations between the IBD haplotype EBV of the full and reduced data sets were quite low. This indicates that effects of IBS haplotypes had higher predictive ability than IBD haplotypes.

Cows. Figure 1A and 1B show, respectively, the incidence of CM1 and CM2 for 10%-quantiles of the EBV for CM1 and CM2 when all data were used. The figures show a large difference in incidence between the cows with the highest (unfavorable) and the lowest (favorable) EBV. For CM1 (Figure 1A), differences between methods were very small. For CM2 (Figure 1B), however, MABLUP-IBS had a higher incidence for cows with the highest EBV than CONBLUP and MABLUP-IBD. In this case, the cows' own phenotypes were used to estimate the EBV. In Figure 2, the incidence is shown for cross-validation cows, whose phenotypes were set to missing in the breeding value estimation using CONBLUP or MABLUP-IBS. Although the variation in incidence of CM1 and CM2 is much smaller than in Figure 1, Figure 2 shows that EBV were predictive for mastitis incidence, indicating that cows with

Table 3. Variance components (SE in parentheses) for a conventional BLUP model and for MABLUP-IBD and MABLUP-IBS for 5 selected QTL regions^{1,2}

Variance	SCS			CM1			CM2		
	CONBLUP	MABLUP-IBD	MABLUP-IBS	CONBLUP	MABLUP-IBD	MABLUP-IBS	CONBLUP	MABLUP-IBD	MABLUP-IBS
Polygenic	4.51E-03 (6.21E-04)	4.26E-03 (6.32E-04)	4.44E-03 (6.21E-04)	3.46E-04 (7.27E-05)	2.78E-04 (7.86E-05)	3.46E-04 (7.31E-05)	3.69E-04 (7.39E-05)	3.44E-04 (7.69E-05)	3.75E-04 (7.46E-05)
QTL1		6.58E-06 (3.29E-05)	B ³		B	B		B	B
QTL2		B	B		B	B		B	B
QTL3		B	B		B	B		B	B
QTL4		5.85E-05 (7.80E-05)	B		2.17E-05 (1.73E-05)	B		B	B
QTL5		B	4.59E-07 (1.53E-05)		2.37E-06 (1.13E-05)	1.93E-06 (4.20E-06)		1.22E-05 (1.40E-05)	5.06E-06 (5.38E-06)
Residual	1.72E-01 (8.57E-02)	3.34E-05 (3.97E-05)	1.98E-05 (3.10E-05)		2.65E-06 (4.90E-06)	5.84E-07 (3.24E-06)		5.63E-06 (6.12E-06)	4.66E-06 (5.49E-06)
		1.78E-01 (8.45E-02)	1.77E-01 (8.55E-02)	1.11E-01 (1.43E-02)	1.14E-01 (1.41E-02)	1.10E-01 (1.43E-02)	8.05E-02 (1.13E-02)	7.87E-02 (1.12E-02)	7.67E-02 (1.12E-02)

¹See Table 1 for locations of QTL. SCS = lactation-average SCS; CM1 = clinical mastitis between d -15 and 50; CM2 = clinical mastitis between d 51 and 300. CONBLUP = conventional BLUP; MABLUP-IBD = marker-assisted breeding value estimation with identical-by-descent haplotypes; MABLUP-IBS = marker-assisted breeding value estimation with identical-by-state haplotypes.

²Variance components for polygenic and QTL effects were a quarter of the polygenic and QTL variances due to using daughter yield deviations.

³Variance component was bounded to keep the estimate positive; estimate was effectively zero.

higher EBV had higher incidence of mastitis. The difference between CONBLUP and MABLUP-IBS was, however, negligible, indicating that both methods had similar predictive ability for mastitis incidence. When the incidence of CM1 and CM2 was calculated per 10%-quantile of haplotype EBV, the difference in incidence of CM1 and CM2 between cows with the highest and lowest EBV was much smaller than for total EBV, but still highly significant ($P < 0.001$; Student's t -test), indicating that the haplotypes capture QTL effects associated with clinical mastitis.

DISCUSSION

The aim of this study was to demonstrate marker-assisted breeding value estimation for mastitis resistance in Finnish Ayrshire cattle, combining data of genotyped and ungenotyped animals using previously identified QTL regions. Furthermore, we compared marker-assisted EBV using either IBD or IBS haplotypes to conventional EBV and investigated the predictive value of these EBV. With both methods, it was possible to deal with ungenotyped and genotyped animals in a single evaluation. Therefore, both methods are able to generate benefits, especially for genotyped animals. In this particular situation, marker-assisted breeding value estimation performed, however, similar to or worse than conventional breeding value estimation.

The selection of the markers and QTL regions was based on previous research, which made use in part of the same data. The focus of our study was marker-assisted breeding value estimation, and therefore we focus here only on differences and similarities by means of analyzing the estimated variances per QTL region. The QTL on BTA9 at ~74 cM explained hardly any genetic variance, even when it was the only QTL in the model, as in Sahana et al. (2008). Sahana et al. (2008) found that this QTL was not significant in Finnish Ayrshire cattle, whereas it was significant in Danish Red or in analyses across Nordic red breeds. In this study, we found that QTL3 (~17 to 18 cM on BTA11) explained a substantial proportion of genetic variance in CM1 when using IBD haplotypes, which was in agreement with Schulman et al. (2009), who found a sharp peak at 17.8 cM. Furthermore, QTL4 (~62 cM on BTA11) explained a considerable proportion of genetic variance in CM2 with IBD haplotypes, which confirmed the peaks found at ~62 cM in Schulman et al. (2009), but was in disagreement with Lund et al. (2007), who found only a peak for SCS.

The IBS haplotypes explained much less genetic variance than the IBD haplotypes. It might be that IBD haplotypes captured family-specific effects beyond

Table 4. Estimates of variance components using identical-by-descent (IBD) or identical-by-state (IBS) haplotypes as a proportion to the total genetic variance for 5 selected QTL regions¹

Method	Trait ²	QTL1	QTL2	QTL3	QTL4	QTL5	Total
IBD	SCS	0.003	0.000	0.026	0.000	0.015	0.044
	CM1	0.000	0.000	0.131	0.014	0.016	0.161
	CM2	0.000	0.000	0.000	0.064	0.030	0.094
IBS	SCS	0.000	0.000	0.000	0.000	0.009	0.009
	CM1	0.000	0.000	0.000	0.011	0.003	0.014
	CM2	0.000	0.000	0.000	0.026	0.024	0.049

¹See Table 1 for locations of QTL.²SCS = lactation-average SCS; CM1 = clinical mastitis between d -15 and 50; CM2 = clinical mastitis between d 51 and 300.

pure QTL effects. Because the numerator relationship matrix and the IBD matrix were partly overlapping, especially with poor or missing marker information, disentangling polygenic and QTL effects might be challenging. Overestimation of QTL effects due to population stratification is a generic problem in QTL mapping (e.g., Kennedy et al., 1992; Martinez et al., 1999; Wang and Elston, 2005). Neuner et al. (2009) showed with simulations that the QTL variance is slightly overestimated when using daughter yield deviations with an IBD model. Nevertheless, with both methods, the 5 QTL used explained a small proportion of the genetic variance, and therefore SCS and clinical mastitis are likely to be affected by many genes, making genome-wide evaluation approaches more effective than marker-assisted breeding value estimation with a few putative QTL regions. Cross-validation studies for the Nordic red dairy cattle have shown that genome-wide evaluation yields good accuracy for udder health, but the accuracy is smaller than in Holstein cattle (Brøndum et al., 2010).

In this study, we performed 2 types of cross-validation: one based on bulls and one based on cows. Cross-validation with bulls is intended to give the predictive

ability of EBV of young bulls before entering progeny testing, but clearly, in this case, the sample size of 104 bulls was quite small. The cross-validation showed that IBD haplotype effects were estimated with much lower accuracy than IBS haplotypes, which caused the lower correlation obtained for total EBV compared with CONBLUP. For all models, correlations between EBV of the reduced set and the full data set were inflated, because both EBV relied heavily on pedigree information. The cross-validation with cows showed that EBV are predictive of mastitis incidence on a phenotypic level, but the conventional breeding value estimation did not differ from marker-assisted breeding value estimation with IBS haplotypes in levels of incidence based on 10%-quantiles of EBV. This might be because the IBS haplotypes explained only a small proportion of the genetic variance. On the other hand, it shows that marker-assisted breeding value estimation with IBS haplotypes is quite robust and should at least give increased accuracy for genotyped animals, such as occurred for CM2 in genotyped bulls. Based on the simulation in Mulder et al. (2010a), it is expected that MABLUP-IBS would increase the accuracy of EBV for genotyped animals when the IBS haplotypes explain 10

Table 5. Correlations between conventional EBV (Con) and marker-assisted total EBV (Tot), polygenic EBV (Pol), and haplotype EBV (Hap) for bulls¹ and cows for SCS, CM1, and CM2 using identical-by-descent or identical-by-state haplotypes

Method ²	Trait ³	Bulls			Cows		
		Con-Tot	Con-Pol	Con-Hap	Con-Tot	Con-Pol	Con-Hap
MABLUP-IBD	SCS	1.000	0.996	0.434	0.998	0.996	0.710
	CM1	0.987	0.986	0.337	0.986	0.979	0.682
	CM2	0.948	0.855	0.719	0.940	0.851	0.753
MABLUP-IBS	SCS	0.999	0.999	0.024	1.000	1.000	0.106
	CM1	0.995	0.997	0.015	0.998	0.998	0.131
	CM2	0.992	0.996	0.078	0.997	0.997	0.111

¹Based on 341 bulls genotyped for all markers with a reliability of 0.80 or higher for SCS.²MABLUP-IBD = marker-assisted breeding value estimation with identical-by-descent (IBD) haplotypes; MABLUP-IBS = marker-assisted breeding value estimation with identical-by-state (IBS) haplotypes.³SCS = lactation-average SCS; CM1 = clinical mastitis between d -15 and 50; CM2 = clinical mastitis between d 51 and 300.

Figure 1. The incidence of clinical mastitis in the first part of lactation (CM1) for each 10% quantile of the EBV for CM1 (panel A) and the incidence of mastitis in the last part of lactation (CM2) for each 10% quantile of the total EBV for CM2 (panel B) using all data for conventional BLUP (CONBLUP) and marker-assisted BLUP either with identical-by-descent (MABLUP-IBD) or identical-by-state haplotypes (MABLUP-IBS).

to 20% of the genetic variance. The limited or absent increase in predictive ability for ungenotyped animals is in agreement with the simulation results in Mulder et al. (2010a). In addition, Gengler et al. (2008) found no increase in accuracy for SCS with gene-assisted breeding value estimation when using a candidate gene that explained a very small amount of genetic variance.

Many countries are currently implementing genome-wide evaluations in dairy cattle, with the advantage that high-density SNP chips capture larger amounts of genetic variance than traditional QTL approaches. Furthermore, genome-wide evaluation eliminates the need to find QTL, because the method assumes either that all SNP contribute equally to the genetic variance (GBLUP) or that the model decides itself which SNP will have large and which will have small effects (e.g.,

Figure 2. The incidence of clinical mastitis in the first part of lactation (CM1) for each 10% quantile of the EBV for CM1 (panel A) and the incidence of mastitis in the last part of lactation (CM2) for each 10% quantile of the total EBV for CM2 (panel B) for cross-validation cows, which phenotype was excluded from the data. EBV were calculated using conventional BLUP (CONBLUP) or marker-assisted BLUP using identical-by-state haplotypes (MABLUP-IBS).

BayesB or BayesC approaches; Meuwissen et al., 2001; Gianola et al., 2009). However, marker-assisted breeding value estimation for several QTL may still be of interest for some species in which genotyping animals by using high-density SNP chips is too expensive. Another application would be for cases in which a major QTL exists and where a combination of GBLUP and the major gene effect modeled separately is as effective as the BayesB or BayesC approach.

We showed the feasibility of estimating marker-assisted breeding values in a large dairy cattle population that consisted of genotyped and ungenotyped animals. A model with IBD haplotypes and IBD matrices was computationally more demanding in many aspects than a model using IBS haplotypes without the use of additional IBD matrices. First, making the IBD matrix

Table 6. Cross-validation for bulls born in 2002 and 2003 with all genotypes and with a reliability of at least 0.80 for SCS when using all data with conventional breeding value estimation (CONBLUP) and marker-assisted breeding value estimation using identical-by-descent or identical-by-state haplotypes (MABLUP-IBD and MABLUP-IBS, respectively)¹

Method	Type of EBV	Trait ²		
		SCS	CM1	CM2
CONBLUP	Conventional	0.512	0.488	0.569
MABLUP-IBD	Haplotype	0.405	0.667	0.433
	Polygenic	0.533	0.455	0.520
	Total	0.502	0.485	0.418
	Haplotype	0.928	0.983	0.971
MABLUP-IBS	Polygenic	0.516	0.491	0.578
	Total	0.506	0.488	0.574

¹Correlations between conventional, haplotype, polygenic, and total EBV of the full data set and the reduced data set.

²SCS = lactation-average SCS; CM1 = clinical mastitis between d -15 and 50; CM2 = clinical mastitis between d 51 and 300.

is memory intensive. To limit memory, IBD matrices were constructed using a 1-dimensional array and an efficient search algorithm to limit computing time. Second, convergence of the marker-assisted breeding value estimation with an IBD matrix was slow. This was due to the model having 5 very similar matrices for the ungenotyped cows, leading to high correlations between effects and thus resulting in slow convergence. Furthermore, many more effects need to be estimated with IBD haplotypes than with IBS haplotypes, which may lead to the lower accuracy of the total EBV than with IBS haplotypes. Therefore, using IBS haplotypes seems to be much more robust and practically feasible to implement in large populations.

CONCLUSIONS

In this study, we showed that marker-assisted breeding value estimation using IBD or IBS haplotypes is feasible even for large populations. The 5 QTL explained only a small proportion of genetic variance, especially with IBS haplotypes. The estimated QTL variances when using IBD haplotypes might have been inflated by capturing family-specific effects. The use of IBD haplotypes gave more reranking of bulls and cows than use of IBS haplotypes. Cross-validation showed no increase in predictive ability when using IBS haplotypes compared with conventional breeding value estimation, whereas a decrease in predictive ability was observed with IBD haplotypes. Furthermore, computing time was lower and convergence was better with IBS haplotypes than with IBD haplotypes. For mastitis resistance evaluation in Finnish Ayrshire, conventional breeding value estimation would be advocated because of the lack in improvement of accuracy and predictive ability when using marker-assisted breeding value estimation. However, it is expected that in situations where IBS haplotypes would explain 10 to 20% of the genetic variance, marker-assisted breeding value estimation with IBS haplotype may yield higher accuracy and predictive ability than conventional breeding value estimation.

ACKNOWLEDGMENTS

The work was financially supported by CRV (Arnhem, the Netherlands), Hendrix Genetics (Boxmeer, the Netherlands), IPG (Beuningen, the Netherlands), and the European Commission, within the 6th Framework project SABRE, contract No. FOOD-CT-2006-016250. The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information. Mario Calus (Animal Breeding and Genomics Centre,

Wageningen UR Livestock Research, Wageningen, the Netherlands) is acknowledged for helpful suggestions during the study. Faba Breeding Finland is kindly acknowledged for providing the data.

REFERENCES

- Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbrandtsen, W. F. Fikse, and M. S. Lund. 2010. Investigation of the reliability of genomic selection using combined reference data of the Nordic red populations. *Commun. 234 in Proc. 9th World Congr. Genet. Appl. Livest. Prod.*, Leipzig, Germany. Gesellschaft für Tierzuchtwissenschaften e. V., Giessen, Germany.
- Cole, J. B., P. M. VanRaden, J. R. O'Connell, C. P. Van Tassell, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and G. R. Wiggins. 2009. Distribution and location of genetic effects for dairy traits. *J. Dairy Sci.* 92:2931–2946.
- de Haas, Y., W. Ouweltjes, J. Ten Napel, J. J. Windig, and G. De Jong. 2008. Alternative somatic cell count traits as mastitis indicators for genetic selection. *J. Dairy Sci.* 91:2501–2511.
- Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* 21:467–477.
- Fikse, W. F., and G. Banos. 2001. Weighting factors of sire daughter information in international genetic evaluations. *J. Dairy Sci.* 84:1759–1767.
- Gengler, N., S. Abras, C. Verkenne, S. Vanderick, M. Szydlowski, and R. Renaville. 2008. Accuracy of prediction of gene content in large animal populations and its use for candidate gene detection and genetic evaluation. *J. Dairy Sci.* 91:1652–1659.
- Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal* 1:21–27.
- Gianola, D., G. De los Campos, W. G. Hill, E. Manfredi, and R. L. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics* 183:347–363.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2006. ASReml User Guide Release 2.0. VSN International Ltd., Hemel Hempstead, UK.
- Haldane, J. B. S. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* 8:299–309.
- Johansson, K., S. Eriksson, J. Pösö, M. Toivonen, U. Sander-Nielsen, J.-Å. Eriksson, and G. Pedersen Aamand. 2006. Genetic evaluation of udder health traits for Denmark, Finland and Sweden. *Interbull Bull.* 35:92–96.
- Kennedy, B. W., M. Quinton, and J. A. M. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* 70:2000–2012.
- Khatkar, M. S., P. C. Thompson, I. Tammen, and H. Raadsma. 2004. Quantitative trait loci mapping in dairy cattle: Review and meta-analysis. *Genet. Sel. Evol.* 36:163–190.
- Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756.
- Lund, M. S., B. Guldbrandtsen, A. J. Buitenhuis, B. Thomsen, and C. Bendixen. 2008. Detection of quantitative trait loci in Danish Holstein cattle affecting clinical mastitis, somatic cell score, udder conformation traits, and assessment of associated effects on milk yield. *J. Dairy Sci.* 91:4028–4036.
- Lund, M. S., G. Sahana, L. Andersson-Eklund, N. Hastings, A. Fernandez, N. F. Schulman, B. Thomsen, S. M. Viitala, J. L. Williams, A. Sabry, H. Viinalass, and J. H. Vilkki. 2007. Joint analysis of quantitative trait loci for clinical mastitis and somatic cell score on five chromosomes in three Nordic dairy cattle breeds. *J. Dairy Sci.* 90:5282–5290.
- Martinez, M. L., N. Vukasinovic, and A. E. Freeman. 1999. Random model approach for QTL mapping in half-sib families. *Genet. Sel. Evol.* 31:319–340.

- Meuwissen, T. H. E., and M. E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. *Genet. Sel. Evol.* 28:161–176.
- Meuwissen, T. H. E., and M. E. Goddard. 2001. Prediction of identity by descent probabilities from marker-genotypes. *Genet. Sel. Evol.* 33:605–634.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Mrode, R., and G. J. Swanson. 2004. Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. *Livest. Prod. Sci.* 86:253–260.
- Mrode, R. A. 2005. *Linear Models for the Prediction of Animal Breeding Values*. CABI, Wallingford, UK.
- Mulder, H. A., M. P. L. Calus, and R. F. Veerkamp. 2010a. Prediction of haplotypes with missing genotypes and its effect on accuracy of marker-assisted breeding value estimation. *Genet. Sel. Evol.* 42:10.
- Mulder, H. A., M. Lidauer, I. Strandén, E. A. Mäntysaari, M. H. Pool, and R. F. Veerkamp. 2010b. *MiXBLUP manual*. Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Lelystad, the Netherlands.
- Neuner, S., C. Edel, R. Emmerling, G. Thaller, and K.-U. Götz. 2009. Precision of genetic parameters and breeding values estimated in marker-assisted BLUP genetic evaluation. *Genet. Sel. Evol.* 41:26.
- Rupp, R., and D. Boichard. 2003. Genetics of resistance to mastitis in dairy cattle. *Vet. Res.* 34:671–688.
- Sahana, G., M. S. Lund, L. Andersson-Eklund, N. Hastings, A. Fernandez, T. Iso-Touru, B. Thomsen, S. M. Viitala, P. Sorensen, J. L. Williams, and J. H. Vilkki. 2008. Fine-mapping QTL for mastitis resistance on BTA9 in three Nordic red cattle breeds. *Anim. Genet.* 39:354–362.
- Schrooten, C., H. Bovenhuis, W. Coppieters, and J. A. M. Van Arendonk. 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *J. Dairy Sci.* 83:795–806.
- Schulman, N. F., G. Sahana, T. Iso-Touru, M. S. Lund, L. Andersson-Eklund, S. M. Viitala, S. Varv, H. Viinalass, and J. H. Vilkki. 2009. Fine mapping of quantitative trait loci for mastitis resistance on bovine chromosome 11. *Anim. Genet.* 40:509–515.
- Schulman, N. F., S. M. Viitala, D. J. De Koning, J. Virta, A. Maki-Tanila, and J. H. Vilkki. 2004. Quantitative trait loci for health traits in Finnish Ayrshire cattle. *J. Dairy Sci.* 87:443–449.
- Tier, B., and K. Meyer. 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. *J. Anim. Breed. Genet.* 121:77–89.
- Verbyla, K. L., M. P. L. Calus, H. A. Mulder, Y. De Haas, and R. F. Veerkamp. 2010. Predicting energy balance for dairy cows using high density SNP information. *J. Dairy Sci.* 93:2757–2764.
- Viitala, S. M., N. F. Schulman, D. J. De Koning, K. Elo, R. Kinoshita, J. Virta, A. Maki-Tanila, and J. H. Vilkki. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *J. Dairy Sci.* 86:1828–1836.
- Wang, T., and R. C. Elston. 2005. The bias introduced by population stratification in IBD based linkage analysis. *Hum. Hered.* 60:134–142.
- Windig, J. J., and T. H. E. Meuwissen. 2004. Rapid haplotype reconstruction in pedigrees with dense marker maps. *J. Anim. Breed. Genet.* 121:26–39.
- Windig, J. J., W. Ouweltjes, J. Ten Napel, G. De Jong, R. F. Veerkamp, and Y. De Haas. 2010. Combining somatic cell count traits for optimal selection against mastitis. *J. Dairy Sci.* 93:1690–1701.