Reliability of genomic evaluations in Holstein-Friesians using haplotypes based on the BovineHD BeadChip

G. C. B. Schopen¹ and C. Schrooten
CRV, PO Box 454, 6800 AL Arnhem, the Netherlands

ABSTRACT

The objectives of this study were to make subsets of high-density (HD) loci based on localized haplotype clusters, without loss of genomic information, to reduce computing time compared with the use of all HD loci and to investigate the effect on the reliability of the direct genomic value (DGV) when using this HD subset based on localized haplotype clusters in the genomic evaluation for Holstein-Friesians. The DNA was isolated from semen samples of 548 bulls (key ancestors) of the EuroGenomics Consortium, a collaboration between 4 European dairy cattle breeding organizations and scientific partners. These bulls were genotyped with the BovineHD BeadChip (~777,000 (777K) single nucleotide polymorphisms (SNP); Illumina Inc., San Diego, CA) and used to impute all 30,483 Holstein-Friesians from the BovineSNP50 BeadChip (~50,000 (50K) SNP; Illumina Inc.) to HD, using the BEAGLE software package. The final data set consisted of 30,483 animals and 603,145 SNP. For each locus, localized haplotype clusters (i.e., edges of the fitted graph model) identifications were obtained from BEAGLE. Three subsets [38,000 (38K), 116,000 (116K), and 322,000 (322K) loci] were made based on deleting obsolete loci (i.e., loci that do not give extra information compared with the neighboring loci). A fourth data set was based on 38K SNP, which is currently used for routine genomic evaluation at the Cattle Improvement Cooperative (CRV, Arnhem, the Netherlands). A validation study using the HD loci subsets based on localized haplotype clusters was performed for 9 traits (production, conformation, and functional traits). Error of imputation from 50K to HD averaged 0.78%. Three thresholds (0.17, 0.05, and 0.008%) were used for the identification of obsolete HD loci based on localized haplotype clusters to obtain a desired number of HD loci (38K, 116K, and 322K). On average, 46% (using threshold 0.008%) to 93% (using threshold 0.17%) of HD loci were eliminated. The computing time was about 9 d for 38K loci, 15.5 d for 116K loci, 21 d for 322K loci, and 7.5 d for 38K SNP. The increase in reliability of DGV compared with pedigree-based estimated breeding values for kilograms of protein was similar for 322K and 116K loci (30.7%), but was 1.5 to 2% higher compared with 38K loci and 38K SNP. Averaged over 9 traits, subset 116K loci resulted in a higher increase in reliability compared with 38K loci and 38K SNP. Eliminating obsolete loci enormously decreased the amount of data to be analyzed for genomic evaluations. The more HD loci used in a genomic evaluation, the higher the increase in reliability of DGV. It is possible to increase the reliability of DGV by 1 to 2% compared with the SNP currently used for routine genomic evaluation.

Key words: high-density genotype, localized haplotype cluster, direct genomic value, reliability

INTRODUCTION

Genomic selection is routinely used in dairy cattle breeding programs. Genomic selection refers to genetic improvement of animals through selection based on genomic EBV (GEBV). To determine GEBV, a large number of markers have been used. Individual bulls or cows, or both, can be genotyped for SNP using the BovineSNP50 BeadChip (~50,000 (50K) SNP; Illumina Inc., San Diego, CA; Matukumalli et al., 2009). The Cattle Improvement Cooperative (CRV; Arnhem, the Netherlands) uses genotypes obtained with the BovineSNP50 BeadChip as well as genotypes obtained with a custom-made 50K chip to determine GEBV for individuals. The 50K chips, however, have limitations with regard to genomic selection across breeds (Hayes et al., 2009). To exploit linkage disequilibrium (LD) across breeds, markers need to be spaced less than 10 kbp apart (de Roos et al., 2008), whereas markers on, for example, the BovineSNP50 BeadChip are roughly 50-kbp spaced, on average. In January 2010, the BovineHD BeadChip (~777,000K (777K) SNP) of Illumina Inc. (Matukumalli et al., 2011) became available. This BovineHD BeadChip is potentially very useful for performing genomic selection across breeds. However, recent studies (Harris et al., 2011; Erbe et al., 2012; Pryce et al., 2012) showed that the BovineHD BeadChip is not as good across breeds as many have thought beforehand.
Rather than using SNP, haplotypes, which are in greater LD with QTL and thus have higher power to trace QTL, can be used (Hayes et al., 2007). A disadvantage of using haplotypes is that more effects per locus need to be estimated compared with individual SNP (Hayes et al., 2007). When the number of phenotypes is large enough, however, haplotypes might give more reliable genomic prediction than individual SNP. When 777K high-density (HD) loci based on haplotypes are used for a genomic evaluation, computing time will increase enormously compared with 50K SNP. This can be problematic for the Bayesian method of Meuwissen and Goddard (2004) that is currently used in the routine genomic evaluation of CRV and has greater computer requirements compared with other methods. Identification of obsolete loci (i.e., loci that do not give extra information across animals compared with the neighboring locus) can, therefore, reduce the amount of HD (777K) data to be analyzed in a genomic evaluation, without loss of genomic information. Localized haplotype clusters were proposed to be used in association studies (Browning and Browning, 2007a) and for genomic selection (Druet and Georges, 2010). de Roos et al. (2011) removed redundant loci (i.e., loci that are identical to another locus) based on localized haplotype clusters to save computing time and memory requirements. de Roos et al. (2011) used a fixed number of localized haplotype clusters per locus and a data set with 50K genotypes using the DualPHASE software (Druet and Georges, 2010).

Loci that do not give extra information across animals compared with their neighboring locus after fitting the graphical model in BEAGLE software (Browning and Browning, 2007b) are defined as obsolete loci. Loci may be obsolete due to no recombination from one localized haplotype cluster to another within the genotyped population. If obsolete loci exist, then it is possible to use, for example, only 1 instead of 3 loci to estimate GEBV without loss of genomic information. This means a reduction of the number of HD loci to be analyzed for genomic evaluation.

The objectives of this study were to make subsets of HD loci based on localized haplotype clusters, without loss of genomic information, to reduce computing time compared with the use of all HD loci and to investigate the effect on the reliability of the direct genomic value (DGV) when using this HD subset based on localized haplotype clusters in the genomic evaluation for Holstein-Friesians.

**MATERIALS AND METHODS**

**Animals**

The EuroGenomics Consortium, a collaboration between 4 European dairy cattle breeding organizations and scientific partners (Lund et al., 2011), genotyped 548 bulls (key ancestors) with the BovineHD BeadChip. Using these 548 HD bulls as a reference set, a large data set consisting of 30,483 Holstein-Friesian animals genotyped with the BovineSNP50 BeadChip were imputed to HD using the BEAGLE software package (Browning and Browning, 2007b).

**Phenotypes**

Estimated breeding values of progeny-tested bulls for kilograms of protein, kilograms of milk, kilograms of fat, longevity, calving interval, 56-d nonreturn rate, SCC, feet and legs, and udder were deregressed and used as phenotypes in subsequent analyses. Deregression was performed using the method of VanRaden et al. (2009):

\[
DRP = PA + (EBV - PA) \times \left( \frac{EDC_{parents + progeny}}{EDC_{progeny}} \right),
\]

where DRP is the deregressed proof, PA is the parent average, and EDC is the effective daughter contribution. Effective daughter contributions were derived from reliabilities of EBV using the method of Mäntysaari et al. (2010):

\[
EDC_{parents + progeny} = \frac{k \times REL_{EBV}}{1 - REL_{EBV}},
\]

\[
EDC_{parents} = \frac{k \times REL_{PA}}{1 - REL_{PA}},
\]

\[
EDC_{progeny} = EDC_{parents + progeny} - EDC_{parents},
\]

where REL_{EBV} is the reliability of EBV, REL_{PA} is the reliability of PA EBV, and \( k = \frac{1 - h^2}{h^2} \), where \( h^2 \) is the heritability of the trait.

**Validation of Imputation From 50K to HD**

The 548 HD bulls were used to validate imputation from 50K to HD using BEAGLE software (Browning and Browning, 2007b). Out of the 548 HD bulls, 60 bulls were selected based on the criteria that they did not have any HD genotyped progeny and were used as validation bulls. The reference set for imputation consisted of 548 − 60 = 488 HD bulls, which were imputed to also obtain genotypes for possible missing genotypes in the HD set (Supplemental Command 1, available online at http://dx.doi.org/10.3168/jds.2012-6510).
The HD genotypes of the 60 validation bulls were set to missing, except genotypes of the BovineSNP50 BeadChip. Missing genotypes of validation bulls were imputed (Supplemental Command 2) and imputed genotypes were compared with original HD genotypes of the validation bulls. Based on comparison, the imputation error from 50K to HD was calculated.

**Imputation of 50K Genotyped Holstein-Friesians**

The 50K genotypes were based on version 1 or 2 of a custom 50K SNP chip or on version 1 or 2 of the BovineSNP50 BeadChip (Matukumalli et al., 2009). Genotyping of 548 bulls with the BovineHD BeadChip and subsequent editing resulted in 618,905 SNP distributed across 29 BTA. The SNP edits were applied for call rate (>90%), minor allele frequency (>2.5%), and Hardy-Weinberg disequilibrium (maximum deviation of 0.15 between observed and predicted fraction of heterozygotes). Before imputation, a call rate of >90% also was applied per animal.

For the 30,483 Holstein-Friesian animals genotyped with the different 50K chips, 56,641 SNP present on at least 1 of the used 50K chips and on the BovineHD chip were selected. Remaining HD SNP were imputed using BEAGLE software (Browning and Browning, 2007b). First, the reference population consisting of 548 HD bulls was imputed to also obtain genotypes for possible missing genotypes in the HD set (Supplemental Command 1, available online at http://dx.doi.org/10.3168/jds.2012-6510). Second, all 50K-genotyped animals were imputed, using the 548 HD genotypes as a reference set. Because of memory limits, it was not possible to impute all animals at once. It was decided, therefore, to split the data set into 30 subsets. Each subset consisted of approximately 1,000 50K-genotyped animals and was separately imputed. After imputation of 30 subsets per chromosome, subsets were combined to 1 data set per chromosome.

After imputation of all 30,483 Holstein-Friesian animals, the same SNP edits as described previously were applied. After SNP editing, the final data set consisted of 30,483 Holstein-Friesians with 603,145 SNP distributed across 29 autosomes. The number of SNP per BTA ranged from 11,505 for BTA 25 to 37,845 for BTA 1 (Supplemental Table S1, available online at http://dx.doi.org/10.3168/jds.2012-6510).

**Identification of Obsolete Loci**

High-density SNP genotypes of all Holstein-Friesians were used to construct a directed acyclic graph (DAG; Browning and Browning (2007b); Supplemental Command 3, available online at http://dx.doi.org/10.3168/jds.2012-6510). Based on the DAG, localized haplotype clusters files for each chromosome were made and used to identify obsolete loci. To control the number of localized haplotype clusters, a scale parameter of 4.0 and a shift parameter of 0.2 (default setting of the parameters) were used. These values are used because these values performed well in simulation studies and in real data analyses (Browning and Browning, 2007a, 2008).

Obsolete loci were identified based on localized haplotype clusters after fitting the graphical model in BEAGLE. If the same combinations (i.e., pair) of localized haplotype cluster identification at position x and position x + 1 existed across animals (e.g., localized haplotype cluster 8 at position x and localized haplotype cluster 11 at position x + 1), then loci at these 2 positions were identical and 1 of these 2 loci could be considered obsolete. In addition, the percentage of localized haplotype clusters that differed among all animals was compared at each pair of loci. If this percentage did not exceed a threshold, this locus was identified as an obsolete locus. The thresholds were chosen to obtain a desired number of HD loci. In total, there were 3 reduced HD loci subsets: (1) a subset of 38,000 (38K) loci using a threshold of 0.17% (Red1), (2) a subset of 116,000 (116K) loci using a threshold of 0.05% (Red2), and (3) a subset of 322,000 (322K) loci using a threshold of 0.008% (Red3).

**Genomic Validation**

The effect of using HD information on reliability of DGV was assessed by validations for 3 reduced HD subsets based on localized haplotype clusters (Red1, Red2, and Red3) and 1 data set based on SNP from the routine genomic evaluation of CRV (38k_SNIP; validation using genotypes for 38K SNP). The genomic validation procedure consisted of 4 steps: (1) identification of validation bulls based on the following rules: the bull and sire of bull were genotyped and progeny tested, the bull had no sons that were genotyped and progeny tested, the bull was born between July 1, 2003, and June 30, 2004, and was a black and white Holstein-Friesian bull; (2) phenotypes of validation bulls and all other bulls born after June 30, 2003, were omitted from the data and DGV and pedigree EBV (PEBV) were determined for these omitted bulls in a genomic evaluation; (3) comparison of DGV and PEBV of validation bulls based on their correlations with the DRP; and (4) calculation of reliability of DGV and PEBV.

The model to estimate DGV for 1 trait was derived from the Bayesian multiple QTL model of Meuwissen and Goddard (2004):

\[ y_j = \mu + u_i + \sum_{j} (q_{ij1} + q_{ij2}) v_j + e_i \]  

\[ [1] \]
where $y_i$ is the DRP of bull $i$, $\mu$ is the overall mean, $u_i$ is the random polygenic effect of bull $i$, $n$ is the total number of loci, $v_j$ is the direction vector of effects of haplotypes at locus $j$, $q_{ij1}$ ($q_{ij2}$) is the size of effect for the first (second) localized haplotype cluster identification of animal $i$ at locus $j$, and $e_i$ is the residual for bull $i$. Each DRP was weighted by its EDC. Only this modified model of Meuwissen and Goddard (2004) was used to estimate DGV because this method is also used in the routine genomic evaluation at CRV.

A Markov chain Monte Carlo method using Gibbs sampling was used to obtain posterior estimates for all effects in the model. The scale parameter of a putative QTL at locus $j$ ($v_j$) was sampled from a normal distribution $N(0, \sigma^2)$ if a QTL was present at locus $j$, whereas $v_j$ was sampled from $N(0, \sigma^2/100)$ if no QTL was present at locus $j$. The variance of $v_j$ ($\sigma^2$) was sampled from a scaled inverse chi-squared distribution. The presence of a QTL at locus $j$ was sampled from a Bernoulli distribution with probability equal to

$$P(v_j \mid \sigma^2) \times \Pr_j \over P(v_j \mid \sigma^2) \times \Pr_j + P(v_j \mid \sigma^2/100) \times (1 - \Pr_j),$$

where $P(v_j \mid \sigma^2)$ is probability of $v_j$ from $N(0, \sigma^2)$; in other words,

$$1 \over \sqrt{2\pi\sigma^2} \times e^{-v_j^2 \over 2\sigma^2},$$

and $\Pr_j$ is the prior probability of the presence of a QTL at locus $j$. More details on prior distributions and full conditional distributions can be found in Meuwissen and Goddard (2004) and Calus et al. (2008). The Gibbs sampler ran for 6,000 iterations and 2,000 iterations were removed as burn-in. In total, 4 chains per trait were analyzed; Red3 was only run for kilograms of protein and only 1 chain was analyzed because of memory constraints and computation time.

Pedigree EBV were estimated from the same data using the following model:

$$y_i = \mu + u_i + e_i,$$

where $y_i$ is the DRP of bull $i$, $\mu$ is the overall mean, $u_i$ is the random polygenic effect of bull $i$, and $e_i$ is the residual for bull $i$. Each DRP was weighted by its EDC. Squared correlations ($R^2$) between DRP and both DGV and PEBV were computed and compared with each other to obtain $\Delta R^2$ using the following formula:

$$\Delta R^2 = \frac{R^2_{\text{DGV,DRP}} - R^2_{\text{PEBV,DRP}}}{\text{REL}_{\text{DRP}}},$$

where $R^2_{\text{DGV,DRP}}$ is the squared correlation between DGV and DRP, $R^2_{\text{PEBV,DRP}}$ is the squared correlation between PEBV and DRP, and REL_{DRP} is the average reliability of the DRP of all validation bulls.

Values for $\Delta R^2$ of the reduced HD loci subsets (Red1, Red2, and Red3) were compared with $\Delta R^2$ obtained in a validation using a set of 38K SNP that is used in the routine genomic evaluation, and using the same reference and validation animals as in the other studied alternatives (38k_SNAP).

### RESULTS AND DISCUSSION

#### Imputation Error and Accuracy

In this study, BEAGLE was used to impute all animals from 50K to HD. The error of imputation from 50k to HD averaged 0.78% (Supplemental Table S1, available online at http://dx.doi.org/10.3168/jds.2012-6510). This is lower than the imputation error of <1% reported for the dense part of BTA 14, corresponding to a genome-wide panel of 250,000 SNP (Druet and Georges, 2010) or between different 50K chips (Druet et al., 2010). BEAGLE does not make use of linkage information, as opposed to other software packages such as DAGPHASE (Druet and Georges, 2010) and CHROMIBD (Druet and Farnir, 2011). Mulder et al. (2012), however, showed that BEAGLE had slightly lower accuracies compared with CHROMIBD and DAGPHASE when both parents of the animals were genotyped and had much higher accuracies compared with CHROMIBD and DAGPHASE when the parents were not genotyped. This indicates that linkage information positively affects imputation accuracy, whereas LD is better captured by BEAGLE (Mulder et al., 2012). In addition, at high SNP density and based on simulated data, BEAGLE also gave higher accuracies compared with FastPHASE and multivariate BLUP (Calus et al., 2011). This indicates that Beagle is a powerful software package to use for imputation.

#### Percentage of Obsolete Loci

To obtain a data set of about 38K HD loci, the threshold of 0.17% was used. Using this threshold, the percentage of obsolete HD loci based on localized haplotype clusters averaged 93.4% and ranged from 91.4% for BTA 18 to 95.7% for BTA 20 (Supplemental Table S1, available online at http://dx.doi.org/10.3168/jds.2012-6510). To obtain a data set of more HD loci, the threshold was
The computing time of the genomic evaluation of Red1 was about 9 d, of Red2 was about 15.5 d, of Red3 was about 21 d, and of 38k_SNP was about 7.5 d. The relatively small difference in computing time between Red1 and 38k_SNP can be explained by the use of haplotypes (Red1) instead of SNP genotypes (38k_SNP) or by a difference in SNP genotypes of which haplotype clusters were created for Red1. Estimation of haplotype effects involves more computations than estimation of allele effects; however, these computations only comprise a fraction of the computation time of the entire algorithm. The difference in computation time of genomic evaluation between Red1 and 38k_SNP is, therefore, negligible.

For Red1 and Red2, multiple servers, each containing 4 processors, were used to run all 9 traits simultaneously. Higher memory requirements for Red3 made it impossible to use these servers. To run Red3 for all 9 traits would result in a computing time of 9 times 21 d, which equals 189 d (>6 mo). Computing time of genomic evaluation will increase further when analyzing all HD loci in a genomic evaluation. de Roos et al. (2011) reported that the computing time of the Gibbs sampler used in this study will be 141 d when analyzing 750K SNP for 1 trait for 30,000 animals.

BEAGLE only runs once to obtain the imputed genotypes for the whole data set. In a routine implementation, computing time of BEAGLE to obtain imputed genotypes for newly genotyped animals is relatively limited, making use of the results of the previous BEAGLE run. The FORTRAN script to find obsolete loci runs quickly. The routine genomic evaluation of CRV is based on the method of Meuwissen and Goddard (2004) using Gibbs sampling. For analyzing all HD loci (or a large subset of HD loci), it might be interesting to look at other methods such as genomic BLUP, which are faster and result in similar genomic reliabilities compared with Bayesian methods based on simulation studies (e.g., Zhang et al., 2010; Misztal et al., 2012) or small real data studies (e.g., de Roos et al., 2011; Croiseau et al., 2012).

**Genomic Validations**

We used 259 validation bulls and the genomic validations were performed with 4 subsets: Red1 (38K loci), Red2 (116K loci), Red3 (322K loci), and 38k_SNP. Red3 was only run for kilograms of protein and ΔR² was 30.7%. Independent of the other 3 subsets, values for ΔR² were the lowest for 56-d nonreturn rate and the highest for kilograms of fat (Table 1).

For the production traits, ΔR² was highest in Red2, except for kilograms of fat, for which it was slightly higher in Red1 (Table 1). For half of the non-production traits, Red2 resulted in the highest ΔR², whereas for the other half, Red2 resulted in similar ΔR² compared with Red1 or 38k_SNP. Differences in ΔR² between Red2 and Red1 ranged from −0.3% for cell count to 3.2% for 56-d nonreturn rate and differences in ΔR² between Red2 and 38k_SNP ranged from 0.0% for calving interval to 5.2% for cell count (Figure 1). Averaged over all 9 traits, Red2 resulted in 0.9% higher ΔR² compared with Red1 and 1.7% higher ΔR² compared with 38k_SNP (Figure 1). As expected, Red1 resulted in almost 1% higher ΔR² compared with 38k_SNP, on average (Figure 1). For kilograms of protein, Red2 and Red3 resulted in the highest ΔR² (30.7%) and ΔR² was 2% higher compared with Red1 and 1.5% higher compared with 38k_SNP (Figure 1). Note that Red3 only ran for kilograms of protein and that the ranking might be different when running Red3 for all other traits.

**HD Loci for Genomic Evaluation**

The above results suggest that the use of more HD loci based on localized haplotype clusters in genomic
evaluation positively affects $\Delta R^2$ and thus will result in a higher reliability of DGV. The average number of 45 localized haplotype clusters per locus may result in a low number of animals per localized haplotype cluster, which makes it difficult to estimate an effect. A reduction in the number of localized haplotype clusters per locus may alleviate this problem. It is possible to reduce the number of localized haplotype clusters per locus in BEAGLE by changing the scale and shift parameter to obtain the DAG file. A disadvantage, however, is that more localized haplotype clusters are defined as being identical, whereas in reality, that may not be the case. Furthermore, percentage of obsolete HD loci will be lower (as discussed above). Further research is needed to investigate the optimal number of localized haplotype clusters per locus and the optimal reduced HD subset for genomic evaluations.

**CONCLUSIONS**

This study reports on a method to identify obsolete loci in HD data and to increase the reliability of DGV in Holstein-Friesians using a reduced HD data set. This work is, to our knowledge, the first to describe results of

**Table 1.** Number of reference (ref.) animals, and values for squared correlations [$\Delta R^2$ (%)$^1$] of genomic breeding values for 4 validations using 259 validation bulls born in 2003 and 2004$^2$

<table>
<thead>
<tr>
<th>Trait</th>
<th>No. of ref. animals</th>
<th>Red1</th>
<th>Red2</th>
<th>Red3</th>
<th>38k_SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feet and legs</td>
<td>17,296</td>
<td>20.6</td>
<td>21.0</td>
<td>19.1</td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td>18,536</td>
<td>35.8</td>
<td>35.5</td>
<td>30.3</td>
<td></td>
</tr>
<tr>
<td>Protein (kg)</td>
<td>18,485</td>
<td>28.7</td>
<td>30.7</td>
<td>30.7</td>
<td>29.2</td>
</tr>
<tr>
<td>Milk (kg)</td>
<td>18,484</td>
<td>35.0</td>
<td>36.1</td>
<td>35.2</td>
<td></td>
</tr>
<tr>
<td>Fat (kg)</td>
<td>18,483</td>
<td>44.3</td>
<td>44.2</td>
<td>43.1</td>
<td></td>
</tr>
<tr>
<td>Longevity</td>
<td>18,405</td>
<td>13.5</td>
<td>14.2</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>56-d nonreturn rate</td>
<td>17,729</td>
<td>7.0</td>
<td>10.2</td>
<td>9.5</td>
<td></td>
</tr>
<tr>
<td>Calving interval</td>
<td>18,482</td>
<td>26.9</td>
<td>27.7</td>
<td>27.7</td>
<td></td>
</tr>
<tr>
<td>Udder</td>
<td>18,281</td>
<td>31.7</td>
<td>31.6</td>
<td>29.4</td>
<td></td>
</tr>
</tbody>
</table>

$^1\Delta R^2$ is the difference between the squared correlation between direct genomic value (DGV) and deregressed proofs (DRP) and the squared correlation between pedigree EBV (PEBV) and DRP divided by the average reliability of the DRP of all 259 validation bulls.

$^2$38k_SNP refers to the validation using genotypes for 38K SNP as in routine genomic evaluation; Red1 [38,000 (38K) loci], Red2 [116,000 (116K) loci], and Red3 [322,000 (322K) loci] are validations using localized haplotype clusters of a subset of high-density (HD) loci.

**Figure 1.** Difference in squared correlations [$\Delta R^2$ (%)$^1$] among validations analyzed for all 9 traits using 259 validation bulls born in 2003 and 2004. 38k_SNP refers to the validation using genotypes for 38,000 (38K) SNP as in routine genomic evaluation; Red1 (38K loci) and Red2 [116,000 (116K) loci] are validations using localized haplotype clusters of a subset of high-density (HD) loci. Color version available in the online PDF.
the identification of obsolete HD loci based on localized haplotype clusters and the use of HD subsets based on localized haplotype clusters in a genomic evaluation to investigate the effect on reliability of DGV. Eliminating obsolete loci enormously decreased the amount of data to be analyzed and consequently decreased computer requirements for genomic evaluations. The more HD loci used in a genomic evaluation, the higher the increase in reliability of the DGV. When using 116K HD loci based on localized haplotype clusters, it is possible to increase the reliability of DGV by 1 to 2% compared with the SNP currently used for routine genomic evaluation.

ACKNOWLEDGMENTS

The authors thank all partners of the EuroGenomics Consortium for providing part of the data used in our study.

REFERENCES

Browning, B. L. 2011. BEAGLE 3.3.2 Manual. Department of Medicine, Division of Medical Genetics, University of Washington, Seattle.


