The single-step genomic model has become the golden standard for routine evaluation in livestock species, such as Holstein dairy cattle. The single-step genomic model with direct estimation of marker effects has been proven to be efficient in accurately accounting for millions of genotype records. For diverse applications including frequent genomic evaluation updates on a weekly basis, estimates of the marker effects from the single-step evaluations play a central role in genomic prediction. In this study we focused on exploring the marker effect estimates from the single-step evaluation. Phenotypic, genotypic, and pedigree data were taken from the official evaluation for German dairy breeds in April 2021. A multilactation random regression test-day model was applied to more than 242 million test-day records separately for 4 traits: milk, fat, and protein yields, and somatic cell scores (SCS). Approximately one million genotyped Holstein animals were considered in the single-step genomic evaluations including ~21 million animals in pedigree. Deregressed multiple across-country breeding values of Holstein bulls having daughters outside Germany were integrated into the national test-day data to increase the reliability of genomic breeding values. To assess the stability and bias of the marker effects of the single-step model, test-day records of the last 4 yr were deleted, and the integrated bulls born in the last 4 yr were truncated from the complete phenotypic dataset. Estimates of the marker effects were shown to be highly correlated, with correlations ~0.9, between the full and truncated evaluations. Regression slope values of the marker-effect estimates from the full on the truncated evaluations were all close to their expected value, being ~1.03. Calculated using random regression coefficients of the marker effect estimates, drastically different shapes of the genetic lactation curve were seen for 2 markers on chromosome 14 for the 4 test-day traits. The contribution of individual chromosomes to the total additive genetic variances seemed to follow the polygenic inheritance mode for protein yield and SCS. However, chromosome 14 was found to make an exceptionally large contribution to the total additive genetic variance for milk and fat yields because of markers near the major gene DGAT1. For the first lactation test-day traits, we obtained ~0 correlations of chromosomal direct genomic values between any pair of the chromosomes; no spurious correlations were found in our analysis, thanks to the large reference population. For trait milk yield, chromosomal direct genomic values appeared to have a large variation in the between-lactation correlations among the chromosomes, especially between first and second or third lactations. The optimal features of the random regression test-day model and the single-step marker model allowed us to track the differences in the shapes of genetic lactation curves down to the individual markers. Furthermore, the single-step random regression test-day model enabled us to better understand the inheritance mode of the yield traits and SCS (e.g., variable chromosomal contributions to the total additive genetic variance and to the genetic correlations between lactations).

Key words: single-step model, random regression test-day model, marker, genetic lactation curve, chromosome

INTRODUCTION

Genomic selection using SNP information has revolutionized dairy cattle breeding (VanRaden, 2008) based on the concept of the genomic model (Meuwissen et al., 2001). A single-step genomic BLUP model (ssGBLUP; Aguilar et al., 2010; Christensen and Lund 2010) uses simultaneously all phenotypic, genotypic, and pedigree data to conduct genomic prediction. Therefore, the ssGBLUP model has become the golden standard for routine genomic evaluation in livestock species such as Holstein. In comparison to the ssGBLUP model, a single-step SNP BLUP model (ssSNPBLUP; Liu et al., 2014) estimates additive genetic effects of the SNP markers jointly with all other effects of the single-step model and has been proven to be most efficient for large-scale genomic evaluation in livestock (Vandenplas
et al., 2019, 2021; Alkhoder et al., 2022; Osawa et al., 2022). The SNP effects estimated with the phenotypic data in genotyped animals play a central role for predicting genomic breeding values (GEBV) of particularly young selection candidates without their own performance. Moreover, frequent genomic evaluation such as weekly genomic evaluation in German Holstein requires accurate estimates of the SNP marker effects.

Before the genomic selection, genetic improvement was mainly based on conventional genetic evaluation that focused on optimal modeling of diverse types of phenotype data. For example, test-day milk yields were analyzed with a so-called random regression test-day model (RRTDM, Jamrozik et al., 1997; Schaeffer et al., 2000; Liu et al., 2004). However, the complex conventional model such as RRTDM has not been able to be considered in the current 2-step genomic evaluation which usually used a single-trait or single-effect genomic model (Alkhoder et al., 2022). Oliveira et al. (2019) applied the ssGBLUP model to the test-day traits of Canadian dairy cattle breeds using a multiple-trait RRTDM model for their single-step genomic evaluation. Alkhoder et al. (2022a) conducted a single-step evaluation of test-day traits in German dairy cattle using a ssSNPBLUP RRTDM model. Both studies demonstrated the combined advantages of the single-step model as well as the RRTDM model.

According to the theory of the RRTDM model, additive genetic effects of the SNP markers from the ssSNPBLUP model are expressed in form of random regression coefficients (RRC). For the multiple-lactation RRTDM model (Liu et al., 2004; Alkhoder et al., 2022), every SNP marker received estimates of the RRC effects for each of the lactations, with which SNP effects on a 305-d lactation basis can be calculated thereafter. To our knowledge, little has been investigated in the literature yet about the properties of the SNP effects in the forms of RRC for milk production traits in dairy cattle. Furthermore, we want to learn how the SNP effects change between genomic evaluations across time.

The objectives of this study were (1) to assess the stability and bias of the SNP effect estimates of the ssSNPBLUP model, (2) to evaluate the shape of genetic lactation curves by individual SNP markers, and (3) to compare chromosomal contributions to the additive genetic variance and genetic correlations between lactations.

**MATERIALS AND METHODS**

Institutional Animal Care and Use Committee (IACUC) guidelines were followed in this study. Because the routine collection of test-day records by milk recording agencies in last decades does not harm the dairy cows in any way, IACUC approval was not necessary.

**A Single-Step SNP BLUP Multiple-Lactation RRTDM**

Test-day milk yields (MKG), fat yields (FKG), protein yields (PKG), or SCC from the first 3 lactations of a cow were analyzed with the following RRTDM model for dairy cattle in Germany (Liu et al., 2004):

$$y = X_h h + X_f f + V_p + V_u + e,$$  \[1\]

where $y$ is a vector of the cow’s test-day yields or SCC in the first 3 lactations that have been pre-adjusted for heterogeneous variances (Reents et al., 1998), $h$ is a vector of fixed effects of herd-test-date-parity-milking-frequency associated with her test-day records, $f$ is a vector of fixed lactation curve effects, modeled as regressions on DIM using the Wilmink function (Liu et al., 2004), $p$ is a vector of permanent environmental effects that are expressed as $RRC$, $u$ is a vector of additive genetic effects or breeding values of the cow also as $RRC$, and $e$ is a vector of random residuals. For modeling the effect $u$, normalized orthogonal Legendre polynomials with 3 parameters are used (e.g., for the additive genetic effects):

$$u_i + \sqrt{3}v \times u_2 + \frac{1}{2} \sqrt{5} (3v^2 - 1) \times u_3,$$  \[2\]

with $v = \frac{2(x - 5)}{300} - 1$ representing standardized DIM with values ranging from $-1$ to $1$ and DIM with the range $x = [5,305]$ (Liu et al., 2004), and $u_i$ is the $i$th RRC ($i = 1, 2, 3$) of breeding value as the regression coefficient of variable $v$ on test-day yield or SCC. The RRC $u_1$, $u_2$, and $u_3$ are the intercept, slope, and a quadratic term, respectively. The same Legendre polynomials were used for effect $p$. Design matrix $V$ for either $u$ or $p$ contains the Legendre coefficients for the first 3 lactations of the cow. $X_h$ is the incidence matrix for the fixed herd-test-date-parity-milking-frequency effects associated with this cow’s test-day records $h$; $X_f$ is the design matrix for the fixed lactation curve effects $f$. The joint distribution of the 3 RRC in first 3 lactations for the additive genetic effects of the cow is

$$var(V_u) = \begin{bmatrix}
V_1 & \begin{bmatrix}
\begin{bmatrix}
U_{0,11} & U_{0,12} & U_{0,13}
\end{bmatrix}
\begin{bmatrix}
V_1^T
\end{bmatrix}
\end{bmatrix}
& \begin{bmatrix}
U_{0,12} & U_{0,22} & U_{0,23}
\end{bmatrix}
\begin{bmatrix}
V_2^T
\end{bmatrix}
\end{bmatrix}
\begin{bmatrix}
U_{0,13} & U_{0,23} & U_{0,23}
\end{bmatrix}
\begin{bmatrix}
V_3^T
\end{bmatrix}
\end{bmatrix},$$  \[3\]
where $V_i$ denotes the design matrix for this cow’s test-day records in lactation $i$ ($i = 1, 2, 3$) and $U_{0-ij}$ the genetic covariance matrix of RRC between lactations $i$ and $j$, which is defined as

$$U_{0-ij} = \begin{bmatrix} \sigma_{11-ij}^2 & \sigma_{12-ij} & \sigma_{13-ij} \\ \sigma_{21-ij} & \sigma_{22-ij} & \sigma_{23-ij} \\ \text{symm.} & \sigma_{32-ij} & \sigma_{33-ij} \end{bmatrix},$$

where $\sigma_{kk-ij}^2$ and $\sigma_{kl-ij}$ represent additive genetic variance of $k$th RRC ($k = 1, 2, 3$) and covariance between $k$th and $l$th ($l = 1, 2, 3$) RRC for the lactation pair $i$ and $j$, respectively. Similar joint distribution is also defined for the permanent environmental RRC. The additive genetic and permanent environment effects are correlated between the lactations. However, no residual correlations are assumed between the lactations for the multilactation test-day model.

Table 1 shows genetic parameters of the single-step random regression test-day model used in the genomic evaluation of milk yields for German dairy breeds.

<table>
<thead>
<tr>
<th>Heritability or genetic correlations on or above diagonal</th>
<th>Genetic regression coefficient</th>
<th>Intercept</th>
<th>Slope</th>
<th>Quadratic</th>
<th>305-d lactation</th>
<th>Combined lactation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic regression coefficients of lactation</td>
<td>First</td>
<td>0.32</td>
<td>0.17</td>
<td>−0.51</td>
<td>0.999</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>0.07</td>
<td>0.04</td>
<td>−0.46</td>
<td>0.12</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Quadratic</td>
<td>0.09</td>
<td>0.18</td>
<td>0.04</td>
<td>−0.52</td>
<td>−0.53</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>0.01</td>
<td>0.04</td>
<td>−0.51</td>
<td>−0.01</td>
<td>−0.01</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>0.08</td>
<td>0.12</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Quadratic</td>
<td>0.04</td>
<td>−0.50</td>
<td>−0.50</td>
<td>−0.50</td>
<td>−0.50</td>
</tr>
<tr>
<td>305-d yield of lactation</td>
<td>Lactation</td>
<td>First</td>
<td>0.43</td>
<td>0.85</td>
<td>0.84</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Second</td>
<td>0.55</td>
<td>0.26</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Third</td>
<td>0.52</td>
<td>0.54</td>
<td>0.25</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Combined lactation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.39</td>
</tr>
</tbody>
</table>

According to the ssSNPBLUP model (Liu et al., 2014), GEBV of all genotyped animals, $u_{all}$, sorted by RRC within animal, are further decomposed into 2 components:

$$u_{all} = Zg + a,$$

where matrix $Z$ is a design matrix containing regression coefficients on animals’ genotype data (Liu et al., 2016), and $g$ is a vector of all SNP marker effects also in form of RRC and $a$ is a vector of residual polygenic effects (RPG) for all the genotyped animals, expressed in RRC with the joint distribution:

$$g \sim N(0, (1 - m) \begin{bmatrix} U_{011} & U_{012} & U_{013} \\ U_{022} & U_{023} \\ \text{symm.} & U_{023} \end{bmatrix} \otimes B, \begin{bmatrix} 0 & kA \\ \text{symm.} & U_{023} \end{bmatrix}),$$

where $k$ represents percentage of additive genetic variance not explained by the SNP markers, $\otimes$ denotes the Kronecker product, and matrix $B$ being

$$B = \left( \sum_{j=1}^{m} 2p_j \left( 1 - p_j \right) \right)^{-1} I,$$

where $m$ represents the number of SNP markers, $p_j$ is allele frequency of the marker $j$ of the current population, and $I$ is the identity matrix. The vector $a$ is a vector of residual polygenic effects (RPG) for all the genotyped animals, also expressed in RRC with the joint distribution:

$$a \sim N(0, kA \begin{bmatrix} U_{011} & U_{012} & U_{013} \\ U_{022} & U_{023} \\ \text{symm.} & U_{023} \end{bmatrix}),$$
where $A_{gg}$ represents pedigree relationship matrix for all the genotyped animals. It was assumed that RPG explained 30% of additive genetic variance, $k = 0.3$, for any of the test-day traits (Alkhoder et al., 2022).

A total of 45,613 SNP markers, $m = 45,613$, were selected for genomic evaluation in German Holstein, covering all 31 chromosomes including 29 autosomes and the sex chromosomes. The selection of the SNP marker set was primarily based on the minor allele frequency with a minimum of 0.01 in the German Holstein genotyped population. At the time of setting the SNP marker set, Illumina Bovine SNP 50K chips were used as the standard chip for German genomic evaluation. Table 2 gives numbers of SNP markers selected for all the chromosomes for genomic evaluation of German Holstein.

In the German dairy cattle evaluation (Liu et al., 2004), breeding values on 305-d basis for the 3 yield traits or on an average daily basis for SCS were weighted across 3 lactations and were defined as the official breeding values ($u_{comb}$):

$$u_{comb} = \sum_{l=1}^{3} w_l u_{305l},$$

where $w_l$ represents the weight for lactation $l$, $u_{305l}$ is the 305-d lactation breeding value of lactation $l$, $u_j$ is the breeding value of the $j$th RRC of lactation $l$, and $v_i$ is the standardized day in milk $i$.

The official breeding values expressed on the combined lactation basis were submitted to Interbull’s bull multiple across country evaluation (MACE). Deregressed MACE EBV of bulls from the MACE evaluation are analyzed with a single-trait animal model:

$$y_m = \mu + u_m + e_m,$$

where $y_m$ represents deregressed EBV (DRP) of the bulls for a MACE trait, such as MKG, FKG, PPK, or SCS; $\mu$ is a general mean; $1$ is a vector of 1’s; $u_m$ is a vector of breeding values of the bulls for the MACE trait; and $e_m$ is a vector of residual effects of the bulls for the MACE trait with

$$\text{var}(e_m) = \sigma^2_{em} D,$$

where $\sigma^2_{em}$ is residual variance of the MACE trait, and matrix $D$ is a diagonal matrix containing reciprocal values of effective daughter contribution (EDC) of the bulls expressed on an animal-model basis. For genotyped MACE bulls, their GEBV can be decomposed as

$$u_m = Z_m g_m + a_m,$$

where $g_m$ is a vector of SNP effects for the MACE trait, $Z_m$ is a design matrix containing regression coefficients on the bulls’ genotype data, and $a_m$ is a vector of RPG effects of the bulls. The RPG effects of the genotyped bulls with MACE data have a joint distribution:

$$a_m \sim N\left(0, k\sigma^2_{am} A_{mm}\right),$$

where $A_{mm}$ represents pedigree relationship matrix for the genotyped MACE bulls, and $\sigma^2_{am}$ is additive genetic variance of the MACE trait that is equal to variance of the official breeding values $u_{comb}$. Similar to Equation [6], joint distribution of additive genetic effects of the SNP markers for the MACE trait is

### Table 2. Number of SNP markers by chromosome selected for genomic evaluation in German Holsteins

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2,905</td>
</tr>
<tr>
<td>2</td>
<td>2,357</td>
</tr>
<tr>
<td>3</td>
<td>2,190</td>
</tr>
<tr>
<td>4</td>
<td>2,168</td>
</tr>
<tr>
<td>5</td>
<td>1,868</td>
</tr>
<tr>
<td>6</td>
<td>2,218</td>
</tr>
<tr>
<td>7</td>
<td>1,947</td>
</tr>
<tr>
<td>8</td>
<td>2,100</td>
</tr>
<tr>
<td>9</td>
<td>1,763</td>
</tr>
<tr>
<td>10</td>
<td>1,858</td>
</tr>
<tr>
<td>11</td>
<td>1,936</td>
</tr>
<tr>
<td>12</td>
<td>1,485</td>
</tr>
<tr>
<td>13</td>
<td>1,527</td>
</tr>
<tr>
<td>14</td>
<td>1,546</td>
</tr>
<tr>
<td>15</td>
<td>1,496</td>
</tr>
<tr>
<td>16</td>
<td>1,441</td>
</tr>
<tr>
<td>17</td>
<td>1,391</td>
</tr>
<tr>
<td>18</td>
<td>1,176</td>
</tr>
<tr>
<td>19</td>
<td>1,192</td>
</tr>
<tr>
<td>20</td>
<td>1,385</td>
</tr>
<tr>
<td>21</td>
<td>1,192</td>
</tr>
<tr>
<td>22</td>
<td>1,103</td>
</tr>
<tr>
<td>23</td>
<td>947</td>
</tr>
<tr>
<td>24</td>
<td>1,095</td>
</tr>
<tr>
<td>25</td>
<td>866</td>
</tr>
<tr>
<td>26</td>
<td>948</td>
</tr>
<tr>
<td>27</td>
<td>863</td>
</tr>
<tr>
<td>28</td>
<td>842</td>
</tr>
<tr>
<td>29</td>
<td>942</td>
</tr>
<tr>
<td>30/X chromosome</td>
<td>773</td>
</tr>
<tr>
<td>31/Y chromosome</td>
<td>93</td>
</tr>
<tr>
<td>Total</td>
<td>45,613</td>
</tr>
</tbody>
</table>
\[ g_m \sim N \left[ 0, (1-k) \sigma_m^2 B \right]. \]  

To avoid double counting the contribution of domestic daughters to the bulls evaluated in MACE, adjustments were carried out for both EDC and DRP of the bull (Alkhoder et al., 2022). Because the MACE trait is the German official breeding value expressed on a combined lactation basis, its genetic correlations with the national genetic RRC were derived based on the genetic (co)variance matrix of \( \mathbf{u} \). Genetic correlations for the genetic RRC with the MACE trait (i.e., the combined 305-d lactation) are shown in Table 1.

**Direct Genomic Values Per SNP Marker and Per Chromosome**

For a genotyped animal \( i \), direct genomic breeding value (DGV) for a selected trait is calculated, according to Liu et al. (2016), with

\[
d_i = \sum_{j=1}^{M} \left( n_{ij} - 2p_j \right) g_j, \tag{15}
\]

where \( d_i \) denotes DGV of the animal \( i \) for the chosen trait, \( g_j \) is the additive genetic effect of SNP marker \( j \) describing the substitution effect of allele A, \( p_j \) is the observed frequency of allele A of the \( j \)th marker calculated from the current genotyped population, \( M = 45,613 \) is the total number of fitted SNP markers, and \( n_{ij} = 0, 1, 2 \) is the number of copies of allele A in genotype BB, AB, or AA, respectively. The term \( n_{ij} - 2p_j \) equals \( z_{ij} \) defined in the design matrices of genotyped animals in Equations [5] and [12].

For a specific SNP marker \( j \), this animal’s DGV can be defined, following Liu et al. (2016), as

\[
d_{ij} = \left( n_{ij} - 2p_j \right) g_j, \tag{16}\]

Similarly, for a given chromosome \( c \), we define DGV of this animal \( i \) as

\[
d_{ic} = \sum_{j=M_{c1}}^{M_{cn}} \left( n_{ij} - 2p_j \right) g_j, \tag{17}\]

where \( M_{c1} \) and \( M_{cn} \) represent the starting and ending SNP markers on the chromosome \( c \). For the genotyped animal, \( d_i, d_{ic} \), and \( d_j \) represent genome-wide, chromosome-wide, and SNP marker specific DGV, respectively.

**Breeding Values of the SNP Markers**

Same as the GEBV or DGV of evaluated animals, the SNP marker effects \( \mathbf{g} \) in Equation [5] are expressed in RRC as well (Equation [2]). For a given SNP marker \( j \), its 305-d lactation breeding value for lactation \( l \) is computed as

\[
L_{jl} = c_1 \times g_{1l} + c_2 \times g_{2l} + c_3 \times g_{3l}, \tag{18}\]

where \( L_{jl} \) represents 305-d lactation breeding value of the \( j \)th SNP marker for lactation \( l (l = 1, 2, 3) \), \( g_{kl} \) is the SNP effect of the \( k \)th RRC \( (k = 1, 2, 3) \) of lactation \( l \), and \( c_1 = 305, c_2 = 0, c_3 = 2.273 \) (Liu et al., 2004). For SCS, lactation breeding values are defined as the average of breeding values of DIM 5 to DIM 305, instead of the sum of breeding values of all DIM as for the 3 yield traits. A combined lactation breeding value for SNP marker \( j \) is a weighted average of its 305-d lactation breeding values of the first 3 lactations:

\[
C_j = w_1 \times L_{j1} + w_2 \times L_{j2} + w_3 \times L_{j3}, \tag{19}\]

where \( C_j \) represents the combined lactation breeding value of SNP marker \( j \), \( w_l \) denotes relative weight on lactation \( l \), and \( w_1 + w_2 + w_3 = 1 \). For the 3 yield traits, the 3 lactations had equal weights \( w_1 = w_2 = w_3 = \frac{1}{3} \); whereas for SCS the weights are 0.26, 0.37, and 0.37 for first, second, and third lactation 305-d breeding values, respectively.

Similar to the expression of the SNP effects on a 305-d lactation basis, marker DGV of the diverse expression forms can be expressed on different bases: SNP specific DGV calculated with Equation [16] for a given marker for a lactation, chromosome-wide DGV of a chromosome for lactation \( l \) using Equation [18], and genome-wide DGV on a combined lactation basis with Equation [19].

**Phenotypic, Genotypic, and Pedigree Data for the Single-Step Evaluation**

Phenotypic, genotypic, and pedigree data were obtained from the official April 2021 evaluation for German dairy cattle (Alkhoder et al., 2022). Table 3 summarizes the genotype and phenotype data used for the single-step evaluation. A total of 242,121,126 test-day records from 12,432,940 cows of the breeds, Holstein, Red Dairy Cattle, and Jersey, were analyzed together with 138,770 MACE Holstein bulls that had daughters outside of Germany. The total number of national cows with test-day data and integrated bulls with foreign daughters amounted to 12,571,710. All genotyped Holstein animals, including culled animals, were jointly evaluated with those animals with phenotype data, and the total number of genotyped Holstein...
animals was 949,636 for the April 2021 evaluation. A maximum number of 20 generations was set to trace ancestors of the genotyped or phenotyped animals. Additionally, the oldest bulls with daughters or cows with records were guaranteed to have at least 3 generations of ancestors. The pedigree file for the single-step evaluation contained 20,461,400 real animals and 176 phantom parent groups that were defined according to breed, country of origin, 4 selection paths, and birth year of animals having parents missing. There were 130, 20, and 26 phantom parent groups for the breeds Holstein, Red Dairy Cattle, and Jersey, respectively. Among all the cows with test-day records, 293,724 cows had genotype data. In total, 43,699 genotyped Holstein bulls had daughters either in Germany or in foreign countries included in the MACE evaluation.

A Genomic Validation for the Single-Step Evaluation

Test-day records from last 4 test years, 2017 to 2021, were removed to simulate a genomic evaluation 4 yr ago. Because of the lack of MACE EBV from a truncated MACE evaluation, bull MACE evaluations from April 2021 were used instead for the validation study. The 4 youngest birth years of the bulls were deleted from the MACE data. As an extra step, national daughters of the validation bulls were additionally deleted from the truncated test-day data, if any daughters were left in the truncated test-day dataset. All the genotype records used in the full evaluation were also considered in the truncated genomic evaluation. After the data truncation, 222,634,210 test-day records from 10,903,891 cows and 128,504 bulls with foreign daughters remained in the phenotype data for the validation study (Table 3). The numbers of genotyped bulls with daughters or genotyped cows with test-day records were reduced to 37,283 from 43,699 or to 112,759 from 293,724, respectively. Due to the routine whole-herd female animal genotyping service started in Germany in 2016, the number of genotyped female animals is much higher than that of genotyped male animals. Furthermore, more than 60% genotyped cows with test-day records from the last 4 yr were removed for the validation study. Figure 1 shows the numbers of cows with test-day records or with genotype records in the full or truncated evaluations. Because of the relatively short history of the large-scale genotyping service in Germany, the number of genotyped cows with test-day records was significantly lower in the truncated dataset than in the full dataset, but still at a high level.

RESULTS AND DISCUSSION

Our SNP markers in Table 2 contained a total of 866 markers on the sex chromosomes: 773 on chromosome X and 93 on chromosome Y. This set of SNP markers was set up at the beginning of genomic selection based on a 2-step genomic model and a bull reference population. As argued by Fernando and Grossman (1990), the transmission of markers on the sex chromosomes was not correctly modeled by the pedigree relationship matrix A. As Druet and Legarra (2020) pointed out, a bull does not transmit the nonpseudo-autosomal sex chromosome X to his male progeny, therefore, using the pedigree matrix A to model the transmission of sex chromosomes is not correct. Removal of the SNP markers on the sex chromosomes from routine single-step evaluation is planned for German Holstein. However, the SNP markers on the sex chromosomes were
included in this investigation. Because of most of the genotyped animals being females (Table 3), the effect of the inadequate modeling of the SNP markers on the X chromosomes in the genotyped sires should be limited.

Single-step evaluations for the test-day traits with the full or truncated datasets were conducted using the software package MiX99 (Strandén and Lidauer, 1999), in which the ssSNPBLUP by Liu et al. (2014) was implemented in a special way (Esa Mäntysaari, personal communication). In addition, we applied the software package MiXBLUP (Ten Napel et al., 2020) to the same datasets. Approximately equal execution time was needed for either of the software packages to reach the predefined convergence criteria. However, MiXBLUP needed only a maximum RAM of 135Gb for the full data evaluation, but MiX99 required 358Gb for the same dataset due to the difference in the implementation of the ssSNPBLUP model. Because identical effect estimates were obtained with the 2 software packages, we analyzed only the solutions from MiX99 for further analyses. For the single-step evaluation of trait MKG, the total number of all estimated effects was 336,548,724 with the full dataset (Table 3). The total clock time per round of iteration was 1.63 min using 17 threads on a Linux server with 2 × 10 cores of Intel Xeon CPU E5–2690 v2 @3.00Ghz. Based on the convergence criteria CD (convergence according to differences in solutions) or CR (convergence according to residuals of preconditioned gradients) (Strandén and Lidauer, 1999), reasonably accurate solutions could be obtained within approximately 2,000 rounds of iteration. The single-step model gave higher correlation of the SNP effect estimates between the full and truncated validation datasets than the current multistep genomic model. The regression coefficient of the SNP effect estimates, expressed on the combined lactation basis, from the full on the truncated evaluation was ~1 for the single-step model. Based on the results of the GEBV test on German validation bulls, the single-step model was proven to give unbiased genomic prediction for the 4 test-day traits. More details about the rate of convergence and GEBV validation results can be found in the study by Alkhoder et al. (2022).
Correlation and Regression of the SNP Effect Estimates Between the Evaluations

Between the full and truncated single-step evaluations, estimates of SNP effects or marker breeding values in different expression forms were compared. First, SNP effect estimates on the combined lactation basis (Equation [19]) from the full evaluation were regressed on those from the truncated evaluation. Figure 2 shows observed correlation and regression coefficients of the SNP effect estimates in the form of combined lactation for the 4 test-day traits. The correlations of the SNP effect estimates are high, ranging from 0.895 for trait protein to 0.935 for SCS, despite the large difference in the number of genotyped and phenotyped cows with test-day records and bulls with daughters (Table 3) between the full and truncated evaluations. The regression slopes of the SNP effect estimates are all close to their expected value of 1. Because standard errors (SE) of the regression slope estimates are ~0.02, none of the slope estimates deviate significantly from their expected value. These regression slope coefficients indicate neither an inflation nor a deflation of the SNP effect estimates.

Correlation and Regression of the SNP Effect Estimates in the RRC Form

Second, we compared the SNP effect estimates or marker breeding value estimates also in RRC form. Table 4 shows correlations of the SNP marker effect estimates between the full and truncated evaluation as well as regression slopes of the SNP marker effect estimates of the full on truncated evaluation for the test-day trait MKG. The SNP effect estimates in the RRC form were highly correlated between the 2 evaluations, all being 0.88 or higher. The quadratic term of the Legendre polynomials was the least correlated among the 3 RRC coefficients, with the lowest correlation of 0.88 for the quadratic term of first lactation MKG. The highest correlation was seen for the slope term of the Legendre polynomials among all the 3 terms. Overall, the SNP effect estimates of first lactation were correlated between the 2 evaluations with the lowest correlation value among all the 3 lactations, possibly caused by the largest increase in the number of genotyped cows with first lactation test-day records between the full and truncated evaluations. Regarding the regression slopes of the SNP effect estimates in the RRC form, their estimates ranged from 0.983 for the quadratic term of the Legendre polynomials for first lactation to 1.032 for the intercept term for second lactation MKG. In general, all the SNP effect estimates in the RRC form were unbiased, because of the SE of the regression slope being ~0.02. It appeared that the quadratic term of the SNP effect estimates has lower regression slope value than the other 2 terms, all slightly below 1. Similar results of the SNP effect estimates in the RRC form were found also for the other test-day traits, FKG, PKG and SCS, with little difference to those for the trait MKG.

Figure 2. Correlation and regression values of the SNP effect estimates, expressed on a combined 305-d lactation basis, of the full and truncated single-step evaluations for the 4 test-day traits.
Third, we examined the consistency of the SNP effect estimates or marker breeding value estimates on the 305-d lactation basis for the trait MKG. The correlations of the SNP effect estimates between the full and truncated evaluations were between 0.89 for first lactation to 0.92 for the second lactation, being close to the correlations of the SNP effect estimates in the RRC form. The combined lactation SNP effect estimates had a correlation of 0.91 between the full and truncated evaluations. Little difference in the SNP effect estimates in the form of 305-d lactation existed in the other traits, FKG, PKG, and SCS, as the trait MKG.

Genetic correlations between the 305-d lactation and the RRC terms in Table 1 were close to 1 for the intercept term, very low for the slope term and negative for the quadratic term of the Legendre polynomials. As expected, we observed very similar correlation patterns between the 305-d lactation and the RRC forms of the SNP effect estimates from the full and truncated evaluations. The similar correlation patterns between the true genetic correlations and the observed ones indicate that the SNP effect estimates in the RRC form or 305-d lactations are reliably estimated.

Table 4 also shows the regression slope values of the SNP effect estimates in the 305-d lactation form from the full evaluation on the truncated evaluation are ~1.03 for each of the 3 lactations. On the combined 305-d lactation basis, the regression slope of the SNP effect estimates is 1.031. All the regression slope values suggest an unbiased estimation of the SNP effect estimates for the trait MKG, which was the case for all the other traits.

With more phenotypic data added to single-step evaluation, variance of the SNP effect estimates or the marker breeding value estimates is expected to increase. Table 4 also shows the ratio of standard deviations (SD) of the SNP effect estimates, in either RRC or 305-d lactation forms, of the full evaluation over the truncated evaluation. We can see that there is a 14% increase in SD of the SNP effect estimates for the combined 305-d lactation, and a 15%, 13%, and 12% increase for the first, second, and third lactations, respectively. Looking at the SNP effects in the RRC form, we observe that the variance increase is largest for the intercept term, followed by the slope and quadratic term of the Legendre polynomials. The positive correlation of the SNP effect estimates between the evaluations and the higher variance of the SNP effect estimates in the full evaluation mean that the SNP markers of larger effect will receive even larger estimates over time, and those of small effect will deviate further from zero when more phenotypic data will be added to the single-step evaluation.

**Marker-Specific Genetic Lactation Curves**

Oliveira et al. (2019) demonstrated that the ssGB-LUP model was able to model the genetic lactation curves for GEBV of individual animals. Comparing to
the ssGBLUP model that can also estimate SNP effects by back-solving GEBV, our ssSNPBLUP (Liu et al., 2014) enables direct modeling of not only animal GEBV but also the SNP effects in genetic lactation curves using the Legendre polynomials. Figure 3 and Figure 4 show genetic lactation curves for 2 selected SNP markers, both near the DGAT1 gene on chromosome 14, one with the largest positive effect and the other with the largest negative effect on trait MKG from the full single-step evaluation, respectively.

For the SNP marker, ARS-BFGL-NGS-4939 with a frequency of allele A being 0.73 and a large positive effect of allele A on trait MKG, genetic lactation curve shapes appear to differ between the trait MKG or PKG in concave form on one and FKG or SCS in convex form on the other. But no difference in the shape of the genetic curves is observed across the first 3 lactations for any of the traits. This SNP marker seems to have a biologically favorable effect on MKG and PKG and unfavorable effect on FKG and SCS.

Figure 4 shows genetic lactation curves of the SNP marker near the DGAT1 gene on chromosome 14, Hapmap30381-BTC-005750 with a frequency of allele A being 0.81 and a large negative effect of allele A on trait MKG. In contrast to Figure 3 for the SNP marker with a large positive effect on MKG, this SNP marker has a convex form of genetic lactation curve for traits MKG and PKG but a concave form for FKG and SCS. No difference in the shape of the genetic curves is observed across the lactations for any of the traits.

Furthermore, we compared the genetic lactation curves of the 2 SNP markers between the full and truncated evaluation. No change in the curve shape was seen in any of the traits or lactations. However, the 305-d lactation breeding values of the 2 SNP markers increased from the truncated to the full evaluation with more phenotype data added. This is indeed expected: the SNP effect estimates differentiated more from one another when more phenotype data were used in the genomic evaluation.
We also examined genetic lactation curves for 2 genotyped female animals, one with extremely high and the other with extremely low GEBV, for trait MKG on the combined lactation basis. At the chromosomal level, the superior female animal had clearly higher GEBV average across all chromosomes than the inferior one, especially for chromosome 14. The GEBV curve of first lactation for the superior female had a large positive intercept term but a concave form for chromosome 14, whereas the inferior female animal had a GEBV lactation curve, for chromosome 14, with a large negative intercept term and in convex form for MKG in first lactation.

Additive Genetic Variances by Chromosomes

For all 293,724 genotyped cows with test-day records and 43,699 genotyped bulls with daughters in the full single-step evaluation, we computed, using Equation [17], their chromosomal DGV of each chromosome for all the 3 lactations of the 4 test-day traits. Then we calculated variances of the chromosomal DGV in the genomic reference population for each trait and lactation combination. Figure 5 shows proportions of additive genetic variances contributed by all the chromosomes for first 3 lactations and combined lactation of the 4 test-day traits, calculated as the variance of chromosome DGV divided by their sum. Due to the low number of SNP markers, 93, on the sex chromosome Y (Table 2), we plotted variance contributions only for 29 autosomes and the sex chromosome X labeled as chromosome 30.

Because of the assumption of equal SNP marker variances under the ssSNPBLUP model, the additive genetic variances contributed by chromosomes in Figure 5 should be highly correlated with the numbers of SNP markers on the chromosomes (Table 2). This can indeed be seen for trait PKG or SCS. However, for traits MKG and FKG heavily influenced by the major gene DGAT1, chromosome 14 harboring DGAT1 seems to contribute a significantly higher proportion of additive genetic variance (ca. 15% to 25% for FKG and 25% to 30% for MKG) in the first 3 lactations and the combined
lactation. Chromosome 14 contributes the lowest percentage of additive genetic variance for first lactation MKG (MKGlact1) and FKG (FKGlact1) among the 3 lactations. Chromosome 5 also appears to explain more additive genetic variances than expected based on the number of SNP markers for both traits MKG and FKG, likewise chromosome 6 for trait MKG and chromosome 2 for FKG. Traits PKG and SCS seem to have a more polygenic inheritance than MKG or FKG (i.e., additive genetic variances contributed by the chromosomes are in proportion to the numbers of SNP markers on the chromosomes). However, chromosomes 1, 3, 5, 6, 14, or 15 seem to contribute a little bit more than expected from the numbers of SNP markers, above 4% of the total additive genetic variance each for traits PKG or SCS.

Correlations of DGV Between Chromosomes

The genomic model by Meuwissen et al. (2001) treats all the SNP markers theoretically in the same way by not distinguishing between closely linked markers on the same chromosome and SNP markers located on different chromosomes. When the number of genotyped and phenotyped animals is not sufficiently large enough, spurious correlations might be shown up between the SNP marker effect estimates, even for markers located on different chromosomes. To verify if the SNP effect estimates are not correlated between the markers located on different chromosomes, we show in Figure 6 the correlations of chromosomal DGV of any chromosome with all the other 30 chromosomes for first lactation of the 4 traits based on the reference animals from the full single-step evaluation. The correlation of chromosomal DGV of any chromosome with itself has been excluded. It can be seen in Figure 6 that all the correlations of chromosomal DGV between any pair of chromosomes are between −0.08 and 0.08 for all the traits in first lactation. No systematic pattern in the correlation structure can be identified in Figure 6. These low correlations of chromosomal
DGV between the chromosome pairs demonstrate no spurious association of the SNP effect estimates of the markers located on different chromosomes. Additionally, ~0 correlations between the chromosomal DGV were also found for the later lactations of the 4 test-day traits.

**Correlations of Chromosomal DGV Between Lactations**

For first 3 lactations of trait MKG, we calculated correlations of chromosomal DGV between any pair of the lactations using the reference animals from the full single-step evaluation. Because of the small number of SNP markers on chromosome Y, we did not consider this sex chromosome in this calculation. Figure 7 shows the between-lactation correlations of chromosomal DGV for each chromosome. At the genome level, genetic correlation between first and second lactation is 0.85 for MKG (Table 1), and we can detect a considerable variation in the genetic correlations between the 2 lactations on the chromosome level (lact1–2, blue line), ranging from 0.79 for chromosome 22 to 0.98 for chromosome 14. However, average of the lactation correlations across all chromosomes is close to the assumed genetic correlation of 0.85 on the genome level. Between first and third lactations we observe a slightly higher variation in the between-lactation correlations (lact1–3, red line) among all the chromosomes than between first and second lactations. On average, the chromosomal correlations between first and third lactations are a little lower than between first and second lactations, which is also the case for the genetic correlation between lactations on the genome level shown in Table 1. Among all the chromosomes, we see that chromosome 14 has the highest correlation of chromosomal DGV between first and third lactations. For the pair of second and third lactations, genetic correlation is assumed to be 0.97 on the genome level (Table 1), and we see in Figure 7 that the correlations of chromosomal DGV between the 2 lactations (lact2–3, green line) are all between 0.95 and 0.99. The highest correlation between second and third lactation MKG is found again for chromosome 14, likely caused by the major gene DGAT1.

---

Figure 6. The direct genomic breeding value correlations of a chromosome with all the remaining 30 chromosomes for the 4 test-day traits in first lactation. MKG = milk yields; PKG = protein yields; FKG = fat yields.
CONCLUSIONS

The marker effects of the single-step random regression test-day model were proven to be accurate and unbiased, via a genomic validation study with a truncated dataset. Marker specific genetic lactation curves were shown to differ in the shape of lactation curves not only between the markers but also between the test-day traits. Protein yield and somatic cell scores seemed to have a pure polygenic inheritance mode, whereas chromosome 14 made a much larger contribution to the total additive genetic variance for milk and fat yields. Chromosomal direct genomic values were almost uncorrelated between any pair of chromosomes. The chromosomes seemed to differ evidently in the correlation of direct genomic values between the lactations. Combining properties of the single-step SNP BLUP model with the random regression test-day model allowed us to explore the SNP effect estimates for a more accurate genomic selection at the level of chromosomes or markers.

ACKNOWLEDGMENTS

This study received no external funding. Esa Mäntysaari and Ismo Strandén from the Natural Resources Institute Finland (Luke, Helsinki, Finland) and Jeremie Vandenplas and Jan ten Napel from University of Wageningen (Wageningen, the Netherlands) are greatly thanked for their help in using the software packages MiX99 and MiXBLUP, respectively. The authors have not stated any conflicts of interest.

REFERENCES


Figure 7. Correlations of chromosomal direct genomic breeding values between any pair of first 3 lactations (Lact) for the trait milk yield.


ORCIDS

H. Alkhoder https://orcid.org/0000-0001-5956-9355
Z. Liu https://orcid.org/0000-0002-7112-0320
R. Reents https://orcid.org/0000-0003-1481-7436