# Single-step prediction of genomic breeding value in a small dairy cattle population with strong import of foreign genes

J. Přibyl[1], J. Haman[1], T. Kott[1], J. Přibylová[1], M. Šimečková[1], L. Vostrý[1], L. Zavadilová[1], V. Čermák[2], Z. Růžička[2], J. Šplíchal[2], M. Verner[2], J. Motyčka[3], L. Vondrášek[3]

[1]Institute of Animal Science, Prague-Uhříněves, Czech Republic
[2]Czech Moravian Breeding Corporation, Prague, Czech Republic
[3]Holstein Cattle Breeders Association of the Czech Republic, Prague, Czech Republic

**ABSTRACT**: The breeding value (EBV) of Holstein cattle milk performance from the first lactation was evaluated using a regular Animal Model or by Single-Step Prediction of the genomic breeding value (GEBV). A total of 838 bulls were genotyped using the Illumina BovineSNP50 BeadChip V2. Two overlapping sets of milk performances were evaluated: calving years 1991–2004, with 729 341 lactations and 1 394 487 animals in the pedigree and calving years 1996–2009, with 808 436 lactations and 1 487 608 animals in the pedigree. The older data set included 526 genotyped bulls, in which the daughters' milk performance was known for 210 individuals. All of the genotyped animals were included in the newer data set. Of the young genotyped bulls from the older set, 279 had more than 50 daughters with performance records in the newer set. Genomic relationship matrices (**G**) were constructed from the allele frequencies of the current genotyped population or by assuming a constant value of 0.5 for all loci. Using current allele frequencies, the correlation of **G** with the pedigree relationship (**A**) was 0.74, while it was 0.77 when the constant value was used. **G** was blended with **A** with weights of 80 or 99%. The average EBV of the genotyped bulls exceeded the mean EBV of the entire population by 3 SD. Although the number of reference bulls was small, genotyping resulted in an increase of approximately 0.05 in the correlation of the GEBV of young bulls with their results after progeny testing. Only small differences in correlations were found in dependency on the methods used for the determination of **G** and in dependency on the weight used in blending **G** with **A**. Both EBV and GEBV in the older set showed higher correlations with the GEBV of the newer set than the EBV of the newer set.

**Keywords**: genomic breeding value; single-step prediction; first lactation; genomic relationship; genetic trend

The main goal of the inclusion of molecular genetic information (MG) in the evaluation of breeding animals is to evaluate young animals without performance records.

The inclusion of MG was first carried out based on the relationship between several genetic markers with QTLs (Přibyl, 1995). This information was combined with the polygenic breeding value (EBV) determined by a BLUP Animal Model. A large number of genetic markers can be determined due to the development of laboratory techniques that influence the methods of EBV prediction. A large number of partial regressions are estimated for a given trait based on many SNP genetic markers. These partial regressions are summed into one total criterion used for the animal's selection for breeding. This criterion predicts the direct genetic value (DGV) and in combination with the

Supported by the Ministry of Agriculture of the Czech Republic (Project No. QI111A167).

residual polygenic EBV, the genomic breeding value (GEBV) of the evaluated animal (Meuwissen et al., 2001). Some SNPs have a negligible effect on the studied traits, and errors in genotyping exist, too. Therefore, methods to eliminate improper SNPs are employed (Wiggans et al., 2009; Verbyla et al., 2010). In spite of the high density of genome coverage by a large number of genetic markers, it is assumed that genetic markers do not fully reflect additive genetic variability. Calculation of the GEBV also implies overestimation. Both matters could be solved by blending **G** with **A** (explained below), assigning an appropriate weight of **G** (or inv(**G**)) from 70 to 100% (VanRaden et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010).

We are working with populations undergoing permanent development, while genotypic responses to permanently changing husbandry conditions vary, too. Therefore, it is necessary to continuously estimate regression coefficients for the SNP markers obtained for cattle in maximally two-year periods (Schaeffer, 2006).

SNP genetic markers can also be used for the determination of the realised genomic relationship matrix **G** (Guo, 1996). In the construction of **G** matrix, deviations from the allele frequencies of the original non-selected population are employed. These frequencies are very difficult to determine and several approximate methods must be applied for that purpose (Forni et al., 2011). Data on genomic relationships can be combined with data on pedigree relationships (Bömcke et al., 2009; Legarra et al., 2009; Christensen and Lund, 2010).

VanRaden (2008) demonstrated that the same GEBV evaluation of animals can be obtained using the above-mentioned regression coefficients or with **G** matrix. Calculation of the GEBV by means of **G** matrix is much simpler, but inversion of **G** is not possible in every case, for example, with the presence of identical twins, etc. Thus, it is frequently modified by blending **G** with a small proportion of the pedigree relationship **A**, which practically does not influence the results (VanRaden, 2008; Aguilar et al., 2010; Christensen and Lund, 2010).

The methods used for GEBV prediction were based on sufficiently large reference sets of sires with highly reliable breeding values that represent the evaluated populations in which the above-mentioned regression coefficients were calculated. The value of young genotyped animals was subsequently determined based on assignment of these regression coefficients to the same SNP in young animals. The value of young animals was also determined by means of **G** matrix on the basis of the relationship between old and young genotyped animals (Hayes et al., 2009; VanRaden et al., 2009). These were consecutive multistep calculations, and each step required fulfillment of a number of conditions and error-free input parameters, if possible. In addition, comparison of genotyped and non-genotyped animals on a common scale is not easy, because genotyped animals are only a biased sampling of the entire population with a different mean and variability. Additionally, calculations of the EBV of genotyped animals and their non-genotyped herd mates and other relatives in the basic population are mutually influenced (Ducrocq, 2011).

The difficulties described above can be overcome by employing a single-step prediction (SSP) of GEBV (Misztal et al., 2009; Christensen and Lund, 2010) that works directly with entire (unselected) basic sets of performance records. The advantages of the SSP are more pronounced for traits with lower heritabilities (Chen et al., 2011) and can be used in multiple trait (MT) analysis for the prediction of the GEBV of several traits (Aguilar et al., 2011). An SSP approach is based on typical methods using linear mixed models in large populations at a national level (Zavadilová et al., 2005, 2009; Krejčová et al., 2008; Vostrý et al., 2008; Komprej et al., 2009), and it immediately follows all processed and programmed national-scale procedures for the genetic evaluation of farm animals (Plemdat, 2011).

The development of the methodical principles of GEBV prediction has been summarised by several authors (Hayes et al., 2009; VanRaden et al., 2009; Přibyl et al., 2010; Poschadel and Mayer, 2011). New applications of SSP have been reported by Misztal et al. (2010), Aguilar et al. (2010, 2011), Chen et al. (2011), Forni et al. (2011), and Tsuruta et al. (2011).

The Czech population of approximately 200 000 Holstein cows has a large proportion of inseminations of foreign origin, mostly with either direct parentage from the USA or descended through a few generations from U.S. cattle. This permanent immigration of foreign genes influences the relationships within the domestic cattle population and the determination of EBV.

The objective of the present study was to assess the SSP of the GEBV in a small population of Holstein cattle with a small number of genotyped reference sires.

## MATERIAL AND METHODS

Data from the first lactations of Holstein cattle were used in this study. For simplicity, the following lactation model was employed:

performance = HYS of calving + calving age + (calving age)$^2$ + days open + (days open)$^2$ + animal + random error

where:

HYS = contemporary group within a herd in a three-month calving period (fixed effect);

calving age, days open = curvilinear regressions (fixed effects);

animal = random effect within a relationship matrix

Based on the model equation, a set of normal equations is constructed as follows:

$$\begin{bmatrix} \mathbf{X`X} & \mathbf{X`Z} \\ \mathbf{Z`X} & \mathbf{X`Z} + k\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X`}\boldsymbol{Y} \\ \mathbf{Z`}\boldsymbol{Y} \end{bmatrix}$$

where:

$\boldsymbol{Y}$ = vector of the performance records (dependent variable)

$\mathbf{X}, \mathbf{Z}$ = matrices of explanatory variables assigning performance records to fixed and random effects, respectively

$\boldsymbol{b}, \boldsymbol{u}$ = estimated unknown vectors of fixed and random effects

$k$ = ratio of the variances of residual and random genetic effects

$\mathbf{A}$ = pedigree additive relationship matrix among the animals, covering ancestral generations

Two overlapping sets of performances were evaluated; each set covers 14 calving years. According to the quantity of production record, both the sets have similar predictive power:

(I) calving years 1991–2004, with 729 341 lactations and 1 394 487 animals in the pedigree, comprising 4 generations of ancestors;

(II) calving years 1996–2009, with 808 436 lactations and 1 487 608 animals in the pedigree, comprising 4 generations of ancestors.

Set (II) forwards the first one by 5 years; cows calved from 1996 till 2004 are included in both sets. Sets therefore differ genetically according to a realised genetic trend of 5 years. The average EBV of a group of genotyped bulls is compared with entire sets. Therefore, the difference between the EBV averages of the same genotyped bulls included in both sets from the average EBV of both sets including all animals should be smaller in set II.

The 842 bulls in the two pedigrees were genotyped using the Illumina BovineSNP50 BeadChip V2 (Illumina Inc., San Diego, USA). There are many outlooks for restricting loci and animals according to the quality of genotyping (Wiggans et al., 2009; Verbyla et al., 2010). We aimed at using the highest number of genotyped bulls. Bulls with genotype call rates < 95% were excluded from further evaluation (4 bulls). SNPs were discarded if more than 5% of SNP calls were missing. The total number of SNPs used for further analyses was 53 906. SNPs with a GC-score < 0.2 were considered missing. Therefore, a total of 838 bulls were used.

The older set (I) included 526 genotyped bulls composed of 316 young bulls and 210 bulls for which the daughters' milk performance was known (84 daughters per bull on average). The newer set (II) included all genotyped bulls. Of the young genotyped bulls from the older set, 279 had more than 50 daughters (77 per bull on average), and 80 had more than 80 daughters (110 per bull on average) with performance records in the newer set. Furthermore, new young genotyped bulls appeared in set II.

Regular EBVs without genomic information were determined using an Animal Model method, and the GEBV was determined by SSP (Aguilar et al., 2010; Christensen and Lund, 2010). For calculation of the GEBV by SSP, the inversion of the pedigree relationship matrix $\mathbf{A}^{-1}$ was substituted in the above set of equations by $\mathbf{H}^{-1}$ matrix:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{F} \end{bmatrix}$$

where:

$\mathbf{H}$ = corrected relationship matrix (Aguilar et al., 2010)

[..] = the same size as $\mathbf{A}$, with partitioning into 4 parts

$\boldsymbol{F}$ = correction for the genomic relationships by means of $\mathbf{G}$ matrix only for genotyped animals

$$\boldsymbol{F} = ((1 - w)\mathbf{G} + w\,\mathbf{A}_{22})^{-1} - \mathbf{A}_{22}{}^{-1}$$

where:

w = weight (0 to 1) of the pedigree relationship (Christensen and Lund, 2010)

$\mathbf{A}_{22}$ = additive pedigree relationship matrix only for genotyped animals

Matrix $\mathbf{A}_{22}$ is cut out of matrix $\mathbf{A}^*$. Matrix $\mathbf{A}^*$ is constructed on the basis of genotyped animals and four generations of their ancestors.

The genomic relationship matrix **G** is constructed according to Forni et al. (2011):

$$\mathbf{G} = \frac{(\mathbf{T} - \mathbf{Q})(\mathbf{T} - \mathbf{Q})`}{trace\,((\mathbf{T}-\mathbf{Q})(\mathbf{T} - \mathbf{Q})`)/n}$$

where:

**T** = matrix of the occurrence of an allele SNP of size *n, m* for *n* genotyped animals (*n* = 526 or 838) and *m* loci of genetic markers (*m* = 53 906), and the matrix elements are the numbers of second alleles in a given animal at locus (*j*), having the values < 0, 1, 2 >

**Q** = matrix of the means of SNP occurrences across particular loci, with column (*j*) containing the values $2.q_j$ ($q_j$ = average frequency of the second allele for SNP in column (*j*) in the original non-selected population)

Usually, the allele frequency in the original non-selected populations is difficult to estimate, so different alternatives are used for the construction of **Q**. The average allele frequencies of the second allele in the current genotyped population (**G**of) or an assumed constant value of 0.5 for all loci (**G**05) were used here (Forni et al., 2011).

As in previous reports (VanRaden et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010), the genomic relationships were assigned weights of 99 or 80% (w = 0.01 or 0.20, respectively). A heritability value of 0.25 was assumed. For simplification of the calculation, heterogeneous variance was not considered.

EBV and GEBV were determined in both sets (I and II), and correlations among the results were calculated with the aim of comparing the EBV to the GEBV. This corresponds to the 279 young bulls in set (I) with the daughter information in set (II).

BLUPF90 family programs (Misztal et al., 2002) and the DMU5 module of the DMU program (Madsen and Jensen, 2008) were used for the EBV and GEBV predictions.

## RESULTS

The pedigree relationships of the genotyped animals, $\mathbf{A}_{22}$, were compared with the realised genetic relationship, **G** (Table 1). The pedigree matrix of the relationship of the genotyped sires $\mathbf{A}_{22}$ has diagonal elements in the range of 1 to 1.16, with a standard deviation of 0.020, which indicates a maximum inbreeding of 16%. Average inbreeding is 2.3%. The off-diagonal elements have values from 0 to 0.632, which shows a maximum animal relationship of 63.2%. The average pedigree relationship among all genotyped animals is 5.2%, and the standard deviation of the relationship is 3.9%.

Table 1. Relationships and their differences in 838 genotyped animals according to pedigree ($\mathbf{A}_{22}$) and genomic (**G**) relationship matrices

| | Mean | SD | Minimum | Maximum | Correlation of relationship matrices |
|---|---|---|---|---|---|
| **Diagonal elements** (inbreeding) | | | | | |
| $\mathbf{A}_{22}$ | 1.023 | 0.020 | 1 | 1.160 | |
| **G**of | 1.000 | 0.040 | 0.839 | 1.238 | |
| **G**05 | 1.000 | 0.014 | 0.941 | 1.066 | |
| Difference $\mathbf{A}_{22}$ – **G**of | 0.023 | 0.039 | –0.177 | 0.184 | 0.306 |
| Difference $\mathbf{A}_{22}$ – **G**05 | 0.023 | 0.018 | –0.027 | 0.109 | 0.521 |
| Difference **G**of – **G**05 | 0.000 | 0.036 | –0.102 | 0.202 | 0.473 |
| **Off-diagonal elements** (relationship) | | | | | |
| $\mathbf{A}_{22}$ | 0.052 | 0.039 | 0.000 | 0.632 | |
| **G**of | –0.001 | 0.043 | –0.132 | 0.580 | |
| **G**05 | 0.580 | 0.022 | 0.495 | 0.840 | |
| Difference $\mathbf{A}_{22}$ – **G**of | 0.053 | 0.030 | –0.275 | 0.402 | 0.737 |
| Difference $\mathbf{A}_{22}$ – **G**05 | –0.528 | 0.026 | –0.685 | –0.207 | 0.774 |
| Difference **G**of – **G**05 | –0.582 | 0.028 | –0.671 | –0.231 | 0.842 |

**G**of = determined on the basis of the frequencies of $q_j$ in the studied current population, **G**05 = determined at a frequency of $q_j$ = 0.5, $q_j$ = average frequency of the second allele for SNP in column (*j*) in the original non-selected population

The mean of the diagonal elements of the standardised genomic relationship matrix (**G**of) is 1, with a range of 0.839 to 1.238 and a standard deviation of 0.040. The range is markedly higher than in the pedigree relationship, and values lower than 1 also occurred. The off-diagonal elements range from –0.132 to 0.580, with a mean close to 0 and a standard deviation of 0.043. The variability of the genomic relationship is higher than that of the pedigree relationship, and negative values also occurred (the animals are to a certain degree opposite in relationship). The mean difference between $\mathbf{A}_{22}$ and **G**of is 2.3% in the diagonal elements, with a standard deviation of 3.9% and a range from –0.177 to 0.184. In the off-diagonal elements, the mean difference is 5.3%, with a standard deviation of 3.0% and a range from –0.275 to 0.402, which indicates a change in the relationship of up to 40.2%. According to $\mathbf{A}_{22}$ and **G**of, the correlation between the diagonal elements (inbreeding) is 0.306, while the correlation between the off-diagonal elements (relationship) is 0.737. Both correlation coefficients are highly statistically significant, but the above values indicate large differences between the pedigree relationship and the realised relationship in some animals.

The values of **G**05 are more similar to the values of $\mathbf{A}_{22}$, and the correlations of the diagonal and off-diagonal elements are 0.521 and 0.774, respectively. The variability in relationship is also lower than in **G**of. The crucial difference lies in the mean value of the off-diagonal elements, which is biased from the expected value (difference –0.528), in agreement with the findings of Forni et al. (2011).

The relationship matrices **G**of and **G**05 exhibit correlations of diagonal and off-diagonal elements of 0.473 and 0.842, respectively. There are also substantial differences between some elements of these matrices, and the standard deviations of the differences are similar to the standard deviations of the differences in $\mathbf{A}_{22}$. In general, the elements in **G**05 matrix are more balanced than in **G**of matrix.

The evaluated sets I and II cover a total period of 19 years. Over this period, there was a marked increase in milk performance and in variability (Table 2). Sets I and II overlap, although there is a difference in the means of the older and the newer set attaining to 1239 kg of milk per lactation, and the standard deviation of the performance records increased from 1768 to 1832 kg. After adjustment by the least-squares method (GLM/SAS) for fixed effects, the residual standard deviations were 979 and 1167 kg for sets I and II, respectively. The determination coefficients of the model are 72% for set I and 62% for set II. In the newer set, fixed effects explained the lower proportion of variability.

Table 2. Mean values and variability (kg of milk per lactation)

| | Set I | | | Set II | | | Difference |
|---|---|---|---|---|---|---|---|
| | *n* | mean | SD | *n* | mean | SD | II – I |
| Milk performance | 729 341 | 5847 | 1768 | 808 436 | 7086 | 1832 | 1 239 |
| Adjustment for fixed effects | | | 979 | | | 1167 | |
| EBV all | 1 394 487 | 0 | 553 | 1 487 608 | 0 | 745 | 0 |
| EBV 279 | 279 | 1681 | 442 | 279 | 1377 | 603 | –304 |
| **G**of GEBV 279 | 279 | 1745 | 530 | 279 | 1377 | 606 | –368 |
| **G**of GEBV 279 0.8 | 279 | 1751 | 524 | 279 | 1378 | 609 | –373 |
| **G**05 GEBV 279 | 279 | 1593 | 364 | 279 | 1354 | 524 | –239 |
| **G**05 GEBV 279 0.8 | 279 | 1621 | 381 | 279 | 1364 | 549 | –257 |
| EBV 80 | 80 | 1709 | 489 | 80 | 1394 | 626 | –315 |
| **G**of GEBV 80 | 80 | 1777 | 607 | 80 | 1397 | 628 | –380 |
| **G**of GEBV 80 0.8 | 80 | 1778 | 600 | 80 | 1398 | 631 | –380 |
| **G**05 GEBV 80 | 80 | 1628 | 422 | 80 | 1378 | 565 | –250 |
| **G**05 GEBV 80 0.8 | 80 | 1653 | 436 | 80 | 1385 | 586 | –268 |
| EBV Proven bulls | 1130 | 959 | 616 | 1130 | 622 | 628 | –337 |

279 = bulls with > 50 daughters (average 77) in set II, 80 = bulls with > 80 daughters (average 110) in set II, 0.8 = weight of genomic relationship of 80% (w = 0.20), proven bulls = with > 80 daughters in both sets (averages of 164 and 168 daughters)

The standard deviation of the EBV of all the animals is 553 kg in set I and 745 kg in set II.

The 1130 proven bulls with a sufficient quantity of daughters have average EBVs of 959 and 622 kg in sets I and II, respectively (Table 2). The difference of 337 kg (last column in Table 2) corresponds to a genetic gain of 66.14 kg per year. The standard deviations within this group are similar in both sets, with values of 616 and 628 kg, respectively.

For the 279 young genotyped bulls in set I without daughter performance data, the mean (pedigree) EBV was 1681 kg, approximately 3 SD of EBV above the average of the entire set, and 722 kg above the average of proven bulls. These 279 bulls have the SD of 442 kg, which is less than in the entire set. After progeny testing, the same bulls have 50 daughters or more in set II, and their mean EBV is 1377 kg. Their mean decreased by 304 kg over 5 years in relation to the entire set, which is associated with the genetic trend in the population, but could also be related to the bias in the estimation of parent average EBV in dataset I. This decrease is lower than in proven bulls. The average of 279 bulls is 755 kg above the average of proven bulls in set II, which is more than in set I. This indicates some undervaluation of pedigree EBV in set I. The subset of 279 bulls has the SD of 603 kg in set II. This value is still lower than that of the entire set II. In both cases, the SD of the EBV of the 279 bulls is approximately 80% of the SD of the particular evaluated population in a given period.

GEBVs were calculated using an Animal Model with the relationship matrix **H**. The results depend on the method of **G** matrix construction employed. Variability is higher within **G**of than within **G**05, which was reflected both in the mean GEBV of the entire group of bulls and in their variability. When **G**of is used, the means and standard deviations of EBV and GEBV are basically identical in set II. In set I, where these bulls have only pedigree information, the mean GEBV is higher by 64–70 kg than the mean EBV. Comparison with proven bulls indicated some overvaluation (the higher values of 368 and 373 kg in the last column in Table 2), but this should be judged carefully. Production and connected variability increase with time, and the progeny of young bulls comes in a new period with higher variability, which is also manifested in genetic variability. When using **G**05, the means and standard deviations of GEBV are lower in both sets compared to EBV.

Weight of the genomic relationship in $\mathbf{H}^{-1}$ matrix of 99% or 80% did not exert any influence on GEBV prediction.

In the evaluation of bulls having more than 80 daughters in set II in the progeny test, the values of the means and standard deviations of GEBV are higher in both sets. However, the relationships among the results are similar to those observed in the evaluation of the 279 bulls.

The correlations between EBV and GEBV for the 279 genotyped sires within the sets are high (Table 3). In set II, after progeny testing, all values are close to 1, while in set I, the correlations between the pedigree EBV and GEBV are in the range of 0.92–0.97. The correlations between the different approaches to **G** construction are also close to 1 in set I.

Table 4 shows the correlations between the values before and after progeny testing for the 279 bulls. A relatively low correlation of approximately 0.5 was calculated between the EBVs for sets I and II. The GEBVs in set I basically show the same correlations with EBV in set II, being 0.54 for the genomic relationship **G**of and 0.52 for the genomic relationship **G**05, regardless of the impact of genomic weights in $\mathbf{H}^{-1}$. The increases in the correlation compared to EBV for **G**of and **G**05 are approximately 0.05 and 0.02, respectively. Higher correlations were determined between GEBV in set I and between GEBV in set II, up to a value of 0.622. Both EBV and GEBV in set I show higher correlations with GEBV in set II than with EBV in set II. This may imply that the GEBV in set II is closer to the actual

Table 3. Correlations between EBV and GEBV in set I (above diagonal) and in set II (below diagonal) for 279 genotyped bulls according to method of evaluation

| | EBV | **G**of | **G**of 0.8 | **G**05 | **G**05 0.8 |
|---|---|---|---|---|---|
| EBV | | 0.921 | 0.935 | 0.926 | 0.965 |
| **G**of | 0.995 | | 0.998 | 0.971 | 0.969 |
| **G**of 0.8 | 0.997 | 1 | | 0.968 | 0.972 |
| **G**05 | 0.983 | 0.994 | 0.992 | | 0.992 |
| **G**05 0.8 | 0.994 | 0.998 | 0.998 | 0.997 | |

Table 4. Correlations between EBV and GEBV for 279 genotyped bulls before (set I) and after (set II) progeny testing according to method of evaluation

| | I EBV | I **G**of | I **G**of 0.8 | I **G**05 | I **G**05 0.8 |
|---|---|---|---|---|---|
| II EBV | 0.495 | 0.544 | 0.543 | 0.521 | 0.519 |
| II **G**of | 0.506 | 0.572 | 0.570 | 0.549 | 0.542 |
| II **G**of 0.8 | 0.504 | 0.567 | 0.565 | 0.544 | 0.538 |
| II **G**05 | 0.548 | 0.622 | 0.619 | 0.600 | 0.591 |
| II **G**05 0.8 | 0.531 | 0.594 | 0.592 | 0.572 | 0.567 |

Table 5. Correlations between EBV and GEBV for 80 genotyped bulls with higher reliabilities before (set I) and after (set II) progeny testing according to method of evaluation

| | I EBV | I **G**of | I **G**of 0.8 | I **G**05 | I **G**05 0.8 |
|---|---|---|---|---|---|
| II EBV | 0.537 | 0.587 | 0.587 | 0.569 | 0.566 |
| II **G**of | 0.550 | 0.610 | 0.609 | 0.593 | 0.586 |
| II **G**of 0.8 | 0.548 | 0.606 | 0.606 | 0.589 | 0.583 |
| II **G**05 | 0.591 | 0.651 | 0.651 | 0.635 | 0.628 |
| II **G**05 0.8 | 0.573 | 0.628 | 0.628 | 0.611 | 0.607 |

genetic value (it exhibits lower noise) than the other predictions. Thus, it seems more appropriate to compare the other approaches only with the GEBV in set II. It is crucial to compare the values shown in the particular row in Table 4 with those in the first column when deviations reach the value of 0.074.

Table 5 shows the correlations between EBV and GEBV for the 80 bulls with higher numbers of daughters. Compared to Table 4, the values in Table 5 are higher by approximately 0.04. The relationships among the results shown in Table 5 are similar to those in Table 4.

The correlation results for **G**of are higher by approximately 0.02 than those of **G**05. The off-diagonal elements in the relationship matrix **G**05 are biased (Table 1). After the mean of the off-diagonal elements in **G**05 was shifted to zero, both the average GEBV and the standard deviations (Table 2) and correlation coefficients (Tables 4, 5) were similar to the results in the relationship matrix **G**of. But the convergence of the iterative calculation was very poor, which agrees with the work of Misztal et al. (2010), stating in this case non-positive definite matrix. The convergence in our case partly improved by blending **G** with $\mathbf{A}_{22}$, putting a weight of **G** at 80%.

## DISCUSSION

Four generations of ancestors were used for the construction of the pedigree relationship, which is a common practice in the sale of breeding animals and in registration in herd books. For the genomic relationship construction, MG data on the frequencies of alleles at particular loci were either used according to the set of currently evaluated bulls (**G**of), or the same frequency of both alleles was substituted (**G**05). The latter approach is based on the assumption that the frequencies of alleles in the original non-selected population were identical at all loci (VanRaden, 2008; Misztal et al., 2010; Forni et al., 2011).

The correlation between pedigree relationship **A** and **G**of is 0.737, but great changes occurred in particular animals. Values in **G**05 have closer relationships to **A** (0.774). The correlation between the two approaches used for the construction of **G** matrices is 0.842 (Table 1), but there are large differences in particular animals.

The means and variabilities of EBV and GEBV are almost identical in set II (Table 2) because offspring represent a significant portion of the information sources used to determine both EBV and GEBV. However, EBV in set I is based only on pedigree data. The mean GEBV of the bulls in set I exceeds the mean and variability of all animals to a greater extent compared to the mean EBV. In order to reduce inflation, Aguilar et al. (2010) and Misztal et al. (2010) applied a weight from 0.7 to 1 to $w^*(\text{inv}(\mathbf{G}) - \text{inv}(\mathbf{A}_{22}))$ or to $\text{inv}(\mathbf{A}_{22})$ only. In our case, a difference in weight of the genomic relationship of 80% or 99% did not influence the means and variabilities of the GEBVs of genotyped animals.

Both weightings of **G** in set I brought about a similar shift of the average GEBV compared to EBV and higher variability. This would imply overvaluation of a genetic trend. **G**05 in set I showed lower means and considerably lower variability. However, the off-diagonal elements of the relationship in this matrix were greatly biased. After the mean was shifted to zero, the results were similar to those obtained using **G**of matrix. For the correct construction of **G**, the SSP approach results in higher reliability than the other methods (Aguilar et al., 2010).

The correlation between the EBVs of the 279 genotyped bulls in sets I and II is low. Other authors have reported higher values (Hayes et al., 2009; VanRaden et al., 2009). This may be caused by the high number of imported foreign sires, and thus by the indirect relationship with the evaluated domestic population and the small reference population of only 210 proven genotyped bulls in set I. The only evaluated trait was first lactation, which is a weaker information source than is used in routine predictions of EBV. The use of bulls with higher EBV reliabilities in set II (though only 80 bulls were employed in this case) resulted in an increase in all correlations (Table 5). The GEBV in set I display by 0.05 higher correlation with the EBV in set II than the EBV in set I. In such a small reference population, this increase is at the lower threshold of the data presented in the literature (Hayes et al., 2009; VanRaden et al., 2009; Aguilar et al., 2010).

The results of GEBV could also be influenced by the use of only minor restrictions for eliminating improper SNPs. In other studies, other criteria are also used at different thresholds, including Hardy-Weinberg equilibrium, minor allele frequency, position on sex chromosomes or unknown position and GenTrain score. These reduce the number of SNPs used in further computations by thousands (Wiggans et al., 2009; Verbyla et al., 2010).

The GEBV in set I exhibits more favourable correlations with the GEBV in set II, with an increase of approximately 0.03, compared to correlations between the two EBVs. The EBV in set I also displays a tendency of higher correlations with GEBV in set II than with the EBV in set II.

## CONCLUSION

Despite using a small number of proven reference bulls, their genotyping and the genotyping of young bulls led to an increase in the correlation of the GEBVs of young bulls with their results after progeny testing.

The pedigree EBV of young bulls exerts slight undervaluation and their GEBV slight overvaluation in our case.

It is advisable to compare methods on the basis of correlations with GEBV rather than with EBV after progeny testing.

In accordance with what is reported in the literature, attention should be paid to the methods of **G** construction used.

## Acknowledgement

We are grateful to Dr. Per Madsen, University of Aarhus, Denmark, and to Prof. Ignacy Misztal, University of Georgia, USA, for methodical support and for provision of computer programs. We gratefully acknowledge the helpful comments of anonymous reviewers.

## REFERENCES

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., Lawlor T.J. (2010): Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science, 93, 743–752.

Aguilar I., Misztal I., Tsuruta S., Wiggans G.R., Lawlor T.J. (2011): Multiple trait genomic evaluation of conception rate in Holsteins. Journal of Dairy Science, 94, 2621–2624.

Bömcke E., Soyeurt H., Szydlowski M., Gengler N. (2009): How to combine pedigree and marker information into a single estimator for the calculation of relationships? In: Proc. 60th Annu. Mtg. European Association for Animal Production (EAAP). Barcelona, Spain.

Chen C.Y, Misztal I., Aguilar I., Tsuruta S., Meuwissen T.H.E., Aggrey S.E., Wing T., Muir W.M. (2011): Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. Journal of Animal Science, 89, 23–28.

Christensen O.F., Lund M.S. (2010): Genomic prediction when some animals are not genotyped. Genetics Selection Evolution, 42, 2.

Ducrocq, V. (2011): Where are we heading to with genomics in dairy cattle evaluations? Interbull Technical Workshop: Establishing the framework for international genomic evaluations. 27th–28th Feb. 2011, Guelph, Canada.

Forni S., Aguilar I., Misztal I. (2011): Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genetics Selection Evolution, 43, 1.

Guo S.W. (1996): Variation in genetic identity among relatives. Human Heredity, 46, 61–70.

Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. (2009): Invited review: Genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science, 92, 433–443.

Komprej A., Gorjanc G., Kompan D., Kovač M. (2009): Covariance components by a repeatability model in Slovenian dairy sheep using test-day records. Czech Journal of Animal Science, 54, 426–434.

Krejčová H., Přibyl J., Přibylová J., Štípková M., Mielenz N. (2008): Genetic evaluation of daily gains of dual-purpose bulls using a random regression model. Czech Journal of Animal Science, 53, 227–237.

Legarra A., Aguilar I., Misztal I. (2009): A relationship matrix including full pedigree and genomic information. Journal of Dairy Science, 92, 4656–4663.

Madsen P., Jensen J. (2008): DMU – a package for analysing multivariate mixed models. Version 6, release 4.7. 33 pp. Available from http://dmu.agrsci.dk

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

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H. (2002): BLUPF90 and related programs (BGF90). 7th WCGALP, 19th–23rd Aug. 2002, Montpellier, France.

Misztal I., Legarra A., Aguilar I. (2009): Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. Journal of Dairy Science, 92, 4648–4655.

Misztal I., Aguilar I., Legarra A., Tsuruta S., Johnson D.L., Lawlor T.J. (2010): A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation. 9th WCGALP, 1st–6th Aug. 2010, Leipzig, Germany.

Plemdat (2011): Available from www.plemdat.cz.

Poschadel N., Mayer M. (2011): Realisirte Verwandtschaftsmatrizen und Zuchtwerschätzung für die genomische Selektion – eine Übersicht. Züchtungskunde, 83, 167–181.

Přibyl J. (1995): A way of using markers for farm animals selection. Živočišná výroba, 40, 375–382. (in Czech)

Přibyl J., Řehout V., Čítek J., Přibylová J. (2010): Genetic evaluation of dairy cattle using a simple heritable genetic ground. Journal of the Science of Food and Agriculture, 90, 1765–1773.

Schaeffer L.R. (2006): Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics, 123, 218–223.

Tsuruta S., Misztal I., Aguilar I., Lawlor T.J. (2011): Multiple-trait genetic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. Journal of Dairy Science, 94, 4198–4204.

VanRaden P.M. (2008): Efficient methods to compute genomic predictions. Journal of Dairy Science, 91, 4414–4423.

VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F., Schenkel F.S. (2009): Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science, 92, 16–24.

Verbyla K.L., Calus M.P., Mulder H.A., de Haas Y., Veerkamp R.F. (2010): Predicting energy balance for dairy cows using high-density single nucleotide polymorphism information. Journal of Dairy Science, 3, 2757–2764.

Vostrý L., Přibyl J., Jakubec V., Veselá Z., Majzlík I. (2008): Selection of a suitable definition of environment for the estimation of genotype × environment interaction in the weaning weight of beef cattle. Czech Journal of Animal Science, 53, 407–417.

Wiggans G.R., Sonstegard T.S., VanRaden P.M., Matukumalli L.K., Schnabel R.D., Taylor J.F., Schenkel F.S., Van Tassell C.P. (2009): Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science, 2, 3431–3436.

Zavadilová L., Němcová E., Štípková M., Bouška J. (2009): Relationships between longevity and conformation traits in Czech Fleckvieh cows. Czech Journal of Animal Science, 54, 385–394.

Zavadilová L., Jamrozik J., Schaeffer L.R. (2005): Genetic parameters for test day model with random regressions for production traits of Czech Holstein cattle. Czech Journal of Animal Science, 50, 142–154.

Received: 2011–08–01
Accepted after corrections: 2011–12–02

*Corresponding Author*

Prof. Josef Přibyl, Institute of Animal Science, K Netlukám 962, 104 00 Prague 10-Uhříněves, Czech Republic
Tel. +420 267 009 649, e-mail: pribyl.josef@vuzv.cz