Acta fytotechn zootechn, 19, 2016(3): 74–78
http://www.acta.fapz.uniag.sk

Original Paper

GWAS in practical cattle breeding in the Czech Republic – single step method, genetic progress

Josef Přibyl[1]*, Jiří Bauer[1], Václav Čermák[2], Daniela Fulínová[2], Jiří Motyčka[3], Petr Pešek[1], Jana Přibylová[1], Jiří Šplíchal[2], Hana Vostrá Vydrová[1], Luboš Vostrý[1], Ludmila Zavadilová[1]
[1]*Institute of Animal Science, Praha-Uhříněves, Czech Republic*
[2]*Czech Moravian Breeders' Corporation, Inc., Hradištko, Czech Republic*
[3]*Holstein Cattle Breeders Association of the Czech Rep., Hradištko, Czech Republic*

Article Details: Received: 2016-04-19 | Accepted: 2016-05-30 | Available online: 2016-08-31

dx.doi.org/10.15414/afz.2016.19.03.74–78

Development of genetic evaluation of animals is permanent process. It was going from estimated breeding value (EBV) calculated by CC-test, across a BLUP – animal model and RR-TDM, to the genomic enhanced breeding value (GEBV) using genetic markers. Methods of genetic evaluation become a part of marketing strategies of insemination companies. Therefore all countries and association of breeders seek to be compatible with others. Now we are in a period of massive global implementation of genomic evaluation, which combines traditional BLUP with huge quantity of genetic SNP markers. Multi-step procedures are now usual in practice, which work with deregressed proofs. Development of methods attained to the single-step procedure (ssGBLUP) which overcomes some difficulties of previous methods, improves reliabilities of evaluation and compares all animals, genotyped and ungenotyped, in entire nation-wide population. Genomic evaluation influence above all young genotyped animals. In Czech Republic single-step procedure is routinely used for national evaluation of milk, linear type traits, reproduction and longevity. GEBVs are accompanied by genomic reliabilities. Genetic trends over last 20 years are in some traits different for genomic evaluation compared to traditional BLUP evaluation, although input data and genetic parameters (heritability) are the same and genotyped animals were only small proportion from entire evaluated population. Differences in genetic trends increase mainly in new batches of animals. Reason of it could be in the changed variability of breeding values and "genomic correction" of relationship between animals, which is expanded from genotyped animals to others individuals in a population.

Keywords: genomic breeding value, single-step, genomic relationship, genetic trend, SNP

1 Introduction

Development of genetic evaluation of farm animals

Evaluation of animals is closely connected to organization of tests and collecting data of production recording and statistical procedures of data evaluation. In evaluation statistical association of phenotype (Y) with genotype is utilized. Genetic value of animal is predicted by estimated breeding value (EBV), usually with help of linear models. Essence of evaluation is the conditional distribution of EBV, if we know Y (p(EBV | Y)). Known phenotype (Y) is directly from evaluated animals or from relatives. Link of Y with related evaluated animal is according similarity of pedigree or similarity of genome.

In Table 1 are some broadly defined epochs connected with breeding and genetic evaluation. All mentioned points are connected with development and application of statistical procedures (including Mendel). CC – test starts global calculation of EBV in a large populations. In 2010 starts global application of genome enhanced breeding value (GEBV). Applications of methodologies are strongly supported by capacity of computers and development of programming strategies.

Accuracy of animal evaluation depends on quantity of information. On Figure 1 are simulated reliabilities of sire evaluations according three sources of information. Heritability 0.25 and genomic information adequate to 15 progenies was used in a simulation. For young animals there are different starting points of reliabilities, if only progeny, or progeny with pedigree, or progeny with pedigree and genomic information are used. In a case of high number of progeny, reliabilities are for all three possibilities similar. Therefore genomic information is valuable mainly for selection of young animals without their own phenotype with goal of using them in breeding more early.

EBV of young animals is derived from related reference animals with production records. In a case of genomic selection the reference population usually consists

***Corresponding Author:** Josef Přibyl, Institute of Animal Science, Přátelství 815, Praha-Uhříněves 104 00, Czech Republic, e-mail: pribyl.josef@vuzv.cz

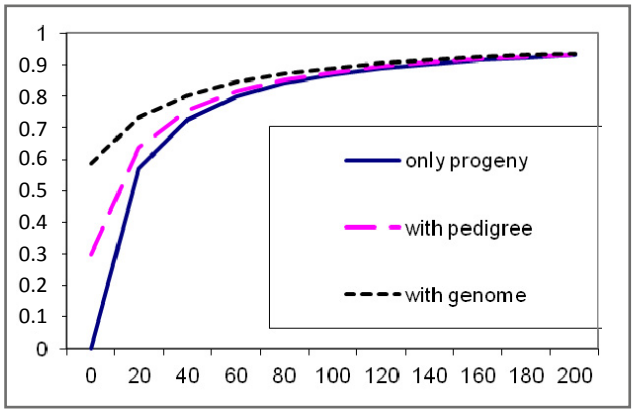

Figure 1 Reliability of sire evaluation according No. of progeny

of progeny tested bulls with high reliability of EBV. Reliability of prediction of young animals depends on the size of reference population and density (volume) of genetic markers for each animal (VanRaden et al., 2011).

Table 1 Application of genetic evaluation procedure (broad epochs)

Time period	Event
18/19 century	Sheep progeny testing, Moravia
1866	G. J. Mendel
1900	Production recording
1920	R. A. Fisher, S. Wright
1950	CC (EBV)
1980	BLUP
1990	AM
2000	RR-TDM (Slovakia 1997, first in the world*)
2005	Survival Kit
2010	GWAS (GEBV)
2015	ssGBLUP
2020	?

* Candrák et al. (1997)

The aim of the paper was to analyze results of GEBV evaluation of Holstein population on national data.

2 Material and methods

2.1 GEBV prediction

Traditional EBV performed by BLUP is step by step globally substituted by GEBV. Genetic chips with large quantity of genetic SNP markers are used. In cattle breeding 50K chips from Illumina are popular (several versions), which has for each animal around 54,000 markers. These markers are implanted into statistical procedures of BLUP or Bayesian methodology.

Two basic techniques are used – multi-step procedures (Meuwissen et al., 2001) and single-step procedures (Misztal et al., 2009).

2.2 Multi-step GEBV

In traditional multi-step procedure regression coefficients are estimated for each marker according to linear model:

$$\mathbf{DYD} = \mathbf{Xb} + \mathbf{T_{1v}} + \mathbf{e} \qquad (1)$$

where:

DYD – vector of pseudo-phenotype data of daughter yield deviations, or deregressed proofs (*DRP*), for m_1 animals

X – design matrix for fixed effect

b – estimated unknown vector, usually only one constant

$\mathbf{T_1}$ – matrix ($m_1 \times n$) of values for n SNP loci and m_1 animals in a referenced population, with values <0,1,2> according numbers of second allele (gene content)

v – estimated unknown vector of "genetic" regression coefficients for *n* loci

e – random error, with values (weights) according reliabilities of **DYD**

Regression coefficient *s v* are used for prediction of values of young animals:

$$\mathbf{DGV} = \mathbf{T_{2v}} \qquad (2)$$

where:

DGV – vector of direct genetic values of m_2 young animals

$\mathbf{T_2}$ – matrix ($m_2 \times n$) of values for *n* SNP loci and m_2 young animals

Even though number of genetic markers is large, they do not explain total genetic variance. Therefore residual polygenic effect (parents average) is added. Genomic breeding value is then:

$$GEBV = k1 \cdot \mathbf{DGV} + k_2 \cdot \mathbf{u^*} \qquad (3)$$

where:

k_1, k_2 – weights in selection index

u* – vector of residual polygenic effect

2.3 GBLUP method

Usual breeding value u has covariance $\mathbf{var(u)} = \mathbf{A} \times \sigma^2_u$. Similar covariance should produce $\mathbf{DGV} = \mathbf{Tv}$.

$$\mathbf{var\ (Tv)} = \mathbf{Tv} \cdot \mathbf{v'T'} = \mathbf{T} \cdot \mathbf{T'} \cdot \sigma^2_v = \mathbf{G} \cdot \sigma^2_u \qquad (4)$$

where:

A – pedigree additive relationship matrix
σ^2_u – genetic variance of evaluated trait
σ^2_v – genetic variance of SNP locus manifested in evaluated trait
G – realized genomic relationship matrix. Practical calculation of **G** follows form

$$G = \frac{[T - \mathrm{Q}]\mathrm{D}[T - \mathrm{Q}]'}{2\sum_i q_i (1 - q_i)} \quad (5)$$

where:
Q – matrix with columns of average frequencies of second allele q_i for *n* loci (*i*) in a basic population of founders
D – (possible) diagonal matrix with weights for loci

Genomic relationship matrix **G** expresses real relationship according genetic information. Pedigree relationship matrix **A** is only according evidence of breeders and can differ from **G** due to Mendelian sampling. Matrix **G** is centralized and scaled, so that **A** and **G** should be in the same scale. There are several modifications for **G** calculation and scaling.

G can be substituted into (1). It covers both proven and young genotyped animals. Method is known as GBLUP (VanRaden, 2008):

$$\mathbf{DYD} = \mathbf{Xb} + \mathbf{Z} \cdot \mathbf{DGV} + \mathbf{e} \quad (6)$$

where:
Z – design matrix for random effect **DGV** connecting animals with theirs **DYD**. DGVs are connected with relationship matrix **G**

Calculations according (1 and 2) and (6) produce the same results.

2.4 Single-step GEBV (ssGBLUP)

Development of GBLUP has focused on inclusion of **G** into BLUP calculation on a whole data in a national scale and evaluate jointly all genotyped and ungenotyped animals. Evaluation is according model equation:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{Z} \cdot \mathbf{u} + \mathbf{e} \quad (7)$$

where:
Y – vector of production records
X – design matrix for all fixed effect
b – estimated unknown vector of fixed effects
Z – design matrix for random effect connecting production records with animals
u estimated unknown vector of random effect

In usual BLUP-animal model random effect **u** (EBV) is connected with pedigree relationship matrix **A**. If genetic markers are used, then for genotyped animals are constructed genomic relationship matrix **G**. Both matrices are combined in ssGBLUP into matrix **H**, which substitute **A** in (7). Because both polygenic and genetic markers effects are used simultaneously, calculation produces directly GEBVs (**u** = GEBV). When solving systems of equations following from (7), inverses of **A** in BLUP and inverse of **H** in ssGBLUP are used. Though construction of **H** is not easy, the construction of $\mathbf{H^{-1}}$ is feasible (Legarra et al., 2009):

$$\mathrm{H}^{-1} = \mathrm{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \lambda\left(G^{-1} - \mathrm{A}_{22}^{-1}\right) \end{bmatrix} \quad (9)$$

where:
H – pedigree-genomic relationship matrix
$\mathbf{A_{22}}$ – part of pedigree relationship connected with genotyped animals only (subtraction avoids double counting of relationship)
λ – weight of genomic in relation to polygenic effect (for fine tuning more parameters τ, ω are used)

Some technical limitation has been inversion $\mathbf{G^{-1}}$, if number of genotyped animals was large. Recently recursive algorithm was developed, which allows practically unlimited volume of animals in **G** (Fragomeni et al., 2015). Masuda et al. (2016) used in single-step evaluation more than ½ million of genotyped animals.

Advantages of ssGBLUP:
- Overcomes bias from preselection of genotyped animals.
- All nation-wide phenotypes used in evaluation.
- Direct comparison of all animals, genotyped + ungenotyped.
- Due to genetic markers corrected EBVs also of ungenotyped animals.
- Higher reliability of evaluation.

3 Results and discussion

3.1 GEBV of Holstein in the Czech Republic

In 2015 ssGBLUP was validated by Interbull like official national procedure. From the all sires, which had daughter after year 1995, ½ was genotyped and used in genomic evaluation. Totally about 7,000 animals was genotyped, from which ½ was used. Majority of animals were genotyped by 50 K Illumina chip. SNP data were edited according relation to domestic population, MAF, No. of loci per animal, No. of animals per locus, big error of prediction of old reference bulls (outliers), big discrepancy of relationship $\mathbf{A_{22}} \times \mathbf{G}$ and genes proportion of Holstein. After editing it were used about 43,000 SNP markers/animal. Correlations of elements $\mathbf{A_{22}}$ with **G** was approximately 0.72. Weight for λ in (8) was used $\lambda = 0.80$ (20% polygenic and 80% genomic effects).

Genomically was by ssGBLUP (Misztal et al., 2009; Christensen and Lund, 2010) evaluated milk production according RR-TDM, conformation according Animal model and reproduction according Animal model with paternal effect. Modification, "blending" ssGBLUP (Gao et al., 2012), was used for longevity, where input data for genomic calculation were deregressed proofs (DRP) of all proven bulls from Survival Kit with weights according reliabilities. GEBVs are accompanied by reliabilities calculated according modifications of Misztal et al. (2013). Recent results of genomic evaluation on Czech data are in Bauer et al. (2014, 2015), Pešek et al. (2015), Přibyl et al. (2014, 2015) and Zavadilová et al. (2014).

By ssGBLUP are jointly evaluated all animals in population. In Figure 2 are distributions of EBV/GEBVs of bulls for milk production. In this partial study were totally 7,603 bulls, from which 2,662 were genotyped (2,180 proven, 482 young). Distribution of all three groups is approaching normal distribution. GEBVs of old proven genotyped bulls covers practically whole interval of EBVs of all bulls. This was the intention of genotyping to have in reference population wide stretch and not only positive selected bull. The average of young genotyped bulls has positive deviation from both groups of proven bulls. Between top bulls are representatives of all three groups of bulls.

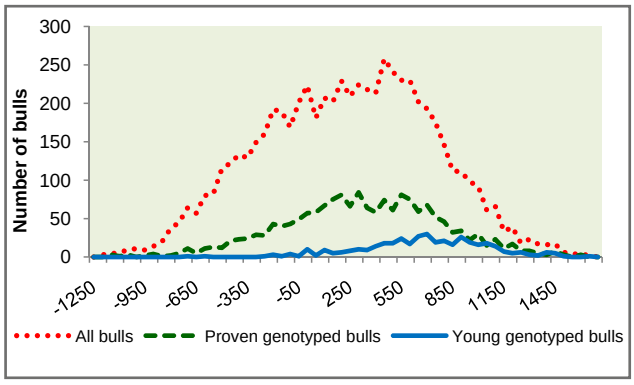

Figure 2 Distribution of EBV/GEBV of genotyped and ungenotyped bulls

Genotyping influences only little the evaluation of old proven animals. Biggest influences are for young genotyped animals. In Figure 3 are reliabilities of milk production according RR-TDM for genotyped bulls. Three trails are presented – (a) for traditional BLUP, (b) for ssGBLUP according only the first lactation and (c) ssGBLUP for multi trait join evaluation of three first lactations. For majority of proven bulls are reliabilities of all three possibilities similar, around 0.90. But for young genotyped bulls there are differences. For (a) majority of young bulls has reliability around 0.30, for (b) around 0.53 and for (c) is shifted to around 0.65. Improvement was considerable.

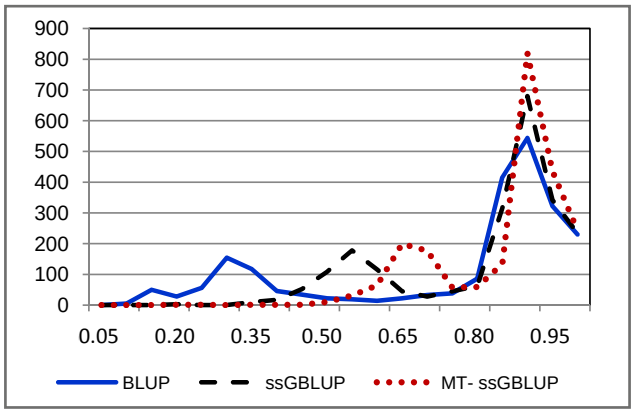

Figure 3 Numbers of genotyped bulls according reliability

3.2 Genetic trends

In an evaluation of milk production about 20 mil. of test-day records in three first lactations for 1.1 mil. of cows was included. Including pedigree over 2 mil. of animals is in the evaluation. In ssGBLUP evaluation are included genotypes for about 3 thousands of bulls. This is very low proportion of the total population, but bulls are important in relationship matrix **A**, connecting the pedigree of many animals. In spite of low number of genotyped animals, genotyping influences the entire population.

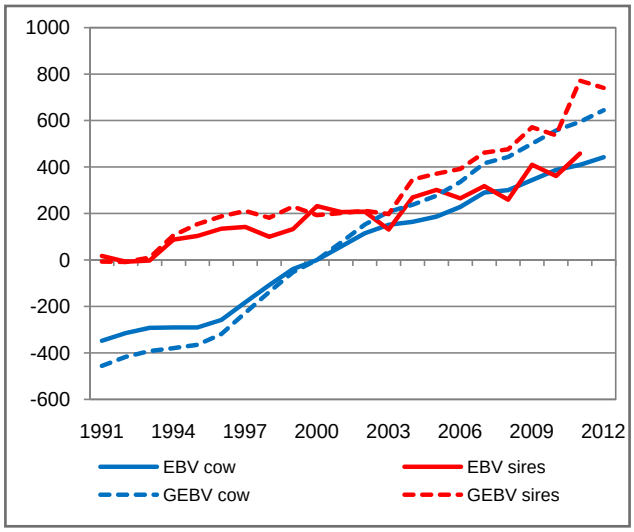

Figure 4 Genetic trends of milk for bulls and cows according EBV and GEBV

As an example in Figure 4 the genetic trend over 20 years of milk production for bulls and cows according EBV and GEBV is shown. For both, bulls and cows, GEBV evaluation resulted higher genetic trend. Therefore GEBV influenced also comparison of different batches of animals. Biggest differences are for youngest animals. Reason can be higher variance of GEBV than EBV and influence of genotyping on ungenotyped animals (Christensen and Lund, 2010).

In a Figure 5 is genetic trend of maternal fertility (pregnancy rate). These are relative breeding values. Genetic level of females according GEBV is going practically to the year 2009 invariably down. Perhaps from this year begins positive trend. In bulls negative trend is not seen from year 2000. But influence of bulls on daughters depends on which bulls were heavily used in insemination. Probably not bulls with good maternal fertility. There is big difference in trend of bulls according EBV and GEBV. Real values according GEBV are worse.

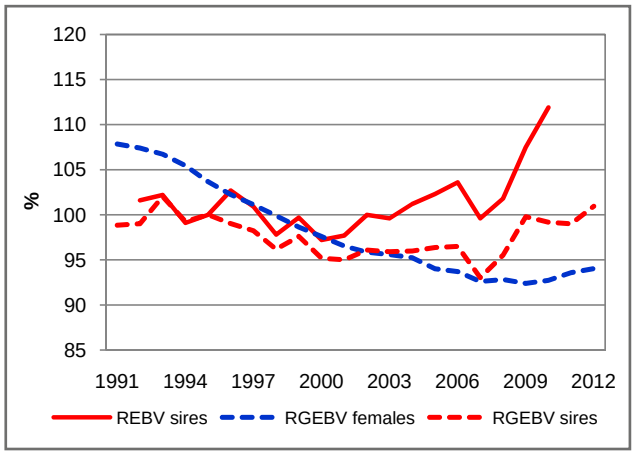

Figure 5 Genetic trends of relative maternal fertility for bulls and females according EBV and GEBV

4 Conclusions

Single-step procedure, which works with entire population, is a logical development of previous methods of genetic evaluation of animals. It produces more accurate breeding values for evaluated young genotyped animals. Genotyping influences the evaluation of genotyped and also of other animals in entire population which is manifested in changes in genetic trend in comparison with evaluation according BLUP method.

Acknowledgments

We are grateful to P. Madsen, Univ. of Aarhus, Denmark and I. Misztal, Univ. of Georgia, USA, for methodological support and for provision of computer programs. Supported by the Ministry of Agriculture of the Czech Republic, Prague (Projects No. QJ1510139 and MZE RO0716).

References

BAUER, J. et al. (2014) Approximation of the reliability of single-step genomic breeding values for dairy cattle in the Czech Republic. *Anim. Sci. Papers and Reports,* vol. 32, pp. 301–306.

BAUER, J., PŘIBYL, J. and VOSTRÝ, L. (2015) Contribution of domestic production records and Interbull EBV on approximate reliabilities of single-step genomic breeding values in dairy cattle. *Czech J. Anim. Sci.,* vol. 60, pp. 263–267.

CANDRÁK, J., KADLEČÍK O. and SCHAEFFER L.R. (1997) The use of test-day model for Slovak cattle populations. In: *Proc. 48th Annual Meeting of the European Association for Animal Production, Vienna, Austria, August,* pp. 25–28.

CHRISTENSEN, O.F. and LUND, M.S. (2010) Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.,* vol. 42, pp. 2. doi: http://dx.doi.org/ 10.1186/1297-9686-42-2

FISHER, R.A. (1918) The correlation between relatives in the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.,* vol. 52, pp. 399–433.

FRAGOMENI, B.O. et al. (2015) Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. *J. Dairy Sci.,* vol. 98, pp. 4090-4094. doi: http://dx.doi.org/10.3168/jds.2014-9125

GAO, H. et al. (2012) Comparison on genomic predictions using three GBLUP methods and two single step blending methods in the Nordic Holstein population. *Genet. Sel. Evol.,* vol. 44, pp. 8. doi: http://dx.doi.org/10.1186/1297-9686-44-8

LEGARRA A., AGUILAR I. and MISZTAL, I. (2009) A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.,* vol. 92, pp. 4656–4663.
doi: http://dx.doi.org/ 10.3168/jds.2009-2061

MASUDA, Y. et al. (2016) Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. *J. Dairy Sci.,* vol. 99, pp. 1968–1974.
doi: http://dx.doi.org/10.3168/jds.2015-10540

MENDEL, G. J. (1866) Versuche über Pflanzen-Hybriden. Verh. Naturforsch. *Ver. Brünn.,* vol. 4, pp. 3–47 (1901, *J. R. Hortic. Soc.,* 26, pp. 1–32).

MEUWISSEN, T.H.E., HAYES, B.J. and GODDARD, M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics,* vol. 157, pp. 1819–1829.

MISZTAL, I., LEGARRA A. and AGUILAR, I. (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.,* vol. 92, pp. 4648–4655. doi: http://dx.doi.org/10.3168/jds.2009-2064

MISZTAL, I. et al. (2013) Methods to approximate reliabilities in single-step genomic evaluation. *J. Dairy Sci.,* vol. 96, pp. 647–654. doi: http://dx.doi.org/10.3168/jds.2012-5656

PEŠEK, P., PŘIBYL, J. and VOSTRÝ, L. (2015) Genetic variances of SNP loci for milk yield in dairy cattle. *J. Appl. Genet.,* vol. 56, pp. 339–347. doi: http://dx.doi.org/10.1007/s13353-014-0257-2

PŘIBYL, J. et al. (2014) Domestic and Interbull information in the single step genomic evaluation of Holstein milk production. *Czech J. Anim. Sci.,* vol. 59, pp. 409–415.

PŘIBYL, J. et al. (2015) Domestic estimated breeding values and genomic enhanced breeding values of bulls in comparison with their foreign genomic enhanced breeding values. *Animal,* vol. 9, pp. 1635–1642. doi: http://dx.doi.org/10.1017/S1751731115001044

RADEN, Van, P.M. (2008) Efficient methods to compute genomic predictions. *J. Dairy Sci.,* vol. 91, pp. 4414–4423. doi: http://dx.doi.org/10.3168/jds.2007-0980

RADEN, Van, P.M. et al. (2011) Genomic evaluations with many more genotypes. *Genet. Sel. Evol.,* vol. 43, pp. 10. doi: http://dx.doi.org/10.1186/1297-9686-43-10

WRIGHT, S. (1921) Systems of mating. *Genetics,* vol. 6, pp. 111–178.

ZAVADILOVÁ, L. et al. (2014) Single-step genomic evaluation for linear type traits of Holstein cows in Czech Republic. *Anim. Sci. Papers and Reports,* vol. 32, pp. 201–208.