

**This is an electronic reprint of the original article.**

**This reprint *may differ* from the original in pagination and typographic detail.**

**Author(s):** Andrei A. Kudinov, Esa A. Mäntysaari, Timo J. Pitkänen, Ekaterina I. Saksa, Gert P. Aamand, Pekka Uimari and Ismo Strandén

**Title:** Single-step genomic evaluation of Russian dairy cattle using internal and external information

**Year:** 2021

**Version:** Published version

**Copyright:** The Author(s) 2021

**Rights:** CC BY-NC-ND 4.0

**Rights url:** <http://creativecommons.org/licenses/by-nc-nd/4.0/>

**Please cite the original version:**

Kudinov, A. A., Mäntysaari, E. A., Pitkänen, T. J., Saksa, E. I., Aamand, G. P., Uimari, P., & Strandén, I. (2021). Single-step genomic evaluation of Russian dairy cattle using internal and external information. *Journal of Animal Breeding and Genetics*, 00, 1– 12.

<https://doi.org/10.1111/jbg.12660>

All material supplied via *Jukuri* is protected by copyright and other intellectual property rights. Duplication or sale, in electronic or print form, of any part of the repository collections is prohibited. Making electronic or print copies of the material is permitted only for your own personal use or for educational purposes. For other purposes, this article may be used in accordance with the publisher's terms. There may be differences between this version and the publisher's version. You are advised to cite the publisher's version.

# Single-step genomic evaluation of Russian dairy cattle using internal and external information

Andrei A. Kudinov<sup>1,2,3</sup>  | Esa A. Mäntysaari<sup>1</sup>  | Timo J. Pitkänen<sup>1</sup>  |  
Ekaterina I. Saksa<sup>3</sup> | Gert P. Aamand<sup>4</sup>  | Pekka Uimari<sup>2</sup>  | Ismo Strandén<sup>1</sup> 

<sup>1</sup>Natural Resources Institute Finland (Luke), Jokioinen, Finland

<sup>2</sup>Department of Agricultural Science, University of Helsinki (UH), Helsinki, Finland

<sup>3</sup>Russian Research Institute for Farm Animal Genetics and Breeding – Branch of the L.K. Ernst Federal Science Center for Animal Husbandry (RRIFAGB), St. Petersburg, Russian Federation

<sup>4</sup>Nordic Cattle Genetic Evaluation (NAV), Aarhus, Denmark

## Correspondence

Andrei A. Kudinov, Natural Resources Institute Finland (Luke), Jokioinen, Finland.  
Email: andrei.kudinov@luke.fi

## Abstract

Genomic data are widely used in predicting the breeding values of dairy cattle. The accuracy of genomic prediction depends on the size of the reference population and how related the candidate animals are to it. For populations with limited numbers of progeny-tested bulls, the reference populations must include cows and data from external populations. The aim of this study was to implement state-of-the-art single-step genomic evaluations for milk and fat yield in Holstein and Russian Black & White cattle in the Leningrad region (LR, Russia), using only a limited number of genotyped animals. We complemented internal information with external pseudo-phenotypic and genotypic data of bulls from the neighbouring Danish, Finnish and Swedish Holstein (DFS) population. Three data scenarios were used to perform single-step GBLUP predictions in the LR dairy cattle population. The first scenario was based on the original LR reference population, which constituted 1,080 genotyped cows and 427 genotyped bulls. In the second scenario, the genotypes of 414 bulls related to the LR from the DFS population were added to the reference population. In the third scenario, LR data were further augmented with pseudo-phenotypic data from the DFS population. The inclusion of foreign information increased the validation reliability of the milk yield by up to 30%. Suboptimal data recording practices hindered the improvement of fat yield. We confirmed that the single-step model is suitable for populations with a low number of genotyped animals, especially when external information is integrated into the evaluations. Genomic prediction in populations with a low number of progeny-tested bulls can be based on data from genotyped cows and on the inclusion of genotypes and pseudo-phenotypes from the external population. This approach increased the validation reliability of the implemented single-step model in the milk yield, but shortcomings in the LR data recording scheme prevented improvements in fat yield.

## KEYWORDS

dairy cattle, genomic prediction, multi-country genomic evaluation, single-step GBLUP

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of Animal Breeding and Genetics* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Genomic information has been successfully used in predicting dairy cattle breeding values during the last decade (VanRaden, 2020). In the original genomic evaluation approach, breeding values of the candidate animals were predicted using information derived from the genotyped animals of the reference population (Meuwissen et al., 2001). Multiple studies have shown that the reliability of genomic prediction depends on its size and structure (Goddard, 2009; Goddard & Hayes, 2009). Large commercial dairy breeding schemes initiated their reference populations by genotyping all progeny-tested bulls and some elite cows. However, this approach is challenging in small populations because only a few progeny-tested bulls may be available, and historical DNA samples of the bulls are not available.

The obvious recipe to attain sufficient prediction accuracy in small populations with a limited number of progeny-tested bulls is to increase the reference population with genotyped cows (Ding et al., 2013; Li et al., 2014). This approach increases genotyping costs because the low reliability of estimated breeding value (EBVs), typical to cows, requires a larger number of genotyped animals to gain the same accuracy as when using progeny-tested bulls that typically have highly reliable EBVs (Daetwyler et al., 2008). In addition, the heritability of a trait affects the optimal size of the reference population; the lower the heritability, the more genotyped animals are needed in the reference population (Goddard & Hayes, 2009). Furthermore, genomic prediction accuracy depends on the model used in genomic prediction. The single-step genomic BLUP (ssGBLUP) approach (Aguilar et al., 2010; Christensen & Lund, 2010) may yield more accurate genomic predictions than the two-step approach when the population has a limited number of genotyped animals (Christensen et al., 2012; Song et al., 2019).

Low genomic prediction reliability in a population with a limited number of genotyped and progeny-tested animals can be enhanced by including data from external sources (Přibyl et al., 2013; VanRaden, 2012). Thus, a joint reference population can be created where countries benefit uni- or bilaterally from the data sharing. Several reported examples of EBVs and genomic data exchange between countries have shown significant increases in the reliability of genomic predictions (Jorjani et al., 2012; Lund et al., 2011; Ma et al., 2014). Even though several countries routinely make joint traditional and genomic evaluations (e.g. Denmark, Finland and Sweden) (Lidauer et al., 2015), most dairy evaluation systems are unwilling to share recorded data from cows and will only disseminate EBV from internal evaluations. In such circumstances, the inclusion of foreign bull EBVs with corresponding



**FIGURE 1** Map. The northeastern part of the Baltic sea. The Leningrad region of the Russian Federation is highlighted with a dark grey colour (the plot was created in R with the ggplot2 software package, Wickham, 2016)

reliabilities into national or internal reference populations has become a common practice (Přibyl et al., 2013; Vandenplas et al., 2014). EBVs of foreign genotyped bulls can be attained from the multi-trait across-country evaluations (MACE, Interbull, Uppsala, Sweden).

Several methods to include external EBV into internal evaluations have been developed (Bonaiti & Boichard, 1995; Luštrek et al., 2021; Přibyl et al., 2013; Täubert et al., 2000; Vandenplas et al., 2014; VanRaden, 2001, 2012). Vandenplas et al. (2014) described a unified approach for combining external EBV with internal data and pedigree information with further extension to genomic information (Vandenplas et al., 2017). The blended information from multiple sources was shown to be free from double counting the internal information. The method avoids the overestimation of reliabilities and can be used in genomic prediction models to include external data.

Contemporary comparison is the current official dairy bull evaluation method in Russia, but state-of-the-art animal model evaluations have already been proposed (Kudinov et al., 2017, 2018). In 2016, the Leningrad Region (LR) (Figure 1) Committee on Agriculture and Fishery (Saint Petersburg, Russia) initiated a research and development project to apply genomic evaluations of production traits using BLUP methodology. The region's largest dairy cattle population consists of animals with an admixture of Holstein (HOL) and Russian Black & White (RBW)

breeds kept in 49 herds, with an average herd size of approximately 1,000 cows (Kudinov et al., 2018). Because the number of progeny-tested bulls was low genotyping cows was the only way to increase the reference population. During the period 2015–2017, the starting pool of genotyped animals was created from 427 bulls and 1,080 cows. A small number of reference animals expectedly would limit genomic prediction reliability for candidate animals. Thus, it was proposed to improve reliability when including genomic and pseudo-phenotypic information from the neighbouring Danish, Finnish and Swedish Holstein (DFS) populations.

The aim of this study was to test the feasibility of ssGBLUP for HOL and RBW cattle in the LR with only a small number of genotyped cows and bulls. We also tested the effect of including genomic and pseudo-phenotypic information from HOL bulls from DFS on the prediction ability of the genomic model. Single-step genomic evaluations were computed using three scenarios: (a) phenotypes and genotypes of LR animals only, (b) including additional bull genotypes from the DFS and (c) further adding external MACE EBVs at the DFS scale.

## 2 | MATERIALS AND METHODS

### 2.1 | Leningrad region data

Phenotypic data, as described in Kudinov et al. (2018), included 363,833 records of 305 days of milk and fat yields from 159,069 highly related HOL and RBW cows in 49 breeding herds. Some animals had recorded up to the fifth lactation. The pedigree included 221,001 animals born between 1960 and 2015. For variance component estimation, the data were truncated to include records from the first three lactations only due to the small number of records in later lactations. Correspondingly the pedigree was pruned to only include informative animals for the truncated data. The final data included 319,509 observations and 206,356 pedigree animals.

### 2.2 | Nordic data

The full ancestral pedigree was traced for bulls present in both the LR and the joint Nordic cattle genetic evaluations (NAV; Denmark, Finland, Sweden). The information of the 486 bulls was extracted from MACE EBVs published by Interbull in December 2018 at the NAV scale. The reliabilities of MACE EBV and values of the LR model heritabilities were used to derive effective daughter contributions (EDC) using reverse reliability estimation, as

described in Taskinen et al. (2014). Using the calculated EDC and full pedigree information, the MACE EBVs were converted into deregressed daughter performances (DRP) using a matrix deregression procedure (Jairath et al., 1998; Strandén & Mäntysaari, 2010).

### 2.3 | Genotypes

The LR data included single nucleotide polymorphism (SNP) marker genotypes from 1,080 cows and 427 bulls provided by repositories of the Russian Research Institute of Farm Animal Genetics and Breeding and LLC Laboratory Genome (Saint-Petersburg, Russia). Both Illumina BovineSNP50v2 and IDBv3 arrays (Illumina, San-Diego, USA) were used for genotyping. Genotyped cows were from 13 LR herds. The average (*SD*) number of samples per herd was 82 (21). The DFS data had 414 bull genotypes from Illumina BovineSNP50 chip provided by NAV. The DFS genotypes were imputed and had passed quality control in the official NAV HOL genomic evaluations (<https://www.nordicebv.info/>). The LR and DFS genotypes were synchronized to have identical reading patterns (i.e. coding). Imputation was performed to unify genotypes and fill-in missing markers. Quality control of genotypes was performed using the following criteria: call rate >95% and minor allele frequency >5%. After processing 43,194 SNP markers remained for the genomic prediction.

### 2.4 | Validation of model fit

Two reduced data sets were created for calculating the validation reliability and bias of genomic enhanced breeding values (GEBV). For bull validation, the milk and fat records from the last four production years (2012–2015) were removed from the data, as described in Mäntysaari et al. (2010). An exception was made for genotyped cows that were not closely related to the validation bulls (i.e. not daughters, granddaughters, or sibs) and represented contemporary groups (herd – year – season) with at least five animals. The data records from these cows were kept in order to avoid exhausting the training set. The bull validation test set included 48 bulls with EDCs greater than 20 in the full data, but EDC = 0 in the reduced data. For the cow validation test, records from the last production year (2015) were excluded. There were 221 test cows which had no records in the reduced data but at least one record in the full data.

The full data were used to calculate daughter yield deviations (DYDs) for the bulls and yield deviations (YDs)

for the cows (VanRaden & Wiggans, 1991) using a corresponding ssGBLUP model. Bias was estimated by (G)EBV overdispersion, that is the regression coefficient  $b_1$  in the validation regression model (D)YD =  $b_0 + b_1$  GEBV, and by the average difference between GEBV and (D)YD. The DYD observations for bull  $i$  were weighted using  $k_i$ , calculated as  $k_i = \frac{EDC_i}{EDC_i + \lambda_1}$ , where  $\lambda_1 = (4 - h^2)/h^2$ . The YD observations for cow  $j$  were weighted using parameter  $k_j$ , calculated as  $k_j = \frac{ERC_j}{ERC_j + \lambda_2}$ , where  $\lambda_2 = (1 - h^2)/h^2$ , and ERC<sub>*j*</sub> is the effective record contribution (Přibyl et al., 2013) of cow  $j$ . Validation reliability ( $R^2$ ) was calculated as the squared correlation between (D)YD and the reduced data GEBV divided by mean  $k_j$ . The within-herd heritability was calculated using the formula  $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2)$ , where  $\sigma_a^2$ ,  $\sigma_{pe}^2$  and  $\sigma_e^2$  are genetic, permanent environment and residual variances respectively.

## 2.5 | Statistical model

The repeatability animal model presented in Kudinov et al. (2018) was modified by including a herd-by-sire interaction random effect (Wiggans et al., 1988) and a respective variance component ( $\sigma_{hs}^2$ ). Variance components and breeding values were estimated using the model:

$$\mathbf{A}_{\text{UPG}}^{-1} = \begin{pmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & -(\mathbf{A}^{11}\mathbf{Q}_1 + \mathbf{A}^{12}\mathbf{Q}_2) \\ \mathbf{A}^{21} & \mathbf{A}^{22} & -(\mathbf{A}^{21}\mathbf{Q}_1 + \mathbf{A}^{22}\mathbf{Q}_2) \\ -(\mathbf{Q}'_1\mathbf{A}^{11} + \mathbf{Q}'_2\mathbf{A}^{21}) & -(\mathbf{Q}'_1\mathbf{A}^{12} + \mathbf{Q}'_2\mathbf{A}^{22}) & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{pmatrix},$$

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{p} + \mathbf{Z}_3\mathbf{h} + \mathbf{e},$$

where  $\mathbf{y}$  is a vector of milk or fat yield records,  $\mathbf{b}$  is a vector of the fixed effects,  $\mathbf{a} \sim \mathbf{N}(0, \mathbf{A}\sigma_a^2)$  and  $\mathbf{p} \sim \mathbf{N}(0, \mathbf{I}\sigma_{pe}^2)$  and  $\mathbf{h} \sim \mathbf{N}(0, \mathbf{I}\sigma_{hs}^2)$  are vectors of random animal breeding values, permanent environmental and herd-by-sire interaction effects,  $\mathbf{X}$  is a design matrix relating fixed effects to the records,  $\mathbf{Z}_1$ ,  $\mathbf{Z}_2$  and  $\mathbf{Z}_3$  are design matrices relating the random effects to the records, and  $\mathbf{e} \sim \mathbf{N}(0, \mathbf{I}\sigma_e^2)$  is a vector of the residual effects. Matrix  $\mathbf{A}$  is the pedigree-based relationship matrix and  $\mathbf{I}$  are identity matrices.

Days open by age of calving by lactation and herd-year-season were the fixed effects (Kudinov et al., 2018). No breed effect was used in the model, as active crossbreeding of RBW cows with HOL bulls began in the late 1970s, before the data sampling point. The 218 originally proposed unknown parent groups (UPGs; Kudinov et al., 2018) were revised and downscaled to 54, due to a low number of observations per group. Rearranged groups were based on

origin (domestic or foreign), selection path and 5-year time intervals.

## 2.6 | Genomic evaluation

The mixed-model equations (MME) of the original ssGBLUP (Aguilar et al., 2010; Christensen, 2010) included a joint relationship matrix  $\mathbf{H}$  and its inverse:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{pmatrix},$$

where  $\mathbf{G}$  is the genomic relationship matrix,  $\mathbf{A}$  is the pedigree-based relationship matrix and  $\mathbf{A}_{22}$  is a subblock of the  $\mathbf{A}$  matrix including genotyped animals only. The UPGs were accounted in the augmented inverse relationship matrix ( $\mathbf{A}_{\text{UPG}}^{-1}$ ), as presented in (Matilainen et al., 2018; Misztal et al., 2013):

$$\mathbf{H}^{-1} = \mathbf{A}_{\text{UPG}}^{-1} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mathbf{B} & -\mathbf{B}\mathbf{Q}_2 \\ 0 & -\mathbf{Q}'_2\mathbf{B} & \mathbf{Q}'_2\mathbf{B}\mathbf{Q}_2 \end{pmatrix},$$

where  $\mathbf{B} = \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}$ ,

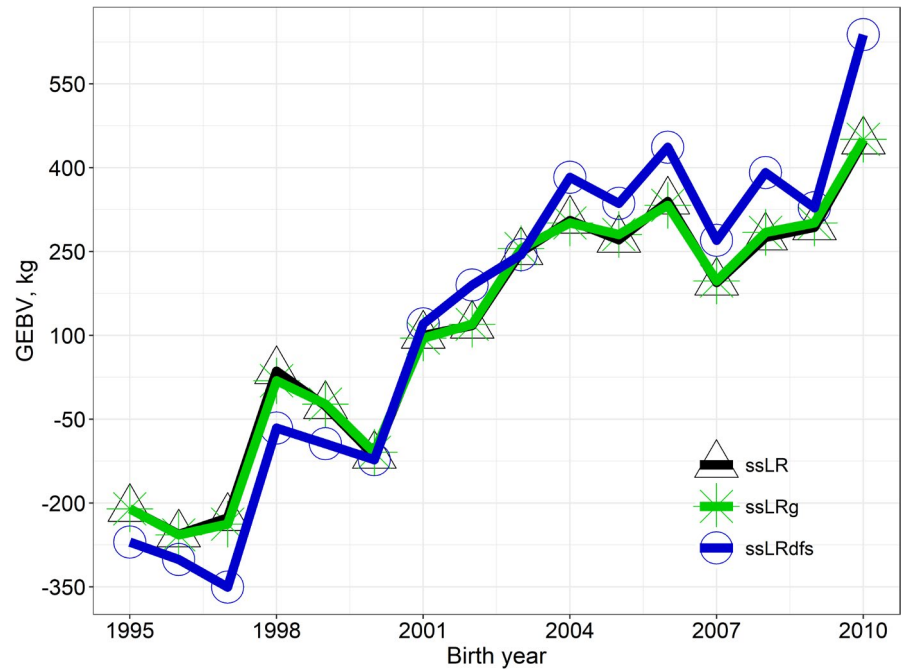
$\mathbf{Q}$  includes the proportions of contributions each animal receives from the UPG,  $\mathbf{Q}_1$  and  $\mathbf{Q}_2$  are the submatrices of  $\mathbf{Q}$  corresponding to the non-genotyped and genotyped animals, respectively, and  $\mathbf{A}^{ij}$  is the submatrix of  $\mathbf{A}^{-1}$ , with a superscript ( $i$  or  $j$ ) value of 1 for the non-genotyped and a value of 2 for the genotyped animals. Inbreeding coefficients were used in the calculations of the inverse pedigree-based relationship matrices  $\mathbf{A}^{-1}$  and  $\mathbf{A}_{22}^{-1}$ .

We assumed that genotypes could describe 90% of the genetic variance, and thus the genomic relationships were regressed towards the pedigree relationships as:

$$\mathbf{G} = s_t(1 - w)\mathbf{G}_{05} + w\mathbf{A}_{22},$$

where  $w$  represents the residual polygenic proportion 0.1, and  $\mathbf{G}_{05} = 2 \left( \frac{\mathbf{M}_{101}\mathbf{M}'_{101}}{m} \right)$ , with  $\mathbf{M}_{101}$  as an  $n$  by  $m$  marker matrix with the genotypes coded by  $\{-1, 0, 1\}$ ,  $m$  is the number of SNP markers,  $n$  is the number of genotyped animals, that is assuming allele frequencies = 0.5. The scaling factor  $s_t = \left( \frac{\text{trace}(\mathbf{A}_{22})}{\text{trace}(\mathbf{G}_{05})} \right)$  was used to assure that the average of the

**FIGURE 2** Average genomic breeding value (GEBV) of bulls by birth year for milk yield (kg). Black line with triangles (ssLR) denotes the ssGBLUP model using Leningrad region (LR) phenotypes and genotypes; green line with snowflakes (ssLRg) denotes ssGBLUP using LR phenotypes, genotypes and Nordic (DFS) genotypes; blue line with circles (ssLRdfs) denotes ssGBLUP using LR phenotypes and genotypes, and DFS genotypes and deregressed EBVs (DRPs)



diagonals of the genomic relationship matrix is equal to the average of the diagonal of the  $\mathbf{A}_{22}$  matrix.

## 2.7 | Integration of external information

For integration of MACE DFS information to the LR evaluation, we used the method presented in Vandenplas et al. (2014), Vandenplas et al. (2017). For notational simplicity, we present the MME with only the animal breeding value and the fixed effects:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{H}^{-1}\sigma_a^{-2} + \mathbf{D}_N^{-1}\sigma_e^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{D}_N^{-1}\sigma_e^{-2}\hat{\mathbf{a}}_N \end{bmatrix},$$

where subscript N pertains to the DFS MACE evaluation, the diagonal matrix  $\mathbf{D}_N^{-1}$  has the effective record contribution (ERC) increase for the bulls due to the DFS information and zero for the cows,  $\mathbf{R} = \mathbf{I}\sigma_e^2$  is the residual (co)variance matrix and  $\hat{\mathbf{a}}_N$  is solution vector of breeding values, and  $\hat{\mathbf{Y}}_N$  has the DRP from the DFS MACE evaluation. Both the  $\mathbf{D}_N^{-1}$  and  $\hat{\mathbf{a}}_N$  were padded by zeros for animals in pedigree. The ERC were back solved from reliabilities in DFS evaluations with the reversed reliability estimation as in Pitkänen et al. (2018). Separate herd-year-season fixed effect class was used for the bull pseudo-observations to reflect the different base of DFS compared to the LR evaluations. The LR is not part of MACE evaluation, thus external data were free from internal data and no actions to avoid double counting of information was needed.

## 2.8 | Evaluation scenarios

The LR ssGBLUP evaluations were implemented and tested using three scenarios. In the first scenario, named ssLR, only LR phenotypic and genomic data were used. In the second scenario, named ssLRg, ssLR was upgraded with DFS bull genotypes. In the third scenario (ssLRdfs), the ssLR was upgraded with the DFS genotypes and the DRPs from the MACE evaluations. Thus, ssLRg included more genomic information than ssLR, while ssLRdfs included more phenotypic information than ssLRg.

## 2.9 | Software

Pedigree pruning, calculation of inbreeding coefficients and relationship submatrix  $\mathbf{A}_{22}$  were performed using the RelaX2 program (Strandén & Vuori, 2006). Variance components were estimated with restricted maximum likelihood (REML) (Patterson & Thompson, 1971) in DMU software (Madsen et al., 2010) using AI-REML algorithm. Unification of the DNA arrays and imputation of missing alleles were performed using FImpute v. 2.2 software (Sargolzaei et al., 2014). The  $\mathbf{G}^{-1}$  and  $\mathbf{B}$  matrices were computed using the HGINV v. 0.87 program (Strandén & Mäntysaari, 2018). The EDC, ERC and DRP, and finally the (G)EBV and (D)YD computations were performed in MiX99 software (Strandén & Lidauer, 1999).

### 3 | RESULTS

The estimated genetic ( $\sigma_a^2$ ), herd-sire ( $\sigma_{hs}^2$ ), permanent environment ( $\sigma_{pe}^2$ ) and residual ( $\sigma_e^2$ ) variance components with respective SEs for milk yield were  $330,735 \pm 6,571$ ,  $80,532 \pm 3,023$ ,  $274,195 \pm 8,741$  and  $955,257 \pm 3,352$  respectively. For the fat yield, the estimated variance components were  $451 \pm 9$  ( $\sigma_a^2$ ),  $118 \pm 4$  ( $\sigma_{hs}^2$ ),  $300 \pm 11$  ( $\sigma_{pe}^2$ ) and  $1,393 \pm 5$  ( $\sigma_e^2$ ). Thus, the estimated heritability  $0.21 \pm 0.005$  was the same for the milk and fat yields.

The average GEBV of milk yield (kg) by the birth year of bulls with EDCs above 20 for the three ssGBLUP models is presented in Figure 2. Each GEBV trend was adjusted for the same base level by centring the mean GEBV of all cows born in 2010 to be zero. The genetic trends from ssLR and ssLRg had similar shapes, with an average annual genetic progress of 40 kg in 1995–2010. ssLRdfs showed a larger annual genetic trend (60 kg). Similar patterns were observed for the fat yield trends; the estimated annual genetic gains were 1.2 kg for ssLR and ssLRg and 1.9 kg for ssLRdfs (Figure 3).

Genetic trends in milk and fat yields for the cows are presented in Figures 4 and 5 respectively. The average annual predicted genetic gain in milk yield was identical using either ssLR or ssLRg (50 kg), but the annual genetic gain was 55 kg when using ssLRdfs. For fat yield, the annual genetic gain based only on LR data was 1.7 kg compared to a genetic gain of 1.9 kg obtained with an augmented data.

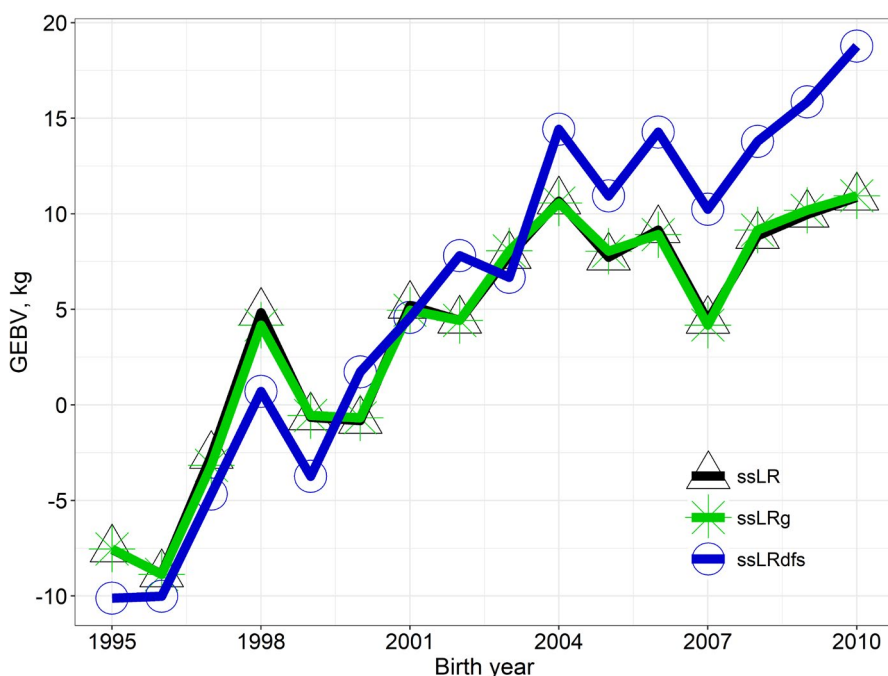
Table 1 shows the validation statistics of GEBV of milk yield based on DYD and YD for bulls and cows respectively.

The highest validation reliability  $R^2$  was observed for ssLRdfs: 0.30 for bulls and 0.42 for cows. Including DFS genotypes (ssLRg) in bull validation did not increase  $R^2$  compared to ssLR. For cows,  $R^2$  was higher in ssLR (0.38) than in ssLRg (0.36). Regression coefficients ( $b_1$ ) of DYD on GEBV for ssLR and ssLRg were similar and below one (0.78 and 0.80). ssLRdfs gave the lowest  $b_1$  (0.58). For cows, the highest  $b_1$  was from ssLR (1.69) and the lowest (1.14) from ssLRdfs.

Results of a linear regression of fat yield (D)YD on GEBV are given in Table 2. For bulls, the highest validation reliability (0.18) was obtained with both ssLRg and ssLRdfs. The difference in  $R^2$  values between ssLR and the other models was 0.01. For cows, the increased amount of foreign information decreased validation reliability; the highest  $R^2$  of 0.41 was achieved with LR data only (ssLR). The  $R^2$  value reduced by 0.07 units in ssLRg and an additional 0.13 units in ssLRdfs. For fat yield,  $b_1$  was larger than that obtained from the milk yield. For bulls,  $b_1$  values were 0.64 and 0.68 for ssLR and ssLRg, respectively, but lower for ssLRdfs (0.41). For cows,  $b_1$  values were above one for ssLR and ssLRg (1.86 and 1.67 respectively). In ssLRdfs,  $b_1$  was below one (0.89).

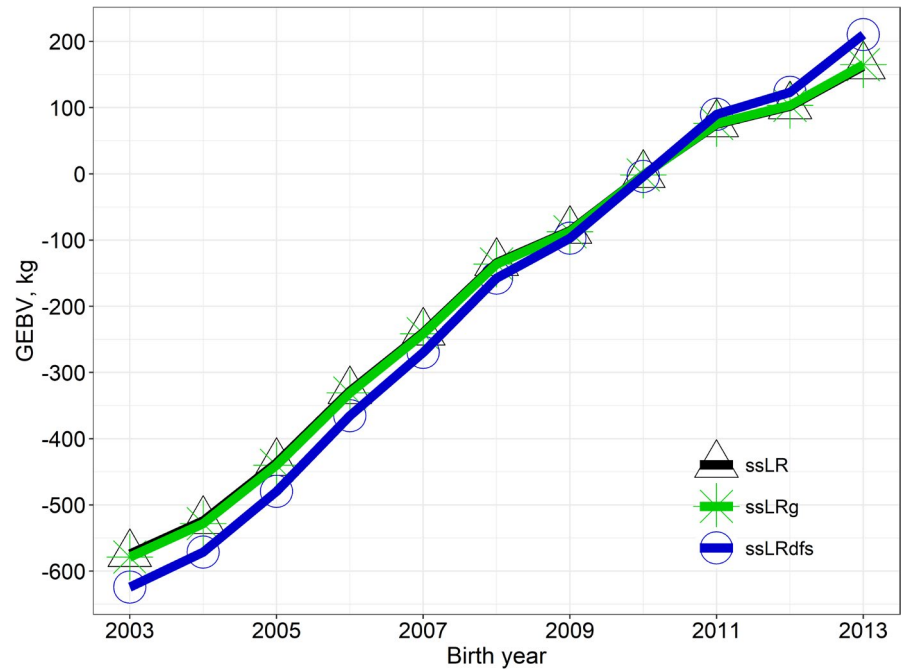
### 4 | DISCUSSION

Genomic selection has been successfully applied in the animal breeding of various species (Stock & Reents, 2013). A particularly large impact has been on the dairy cattle industry, where genomic prediction has reduced the generation interval by approximately 2.6 years and has increased

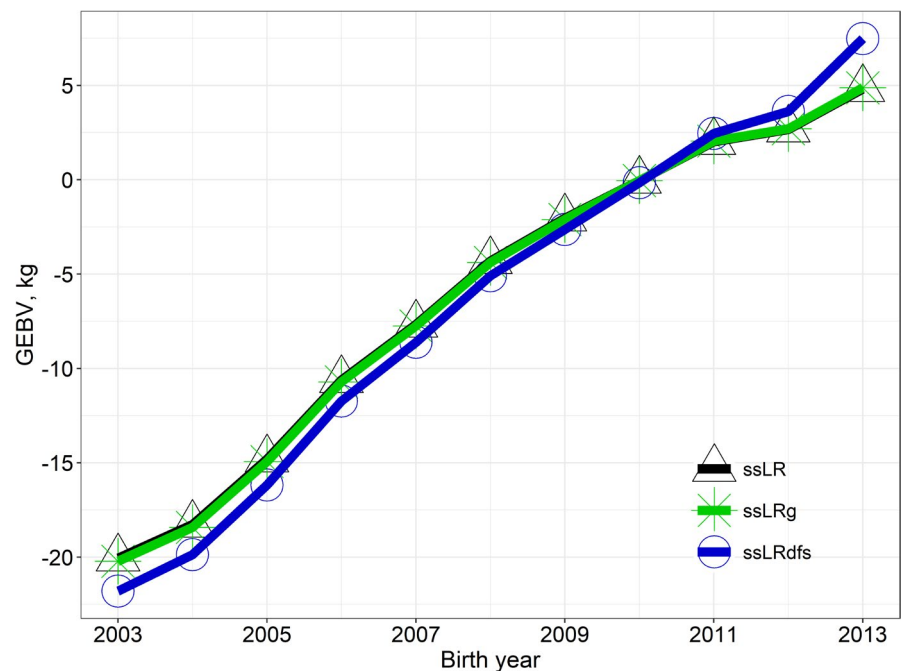


**FIGURE 3** Average genomic breeding value (GEBV) of bulls by birth year for fat yield (kg). Black line with triangles (ssLR) denotes ssGBLUP model using Leningrad region (LR) phenotypes and genotypes; green line with snowflakes (ssLRg) denotes ssGBLUP using LR phenotypes, genotypes and Nordic (DFS) genotypes; blue line with circles (ssLRdfs) denotes ssGBLUP used LR phenotypes and genotypes, and DFS genotypes and deregressed EBVs (DRPs)

**FIGURE 4** Average genomic breeding value (GEBV) of cows by birth year for milk yield (kg). Black line with triangles (ssLR) denotes ssGBLUP model using Leningrad region (LR) phenotypes and genotypes; green line with snowflakes (ssLRg) denotes ssGBLUP using LR phenotypes, genotypes and Nordic (DFS) genotypes; blue line with circles (ssLRdfs) denotes ssGBLUP using LR phenotypes and genotypes, and DFS genotypes and deregressed EBVs (DRPs)



**FIGURE 5** Average genomic breeding value (GEBV) of cows by birth year for fat yield (kg). Black line with triangles (ssLR) denotes ssGBLUP model using Leningrad region (LR) phenotypes and genotypes; green line with snowflakes (ssLRg) denotes ssGBLUP using LR phenotypes, genotypes and Nordic (DFS) genotypes; blue line with circles (ssLRdfs) denotes ssGBLUP using LR phenotypes and genotypes, and DFS genotypes and deregressed EBVs (DRPs)



the number of candidate bulls at AI stations up to 70% (Mäntysaari et al., 2020; VanRaden, 2020). Despite the attractiveness and benefits of genomic selection, it cannot be implemented in all cattle populations due to small population sizes, low numbers of progeny-tested bulls, or other limited resources.

In our study, we implemented the ssGBLUP model for the admixed population of HOL and RBW cattle from LR with a limited number of progeny-tested bulls. The local genotyped animals were mostly cows. To improve prediction accuracy, we added bull genotypes from the

neighbouring HOL DFS population and finally increased the number of progeny-tested animals in the reference population by integrating external deregressed NAV-scale MACE EBVs.

The set of reference animals in LR was expected to be too small to perform genomic prediction with prediction reliability as high as that reported for large USA HOL population (Wiggans et al., 2017). One approach to enlarging a reference population is to include genomic data from an external population. However, the success of this approach depends on the availability of phenotypic data for



Model <sup>a</sup>	Validation animals					
	Bulls (42 animals)			Cows (221 animals)		
	<i>E</i> (GEBV-DYD) <sup>b</sup>	2 * <i>b</i> <sub>1</sub>	<i>R</i> <sup>2</sup>	<i>E</i> (GEBV-YD)	<i>b</i> <sub>1</sub>	<i>R</i> <sup>2</sup>
ssLR	529	0.78	0.21	65	1.69	0.38
ssLRg	557	0.80	0.21	91	1.55	0.36
ssLRdfs	748	0.58	0.30	113	1.14	0.42

Note: Genomic enhanced breeding values (GEBV) and (daughter) yield deviations (D)YD were from the validation animals.

<sup>a</sup>ssLR = model with Leningrad region genomic and phenotypic data, ssLRg = ssLR and Nordic (DFS) genomic data, ssLRdfs = ssLRg and DFS bulls EDCs.

<sup>b</sup>*E* (GEBV-DYD) = difference between GEBV and DYD, *b*<sub>1</sub> = regression coefficient, *R*<sup>2</sup> = validation reliability

TABLE 1 Bull and cow validation results of milk yield by the three single-step GBLUP models in the Leningrad Region Russian Black & White and Holstein population

Model <sup>a</sup>	Validation animals					
	Bulls (42 animals)			Cows (217 animals)		
	<i>E</i> (GEBV-DYD) <sup>b</sup>	2 * <i>b</i> <sub>1</sub>	<i>R</i> <sup>2</sup>	<i>E</i> (GEBV-YD)	<i>b</i> <sub>1</sub>	<i>R</i> <sup>2</sup>
ssLR	18	0.64	0.17	6	1.86	0.41
ssLRg	19	0.68	0.18	7	1.67	0.34
ssLRdfs	27	0.41	0.18	7	0.89	0.21

Note: Genomic enhanced breeding values (GEBV) and (daughter) yield deviations (D)YD were from the validation animals.

<sup>a</sup>ssLR = model with Leningrad region genomic and phenotypic data, ssLRg = ssLR and Nordic (DFS) genomic data, ssLRdfs = ssLRg and DFS bulls EDCs.

<sup>b</sup>*E* (GEBV-DYD) = difference between GEBV and DYD, *b*<sub>1</sub> = regression coefficient, *R*<sup>2</sup> = validation reliability

TABLE 2 Bull and cow validation results of fat yield from the three single-step GBLUP models in the Leningrad Region Russian Black & White and Holstein population

these animals in the internal population and the genetic distance between the pooled and the target populations (Lund et al., 2014). Based on our results, a gradual upgrade of the data by adding genotypes of bulls with genetic ties to both data sets did not sufficiently increase the size of the reference population. However, we did not include EBVs of bulls with ERCs less than 20 in the DFS data. Nonetheless, including 414 DFS genotypes improved the prediction accuracy of the milk yield for bulls. Comparable to our study, correlation between milk yield DGVs and DRPs was 0.26 in a small Chinese HOL reference population with 85 genotyped bulls and 2,862 genotyped cows (Ma et al., 2014). Correlation of unweighted DYDs and GEBVs was higher in current study, the lowest value (0.38) was attained using LR data and genotypes only.

The highest *R*<sup>2</sup> in the milk yield was obtained with a model where DRP pseudo-phenotypes of DFS bulls derived from Interbull MACE evaluations were blended with LR data. Similarly, in Přebyl et al. (2013), *R*<sup>2</sup> increased after Interbull DRP pseudo-phenotypes were blended into Czech HOL genomic evaluations. However, integration of the traditional MACE evaluations should be performed

with caution, as genomic pre-selection bias has been observed in bulls born after 2009 (Patry & Ducrocq, 2011).

The favourable effect of including DFS information on *R*<sup>2</sup> in the milk yield was not observed in the fat yield. The phenotypic data recording pitfalls in fat yield, such as variability in milk analysis systems (Kudinov et al., 2017), may be a potential explanation for this. Including the external pseudo-phenotypes caused bias in the unstable prediction approach, leading to discrepancy between the expected and observed breeding values. From this viewpoint, it is important to understand whether various sources of phenotypic data are equally relevant and accurate before using them in the blending procedure. When recording errors are more prominent in the internal data than in the external data, a bias similar to what we observed in fat yield is also expected in the prediction. The other explanation for the reduced accuracy of the blended method is the validation test practice not fitting the ssLR model perfectly. As shown by Legarra and Reverter (2018), a reciprocal of the size of contemporary groups may generate upwards bias in the *R*<sup>2</sup> due to decreasing size of contemporary groups. Hence, the bias may have reduced after the integration of DFS data. We computed (results not presented) reliability

and linear regression of GEBVs estimated using full and reduced data (Legarra & Reverter, 2018). Obtained  $R^2$  increased along with increment of the data, both in milk and fat yields. However, any validation results from bulls must be considered with caution as only 42 candidates were used.

Advantages of ssGBLUP can be seen when the number of genotyped animals is too modest to allow for accurate estimates from the multi-step genomic prediction (Amaya-Martínez et al., 2020; Christensen et al., 2012). Multi-step genetic evaluations with integrated external information may create biased predictions due to an extra step used for blending the external information (Guarini et al., 2019). In such a case, using ssGBLUP provides less biased prediction due to the simultaneous use of genomic and pedigree-based information (Přibyl et al., 2013). The limitations of ssGBLUP mostly concern the compatibility of the **A** and **G** matrices, originating from an inconsistency in the base population definition in the genotyped and ungenotyped animals (Legarra et al., 2014). Proper consideration of UPG in ssGBLUP is important for compatibility between the **A** and **G** matrices. Pedigree completeness is critical to the compatibility between the **G** and **A**<sub>22</sub> matrices (Misztal et al., 2010, 2013). A promising way to solve the compatibility of the **G** and **A** matrices is to fit the ancestral structure of the population by so-called metafounders, as presented by Legarra et al. (2015). The method per se represents a fusion of Christensen's idea (Christensen et al., 2012) to construct **G** with 0.5 allele frequencies and an extension of the pedigree by related and inbred pseudofounders (Legarra et al., 2015). These relationships are accounted through a Gamma (**Γ**) matrix. The method has provided promising results when used in multiple breed pedigrees (Xiang et al., 2017) and simulated data (Garcia-Baccino et al., 2017). However, implementation of the metafounder approach may be challenging when a population has breeds with high admixture (Kudinov et al., 2020). We tested the metafounder approach for the population in our current study but did not observe improvement in validation reliability. We used nine metafounders to describe the ancestral structure of the population. All diagonal elements of the **Γ** matrix, along with the off-diagonal elements, were very close to each other, except for the highly different old RBW groups. The absence of validation reliability improvement with the metafounder approach may be due to the small number of genotyped animals, most of which were born during the last two decades. The self-relationships of metafounders associated with old birth years were mostly computed using sporadic genotypes of historic RBW and Dutch HOL bulls. The metafounders approach may become attractive when the number of genotype animals increases noticeably.

The total genetic effect in the ssGBLUP model depends on pedigree (**A**) and genomic relationships (**G**). The weight of genomic information in ssGBLUP can be changed by including a polygenic proportion variable ( $w$ ), which weighs variation due to markers and pedigree information in the genomic relationship matrix. Commonly used values for  $w$  range from 10% to 30% (Ma et al., 2014; Matilainen et al., 2018; Přibyl et al., 2013). We tested both 10% and 30% for  $w$ , and observed only a small difference in the cross-validation results. Thus, only results with the 10% proportion are presented.

A random herd by sire interaction effect applied in the model is the same as that originally used by Kudinov et al. in LR data (Kudinov et al., 2018). A random herd by sire interaction effect is also used in routine evaluations in the USA (VanRaden & Wiggans, 1991). The herds in LR are large, but their management and the origin of breeding animals may vary significantly. We observed that imported semen from the top North American bulls were used in only a fraction of the LR farms. Interactions between sire and herd may occur in such situations or when the best bulls are only used in a few top herds (Dimov et al., 1995). Several Animal model and ssGBLUP test runs performed using the model without the herd-by-sire random effect showed lower reliability and higher bias in EBV prediction. The natural future direction to improve prediction accuracy in the LR data is to use test-day records instead of 305-day records and potentially even consider balancing investment into phenotyping and genotyping (Obšteter et al., 2021).

## 5 | CONCLUSIONS

Single-step genomic prediction was successfully implemented for the LR data. The reference population included more genotyped cows than bulls because the number of progeny-tested bulls was low in the LR. Including EBVs and genotypes from the Nordic HOL population into the LR ssGBLUP evaluation created one of the largest dairy reference population among Russian regions. This joint reference population improved the prediction accuracy in the milk yield but not in the fat yield. The prediction accuracy of breeding values can be improved through better recording of phenotypes and pedigrees, and by drastically increasing the number of genotyped cows and progeny-tested bulls.

## ACKNOWLEDGEMENTS

The research is part of a research agreement between the Russian Research Institute of Farm Animals Genetics and Breeding (RRIFAGB), University of Helsinki and Natural Resources Institute Finland (LUKE) in 2015–2017.

Funding from the Committee for the Agro-Industrial and Fishery complex of Leningrad region (Russia) is acknowledged. Work on the collaborative evaluation of the Leningrad region and Nordic data was funded by Viking Genetics (Randers, Denmark). We thank Prof. Kirill V. Plemyashov (Russian Research Institute of Farm Animals Genetics and Breeding) for support and establishment of the international collaboration between RRIFAGB and LUKE. Adjunct prof. Jarmo Juga (University of Helsinki) for discussions concerning the implementation of the single-step model. Nikita A. Loskutov (LLC Plinor) for providing and pre-processing the Leningrad region raw data. RRIFAGB (Saint-Petersburg, Russia), LLC Laboratoria Genome (Saint-Petersburg, Russia), Nordic Cattle Genetic Evaluation (Aarhus, Denmark) and Viking Genetics (Randers, Denmark) for providing genotypes and data.

### CONFLICT OF INTEREST

The authors declare that they have no competing interests.

### AUTHOR CONTRIBUTIONS

AAK was responsible for data editing, all data analyses and writing. The experiment was designed by EAM and IS. IS, EAM and PU were responsible for supervision, paper writing and interpretation of results. TJP developed an algorithm to prepare Nordic data for blending. EIS was responsible for Leningrad region data and genotypes, and for contact with Leningrad region farmers and breeding organizations. GPA was responsible for Nordic data and genotypes. PU, EIS and TJP provided input on data editing and the used methods. All authors read and approved the final manuscript.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from RRIFAGB, LLC Laboratoria Genome, VikingGenetics and NAV, but restrictions apply to the availability of these data, which were used under license for the current study, and therefore are not publicly available.

### ORCID

Andrei A. Kudinov  <https://orcid.org/0000-0002-7811-576X>

Esa A. Mäntysaari  <https://orcid.org/0000-0003-0044-8473>

Timo J. Pitkänen  <https://orcid.org/0000-0002-2275-5338>

Gert P. Aamand  <https://orcid.org/0000-0003-1885-3290>

Pekka Uimari  <https://orcid.org/0000-0002-5349-3932>

Ismo Strandén  <https://orcid.org/0000-0003-0161-2618>

### 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*(2), 743–752. <https://doi.org/10.3168/jds.2009-2730>
- Amaya-Martínez, A., Martínez-Sarmiento, R., & Cerón-Muñoz, M. (2020). Genetic evaluations in cattle using the single-step genomic best linear unbiased predictor. *Ciencia Y Tecnología Agropecuaria*, *21*(1), 19–31. [https://doi.org/10.21930/rcta.vol21\\_num1\\_art:1548](https://doi.org/10.21930/rcta.vol21_num1_art:1548)
- Bonaiti, B., & Boichard, D. (1995). Accounting for foreign information in genetic evaluation. *Interbull Bulletin*, *11*, 4.
- Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, *42*, 2. <https://doi.org/10.1186/1297-9686-42-2>
- Christensen, O. F., Madsen, P., Nielsen, B., Ostensen, T., & Su, G. (2012). Single-step methods for genomic evaluation in pigs. *Animal*, *6*(10), 1565–1571. <https://doi.org/10.1017/S1751731112000742>
- Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*, *3*(10), e3395. <https://doi.org/10.1371/journal.pone.0003395>
- Dimov, G., Albuquerque, L. G., Keown, J. F., Van Vleck, L. D., & Norman, H. D. (1995). Variance of interaction effects of sire and herd for yield traits of holsteins in California, New York, and Pennsylvania with an animal model. *Journal of Dairy Science*, *78*, 939–946. [https://doi.org/10.3168/jds.S0022-0302\(95\)76709-1](https://doi.org/10.3168/jds.S0022-0302(95)76709-1)
- Ding, X., Zhang, Z., Li, X., Wang, S., Wu, X., Sun, D., Yu, Y., Liu, J., Wang, Y., Zhang, Y., Zhang, S., Zhang, Y., & Zhang, Q. (2013). Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. *Journal of Dairy Science*, *96*, 5315–5323. <https://doi.org/10.3168/jds.2012-6194>
- García-Baccino, C. A., Legarra, A., Christensen, O. F., Misztal, I., Pocrnic, I., Vitezica, Z. G., & Cantet, R. J. C. (2017). Metafounders are related to Fst fixation indices and reduce bias in single-step genomic evaluations. *Genetics Selection Evolution*, *49*, 34. <https://doi.org/10.1186/s12711-017-0309-2>
- Goddard, M. E. (2009). Genomic selection: Prediction of accuracy and maximization of long term response. *Genetica*, *136*, 245–257.
- Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, *10*, 381–391. <https://doi.org/10.1038/nrg2575>
- Guarini, A. R., Lourenco, D., Brito, L. F., Sargolzaei, M., Baes, C. F., Miglior, F., Tsuruta, S., Misztal, I., & Schenkel, F. S. (2019). Use of a single-step approach for integrating foreign information into national genomic evaluation in holstein cattle. *Journal of Dairy Science*, *102*(9), 8175–8183. <https://doi.org/10.3168/jds.2018-15819>
- Jairath, L., Dekkers, J. C., Schaeffer, L. R., Liu, Z., Burnside, E. B., & Kolstad, B. (1998). Genetic evaluation for herd life in Canada. *Journal of Dairy Science*, *81*(2), 550–562. [https://doi.org/10.3168/jds.S0022-0302\(98\)75607-3](https://doi.org/10.3168/jds.S0022-0302(98)75607-3)

- Jorjani, H., Jakobsen, J., Hjerpe, E., Palucci, V., & Dürr, J. (2012). Status of genomic evaluation in the Brown Swiss populations. *Interbull Bulletin*, *46*, 46–54.
- Kudinov, A. A., Juga, J., Uimari, P., Mäntysaari, E. A., Strandén, I., Plemyashov, K. V., Saksa, E. I., & Smaragdov, M. G. (2017). Upgrading dairy cattle evaluation system in Russian Federation. *Interbull Bulletin*, *51*, 67–74.
- Kudinov, A. A., Mäntysaari, E. A., Aamand, G. P., Uimari, P., & Strandén, I. (2020). Metafounder approach for single-step genomic evaluations of Red Dairy cattle. *Journal of Dairy Science*, *103*(7), 6299–6310. <https://doi.org/10.3168/jds.2019-17483>
- Kudinov, A. A., Saksa, E. I., Smaragdov, M. G., Juga, J., Uimari, P., Mäntysaari, E. A., & Strandén, I. (2018). Developing a genetic evaluation system for milk traits in Russian black and white dairy cattle. *Agricultural and Food Science*, *27*, 85–95. <https://doi.org/10.23986/afsci.69772>
- Legarra, A., Christensen, O. F., Aguilar, I., & Misztal, I. (2014). Single step, a general approach for genomic selection. *Livestock Science*, *166*, 54–65. <https://doi.org/10.1016/j.livsci.2014.04.029>
- Legarra, A., Christensen, O. F., Vitezica, Z. G., Aguilar, I., & Misztal, I. (2015). Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. *Genetics*, *200*(2), 455–468. <https://doi.org/10.1534/genetics.115.177014>
- Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics Selection Evolution*, *50*, 53. <https://doi.org/10.1186/s12711-018-0426-6>
- Li, X., Wang, S., Huang, J., Li, L., Zhang, Q., & Ding, X. (2014). Improving the accuracy of genomic prediction in Chinese Holstein cattle by using one-step blending. *Genetics Selection Evolution*, *46*, 66. <https://doi.org/10.1186/s12711-014-0066-4>
- Lidauer, M. H., Pösö, J., Pedersen, J., Lassen, J., Madsen, P., Mäntysaari, E. A., Nielsen, U. S., Eriksson, J. A., Johanson, K., Pitkänen, T., Strandén, I., & Aamand, G. P. (2015). Across-country test-day model evaluations for Holstein, Nordic Red Cattle, and Jersey. *Journal of Dairy Science*, *98*(2), 1296–1309. <https://doi.org/10.3168/jds.2014-8307>
- Lund, M. S., de Roos, A. P. W., de Vries, A. G., Druet, T., Ducrocq, V., Fritz, S., Guillaume, F., Guldbandsen, B., Liu, Z., Reents, R., Schrooten, C., Seefried, F., & Su, G. (2011). A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genetics Selection Evolution*, *43*(1), 43. <https://doi.org/10.1186/1297-9686-43-43>
- Lund, M. S., Su, G., Janss, L., Guldbandsen, B., & Brøndum, R. F. (2014). Genomic evaluation of cattle in a multi-breed context. *Livestock Science*, *166*, 101–110. <https://doi.org/10.1016/j.livsci.2014.05.008>
- Luštrek, B., Vandenplas, J., Gorjanc, G., Potočnik, K. (2021). Genomic evaluation of Brown Swiss dairy cattle with limited national genotype data and integrated external information. *Journal of Dairy Science*, *104*(5), 5738–5754. <https://doi.org/10.3168/jds.2020-19493>
- Ma, P., Lund, M. S., Ding, X., Zhang, Q., & Su, G. (2014). Increasing imputation and prediction accuracy for Chinese Holsteins using joint Chinese-Nordic reference population. *Journal of Animal Breeding and Genetics*, *131*(6), 462–472. <https://doi.org/10.1111/jbg.12111>
- Madsen, P., Su, G., Labouriau, R., & Christensen, O. F. (2010). DMU - A package for analyzing multivariate mixed models. I CD communication. In *Proceeding of the 9th WCGALPL Leipzig, Germany*, August 1–6, 2010;732
- Mäntysaari, E. A., Koivula, M., & Strandén, I. (2020). Symposium review: Single-step genomic evaluations in dairy cattle. *Journal of Dairy Science*, *103*(6), 5314–5326. <https://doi.org/10.3168/jds.2019-17754>
- Mäntysaari, E. A., Liu, Z., & VanRaden, P. M. (2010). Interbull validation test for genomic evaluations. *Interbull Bulletin*, *41*, 17–22.
- Matilainen, K., Strandén, I., Aamand, G. P., & Mäntysaari, E. A. (2018). Single step genomic evaluation for female fertility in Nordic Red dairy cattle. *Journal of Animal Breeding and Genetics*, *135*(5), 337–348. <https://doi.org/10.1111/jbg.12353>
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *57*(4), 1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Misztal, I., Aguilar, I., Legarra, A., & Lawlor, T. J. (2010). Choice of parameters for single-step genomic evaluation for type. *Journal of Dairy Science*, *93*, 533.
- Misztal, I., Vitezica, Z. G., Legarra, A., Aguilar, I., & Swan, A. A. (2013). Unknown-parent groups in single-step genomic evaluation. *Journal of Animal Breeding and Genetics*, *130*(4), 253–258. <https://doi.org/10.1111/jbg.12025>
- Obšteter, J., Jenko, J., Gorjanc, G. (2021). Genomic Selection for Any Dairy Breeding Program via Optimized Investment in Phenotyping and Genotyping. *Frontiers in Genetics*, *12*, <https://doi.org/10.3389/fgene.2021.637017>
- Patry, C., & Ducrocq, V. (2011). Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. *Journal of Dairy Science*, *94*, 1011–1020. <https://doi.org/10.3168/jds.2010-3804>
- Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, *58*(3), 545. <https://doi.org/10.1093/biomet/58.3.545>
- Pitkänen, T. J., Koivula, M., Strandén, I., Aamand, G. P., & Mäntysaari, E. A. (2018). Integration of external information into the national multitrait evaluation model. In Scientific Committee (Ed.). *Book of Abstracts of the 69th Annual Meeting of the European Federation of Animal Science. Dubrovnik, Croatia, 27-31 August 2018*. Wageningen Academic Publishers.
- Příbyl, J., Madsen, P., Bauer, J., Příbylová, J., Šimečková, M., Vostrý, L., & Zavadilová, L. (2013). Contribution of domestic production records, Interbull estimated breeding values, and single nucleotide polymorphism genetic markers to the single-step genomic evaluation of milk production. *Journal of Dairy Science*, *96*(3), 1865–1873. <https://doi.org/10.3168/jds.2012-6157>
- Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, *15*, 478. <https://doi.org/10.1186/1471-2164-15-478>
- Song, H., Zhang, J., Zhang, Q., & Ding, X. (2019). Using different single-step strategies to improve the efficiency of genomic prediction on body measurement traits in pig. *Frontiers in Genetics*, *9*, 730. <https://doi.org/10.3389/fgene.2018.00730>
- Stock, K. F., & Reents, R. (2013). Genomic selection: Status in different species and challenges for breeding. *Reproduction in Domestic Animals*, *48*, 2–10. <https://doi.org/10.1111/rda.12201>. PMID: 23962210

- Strandén, I., & Lidauer, M. (1999). Solving large mixed models using preconditioned conjugate gradient iteration. *Journal of Dairy Science*, *82*, 2779–2787. [https://doi.org/10.3168/jds.S0022-0302\(99\)75535-9](https://doi.org/10.3168/jds.S0022-0302(99)75535-9)
- Strandén, I., & Mäntysaari, E. A. (2010). A recipe for multiple trait deregression. *Interbull Bulletin*, *42*, 21–24.
- Strandén, I., & Mäntysaari, E. A. (2018). *HGinv Program. Natural Resources Institute Finland (LUKE)*.
- Strandén, I., & Vuori, K. (2006). Relax2: Pedigree analysis program. In: *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production: 13-18 August 2006*; Belo Horizonte, MG, Brazil, (pp. 27–30).
- Taskinen, M., Mäntysaari, E. A., Aamand, G. P., & Strandén, I. (2014). Comparison of breeding values from single-step and bivariate blending methods. In *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production: August 2014* (pp. 17–22). Vancouver.
- Täubert, H., Simianer, H., & Karras, K. (2000). Blending Interbull breeding values into national evaluations - A new approach. *Interbull Bulletin*, *25*, 53–56.
- Vandenplas, J., Colinet, F. G., & Gengler, N. (2014). Unified method to integrate and blend several, potentially related, sources of information for genetic evaluation. *Genetics Selection Evolution*, *46*, 59. <https://doi.org/10.1186/s12711-014-0059-3>
- Vandenplas, J., Spehar, M., Potocnik, K., Gengler, N., & Gorjanc, G. (2017). National single-step genomic method that integrates multi-national genomic information. *Journal of Dairy Science*, *100*, 465–478. <https://doi.org/10.3168/jds.2016-11733>
- VanRaden, P. M. (2001). Methods to combine estimated breeding values obtained from separate sources. *Journal of Dairy Science*, *84*, 47–55. [https://doi.org/10.3168/jds.S0022-0302\(01\)70196-8](https://doi.org/10.3168/jds.S0022-0302(01)70196-8)
- VanRaden, P. M. (2012). Avoiding bias from genomic pre-selection in converting daughter information across countries. *Interbull Bulletin*, *45*, 29–33.
- VanRaden, P. M. (2020). Symposium review: How to implement genomic selection. *Journal of Dairy Science*, *103*(6), 5291–5301. <https://doi.org/10.3168/jds.2019-17684>
- VanRaden, P. M., & Wiggans, G. R. (1991). Derivation, calculation, and use of National Animal Model Information. *Journal of Dairy Science*, *74*, 2737–2746. [https://doi.org/10.3168/jds.S0022-0302\(91\)78453-1](https://doi.org/10.3168/jds.S0022-0302(91)78453-1)
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>
- Wiggans, G. R., Misztal, I., & VanVleck, L. D. (1988). Implementation of an animal model for genetic evaluation of dairy cattle in the United States. *Journal of Dairy Science*, *71*, 54–69. [https://doi.org/10.1016/S0022-0302\(88\)79979-8](https://doi.org/10.1016/S0022-0302(88)79979-8)
- Wiggans, G. R., Cole, J. B., Hubbard, S. M., Sonstegard, T. S. (2017). Genomic Selection in Dairy Cattle: The USDA Experience. *Annual Review of Animal Biosciences*, *5*(1), 309–327. <https://doi.org/10.1146/annurev-animal-021815-111422>
- Xiang, T., Christensen, O. F., & Legarra, A. (2017). Genomic evaluation for crossbred performance in a single-step approach with metafounders. *Journal of Animal Science*, *95*(4), 1472–1480. <https://doi.org/10.2527/jas.2016.1155>

**How to cite this article:** Kudinov, A. A., Mäntysaari, E. A., Pitkänen, T. J., Saksa, E. I., Aamand, G. P., Uimari, P., & Strandén, I. (2021). Single-step genomic evaluation of Russian dairy cattle using internal and external information. *Journal of Animal Breeding and Genetics*, *00*, 1–12. <https://doi.org/10.1111/jbg.12660>