



Genomic evaluation of Brown Swiss dairy cattle with limited national genotype data and integrated external information

B. Luštrek,^{1*} J. Vandenplas,² G. Gorjanc,^{3,1} and K. Potočnik¹

¹Biotechnical Faculty, University of Ljubljana, 1000 Ljubljana, Slovenia

²Animal Breeding and Genomics, Wageningen University and Research, 6700 AH, Wageningen, the Netherlands

³The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian EH25 9RG, Scotland, United Kingdom

ABSTRACT

This study demonstrated the feasibility of a genomic evaluation for the dairy cattle population for which the small national training population can be complemented with foreign information from international evaluations. National test-day milk yield data records for the Slovenian Brown Swiss cattle population were analyzed. Genomic evaluation was carried out using the single-step genomic best linear unbiased prediction method (ssGBLUP), resulting in genomic estimated breeding values (GEBV). The predominantly female group of genotyped animals, representing the national training population in the single-step genomic evaluation, was further augmented with 7,024 genotypes of foreign progeny-tested sires from an international Brown Swiss InterGenomics genomic evaluation (https://interbull.org/ib/whole_cop). Additionally, the estimated breeding values for the altogether 7,246 genotyped domestic and foreign sires from the 2019 sire multiple across-country evaluation (MACE), were added to the ssGBLUP as external pseudophenotypic information. The ssGBLUP method, with integration of MACE information by avoiding double counting, was then performed, resulting in MACE-enhanced GEBV (GEBVM). The methods were empirically validated with forward prediction. The validation group consisted of 315 domestic males and 1,041 domestic females born after 2012. Increase, inflation, and bias of the GEBV(M) reliability (REL) were assessed for the validation group with a focus on females. All individuals in the validation benefited from genomic evaluations using both methods, but the GEBV(M) REL increased most for the youngest selection candidates. Up to 35 points of GEBV REL could be assigned to national genomic information, and up to 17 points of GEBVM REL could

additionally be attributed to the integration of foreign sire genomic and MACE information. Results indicated that the combined foreign progeny-tested sire genomic and external MACE information can be used in the single-step genomic evaluation as an equivalent replacement for domestic phenotypic information. Thus, an equal or slightly higher genomic breeding value REL was obtained sooner than the pedigree-based breeding value REL for the female selection candidates. When the abundant foreign progeny-tested sire genomic and MACE information was used to complement available national genomic and phenotypic information in single-step genomic evaluation, the genomic breeding value REL for young-female selection candidates increased approximately 10 points. Use of international information provides the possibility to upgrade small national training populations and obtain satisfying reliability of genomic breeding values even for the youngest female selection candidates, which will help to increase selection efficiency in the future.

Key words: genomic selection, single-step evaluation, forward prediction, genotyped cows, external information

INTRODUCTION

Adequate size and appropriate genetic structure of a training population containing individuals with genomic markers and phenotype information are required to accurately predict genomic estimated breeding values (GEBV) in the process of genomic selection (Meuwissen et al., 2001; Metta et al., 2004; Boichard et al., 2016). Use of genomic selection in dairy cattle breeding programs has spread in recent years, but for countries with small national cattle populations, options for genomic evaluation are limited. Establishment of a large enough training population to accurately estimate genomic marker (SNP) effects is the most limiting factor. Many countries have found an effective solution in international collaboration, forming joint genomic

Received August 18, 2020.

Accepted January 7, 2021.

*Corresponding author: barbara.lustrek@bf.uni-lj.si

evaluations (Lund et al., 2016). Some countries have established bilateral and multilateral genomic evaluations, for example, the German-Austrian evaluation for the Simmental population (Edel et al., 2011), and others have formed international consortia. This is the case for Brown Swiss (**BSW**) cattle populations, connecting France, Germany and Austria, Italy, Slovenia, Switzerland, and the United States into the InterGenomics consortium operated by the Interbull Centre (Jorjani et al., 2010; Zumbach et al., 2010).

InterGenomics for BSW is a multistep genomic evaluation that involves only males and results in InterGenomics GEBV. Some of the involved bulls are progeny-tested, but the majority of them are not. All the bulls are genotyped, and their genomic information is available to all participating countries. The evaluation is carried out using deregressed estimated breeding values (**EBVM**) from progeny-tested sires that are calculated within the Interbull's sire multiple across-country evaluation (**MACE**; Schaeffer, 1994), together with genomic information and international pedigree (Jorjani et al., 2012). On the other hand, MACE evaluation involves only progeny-tested sires and is based on their national pedigree-based EBV.

The most important factors affecting GEBV accuracy are the proportion of genetic variance explained by SNPs, heritability of the trait, and the accuracy of estimated SNP effects (Goddard et al., 2016; Zhang et al., 2019). The latter is directly associated with the size and structure of the training population, as well as the range, quality, and quantity of phenotypic and genomic information of individuals in the training population. Another factor is the genetic relationship between genotyped individuals in the evaluation, where the relationship between individuals in the training population should be minimal, and the relationship between the training and prediction populations should be maximal (Pszczola et al., 2012). Currently, training populations are also complemented with carefully selected females to enlarge the population, to improve its structure and relationship to the prediction population, and to reduce selection bias (Pszczola et al., 2012; Plieschke et al., 2016; Jenko et al., 2017). Unlike the InterGenomics evaluation that uses deregressed EBVM for bulls as pseudophenotypic information, national genomic evaluations would include real national phenotypic information and both female and male genotypes.

Vandenplas et al. (2014, 2017) developed a modification of a single-step genomic best linear unbiased prediction method (**ssGBLUP**; Aguilar et al., 2010) that simultaneously analyses phenotypic, genomic, and pedigree information of genotyped and nongenotyped animals, which enables integration of external informa-

tion. This approach is especially convenient when using foreign sire genetic material, and thus including their information in the evaluation. Participating in both MACE and InterGenomics Interbull evaluations with BSW bulls thereby gives a country access to multinational genomic information as well as MACE information (**EBVM** with reliabilities; **REL**) on its own scale for all of the involved bulls. Integrating external MACE information with the **ssGBLUP** complements national sire information and provides pseudophenotypic information for foreign sires without (or with few) progenies in the national population. As the **ssGBLUP** also results in a genomic breeding value for females, the method is particularly useful for females with only parent average (**PA**) information. Because single-step genomic evaluation already combines all national information (Legarra et al., 2014), the modification also considers potential double counting of the same information (Vandenplas et al., 2017). This ensures a more accurate estimation of genomic breeding values and especially their **REL**.

The aim of this study was to develop and assess the accuracy, bias, and inflation of genomic breeding values and associated **REL** for Slovenian BSW cattle with national data using the **ssGBLUP** method with and without integration of external genomic and MACE information. The focus was on the genomic breeding values and **REL** of the validation group that included young and mostly female domestic genotyped selection candidates. A special emphasis was placed on the female portion of the validation group.

MATERIALS AND METHODS

Data Sets

Data that included national phenotypic and external MACE information, pedigree, and genomic information were established as a full and a reduced data set. Within the reduced data set, the recent phenotypic information was omitted. External MACE information in the reduced data set was selected from the past MACE evaluation to fit the phenotype reduction criteria described below. Pedigree and genomic information data were not reduced (Table 1).

Scenarios

Evaluations and validation were performed in 3 scenarios (Table 1; Figure 1). All evaluations within all scenarios were carried out using the same prediction model, national phenotypic information, and extended national pedigree. Pedigree-based evaluation scenario (**S0**) resulted in EBV with **REL**. Single-step genomic

Table 1. Data used in the full and reduced evaluations within scenarios S0, S1, and S2¹

		Scenario / Evaluation (Data set)					
		S0		S1		S2	
Information	Number ²	Full	Reduced	Full	Reduced	Full	Reduced
Phenotype ³	Females	52,188	37,412				
	Records	1,155,043	747,467				
MACE	Sires	—	—	222	174	7,246	5,462
	EDC >0	—	—	222	174	346	254
	EDC = 0	—	—	—	—	6,900	5,208
Genotype Pedigree ⁴	Animals	—	—	2,166	2,166	9,190	9,190
	Animals		97,989				
	Females		87,058				
	Males		10,931				

¹S0 = national pedigree-based evaluation; S1 = national genomic evaluation with domestic genomic and external multiple across-country evaluation (EBV and associated reliability from Interbull MACE evaluation; MACE) information; S2 = national genomic evaluation with domestic and foreign genomic and external MACE information; full = evaluation or data set that included phenotypic and MACE information, pedigree, and genomic information; reduced = full evaluation or data set with omitted phenotypic information and MACE information from the past MACE evaluation.

²Females = number of females with phenotypic records; records = number of phenotypic records; sires = number of sires with included MACE information in genomic evaluation; EDC = effective daughter contribution; “EDC >0” = number of sires with national EDC >0 in genomic evaluation; “EDC = 0” = number of foreign sires with national EDC = 0 in genomic evaluation; — = information not used in evaluation.

³Common to all scenarios.

⁴Common to all scenarios and evaluations.

evaluation scenarios (**S1** and **S2**), however, were carried out using different genomic and MACE information. Scenario S1 contained only national genomic information and MACE information from sires that had daughters in the national population. Scenario S2 contained genomic and MACE information from S1, as well as foreign sire genomic information from the InterGenomics evaluation and their MACE information. Scenario S1 was designed to test single-step genomic evaluation with the limited size of a national training population, and S2 served to investigate the effect of the augmented national training population. Both genomic single-step evaluation scenarios resulted in GEBV with REL, and MACE-enhanced GEBV (**GEBVM**) with REL (Figure 1).

Evaluations and Validation

To validate genomic evaluations and estimate validation accuracy, evaluations within all 3 scenarios were performed using the full and reduced data sets (Figure 1). Evaluations with different data sets within each scenario are hereinafter referred to as “full evaluation” and “reduced evaluation.” Full evaluations represent the situation where phenotypic information was available for most individuals. Reduced evaluations represent the situation where only pedigree and genomic information was available for young individuals. Validation of the single-step genomic evaluation was performed using the set of domestic genotyped individuals (S1), and

validation of the single-step genomic evaluation with integrated MACE information was carried out using the set of combined national and foreign genotyped individuals (S2).

Data

Phenotypic Information. National phenotypic records for test-day milk yield (kg) recorded between 2004 and 2019 were used for the full evaluation. Data were obtained from the routine November 2019 national pedigree-based evaluation for the Slovenian BSW cattle population. Cows in milk recording were born between 1994 and 2017. On average, they were in second lactation (overall in first through fifth lactations) and had 18.7 kg of milk per test day, on average.

Validation Group. The validation group—the focus population of selection candidates in this study—consisted of 1,356 domestic genotyped (mostly female) individuals born from January 1, 2013, to September 1, 2019 (Table 2). Phenotypic information in the reduced evaluation was determined by discarding all phenotype records of individuals born or recorded after December 31, 2012. Altogether, 35.3% (407,576) of the phenotypic records from the full data set, recorded between 2013 and 2019, were omitted from the reduced evaluation (Table 1). Of these omitted records, 3% (12,133), obtained between 2015 and 2019, represented all own or daughter phenotypic information for individuals in the validation group. Specifically, these records belonged

Table 2. Number (N) of genotyped individuals in the validation group by sex and birth year, and number of validation females within their birth year by the status of the domestic sire

Birth year	Genotyped individuals in the validation group			Females with domestic sire ¹		
	Females	Males	N	Young bull	PT sire	N
2013	233	21	254	2	151	153
2014	201	35	236	10	137	147
2015	138	52	190	8	74	82
2016	108	46	154	11	50	61
2017	173	53	226	33	101	134
2018	163	72	235	84	36	120
2019	25	36	61	2	0	2
Total	1,041	315	1,356	150	549	699

¹Young bull = sire is not progeny-tested; that is, does not have multiple across-country evaluation (MACE) information in the evaluation; PT = sire is progeny-tested; that is, has MACE information in the evaluation.

to 653 of 1,041 validation cows that were born between 2013 and 2017. On average, these cows were in first lactation and had an average of 22.4 kg of milk per test day. The remaining validation cows had no own phenotypic information. As for males from the validation group, only 24 of them (that were born between 2013 and 2015) had daughter phenotypic information. All of that daughter information was omitted from the reduced evaluation. There were 28 of 653 validation

cows altogether with omitted phenotypic information among their daughters.

The MACE Information. International EBVM and associated REL, representing external information, were obtained from MACE evaluation (Interbull). Sire information was determined by the national effective daughter contribution (EDC) estimated by the Interbull Centre. National EDC information was obtained from the routine December 2019 MACE evaluation.

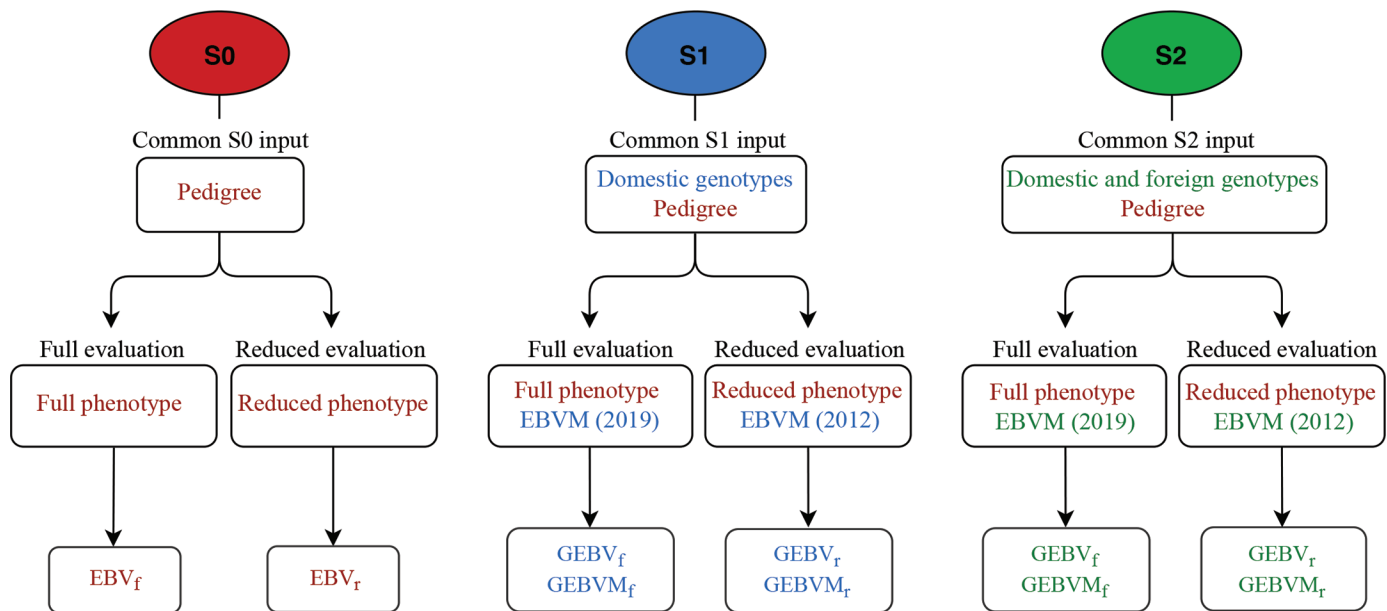


Figure 1. Evaluation scheme. S0 = national pedigree-based evaluation; S1 = national genomic evaluation with domestic genomic and external multiple across-country evaluation (MACE) information; S2 = national genomic evaluation with domestic and foreign genomic and external MACE information; full evaluation = data set that included phenotypic and MACE information, pedigree, and genomic information; reduced evaluation = full data set with omitted phenotypic information and MACE information from the past MACE evaluation; EBVM = MACE estimated breeding values; 2012 or 2019 = the year of routine December MACE evaluation; EBV_f = pedigree-based estimated breeding value from full evaluation; EBV_r = pedigree-based estimated breeding value from reduced evaluation; GEBV_f = genomic breeding value from full evaluation; GEBV_r = genomic breeding value from reduced evaluation; GEBVM_f = MACE-enhanced GEBV from full evaluation; GEBVM_r = MACE-enhanced GEBV from reduced evaluation.

Integrated EBVM and associated REL represented external information for genotyped domestic and foreign sires that contributed with national information into MACE evaluation ($EDC > 0$), and for other foreign-genotyped sires ($EDC = 0$). The national information for sires that contributed to MACE was considered to avoid double counting. Information from the routine December 2019 MACE evaluation was used in the full evaluation along with genetic parameters from the corresponding routine national pedigree-based evaluation. Information from the routine December 2012 MACE evaluation, together with genetic parameters from the corresponding national routine evaluation, was used in the reduced evaluation. The double counting in the reduced evaluation was considered in the same way as in the full evaluation, but with the time difference of 7 yr.

Pedigree. To build the pedigree that was used in all evaluations within all 3 scenarios (Table 1; Figure 1), ancestors of a combined set of genotyped individuals from S2 (9,190) and domestic cows with phenotype records from the full evaluation (52,188) were extracted up to 5 generations back.

Genomic Information

Genotyped Individuals. Each single-step genomic evaluation scenario comprised different numbers of genotyped individuals in the single-step genomic evaluations: scenario S1 included 2,166 domestic genotyped, mostly female, individuals that were further complemented with 7,024 foreign-genotyped progeny-tested sires from the InterGenomics consortium for scenario S2. These sires represented about one-third of the current InterGenomics bull population, as the rest of the InterGenomics population were young bulls. They were selected from the InterGenomics genotype pool based on the presence of MACE information for test-day milk yield in the routine December 2019 MACE evaluation. Selected genotyped sires originated from 10 countries,

mostly from Germany (30.2%), Switzerland (26.1%), and Italy (19%). Together with domestic individuals, they formed a set of 9,190 genotyped individuals for scenario S2 (Table 3). Individuals with genomic information were born between 1981 and 2019.

Genotypes. All 9,190 genotypes were obtained from the December 2019 InterGenomics evaluation. They originally consisted of 55,172 SNPs selected within the Interbull set of SNP markers. Female genotypes were processed as part of this routine. Genomic information quality control, on the individual and marker levels, was performed by the PreGSF90 program (Misztal, 2013). Quality control criteria for marker and individual exclusion were as follows: call rate < 0.90 , minor allele frequency < 0.05 , monomorphic SNPs when minor allele frequency was not equal 0, parent-progeny Mendelian conflicts, and genotype duplicate > 0.95 identical marker genotypes.

Prediction Model

(Co)variance components (Table 4) used in all evaluations within all scenarios were estimated by the VCE program (Groeneveld et al., 2010) based on a pedigree BLUP within the national routine November 2019 evaluation, using the REML method. The routine national pedigree-based univariate multiple-lactation repeatability evaluation model for milk production was used to estimate (co)variance components and breeding values within the pedigree-based (S0) and ssGBLUP genomic evaluations (S1, S2) as follows:

$$\mathbf{y}_i = \mathbf{X}\mathbf{b} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_h\mathbf{h} + \mathbf{Z}_p\mathbf{p} + \mathbf{e},$$

where \mathbf{y}_i represented the vector of observations for test-day milk yield (kg) of individual i ; \mathbf{X} , \mathbf{Z}_a , \mathbf{Z}_h , and \mathbf{Z}_p were the design matrices for fixed and random effects, respectively; \mathbf{b} was the vector of fixed effects [year and month of calving interaction, lactation stage (days) formed using the Ali-Schaeffer curve (Ali and Schaeffer, 1987) nested within parity]; $\mathbf{a} \sim MVN(0, \mathbf{A}\sigma_a^2)$ in the pedigree-based evaluation, and $\mathbf{a} \sim MVN(0, \mathbf{H}\sigma_a^2)$ in the single-step genomic evaluations, were vectors of additive genetic effect where σ_a^2 represented the additive genetic variance, \mathbf{H} represented the combined pedigree-based additive (\mathbf{A}) and genomic (\mathbf{G}) relationship matrices (Aguilar et al., 2010; Christensen and Lund, 2010), and MVN represented multivariate normal distribution; $\mathbf{h} \sim MVN(0, \mathbf{I}\sigma_h^2)$ was the vector of random herd-year effect where \mathbf{I} represented an identity matrix and σ_h^2 represented the variance of herd-year effect;

Table 3. Number of genotyped individuals by sex and origin

Origin	Country of origin	Females	Males
National Foreign	Slovenia	1,589	577
	Germany		2,119
	Switzerland		1,832
	Italy		1,336
	United States		937
	Austria		579
	France		189
	Canada		27
	Other		5
	Total		1,589

$\mathbf{p} \sim MVN(0, \mathbf{I}\sigma_p^2)$ was the vector of random permanent environment effect where σ_p^2 represented the variance of permanent environment effect; and $\mathbf{e} \sim MVN(0, \mathbf{I}\sigma_e^2)$ was the vector of random residual effects where σ_e^2 represented the residual variance.

The inverse of the \mathbf{H} matrix was computed as follows:

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

where \mathbf{A}^{-1} represented the inverse of the additive pedigree relationship matrix, \mathbf{A}_{22}^{-1} represented the inverse of the additive pedigree relationship submatrix for genotyped individuals \mathbf{A}_{22} , and \mathbf{G}^{-1} represented the inverse of the genomic relationship (co)variance matrix \mathbf{G} . The \mathbf{G} matrix was equal to blended $\mathbf{G} = 0.85\mathbf{G}^* + 0.15\mathbf{A}_{22}$, where 0.15 was the proportion of the total genetic variance explained by residual polygenic effects as in Vandenplas et al. (2017). The raw \mathbf{G}^* matrix was created by the first method of VanRaden (2008), as described in Aguilar et al. (2011), and the allele frequencies calculated from the observed genotypes of individuals within specific single-step genomic evaluation scenario (Table 1; Figure 1) were used to center the genotype matrix (\mathbf{W}) as follows:

$$\mathbf{G}^* = \frac{\mathbf{W}\mathbf{W}'}{k},$$

where $\mathbf{W} = \mathbf{M} - 2\mathbf{1}\mathbf{p}'$, where \mathbf{M} was the matrix of SNP marker genotypes for each individual, $\mathbf{1}$ was a vector of ones and \mathbf{p} was the vector of observed allele frequencies, and $k = 2 \mathbf{p}'(\mathbf{1} - \mathbf{p})$ (VanRaden, 2008). Dividing $\mathbf{W}\mathbf{W}'$ by k scaled \mathbf{G}^* to become analogous to the pedigree-based additive relationship submatrix \mathbf{A}_{22} .

Breeding value and associated prediction error variance (PEV) computations were carried out with the

BLUP, ssGBLUP, and the modified ssGBLUP method that integrates external information (Vandenplas et al., 2014) using the BLUPF90 program (Miszta, 2013).

The approach that enables integration of MACE information into a national single-step genomic evaluation is described in Vandenplas et al. (2014, 2017) and Colinet et al. (2018). It uses national phenotypic information (\mathbf{y}_i) in ssGBLUP and integrates MACE information for domestic and foreign-genotyped sires. It propagates foreign information to domestic animals that did not participate in the MACE evaluation. The approach corrects the dependency of MACE information on national information for each genotyped sire. It avoids double counting of information from own phenotypic records by accounting for nonzero sire EDC, and double counting of information from relatives by estimating the effective number of records using the 2-step TSA algorithm described by Vandenplas and Gengler (2012).

Precision of pedigree-based and genomic breeding values was determined by the theoretical prediction reliability that was calculated using the formula

$$REL_i = 1 - \left(\frac{PEV_i}{\sigma_{ai}^2} \right),$$

where PEV_i was the PEV for individual i obtained from the direct inverse of the left-hand-side of the mixed model equations system, and σ_{ai}^2 represented the diagonal element times the additive genetic variance for individual i in the pedigree (\mathbf{A}) or combined pedigree and genomic (\mathbf{H}) (co)variance matrix. Data and results were prepared and analyzed by the SAS/STAT statistical package (version 9.4, SAS Institute Inc.).

Analyses

Principal component analysis was carried out using PLINK v1.9 (Purcell et al., 2007) to estimate the genetic relationship among the genotyped individuals, and GENESIS (Buchmann and Hazelhurst, 2014) was used to display the first 2 principal components graphically. Results of single-step genomic evaluations (GEBV with REL from S1 and GEBVM with REL from S2) and validation were analyzed in depth for validation group (Table 2) along with the results from pedigree-based evaluations (EBV with REL from S0). Full evaluation results are marked with the subscript “f,” and subscript “r” marks the results from the reduced evaluations. Pearson correlation coefficients were estimated between breeding values and between REL obtained from different single-step genomic evaluation scenarios for male and female validation individuals to determine the effect of foreign genomic and MACE information.

Table 4. Estimated (co)variance components from the national routine November 2019 evaluation

Parameter ¹	(Co)variance component estimates
h^2	0.331
c^2	0.152
p^2	0.155
e^2	0.362

¹ h^2 = heritability; c^2 = herd-year; p^2 = permanent environment; e^2 = random residual.

Changes in breeding value reliabilities from the full pedigree-based evaluation and the reduced single-step genomic evaluation with and without MACE information were assessed as the difference between mean reliability values within each birth year by sex of validation individuals. The same analysis was performed for breeding value reliabilities from the full pedigree-based evaluation and the full single-step genomic evaluations. The difference in the mean reliability from full pedigree-based and full genomic evaluations, expressed as a reliability change, was also calculated by sex of validation individuals. Reliability change was further assessed as the difference between mean reliability values from full and reduced single-step genomic evaluations within each birth year for female individuals by status (progeny-tested with MACE information or not) of their sire.

Pearson correlation coefficients were estimated between breeding values obtained from the full and reduced pedigree-based and single-step genomic evaluations for male and female validation individuals to determine the extent of an increase of the breeding value accuracy. Standard errors for correlation coefficients were calculated using the formula $SE(\widehat{\text{corr}}) = \sqrt{(1 - \text{corr}^2)/(n - 2)}$, where corr is the correlation coefficient and n is number of observations (Whitlock and Schluter, 2009).

Inflation and deflation of breeding values were measured for females as a linear regression of values from the full evaluation on the values from the reduced or full evaluation. Specifically, values from the full pedigree-based evaluation were regressed on the values from

the reduced pedigree-based evaluation, as well as on the values from the reduced single-step genomic evaluations with and without MACE information. Values from full genomic evaluations were regressed on the values from the reduced genomic evaluations within the single-step genomic evaluation without MACE information and the single-step genomic evaluation with MACE information. Values from full genomic evaluations were also regressed on the values from the full pedigree-based evaluation. Values from the full single-step genomic evaluation with MACE information were further regressed on the values from the full single-step genomic evaluation without MACE information. Using listed combinations of variables, the regression model was $(G)EBV(M)_f = b_0 + b_1 \times (G)EBV(M)_{f(r)} + e$, where

b_0 was the intercept and b_1 the coefficient of regression. The coefficient of determination (R^2), taken from the linear regression analysis, was used to assess the validation reliability of evaluations. Standard errors for R^2 were estimated using the bootstrap approach described by Efron and Tibshirani (1986) with 5,000 replications of R^2 , from 5,000 bootstrap samples of the same size as the female validation subgroup generated by the PROC SURVEYSELECT procedure of SAS/STAT statistical package using balanced bootstrap sample selection method.

RESULTS AND DISCUSSION

Structure of Genotyped Individuals

Population overlap (Figure 2), demonstrated by principal component analysis for all 9,190 genotyped indi-



Figure 2. First (PCA 1) and second (PCA 2) principal components of the genomic relationship for domestic individuals (SVN) and foreign sires with multiple across-country evaluation (MACE) information; PCA = principal component analysis.

viduals, pointed to the strong interconnectedness of the national and foreign populations. Domestic genotyped individuals are related to both foreign sire subpopulations due to the exchange of genetic material between participating countries. The first principal component explained 13.1%, and the second principal component explained 9.5% of the genotype variability.

To accurately estimate genomic breeding values, a training population traditionally consists of a minimum of 1,000 individuals with genomic and reliable phenotypic information (Calus, 2010) in the form of own phenotypic records, information from progeny performance (Thomassen et al., 2014), or (deregressed) EBV (VanRaden et al., 2009; Jenko et al., 2017). For this reason, the progeny-tested sires are a crucial component of the training population, but their number in small cattle populations is generally low. This is also the case in the Slovenian BSW population. The population counts less than 15,000 active cows, and only a small portion of the population is genotyped, including all domestic progeny-tested sires. A convenient alternative to increase the national training population is the use of genotypes and the available external information of foreign progeny-tested bulls that are genetically connected to the national population.

Correlation Between Genomic Evaluation Scenarios

The $GEBV_f$ REL for validation females and males did not differ among scenarios S1 and S2 as indicated by 0.99 Pearson correlation coefficients. These results suggested that there was no important effect on $GEBV_f$ REL for validation animals from full evaluations when genomic information for 7,024 foreign sires without their MACE information was added to domestic information in the evaluations within scenario S2. Foreign genomic information slightly affected $GEBV_f$ breeding values of validation males (females were less affected), as indicated by correlation of 0.97 (0.99 for females) between genomic scenarios.

Additionally, no differences were found between $GEBVM_f$ REL from scenario S1 and $GEBV_f$ REL from scenarios S1 and S2, as suggested by correlations of 0.99 for both sexes in the validation group. The correlations between these genomic breeding values were also high: 0.97 to 0.99 (0.99) for the males (females). These values suggested that MACE information (or the extent of it) included in the S1 genomic evaluations had no considerable effect on the REL or breeding values. These correlations also indicated the similarly minor effect of foreign genotypes itself without MACE information and confirmed the finding of the previous paragraph.

Correlations between $GEBVM_f$ REL from S2 and $GEBV_f$ REL from S1 were 0.91 (0.99) for validation males (females), whereas the breeding value correlations were 0.83 (0.95) for males (females). Similar correlations were estimated between $GEBVM_f$ from S2 and $GEBV_f$ from S2. Reliability correlations were 0.90 (0.99) for males (females), and breeding value correlations were 0.85 (0.95) for males (females). These correlations suggested that the combined foreign genomic and MACE information in the evaluation affected validation males more than females, and their breeding values more than the reliabilities. Based on the findings in this section, the following results present $GEBV_{f(r)}$ and associated REL from scenario S1, and $GEBVM_{f(r)}$ with associated REL from scenario S2.

Reliability Changes by Sex

The importance of the relationship proximity to the (ancestral) individuals with phenotypic information, as mentioned by Lund et al. (2016), was indicated by the decrease of all mean EBV_f and $GEBV(M)_r$ REL by birth year for both sexes (Figure 3). Figure 3 shows that a greater birth year distance from the phenotype cutoff point in the year 2013, and thus from the ancestors with phenotypic information, yielded a lower REL. This trend was particularly apparent in the reduced genomic evaluations, where the phenotype cutoff point was the same for all validation individuals. The decrease was, however, less extreme in the genomic breeding value REL compared with the EBV_f REL, where natural absence of phenotypic information occurs. Because the test-day milk yield can be recorded after a cow's first calving (at approximately 2 yr of age) and after the first daughter's calving for bulls at 3 to 4 yr of age, validation females born in late 2017 and later had no own phenotypic information in this study, and validation males born in late 2015 and later had no daughter phenotypic information available. As the relationship distance increased, mean EBV_f REL decreased from 0.81 (0.59) for females (males) born in 2013 to 0.29 (0.28) for females (males) born in 2019. Only 9 out of 21 validation males born in 2013 (Table 2) had daughters with phenotypic information in the full evaluations, and consequently these males had lower average (EBV_f) REL compared with validation females born in the same year. The mean $GEBV_r$ REL of females (males) born in 2013 decreased from 0.35 (0.29) to 0.15 (0.16) for individuals born in 2019. The mean $GEBVM_r$ REL decrease was the mildest and almost identical for both sexes, dropping from 0.45 for the individuals born in 2013 to 0.32 for individuals born in 2019.

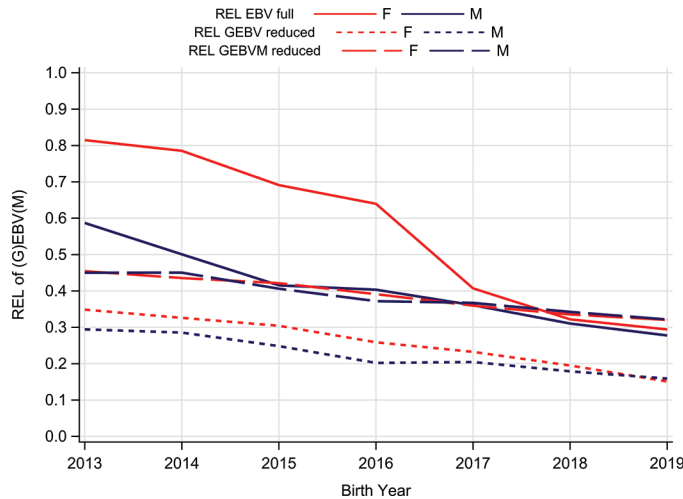


Figure 3. Trend for mean breeding value reliability (REL) from the full pedigree-based evaluation, and for mean genomic breeding value REL from the reduced genomic evaluations for the validation group (Table 2) by sex (EBV = pedigree-based estimated breeding value; GEBV = genomic breeding value; GEBVM = MACE-enhanced GEBV; MACE = multiple across-country evaluation; F = females; M = males).

Regardless of sex, the $GEBV_f$ of a young individual without phenotypic information in this study, estimated by using small national training population in the single-step genomic evaluation, was, on average, almost 10 points higher than predictions based on EBV (Figure 4). In contrast to the report on milk yield by Lee et al. (2020), who estimated 0.47 to 0.52 GEBV REL for females with phenotype, and 0.68 to 0.75 GEBV REL for sires with progeny, we estimated higher mean $GEBV_f$ REL for females with phenotypic information (0.68–0.83). Our estimated mean $GEBV_f$ REL for males with daughter phenotype was somewhat lower, presumably due to the smaller amount of daughter information (0.58–0.65; Figure 4). However, for both young females and young males without phenotypic information, our mean $GEBV_f$ REL, between the range of 0.43 and 0.5 (0.42–0.51) for females (males; Figure 4), were higher than those for young females without phenotypic information (0.34–0.38) and males without progeny (0.35–0.39) reported by Lee et al. (2020). Their heritability for milk yield, ranging between 0.16 and 0.28 for different parities, was lower than ours (0.33); that may explain lower REL values for the younger animals in their study.

The combination of genomic and MACE information can equivalently replace domestic phenotypic information and provide equal or even slightly higher genomic breeding value ($GEBVM_r$) REL for young-female selection candidates much sooner than the pedigree-based evaluation would (Figure 3). The $GEBV_r$ REL trend

suggests that 15 to 35 points of GEBV REL can be assigned to national genomic information, where the value depends on the distance from relatives' phenotypic information. The $GEBVM_r$ REL trend suggests that an additional 11 to 17 points can be assigned to the combination of added sire genomic and MACE information (Figure 3). From the latter, females obtained, on average, 13 points, and males 16 points of $GEBVM_r$ REL. As the $GEBV(M)_r$ REL includes only ancestral phenotypic information, these results can be considered a reliability increase caused by genomic and MACE information. Similarly, the GEBV REL increase can also be estimated by comparison with PA REL. For example, Boison et al. (2017) estimated between a 22.9- and 26.4-point PA to GEBV REL increase for milk yield in Gyr cattle (with heritability 0.28), having ~150 males in the validation group using GBLUP with genotyping panels of diverse density. This outcome is in satisfactory agreement with our $GEBV_r$ REL results, especially because the study of Boison et al. (2017) had a similarly structured and sized training population. Our results are also in line with the report of Hayes et al. (2009), who reviewed the PA-GEBV REL increase for different traits from several countries. The PA to GEBV REL increase for young dairy bulls with no daughter information in New Zealand was between 16 and 33 points, including milk production; the average increase for US and Canadian young Holstein bulls was 23 points, and between 9 and 33 points for Dutch dairy bulls.

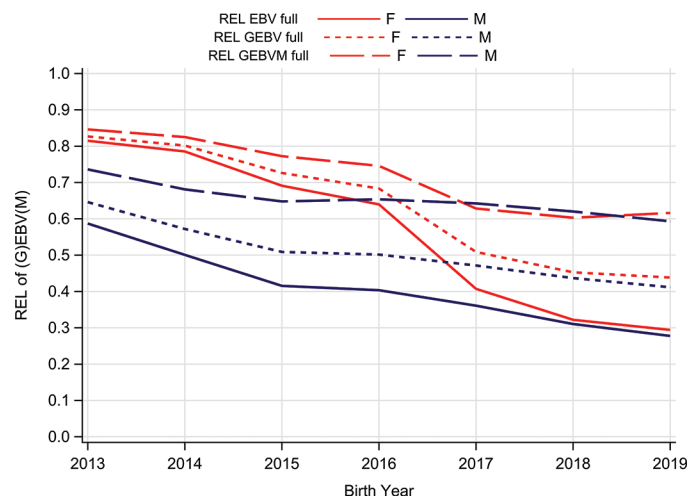


Figure 4. Trend for mean breeding value reliability (REL) from the full pedigree-based evaluation, and for mean genomic breeding value RELs from the full genomic evaluations for the validation group (Table 2) by sex (EBV = pedigree-based estimated breeding value; GEBV = genomic breeding value; GEBVM = MACE-enhanced GEBV; MACE = multiple across-country evaluation; F = females; M = males).

Table 5. Average reliability increase (points) between breeding values from the full (f) pedigree-based and genomic evaluations for the validation group by sex (increase for females in the upper matrix diagonal, increase for males in the lower matrix diagonal)

Item ¹	EBV _f	GEBV _f	GEBVM _f
EBV _f		5	12
GEBV _f	10		7
GEBVM _f	25	15	

¹EBV = pedigree-based estimated breeding value; GEBV = genomic breeding value; GEBVM = multiple across-country evaluation (MACE)-enhanced GEBV.

Difference in Mean Reliability by Sex

When MACE information was included for progeny-tested genotyped sires, a positive deviation of GEBVM_f REL from GEBV_f REL was observed (Figure 4; Table 5). Integration of MACE information into genomic evaluation for a total of 7,246 foreign and domestic sires, avoiding double counting of the same information, brought the highest REL increase for the entire validation group. Reliability increase was especially high for the youngest individuals, as the MACE integration compensated for their lack of phenotypic information. The GEBV_f REL value diminished by each year similarly to EBV_f REL, which was anticipated after only a small number of genotypes linked to the domestic phenotypic information was included in the evaluations, whereas GEBVM_f REL retained a steadier decline (Figure 4).

The main aim of genomic selection is to predict sufficiently reliable estimated breeding values for young selection candidates without own phenotypic information (Meuwissen et al., 2001; Schaeffer, 2006). Reliability of genomic breeding values of the young-genotyped selection candidates without own phenotypic information strongly depends on the number of individuals with genomic information in the training population (Calus, 2010). Reliability of estimated breeding values for such individuals was improved in this study by including combined foreign genomic and external MACE information.

A small training population can negatively affect GEBV REL (e.g., Lund et al., 2011), but that was not the case in the present study. However, the importance of combining national and available foreign information in genomic evaluation has been shown in several publications (Vandenplas et al., 2015; Vandenplas et al., 2017; Andonov et al., 2017; Colinet et al., 2018; Guarini et al., 2019). Combining multiple information sources is especially relevant for small cattle populations with limited national information where all additional information in evaluation is useful.

The largest REL increase was indicated by the difference between EBV_f and GEBVM_f REL in both sexes. As mentioned above, the correlation between GEBV_f and associated REL from S1 and S2 was 0.99; therefore, the EBV_f to GEBVM_f REL increase can be assigned directly to the extensive MACE information that augmented contribution of the genomic relationship matrix (Misztal et al., 2013; Figure 4; Table 5). The GEBVM_f REL increase was more prominent for males.

Lee et al. (2020) obtained an average of 9 points of GEBV REL increase compared with EBV for milk production traits estimated using the ssGBLUP method with ~2,000 genotyped Holstein individuals of both sexes, which corresponded to the difference between EBV_f and GEBV_f REL for males in this study (Table 5). They noted lower REL increase for progeny-tested bulls (4 points) and females with phenotype information (7 points), and higher REL increase for heifers without own and bulls without progeny phenotypic information (13 and 17 points, respectively). Our prediction group consisted of females with and without phenotypic information, and young males with mainly limited or no phenotypic information available within the full evaluations. The structure of our prediction group may explain the somewhat lower mean GEBV_f REL increase compared with EBV_f REL in the present study. Unlike other authors, Vandenplas et al. (2017) integrated the GEBV and genomic information for ~6,000 BSW bulls from InterGenomics evaluation into the ssGBLUP evaluation for milk yield. Only 277 of ~6,000 bulls contributed national information to the MACE evaluation, of which results were integrated into the InterGenomics evaluation. Their genomic REL increased by 10 points (0.97) compared with EBV REL (0.87) for 319 domestic and foreign bulls with phenotyped daughters in the national evaluation. The genomic REL increased by 8 points, on average, for individuals with low amounts of phenotypic information and low EBV REL (<0.25) after the GEBV integration. The same effect was observed in the present study, with a GEBVM_f REL increase by up to 33 points compared with EBV_f REL for animals with low amounts of phenotypic information (Figure 4).

Studies with integration of external information are mostly conducted using Holstein cattle data. For example, Colinet et al. (2013, 2018) carried out ssGBLUP genomic evaluation with MACE integration from 2013 and 2015 evaluations for the Walloon Holstein cattle population with features similar to the Slovenian BSW population (small size, small training population size, and common use of foreign genetic material). In their studies, they integrated the MACE information of ~1,980 and almost 4,000 bulls, of which ~400–700

contributed to the integrated MACE evaluation. They included 2,000 and 5,000 genotyped individuals, respectively. Both studies confirmed the highest GEBV(M) REL increase for the genotyped individuals with the lowest EBV REL. Their milk yield (heritability 0.38 in both studies) GEBV REL increase was 7 points lower than the EBV for young bulls when a smaller amount of domestic data were included in the evaluation (Colinet et al., 2013). Nevertheless, even with a smaller amount of domestic data, a higher number of bulls with EDC >0 and integrated MACE information in the evaluation resulted in a larger increase of GEBVM REL compared with the EBV REL (10 points more), and compared with GEBV REL (17 points more) for young bulls than in the most recent study.

Our average GEBV_f to GEBVM_f REL increase for the youngest individuals was closer to the increase from the recent study (Colinet et al., 2018; 19 points) where they included a greater amount of genomic and MACE information. Guarini et al. (2019) also used Holstein data with genotype information for 33,000 bulls and obtained higher reliabilities integrating MACE information into ssGBLUP than in multiple-step GBLUP evaluation for milking temperament and speed. Similar to Colinet et al. (2013), they integrated April 2013 MACE information into genomic evaluations for around 1,800 genotyped sires, of which ~100 contributed to integrated MACE evaluation. They obtained 9.6-point and 19.3-point EBV to GEBV REL increases with domestic ssGBLUP evaluation for their 2 traits, which is in agreement with Colinet et al. (2013) for bulls with low EBV REL, and with our results for validation males. Guarini et al. (2019) obtained additional 0.2-point and 1.4-point GEBV to GEBVM REL increases, respectively, when they integrated MACE information, which is lower than the rest of the presented and reviewed results in the present study.

Reliability Changes for Females by the Sire Status

The majority of validation females had a domestic sire, and most of them were MACE-evaluated progeny-tested sires (Table 2). The rest of the sires were not progeny-tested and were considered young bulls. The number of progeny-tested sire daughters within each birth year was higher than the number of domestic young bulls' daughters, with the exception of the years 2018 and 2019. There were no daughters of domestic progeny-tested sires born in 2019.

In general, the REL for females from all single-step genomic evaluations was higher when their sire was progeny-tested (Figure 5; Y). These differences can be explained by the different number of sires and their MACE information in the evaluations. The difference

in REL from full evaluations between the daughters of domestic progeny-tested and young bull sires became larger for those born after 2016 when the availability of the phenotypic information naturally reduced. The GEBV(M)_f REL for the domestic progeny-tested sire daughters born in 2018 reached 0.52 (GEBV_f) and 0.64 (GEBVM_f), whereas the REL for the young bull sire daughters born in 2018 and 2019 reached 0.43 (GEBV_f) and 0.58 (GEBVM_f), respectively.

The most considerable difference on account of the sire status in mean GEBV_f REL (10 points) and mean GEBVM_f REL (7 points) was for females born in 2017. For the domestic progeny-tested sire daughters born after 2016, the difference between mean GEBV_r and GEBVM_r REL was 12 points. The full evaluation difference stayed in a similar range between 10 and 12 points (Figure 5). For daughters of young domestic bull sires born in the same years, the difference between mean GEBV_r and GEBVM_r REL ranged between 14 and 17 points, and in the full evaluation, the difference was 13 to 15 points. Colinet et al. (2013, 2018) obtained a 9- to 30-point GEBV REL increase compared with EBV REL, an 11- to 41-point EBV to GEBVM REL increase, and a 1- to 14-point GEBV to GEBVM REL increase for their genotyped individuals of both sexes sired by progeny-tested MACE-evaluated sires. The higher GEBVM REL increase obtained in their 2013 study confirms the critical role of the sires with EDC >0 in the single-step genomic evaluation. Findings from our study for individuals of both sexes without

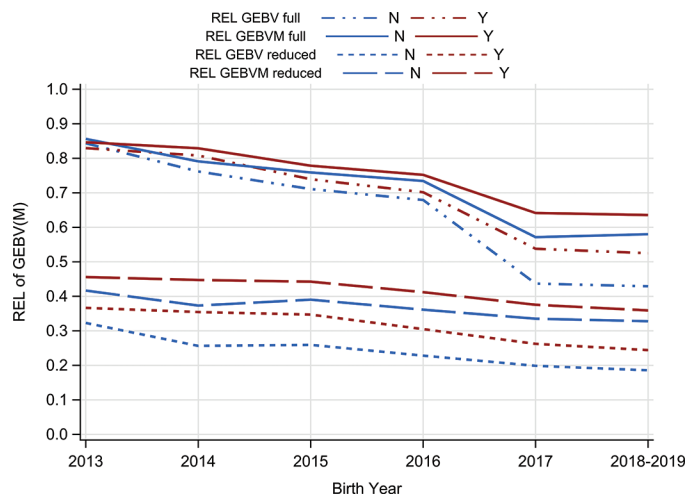


Figure 5. Trend for mean reliability (REL) of genomic breeding values (GEBV, GEBVM) from full and reduced genomic evaluations for females in the validation group (Table 2) by sire progeny-tested status (GEBV = genomic breeding value; GEBVM = multiple across-country evaluation (MACE)-enhanced GEBV; Y = sire is progeny-tested; that is, has MACE information in the evaluation; N = sire is not progeny-tested; that is, does not have MACE information in the evaluation).

Table 6. Correlations (corr) and SE between estimated breeding values from the full (_f) and reduced (_r) evaluations for validation group (Table 2) by sex and birth year

Item ¹	Statistic	Birth year						All
		2013	2014	2015	2016	2017	2018–2019	
Female								
Number of individuals		233	201	138	108	173	188	1,041
EBV _f – EBV _r	Corr	0.38	0.46	0.37	0.28	0.46	0.37	0.40
	SE	0.06	0.06	0.08	0.09	0.07	0.07	0.03
GEBV _f – GEBV _r	Corr	0.39	0.50	0.26	0.25	0.50	0.31	0.39
	SE	0.06	0.06	0.08	0.09	0.07	0.07	0.03
GEBVM _f – GEBVM _r	Corr	0.52	0.53	0.50	0.59	0.70	0.66	0.59
	SE	0.06	0.06	0.07	0.08	0.05	0.06	0.03
GEBVM _f – EBV _f	Corr	0.97	0.97	0.94	0.88	0.80	0.74	0.90
	SE	0.02	0.02	0.03	0.05	0.05	0.05	0.01
Male								
Number of individuals		21	35	52	46	53	108	315
EBV _f – EBV _r	Corr	NS	0.45	0.33	0.43	NS	0.35	0.29
	SE	NS	0.16	0.13	0.14	NS	0.09	0.05
GEBV _f – GEBV _r	Corr	NS	0.49	0.29	NS	0.29	0.34	0.27
	SE	NS	0.15	0.14	NS	0.13	0.09	0.05
GEBVM _f – GEBVM _r	Corr	0.46	0.73	0.58	0.74	0.65	0.74	0.66
	SE	0.20	0.12	0.12	0.10	0.11	0.07	0.04
GEBVM _f – EBV _f	Corr	0.78	0.55	0.78	0.68	0.61	0.69	0.69
	SE	0.14	0.15	0.09	0.11	0.11	0.07	0.04

¹EBV = pedigree-based EBV; GEBV = genomic breeding value; GEBVM = multiple across-country evaluation (MACE)-enhanced GEBV.

accounting for sire status agree with these values; however, the young bull daughters in our study obtained up to a 3-point-higher GEBV to GEBVM REL increase.

These results suggest that regardless of the availability of the recent phenotypic information and domestic sire status, the inclusion of combined abundant foreign sire genomic and MACE information increases the reliability of genomic breeding values for young females by around 10 points. The increase is slightly higher if the sire is a young domestic bull. Domestic and especially progeny-tested sires have more relatives in the domestic population, and thus a stronger association with phenotypic information, which comes to the fore when selection candidates have no own or progeny phenotypic information. Younger bulls have fewer progenies in the population, and thus a weaker link to the phenotypic information. Therefore, their daughters are more affected by the additional MACE information in the evaluation.

Correlation Between Breeding Values

Each correlation presents the effect of a different piece of information on breeding value accuracy. The correlation between breeding values from the full and reduced evaluations is also known as a direct estimate of an increase of the population breeding value accuracy and can be used as information on the ratio of accuracies (Legarra and Reverter, 2018). The correlation between EBV_f and EBV_r showed the effect of recent

own or progeny phenotypic information on the accuracy obtained from the pedigree-based evaluation. The correlation between GEBV_f and GEBV_r showed the effect of recent own or progeny phenotypic information in the presence of genomic information in the single-step genomic evaluation compared with the correlation between pedigree-based evaluations. The correlation between GEBVM_f and GEBVM_r further described the additional effect of the combined recent external MACE information and foreign genomic information included in the single-step genomic evaluation.

Results suggest (Table 6) that available (ancestral) phenotypic information affected the accuracy for the youngest selection candidates to a similar degree that it affected the older individuals in the pedigree-based evaluation, as well as in the single-step genomic evaluation. Overall, the correlations within sex differed among the birth years, but they were not always statistically significant ($P < 0.0001$) for males. The correlations between EBV_f and EBV_r and between GEBV_f and GEBV_r were up to 0.37 for the youngest individuals (Table 6). Decreased correlation between GEBV_f and GEBV_r, compared with EBV_f and EBV_r, indicated extra accuracy on account of national genomic information for individuals born in 2015, 2016, 2018, and 2019. The decrease can be explained by the introduction of the national genomic information in the evaluation.

Results also suggest that the combination of genomic information and MACE sire information from the past MACE evaluation included in the reduced single-step

genomic evaluation contributed to the accuracy of the individuals with the least amount of information in the evaluation to a similar degree as did the inclusion of recent phenotypic and MACE information in the full evaluation. This was shown by the correlation between $GEBVM_f$ and $GEBVM_r$ (Table 6). The combination of national and foreign genotypes and external information considerably increased the correlation between $GEBVM_f$ and $GEBVM_r$ for all birth years and both sexes. The correlation increase was somewhat smaller for females than males, pointing to the higher contribution of the recent combined information for the female breeding value accuracy compared with males.

The contribution of all domestic, foreign, and MACE information to the breeding value accuracy was minor if the females already had sufficient phenotypic information. Nevertheless, the accuracy for the younger females was affected to a higher degree by the additional information. This trend was suggested by the highest correlation between $GEBVM_f$ and EBV_f for all females; however, the correlation decreased for the females born after 2016 (Table 6). The correlation between $GEBVM_f$ and EBV_f for males was lower than for the females. Compared with the pedigree-based evaluation, genomic and MACE information contributed more to the accuracy of genomic breeding values for males than it did for females.

Additional breeding value correlations are presented in Table 7. The correlations between EBV_f and $GEBV_r$ and between EBV_f and $GEBVM_r$ demonstrated the effect of recent phenotypic information (EBV_f) on breeding value accuracy compared with the effect of only genomic information ($GEBV_r$), or genomic information combined with external information ($GEBVM_r$) in the evaluation (Table 7). Correlations for males were not all statistically significant, and no conclusions could be made based on these results.

Presented correlations suggest that for the breeding value accuracy of the young female selection candidates, genomic information is similarly important as the recent and own or progeny phenotypic information. This was indicated by the similar correlations between EBV_f and $GEBV_r$ and between $GEBV_f$ - $GEBV_r$ for females. The correlation between EBV_f and $GEBV_r$ for the youngest males was lower than for the youngest females (Table 7).

For females, the combination of national and foreign genomic and MACE information contributed somewhat less to the accuracy than the national genomic information. The correlation between EBV_f and $GEBVM_r$ was generally higher than between EBV_f and $GEBV_r$ for both sexes, which was expected because MACE information indirectly includes information that is also considered in foreign pedigree-based evaluations (Table

Table 7. Additional correlations (corr) and SE between estimated breeding values from the full (i) and reduced (r) evaluations for the validation group (Table 2) by sex and birth year

Item ¹	Statistic	Birth year						All
		2013	2014	2015	2016	2017	2018–2019	
Female								
Number of individuals		233	201	138	108	173	188	1,041
$EBV_f - GEBV_r$	Corr	0.38	0.51	0.25	0.22	0.41	0.28	0.37
	SE	0.06	0.06	0.08	0.09	0.07	0.07	0.03
$EBV_f - GEBVM_r$	Corr	0.45	0.47	0.39	0.38	0.43	0.34	0.43
	SE	0.06	0.06	0.08	0.09	0.07	0.07	0.03
$GEBV_f - GEBVM_r$	Corr	0.47	0.48	0.38	0.45	0.51	0.44	0.47
	SE	0.06	0.06	0.08	0.09	0.07	0.07	0.03
$GEBVM_f - EBV_r$	Corr	0.34	0.38	0.27	0.21	0.38	0.21	0.31
	SE	0.06	0.07	0.08	0.09	0.07	0.07	0.03
$GEBVM_f - GEBV_r$	Corr	0.38	0.47	0.24	0.24	0.45	0.29	0.36
	SE	0.06	0.06	0.08	0.09	0.07	0.07	0.03
Male								
Number of individuals		21	35	52	46	53	108	315
$EBV_f - GEBV_r$	Corr	NS	NS	0.30	NS	NS	0.21	0.20
	SE	NS	NS	0.13	NS	NS	0.09	0.06
$EBV_f - GEBVM_r$	Corr	NS	NS	0.36	0.38	NS	0.39	0.31
	SE	NS	NS	0.13	0.14	NS	0.09	0.05
$GEBV_f - GEBVM_r$	Corr	NS	0.56	0.33	0.43	0.38	0.54	0.44
	SE	NS	0.14	0.13	0.14	0.13	0.08	0.05
$GEBVM_f - EBV_r$	Corr	NS	NS	NS	0.37	NS	0.28	0.16
	SE	NS	NS	NS	0.14	NS	0.09	0.06
$GEBVM_f - GEBV_r$	Corr	NS	0.46	0.32	NS	NS	0.27	0.19
	SE	NS	0.15	0.13	NS	NS	0.09	0.06

¹EBV = pedigree-based estimated breeding value; GEBV = genomic breeding value; GEBVM = multiple across-country evaluation (MACE)-enhanced GEBV.

7). The correlation between EBV_f and $GEBVM_r$ for males were similar to their correlations between EBV_f and EBV_r , and $GEBV_f$ and $GEBV_r$. For females, the values were generally closer to the correlations between $GEBV_f$ and $GEBV_r$. These results also confirmed the previous finding that the combination of genomic and MACE information can be used as an equivalent substitute for scarce phenotypic information for males as well as females.

The higher correlation between $GEBV_f$ and $GEBVM_r$, especially for males born in 2018 and 2019, suggested that inclusion of foreign genomic information and MACE information from past evaluations contributed less to the accuracy obtained from the single-step genomic evaluation than the recent and own or progeny phenotypic information did when domestic genomic information was already included in the evaluation (Table 7). For females, the correlation between $GEBV_f$ and $GEBVM_r$ generally resembled the correlation between EBV_f and $GEBVM_r$, revealing a somewhat less significant effect of domestic genomic information. The same was true for males born in 2015 and 2016. The correlation between $GEBV_f$ and $GEBVM_r$ for these males was the same as the correlation with EBV_f and EBV_r , indicating that the contribution of foreign genomic information and MACE information from past evaluation to the accuracy of the single-step genomic evaluation is the same as that of recent own or progeny phenotypic information in the pedigree-based evaluation.

Domestic and foreign genomic information combined with MACE information in the single-step genomic evaluation contributes more to accuracy than recent own or progeny phenotypic information. This was suggested by the lower correlation between $GEBVM_f$ and EBV_r compared with EBV_f and EBV_r for both sexes (Table 7).

Recent and own phenotypic information, along with additional foreign genomic and MACE informa-

tion, had no major effect on breeding value accuracy of the females in the single-step genomic evaluation with domestic genomic information. The correlation between $GEBVM_f$ and $GEBV_r$ for females born in 2016 and later was slightly higher than their correlation between EBV_f and $GEBV_r$; however, it was equal or slightly lower for the older females (Table 7). This correlation was also slightly lower compared with the correlation between $GEBV_f$ and $GEBV_r$. For males born in 2018 and 2019, the correlation between $GEBVM_f$ and $GEBV_r$ was lower than the correlation between $GEBV_f$ and $GEBV_r$. It was slightly higher than the correlations between $GEBV_f$ and $GEBV_r$, and EBV_f and $GEBV_r$ for males born in 2015, and it was again slightly lower than the correlation between $GEBV_f$ and $GEBV_r$ for males born in 2014.

Breeding Value Inflation

Results indicated that the inclusion of genomic information alone, or combined with MACE information, increased inflation in the evaluation for validation females (Table 8). Table 8 shows the over-dispersion of breeding values from the reduced evaluation as well as from the pedigree-based evaluation. Biased breeding values can lead to biased selection decisions, especially when selection by truncation on breeding values is the primary method of animal selection. Inflation of the variance of the estimated breeding values, measured as a regression of breeding values from full evaluation on breeding values from reduced evaluation, indicates their over- as well as under-dispersion. When no inflation is present, the expected value of regression coefficient (b) is 1 (Mäntysaari et al., 2010; Legarra and Reverter, 2018). A value of $b < 1$ indicates under-dispersion of breeding values from the full evaluation and over-dispersion of breeding values from the reduced evaluation, and thus points to overestimation of future breeding values. By including more information, genomic evaluation can also be considered as a full evaluation and pedigree-based evaluation as a reduced evaluation. In genomic evaluation, $b < 1$ is commonly observed, and it can indicate under-dispersion of genomic breeding values along with over-dispersion of pedigree-based breeding values (Legarra and Reverter, 2018).

The over-dispersion of the breeding values from reduced evaluations was estimated within the pedigree-based and the single-step genomic evaluations with $b_{(GEBVM)_{f,r}} = 0.60$ to 0.69 , on account of the reduction of phenotypic information (Table 8). Regression coefficient was the same for the $EBV_{f,r}$ and the $GEBV_{f,r}$ (0.60), indicating similar over-dispersion of breeding values from the reduced evaluation in the pedigree-

Table 8. Dispersion of estimated breeding values from the full (f) and reduced (r) evaluation for validation females ($n = 1,041$)

Item	$b \pm SE$	$R^2 \pm SE$
$EBV_{f,r}$	0.60 ± 0.04	0.16 ± 0.02
$GEBV_{f,r}$	0.60 ± 0.04	0.15 ± 0.02
$GEBVM_{f,r}$	0.69 ± 0.03	0.35 ± 0.03
$GEBVM_{f,r}, EBV_f$	0.84 ± 0.01	0.81 ± 0.01
$EBV_f, GEBV_r$	0.59 ± 0.05	0.14 ± 0.02
$EBV_f, GEBVM_r$	0.54 ± 0.04	0.19 ± 0.02
$GEBV_f, EBV_f$	0.92 ± 0.01	0.89 ± 0.01
$GEBVM_f, GEBV_f$	0.91 ± 0.01	0.90 ± 0.01

EBV = pedigree-based estimated breeding value; $GEBV$ = genomic breeding value; $GEBVM$ = MACE-enhanced $GEBV$; b = regression coefficient; R^2 = coefficient of determination.

based as well as in the single-step genomic evaluation. The integration of MACE information reduced the over-dispersion in single-step genomic evaluation by 9 points. When all available information was included in the full single-step genomic evaluation, the GEBVM_f became under-dispersed compared with the EBV_f , which turned out to be over-dispersed ($b_{(\text{GEBVM}_f, \text{EBV}_f)} = 0.84$). Vandenplas et al. (2017) and Guarini et al. (2019), in contrast to our results, found that integration of MACE information in the single-step genomic evaluation for milk yield and milking traits reduces inflation of genomic breeding values and genomic breeding value reliabilities in bulls. However, the results are hard to compare as the regression analyses differ among studies.

Coefficient of determination representing validation reliability (R^2) indicated that phenotypic information itself or captured within MACE information explained most of the variance in all evaluations. When phenotypic information was omitted from the pedigree-based and the single-step genomic evaluation, the variance in the female pedigree-based breeding value was poorly predictable; only between 14 and 19%, as shown in the Table 8. When recent phenotypic information was included in the single-step genomic evaluation, 81% of the variance in $\text{GEBV}(\text{M})_f$ could be predicted from the pedigree-based evaluation (Table 8). The GEBVM also indirectly included foreign phenotypic information through MACE information that is not captured within the pedigree-based evaluation. This foreign information, combined with foreign genomic information, accounted for 19 percentage points of GEBVM_f variance [difference between 81% (GEBVM_f , EBV_f) and total EBV_f variance; Table 8] for validation females not captured by the EBV_f .

The $\text{GEBV}(\text{M})$ from the reduced evaluations were more over-dispersed compared with EBV_f ($b_{(\text{EBV}_f, \text{GEBV}(\text{M})_f)} = 0.54$ to 0.59). The GEBVM_f remained under-dispersed even when regressed on the GEBV_f ($b_{(\text{GEBVM}_f, \text{GEBV}_f)} = 0.91$; Table 8).

The information from the past MACE evaluation only explained an additional 5 percentage points of the variance in EBV_f . In GEBVM_r , the breeding value validation reliability remained higher even when phenotypic information was omitted due to MACE information (19%), compared with the GEBV_r (14%). The greater extent of the GEBV variance that was predictable from the pedigree-based breeding evaluation suggested that national genomic information alone only explained 11 percentage points of the GEBV_f variance for females. Foreign information combined with foreign genomic in-

formation explained an additional 10 percentage points of the GEBVM_f variance for validation females in the single-step genomic evaluation (Table 8).

Prospects

Positive REL changes brought by the national genomic evaluation will likely increase breeders' interests in continuous genotyping. The availability of national genomic information is expected to improve in the next few years. With a greater amount of genomic data over time, even higher reliability increases might be reachable. Further use of foreign genetic material and integration of increasing international information is also expected to additionally affect REL in both sexes and to have a significant effect on the youngest animal categories.

CONCLUSIONS

We performed a national genomic evaluation with and without foreign genomic and external MACE information. This study showed that combining domestic genomic information, foreign progeny-tested sire genomic information, and MACE information in single-step genomic evaluation can be used to offset the lack of national information. This approach of combining information increased the reliability of genomic breeding values for young and old individuals, with a larger effect on young females than on young males. In small dairy cattle populations with limited national genomic information and available international information, the single-step genomic evaluation method that combines information appears more suitable for genetic evaluation.

ACKNOWLEDGMENTS





The authors acknowledge the efforts of the InterGenomics consortium and the use of its results, and thank the Slovenian Brown Swiss cattle breeding organization (Ljubljana, Slovenia) and other InterGenomics partners for providing genomic data used in this study. The financial support for selection related activities of the Ministry of Agriculture, Forestry and Food of Slovenia (Ljubljana, Slovenia) and the Slovenian Research Agency (ARRS; Ljubljana, Slovenia) program group P4-0092 is also gratefully acknowledged. G. Gorjanc acknowledges support from the BBSRC to The Roslin Institute (BBS/E/D/30002275) and The University of Edinburgh's Data-Driven Innovation Chancellor's fellowship. The authors declare that there is no conflict of interest.

REFERENCES

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752. <https://doi.org/10.3168/jds.2009-2730>.
- Aguilar, I., I. Misztal, A. Legarra, and S. Tsuruta. 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* 128:422–428. <https://doi.org/10.1111/j.1439-0388.2010.00912.x>.
- Ali, T. E., and L. R. Schaeffer. 1987. Accounting for covariances among test day milk yields in dairy cows. *Can. J. Anim. Sci.* 67:637–644. <https://doi.org/10.4141/cjas87-067>.
- Andonov, S., D. A. L. Lourenco, B. O. Fragomeni, Y. Masuda, I. Pocrnic, S. Tsuruta, and I. Misztal. 2017. Accuracy of breeding values in small genotyped populations using different sources of external information—A simulation study. *J. Dairy Sci.* 100:395–401. <https://doi.org/10.3168/jds.2016-11335>.
- Boichard, D., V. Ducrocq, P. Croiseau, and S. Fritz. 2016. Genomic selection in domestic animals: Principles, applications and perspectives. *C. R. Biol.* 339:274–277. <https://doi.org/10.1016/j.crvi.2016.04.007>.
- Boison, S. A., A. T. H. Utsunomiya, D. J. A. Santos, H. H. R. Neves, R. Carvalheiro, G. Mészáros, Y. T. Utsunomiya, A. S. do Carmo, R. S. Verneque, M. A. Machado, J. C. C. Panetto, J. F. Garcia, J. Sölkner, and M. V. G. B. da Silva. 2017. Accuracy of genomic predictions in Gyr (*Bos indicus*) dairy cattle. *J. Dairy Sci.* 100:5479–5490. <https://doi.org/10.3168/jds.2016-11811>.
- Buchmann, R., and S. Hazelhurst. 2014. The “Genesis” Manual. Accessed Aug. 5, 2020. <http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf>.
- Calus, M.P.L. 2010. Genomic breeding value prediction: methods and procedures. *Animal* 4:157–164. <https://doi.org/10.1017/S1751731109991352>.
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2. <https://doi.org/10.1186/1297-9686-42-2>.
- Colinet, F. G., J. Vandenplas, P. Faux, S. Vanderick, R. Renaville, C. Bertozzi, X. Hubin, and N. Gengler. 2013. Walloon single-step genomic evaluation system integrating local and MACE EBV. *Interbull Bull.* 47:203–210.
- Colinet, F. G., J. Vandenplas, S. Vanderick, H. Hammami, R. R. Mota, A. Gillon, X. Hubin, C. Bertozzi, and N. Gengler. 2018. Bayesian single-step genomic evaluations combining local and foreign information in Walloon Holsteins. *Animal* 12:898–905. <https://doi.org/10.1017/S1751731117002324>.
- Edel, C., H. Schwarzenbacher, H. Hamann, S. Neuner, K. U. Gotz, and R. Emmerling. 2011. The German-Austrian genomic evaluation system for Fleckvieh (Siemmental) cattle. *Interbull Bull.* 44:152–156.
- Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* 1:54–75. <https://doi.org/10.1214/ss/1177013815>.
- Goddard, M. E., K. E. Kemper, I. M. MacLeod, A. J. Chamberlain, and B. J. Hayes. 2016. Genetics of complex traits: Prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc. Biol. Sci.* 283:20160569. <https://doi.org/10.1098/rspb.2016.0569>.
- Groeneveld, E., M. Kovač, and N. Mielenz. 2010. VCE User’s Guide and Reference Manual, version 6.0. Institute of Farm Animal Genetics.
- Guarini, A. R., D. A. L. Lourenco, L. F. Brito, M. Sargolzaei, C. F. Baes, F. Miglior, S. Tsuruta, I. Misztal, and F. S. Schenkel. 2019. Use of a single-step approach for integrating foreign information into national genomic evaluation in Holstein cattle. *J. Dairy Sci.* 102:8175–8183. <https://doi.org/10.3168/jds.2018-15819>.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443. <https://doi.org/10.3168/jds.2008-1646>.
- Jenko, J., G. R. Wiggans, T. A. Cooper, S. A. E. Eaglen, W. G. L. Luff, M. Bichard, R. Pong-Wong, and J. A. Woolliams. 2017. Cow genotyping strategies for genomic selection in a small dairy cattle population. *J. Dairy Sci.* 100:439–452. <https://doi.org/10.3168/jds.2016-11479>.
- Jorjani, H., J. Jakobsen, E. Hjerpe, and V. D. J. Palucci. 2012. Status of genomic evaluation in the Brown Swiss populations. *Interbull Bull.* 46:46–54.
- Jorjani, H., B. Zumbach, J. Dürr, and E. Santus. 2010. Joint genomic evaluation of BSW populations. *Interbull Bull.* 41:8–14.
- Lee, Y.-M., C.-G. Dang, M. Z. Alam, Y.-S. Kim, K.-H. Cho, K.-D. Park, and J. J. Kim. 2020. The effectiveness of genomic selection for milk production traits of Holstein dairy cattle. *Asian-Australas J. Anim. Sci.* 33:382–389. <https://doi.org/10.5713/ajas.19.0546>.
- Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single step, a general approach for genomic selection. *Livest. Sci.* 166:54–65. <https://doi.org/10.1016/j.livsci.2014.04.029>.
- Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel. Evol.* 50:53. <https://doi.org/10.1186/s12711-018-0426-6>.
- Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbandsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol.* 43:43. <https://doi.org/10.1186/1297-9686-43-43>.
- Lund, M. S., I. Van Den Berg, P. Ma, R. F. Brøndum, and G. Su. 2016. Review: How to improve genomic predictions in small dairy cattle populations. *Animal* 10:1042–1049. <https://doi.org/10.1017/S1751731115003031>.
- Mäntysaari, E., Z. Liu, and P. M. VanRaden. 2010. Interbull validation test for genomic evaluations. *Interbull Bull.* 41:17–22.
- Metta, M., S. Kanginakudru, N. Gudiseva, and J. Nagaraju. 2004. Genetic characterization of the Indian cattle breeds, Ongole and Deoni (*Bos indicus*), using microsatellite markers—a preliminary study. *BMC Genet.* 5:16. <https://doi.org/10.1186/1471-2156-5-16>.
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Misztal, I. 2013. BLUPF90 family of programs. Accessed Jan. 15, 2013. <http://nce.ads.uga.edu/wiki/doku.php>.
- Misztal, I., S. Tsuruta, I. Aguilar, A. Legarra, P. M. VanRaden, and T. J. Lawlor. 2013. Methods to approximate reliabilities in single-step genomic evaluation. *J. Dairy Sci.* 96:647–654. <https://doi.org/10.3168/jds.2012-5656>.
- Plieschke, L., C. Edel, E. C. G. Pimentel, R. Emmerling, J. Bennewitz, and K.-U. Götz. 2016. Systematic genotyping of groups of cows to improve genomic estimated breeding values of selection candidates. *Genet. Sel. Evol.* 48:73. <https://doi.org/10.1186/s12711-016-0250-9>.
- Pszczola, M., T. Strabel, H. A. Mulder, and M. P. L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J. Dairy Sci.* 95:389–400. <https://doi.org/10.3168/jds.2011-4338>.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575. <https://doi.org/10.1086/519795>.
- Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci.* 77:2671–2678. [https://doi.org/10.3168/jds.S0022-0302\(94\)77209-X](https://doi.org/10.3168/jds.S0022-0302(94)77209-X).
- Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218–223. <https://doi.org/10.1111/j.1439-0388.2006.00595.x>.
- Thomasen, J. R., C. Egger-Danner, A. Willam, B. Guldbandsen, M. S. Lund, and A. C. Sørensen. 2014. Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. *J. Dairy Sci.* 97:458–470. <https://doi.org/10.3168/jds.2013-6599>.

- Vandenplas, J., F. G. Colinet, and N. Gengler. 2014. Unified method to integrate and blend several, potentially related, sources of information for genetic evaluation. *Genet. Sel. Evol.* 46:59. <https://doi.org/10.1186/s12711-014-0059-3>.
- Vandenplas, J., F. G. Colinet, G. Glorieux, C. Bertozzi, and N. Gengler. 2015. Integration of external estimated breeding values and associated reliabilities using correlations among traits and effects. *J. Dairy Sci.* 98:9044–9050. <https://doi.org/10.3168/jds.2015-9894>.
- Vandenplas, J., and N. Gengler. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. *J. Dairy Sci.* 95:1513–1526. <https://doi.org/10.3168/jds.2011-4322>.
- Vandenplas, J., M. Spehar, K. Potocnik, N. Gengler, and G. Gorjanc. 2017. National single-step genomic method that integrates multi-national genomic information. *J. Dairy Sci.* 100:465–478. <https://doi.org/10.3168/jds.2016-11733>.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. *Invited review: Reliability of genomic predictions for North American Holstein bulls.* *J. Dairy Sci.* 92:16–24. <https://doi.org/10.3168/jds.2008-1514>.
- Whitlock, M., and D. Schluter. 2009. Correlation between numerical variables. Pages 431–448 in *The Analysis of Biological Data*. Roberts and Co. Publishers.
- Zhang, H., L. Yin, M. Wang, X. Yuan, and X. Liu. 2019. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. *Front. Genet.* 10:189. <https://doi.org/10.3389/fgene.2019.00189>.
- Zumbach, B., H. Jorjani, and J. Dürr. 2010. Brown Swiss genomic evaluation. *Interbull Bull.* 42:44–51.

ORCID

- B. Luštrek  <https://orcid.org/0000-0002-5523-5304>
 J. Vandenplas  <https://orcid.org/0000-0002-2554-072X>
 G. Gorjanc  <https://orcid.org/0000-0001-8008-2787>
 K. Potočnik  <https://orcid.org/0000-0003-0734-5733>