Hybrid One-Step Genomic Evaluation System for the Italian Simmental Breed

L. Degano[1*], G. Jansen[2*], R. Finocchiaro[3*], A. Rossoni[4*] and D. Vicario[1*]
[1*]ANAPRI: Via Ippolito Nievo 19, 33100 Udine (UD)-Italy; tel. +39 043222411; email: vicario@anapri.it
[2*] Dekoppel Consulting, Casale Rovera 10, Chiaverano(TO), Italy
[3*]ANAFI: Via Bergamo 292, 26100 Cremona (CR)-Italy
[4*]ANARB: 204, Localita' Ferlina, 37012 Bussolengo (VR)-Italy

Abstract

The One-step project is a collaboration between ANAFI, ANAPRI and ANARB whose objective is to develop a system of joint genetic-genomic evaluations for bulls, cows and young animals, for all traits. A hybrid one-step procedure was developed in which single trait animal model deregression of domestic cow and bull EBVs as well MACE EBV is followed by single-step genomic evalutation using the deregressed EBV as phenotypes. The system was initially built around SNP-BLUP evaluations, optionally including or excluding cows from the training set, as a transition step and for comparison purposes. There were not big differences between SNPBLUP and ONE-STEP for the productive traits in term of reliability gain, while the difference was larger for SCS and conformation traits. Increase of reliability of GEBV due to the inclusion of the cows in the reference population was limited in general. However, in the future, the inclusion of genotyped cows and genotyped foreign bulls with MACE evaluations is essential for breeds like Italian Simmental that have a relatively few genotypes and many foreign bulls with few or no daughters in Italy.

Key words: One-step, Italian Simmental

Introduction

The Italian ONE-STEP project involving ANAPRI (Italian Simmental), ANARB (Italian Brown) and ANAFI (Italian Holstein) started in 2014. The project aim was to develop a common system for genomic evaluations, capable of handling several trait groups (production, conformation, functional, etc.) taken into consideration in different breeds and analyzed with different models. Table 1 shows the main features of the actual genetic/genomic evaluation for production traits in the three breeds, demonstrating the variation in size of Test Day and pedigree data sets, genetic evaluation models and the number of available genotypes. Initially the project was aimed to develop full single-step genomic evaluations. The differences existing among these populations lead to choosing a hybrid solution using deregressed proofs instead of original phenotypes. The procedure developed can be applied to all populations despite the huge differences. Moreover, it produces all the necessary files for validation Interbull.

The aim of this work was to compare different approaches and the impact of the use of the cows in the genomic evaluation for productive traits, cell somatic count and conformation traits in a small population as Italian Simmental (64,554 cows in 5,163 herds on milk recording in 2015).

Table 1. Consistency of the data-sets used for genetic/genomic evaluation of production traits.

	Italian Simmental	Italian Brown	Italian Holstein
N° Test day	5,300,000	22,000,000	75,000,000
Pedigree	600,000	3,800,000	17,400,000
Genotypes	4,200	21,000	175,000
TD Model	Repeatability TDM	Repeatability TDM	Random Regression TDM
Current official genomic model	Domestic SNPBLUP	Intergenomics	Domestic SNPBLUP + poligenic effect

Material and Methods

The datasets used in this study included 286,311 EBV records for milk production traits, 212,223 for SCS (somatic cell score) and 143,652 for conformation final scores.

The traits considered were:

1. milk yield (h^2=0.193)
2. fat yield (h^2=0.129)
3. protein yield (h^2=0.170)
4. SCS (h^2=12.0%)
5. cow muscling (h^2=21.4%)
6. udder final score (h^2=23.5%).

Production traits were expressed in kg and the others as RBV with mean 100 and standard deviation 12.

Regarding productive traits and SCS (traits 1-4), two kind of EBVs have been considered for deregression, MACE for bulls, if available, and domestic EBVs for bulls with daughters in Italy and cows with phenotypes included in the traditional genetic evaluations.

For conformation traits (5-6), EBVs were from the joint European evaluation of the Simmental populations of Germany, Austria, Czech Republic and Italy. Those populations have been using a harmonized linear system also called "Fleckscore" for many years.

The joint pedigree had 581,521 animals. There were 4,226 edited genotype records (3083 males, 1143 females) from several Illumina SNP chips, mainly 54k. The genotypes came from our routine genotyping program, exchanges with other countries (Germany, Austria, Czech Republic, Switzerland) and research projects (table 2). All genotypes were imputed to a reference genotype including 40,200 selected SNP using the program Pedimpute (Nicolazzi *et al*., 2012).

Table 2. Sources of available genotypes.

Source	Males	Females
International exchange		
Research projects		
ANAPRI activity		

Deregression

In order to reduce computing resources and avoid possible problems due to uneven pedigree depth for genotyped animals (Legarra *et al*., 2014), the pedigree and EBV files were pruned to remove animals of scarce current interest. A starting list of all genotyped animals, young animals and active cows (alive or born in last 8 years with an EBV for at least one trait) was built and two generations of ancestors were added. Only EBV of animals in the pruned pedigree were retained. For milk yield, for example, the reduced population consisted of 124,124 EBVs and 232,668 animals in the pedigree.

Domestic cow and bull EBVs and MACE EBVs and their associated reliabilities were assumed to be expressed on the same scale and genetic base. They were deregressed by a multi-phase iterative procedure assuming an underlying reduced animal model for single records on a single trait. The underlying mixed model equations can be represented as

$$(\mathbf{Z'DZ} + \mathbf{A}^{-1}k)\{EBV_i\} = 2\mathbf{Z'Dy} \qquad (1)$$

where **D** is a diagonal matrix of unknown EDC, **A** is the traditional relationship matrix, k = $(4-h^2)/h^2$ and **y** is a vector of unkown equivalent daughter performance (EDP) values, and **Z** is an incidence matrix relating animals to EDP. Prior to deregression, all EBV were centered by subtracting the mean domestic bull EBV and after deregression the same mean was added to the EDP. Genetic groups were assigned by 5-year birth date groups separately by country of registration (ITA vs non-ITA) and sex of the animal with unknown parent(s).

In the first phase only domestic cow and bull EBV and REL were considered. Note that it was necessary to include domestic bull EBVs in addition to cow EBVs to avoid information loss due to the pruning described above. The diagonal elements D_{ii} (i.e EDC) were solved iteratively using a simplified version of the algorithm of Jamrozik *et al*. (2000), appropriate to a single trait, reduced animal model and computing a set of D_{ii} compatible with the official reliabilities. Only animals with input EBV were assigned non-zero EDC. During iteration any EDC < 3.5, corresponding roughly to a single own record with few contemporaries, was set to 0. Note that recently proven bulls whose daughters were all included in the analysis had 0 EDC because all the information content was assigned to the daughters. Given the set of EDC > 0, the elements of **y** (i.e. the EDP) were solved iteratively to obtain minimal discrepancies in the EBV in equation (1) for these animals. Input EBV were held constant and new EBV for all other animals in the pedigree were computed as a by-product of iteration process, including parent averages for all young animals.

In a second phase, domestic bull EBVs and corresponding reliabilities were replaced by MACE EBVs and reliabilities and MACE EBVs and reliabilities of any foreign bulls in the pedigree were also included. The process for step 1 was repeated but only the EDC and EDP of bulls with a MACE evaluation were updated each iteration, holding constant the EDC and EDP from step one for cows and bulls with only a domestic EBV.

During iteration in phases 1 and 2, some extreme EDP arose due to imperfect compatibility of input EBVs and RELs the assumed underlying model. EDP with extreme deviations from the PA, accounting for reliability of the EDP (EDC/(EDC+k)), were limited to ±3 sd.

In the third and final phase cow EDP were adjusted so as to achieve comparable within-year variance of mendelian sampling solutions between cows and bulls, adjusted for reliability of EDP, similar to the adjustment of Wiggans *et al*. (2012). Finally, holding all EDP and EDC constant, new EBV and reliability were recomputed for the entire pruned pedigree. This process effectively results in whole-population blending of domestic and MACE information. The recomputed EBV and REL were compared to the input EBV and REL by type of EBV (domestic cow and bull EBV or MACE EBV) and birth year and generally showed essentially no time trend differences and very high correlations (>0.97, but typically >0.99 for most trait x birth year combinations).

One-step genomic evaluation

Single-step GBLUP evaluations (Legarra *el al.*, 2014) were computed using model (1) with the addition of an overall mean as the sole fixed effect and with matrix $\mathbf{A}^{-1}$ replaced by matrix $\mathbf{H}^{-1}$ as in (Legarra *et al.*, 2014) with a polygenic component incorporated through $\mathbf{G}_w = 0.9\mathbf{G} + 0.1\mathbf{A}_{22}$.

The BLUPF90 suite of software of I. Misztal and collaborators was used to form the genomic relationship matrix with default scaling parameters and to solve the MME (http://nce.ads.uga.edu/wiki).

SNPBLUP genomic evaluation

For the purposes of comparison, DGV were also estimated with a SNPBLUP model followed by simple blending of the parent average (PA) from the final deregression step, according to relative EDC, to obtain GEBV. In this case the deregression was performed using a subset of the EBVs and pedigree EBVs tracing back from genotyped animals.

Validation

Validation was performed following the guidelines of the Interbull GEBV test (https://wiki.interbull.org/public/CoPAppendixVIII). For production traits and SCS, domestic EBV from reduced data were re-computed removing four years of TD records. For conformation traits the last four birth years of current cow and bull EBVs were removed and for all traits the last four birth years of current MACE EBV were removed. The deregression procedure was applied to this reduced set of EBV and REL and PA and GEBV were computed as described for the full data.

Approximation of reliability of GEBV

Average reliability of GEBV was computed for the validation group of bulls based on their average reliability of EDP, average reliability of PA from the reduced data and R^2 values for the regression of full data EDPs on GEBV or PA from the reduced data, as in VanRaden *et al.* (2009).

Given the limited number of genotypes available and considering that a substantial proportion of animals with genotypes and phenotypes (EPD) from the full data was missing in the reduced data, the average reliability of GEBV of young animals computed from the reduced data was extrapolated according to the increase in a simplified theoretical formula, $S/(S+M_e)$ where S is the sum of reliabilities of EDP for genotyped animals and M_e is an estimate of the number of independently segregating chromosome segments. Here M_e was set to 6000 to provide a reasonable match between theoretical and computed reliabilities in from the reduced data.

Inclusion or exclusion of cows

Four different scenarios have been considered (table 3).

Table 3. Scenarios for GEBV prediction.

Scenario	Model	Reference population
Scenario 1	SNPBLUP	Only Bulls
Scenario 2	SNPBLUP	Bulls & Cows
Scenario 3	ONE-STEP	Only Bulls
Scenario 4	ONE-STEP	Bulls & Cows

For SNPBLUP, cows were included or excluded from the reference population used to estimate SNP effects. For the ONE-STEP approach there is no clear separation between reference and prediction groups so it was necessary to completely remove the genotypes of cows from the computation of **G** to obtain the comparable results.

Results and Discussion

Descriptive statistics

In table 4 descriptive statistics of analyzed EBVs are reported.

Table 4. Descriptive statistics of EBV

	Mean±SD	Min	Max
Milk kg	-447.7±557.4	-2170	2248
Fat kg	-18.1±22.1	-70.0	69.0
Protein kg	-16.1±18.4	-69.3	66.6
SCS	95.8±10.0	50	148
Muscling	101.6±9.6	50	149
Udder	95.2±8.9	50	138

Validation

Concerning cross validation results (table 5 and 6), GEBV reliabilities were greater than PA reliabilities especially in production traits when cows were included in the reference population. This is much clearer when ONE-STEP is used to get GEBV, instead of SNPBLUP. In particular, with the ONE-STEP model including genotyped cows the reliability increased from 44.1% to 46.1% (+2.0%) for milk yield, from 45.5% to 51.7% (6.2%) for fat yield, from 44.9% to 51.9% (6.0%) for protein yield, and from 58.2% to 60.5% (+2.3%) for the SCS. For conformation traits, there was little difference in reliability due to the inclusion of cows in the reference population for muscling (from 73.0% to 74.5%; +1.5%), while for udder the difference was very high (from 60.0% to 70.9%; +10.9%). In muscling and udder the highest values of genomic reliability have been observed (74.5% and 70.9%, respectively); such gain of reliability due to genomic evaluation is very high considering the small dimension of Italian Simmental genotyped population.

Increase of reliability of GEBV due to the inclusion of the cows in the reference population was limited in general. However, in the future, increased availability of cow genotypes will enlarge and keep updated the reference population strongly increasing the quality of genomic evaluation.

Concerning b-values of the regression of EDP on GEBV, there are no large differences between SNPBLUP and ONE-STEP except for muscling where the b-values were less than 1 (0.79 and 0.80) using SNPBLUP whereas is more than 1 (1.11 and 1.12) using ONE-STEP. Including the cow genotypes hardly affected the b-values.

The R^2 values were larger for SCS, muscling and udder, while for the productive traits b-values and R^2 were low. R^2 values increased considerably for GEBV obtained using ONE-STEP rather than SNPBLUP. Undoubtedly, considerable care must be taken in interpreting these results, given the rather limited size of the validation group.

Comparison between GEBV

Table 7 shows some descriptive statistics of GEBV obtained in the different scenarios. Using SNPBLUP, GEBVs, obtained by blending PA and DGV, show a much lower variability than GEBVs from ONE-STEP. This strongly affects ranking of young genomic evaluated candidates, especially comparing them to domestic daughter proven bulls or to foreign genomic bulls.

Concerning correlations between GEBVs obtained through different models, table 8 reports the value for protein yield (values for milk and fat yield are very similar), SCS and the conformation traits. With SNPBLUP, very high correlations (r=0.99) between GEBVs obtained considering or not cows in reference population, have been found for all traits. In case of ONE-STEP, slightly lower correlations have been observed (r=0.96 for protein yield) due to the inclusion or exclusion of cow genotypes. Comparing SNPBLUP vs ONE-STEP the correlation is not high (from 0.81 to 0.83 for protein, from 0.87 to 0.90 for SCS, from 0.89 to 0.91 for muscling and from 0.86 to 0.89 for udder). In this case, that is the Italian Simmental population, with a small number of genotyped animals, moving from a SNPBLUP model to a ONE-STEP model, a strong re-ranking has been observed.

Conclusions

An efficient, flexible and user-friendly pipeline for computing GEBVs for different breeds and different traits has been developed for the main dairy breeds in Italy. This was feasible only by using a hybrid one-step approach based on deregressed EBV instead of true phenotypes.

In small-medium size populations as the Italian Simmental, with a low number of genotyped bulls and without a borderless genomic evaluation, inclusion of genotyped cows in one-step genomic evaluation is expected to give substantial advantage in terms of genomic reliability gain as the number of genotyped cows increases. This will be of increasing importance as number of daughter proven bulls declines due to the use of genomically selected young bulls.

At the application level, it is anticipated that official genomic evaluations in the Italian Simmental will shortly switch from the SNPBLUP to the ONE-STEP model. The increased scale of GEBVs is an important practical aspect, yielding GEBV that are more comparable to conventional EBVs. Validation passed the Interbull GEBV test for the traits evaluated at Interbull (production and SCS).

References

Legarra, A., Christensen, O.F., Aguilar, I. & Misztal, I. 2014. Single Step, a general approach for genomic selection. *Livest. Sci. 166*, 54-65.

Nicolazzi, E.L., Biffani, S. & Jansen, G. 2013. Short communication: imputing genotypes using PedImpute fast algorithm combining pedigree and population information. *J Dairy Sci. 96*, 2649-2653.

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

Wiggans, G.R., VanRaden, P.M. & Cooper, T.A. 2012. Technical note: Adjustment of all cow evaluations for yield traits to be comparable with bull evaluations. *J. Dairy Sci. 95*, 3444-3447.

Acknowledgements

We are grateful to ASR, AGOF, CESTR and SWISS HERDBOOK for providing bulls genotypes and to Ignacy Misztal for providing the publicly available BLUPF90 software used in this study.

Table 5. Validation parameters for GEBV calculated with SNPBLUP.

Reference population: bulls; Model: SNPBLUP										
	N° Animals		h^2	Reliability			EDP/GEB		EDP/PA	
	Reference	Validatio		PA	GEB	Gain	b	R^2	b	R^2
Milk kg	1644	121	19.3	38.3	46.7	8.4	0.67	9.1	0.48	6.0
Fat kg	1646	121	12.9	38.0	49.3	11.3	0.83	16.8	0.64	10.9
Protein	1645	121	17.0	37.9	51.9	14.0	0.86	14.8	0.57	8.0
SCS	1700	120	12.0	36.4	39.2	2.8	1.10	30.0	1.07	29.5
Musclin	1520	95	21.4	37.2	54.0	16.8	0.79	36.8	0.86	39.5
Udder	1523	94	23.5	37.6	65.4	27.8	1.14	42.5	1.45	0.49

Reference population: bulls and cows; Model: SNPBLUP										
	N° Animals		h^2	Reliability			EDP/GEB		EDP/PA	
	Reference	Validatio		PA	GEB	Gain	b	R^2	b	R^2
Milk kg	2170	121	19.3	38.3	48.8	10.5	0.64	8.7	0.48	6.0
Fat kg	2172	121	12.9	38.0	50.9	12.9	0.82	16.4	0.64	10.9
Protein	2171	121	17.0	37.9	52.2	14.3	0.81	13.6	0.57	8.0
SCS	2253	120	12.0	36.4	40.9	4.5	1.17	30.5	1.07	29.5
Musclin	1929	95	21.4	37.2	54.4	17.2	0.80	36.7	0.86	39.5
Udder	1932	94	23.5	37.6	65.5	27.9	1.19	42.7	1.45	0.49

Table 6. Validation parameters for GEBV calculated with ONE-STEP.

Reference population: bulls; Model: ONE-STEP										
	N° Animals		h^2	Reliability			EDP/GEB		EDP/PA	
	Reference	Validatio		PA	GEB	Gain	b	R^2	b	R^2
Milk kg	1655	116	19.3	38.3	44.1	5.8	0.67	15.6	0.64	11.0
Fat kg	1658	116	12.9	38.0	45.5	7.5	0.85	24.9	0.89	19.0
Protein	1655	116	17.0	37.9	44.9	7.0	0.77	23.1	0.82	17.5
SCS	1714	116	12.0	36.5	58.2	21.7	1.20	46.5	1.24	37.0
Musclin	1518	92	21.4	37.5	73.0	35.5	1.11	55.0	1.03	43.8
Udder	1521	92	23.5	37.8	60.0	22.2	1.14	46.0	1.37	0.52

Reference population: bulls and cows; Model: ONE-STEP										
	N° Animals		h^2	Reliability			EDP/GEB		EDP/PA	
	Reference	Validatio		PA	GEB	Gain	b	R^2	b	R^2
Milk kg	2179	116	19.3	38.3	46.1	7.8	0.62	15.4	0.63	11.1
Fat kg	2182	116	12.9	38.0	51.7	13.7	0.79	26.4	0.78	16.6
Protein	2179	116	17.0	37.9	51.9	14.0	0.77	25.6	0.75	15.4
SCS	2265	116	12.0	36.5	60.5	24.0	1.22	48.0	1.23	36.8
Musclin	1926	92	21.4	37.5	74.5	37.0	1.12	56.1	1.03	44.3
Udder	1929	92	23.5	37.8	70.9	33.1	1.26	54.4	1.47	56.0

Table 7. Descriptive statistics of GEBV.

Model	Reference	Milk kg	Fat kg	Protein kg	SCS	Muscling	Udder
SNPBLU	Only bulls	340.1±222.	12.0±9.0	11.7±7.4	102.5±6.4	98.3±7.8	107.7±6.8
SNPBLU	Bulls &	355.3±224.	12.6±8.8	12.5±7.5	102.6±6.1	98.5±7.6	107.4±6.4
ONESTE	Only bulls	439.4±332.	15.2±12.	15.2±10.5	102.9±8.2	99.0±8.1	106.1±7.2
ONESTE	Bulls &	430.1±337.	15.0±13.	14.8±10.6	102.8±8.4	99.0±8.2	106.0±7.2

Table 8. Correlation between GEBV obtained from the different scenarios considered.

Trait	Model	Reference	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Protein	SNPBLU	Only Bulls	1	0.99	0.83	0.81
	SNPBLU	Bulls &		1	0.82	0.81
	ONE-	Only Bulls			1	0.96
	ONE-	Bulls &				1
SCS	SNPBLU	Only Bulls	1	0.99	0.90	0.87
	SNPBLU	Bulls &		1	0.89	0.88
	ONE-	Only Bulls			1	0.97
	ONE-	Bulls &				1
Muscling	SNPBLU	Only Bulls	1	0.99	0.91	0.91
	SNPBLU	Bulls &		1	0.90	0.89
	ONE-	Only Bulls			1	0.99
	ONE-	Bulls &				1
Udder	SNPBLU	Only Bulls	1	0.99	0.89	0.88
	SNPBLU	Bulls &		1	0.87	0.86
	ONE-	Only Bulls			1	0.99
	ONE-STEP	Bulls & Cows				1