Testing different genomic selection scenarios in a small cattle population by simulation

J. Obšteter[1], J. Jenko[2], J. M. Hickey[2] & G. Gorjanc[2]

[1] Agricultural Institute of Slovenia, Department of Animal Science, Hacquetova ulica 17, 1000 Ljubljana, Slovenia
jana.obsteter@kis.si (Corresponding Author)
[2] The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Midlothian, EH259RG, United Kingdom

Introduction

Selection aims to drive genetic improvement of a population. It involves selecting genetically superior individuals for economically important traits as the parents of the next generation. The identification of these individuals can be challenging due to many sources of variation: i) unknown genetic "architecture" for the traits; ii) substantial environmental effects; iii) prevalent sex-limited-expression of traits; iv) recombination and segregation of parents and other ancestors genomes; v) required cooperation of technical services, scientists, and breeders; and vi) limited financial resources. For conventional selection, the selection criteria are estimated breeding values (EBVs), which require own or progeny phenotypes on selection candidates for accurate selection. Consequently, they reduce to the average of the parents' EBVs (parent average, PA) prior to the collection of phenotypes.

The sex-limited-expression of phenotypes has led to multi-stage selection in dairy cattle breeding, which can be outlined in four steps: i) PA-based selection of male calves for performance testing; ii) EBV-based selection of the best young bulls for progeny testing (based on some own performance) iii) EBV-based selection of the best progeny tested bulls for wide-spread use in population; and iv) EBV-based selection of the best dams for insemination with elite bulls to generate a new generation of selection candidates for i). Since the accuracy, intensity, and generation interval differ substantially between selection of female and male parents, Rendel and Robertson (1950) defined four selection paths: dams of dams, dams of sires, sires of dams, and sires of sires.

The collection of genomic data has enabled the implementation of genomic selection in cattle breeding (Meuwissen et al., 2001). It enables estimation of genomic breeding values (gEBVs) for all genotyped animals, even young animals without phenotypes. This provides an opportunity for cattle breeding by reducing the generation interval and increasing the accuracy of early selection decisions. A large reference population of genotyped and phenotyped individuals is required for accurate genomic prediction. To achieve acceptable accuracies, small populations have combined their animals across breeds or across countries to create an international reference population. However, in small populations, the benefit of genomic selection is constrained by limited resources, technical difficulties, such as inadequate number of own individuals for the reference population and distrust among breeders and producers regarding the new technology. These raise a crucial question regarding the optimal use of resources and obtained genomic information to maximize return on investment. The answer to this question is further blurred by many sources of variations in quantitative traits.

Previous studies have mainly explored the effect of different parameters on the accuracy of genomic prediction by considering simplified or generic cattle populations with

discrete generations. Here we present a simulator that can be used to evaluate strategies and the extent to which genomic information should be used in a breeding program. The simulator can model realistic cattle populations with overlapping generations and involves stochastic simulation of the genome, individuals and associated data, estimation of breeding values, selection, and mating. The simulator was used to model the Slovenian Brown Swiss population and to test different scenarios of applying genomic selection in the male selection paths.

Material and methods

Simulation of a cattle breeding program

A simulator of dairy cattle breeding programs was developed. The simulator is driven by user-defined parameters to perform all breeding steps. The simulator was written as a Python wrapper around the AlphaSim (Faux et al., 2016) and blupf90 software (Misztal et al., 2002). In this study, we used this simulator to model a small cattle population of ~30,000 active individuals, ~10,000 of which cows. Initially, we generated sequence data for 10 chromosomes, from which 20,000 SNPs were used as causal loci affecting a trait with a heritability of 0.25. Then, we initiated a dairy cattle breeding program by first randomly mating the population for 20 generations to obtain animals to model overlapping generations. This was followed by 20 generations of burn-in selection with conventional pedigree-based EBVs and then another 20 generations for each of the designed selection scenarios. Breeding values were estimated using conventional pedigree-based BLUP or single-step genomic BLUP (Legarra et al., 2009).

Sire selection scenarios

We tested conventional scenario and four genomic selection scenarios that differed in the use of genomic information in the male selection paths (Table 1). The genomic information was used either for the selection of bulls for progeny testing or for breeding in artificial insemination directly (PT* and GT in Table 1, respectively). For genomic selection, a reference population consisting of ~11,000 cows and 100 progeny tested bulls was assumed. The reference population was updated each generation by replacing the oldest 2000 cows with a random sample of females from the current cow population. In each selection cycle, all male selection candidates for sires were genotyped and added to the reference population.

Table 1. Sire selection criterion by scenario.

	Conventional scenario	**Genomic selection scenarios**			
		A	**B**	**C**	**D**
Sires of sires	PT[1] bulls	PT*[2] bulls	PT[1] bulls	GT[3] bulls	GT[3] bulls
Sires of dams	PT[1] bulls	PT*[2] bulls	GT[3] bulls	PT[1] bulls	GT[3] bulls

[1]PT = progeny tested
[2]PT* = young bulls for progeny testing selected based on genomic breeding values
[3]GT = genomically tested

Sire use strategies

The five sire selection scenarios were tested within three different strategies of sire use. In the original strategy we selected five bulls each year and kept them in use for five years (SU 5/5). In the next strategy we increased the selection intensity by selection only one bull per year

instead of five and kept him in use for five years (SU 1/5). In the last strategy we selected five bulls each year and kept them in use for five years (SU 5/1) which decreased generation interval even further.

Metrics of comparison

The scenarios were compared in terms of genetic gain per year (expressed with mean zero and in units of genetic standard deviation in the first generation of comparison) and efficiency. We defined the efficiency of selection as a regression of the achieved genetic gain on the amount of reduction of the genic standard deviation. We used the genic rather than the genetic standard deviation due to large fluctuations in the latter. When computing efficiency, both genetic gain and the genic standard deviation were standardized by the genic standard deviation in the first generation of comparison. Therefore, this efficiency metric indicates the potential genetic gain in units of genic standard deviation when all variation is converted into gain or lost due to drift. Results are presented as the mean of ten replicates for each scenario on a per generation or cumulative basis.

Results and Discussion

Genetic Gain

In Table 2 we present genetic gains for the tested scenarios within the three sire use strategies. All the values in the table are relative to the conventional scenario of the SU 5/5 strategy. Conventional scenario had the lowest genetic gain in all three sire use strategies. The results show that in this particular population we can achieve a substantial increase in genetic gain, i.e. between 35 and 46%, by using genomic information for a pre-selection step of male calves for progeny testing only (Genomic A). We observe that in all sire use strategies the genetic gain increases with the increasing use of genomic information which is in concordance with the reduction in generation interval. Using genomically tested bulls as sires of dams or sires of sires increases the genetic gain by 50 - 73% and 62 - 69%, respectively. The most comprehensive genomic scenario increased genetic gain by 88 - 123%. Increasing the intensity of selection (SU 1/5) or reducing the generation interval even further (SU 5/1) increased the genetic gain. However, the observed increases were only up to 27% in the SU 1/5 strategy and up to 35% in the SU 5/1 strategy.

Table 2. Genetic gain by sire selection scenario and sire use strategies averaged across ten replicates.

Sire selection strategies: conventional = only progeny tested bulls (no genomic information), gebomic A = genomic information used for selection of young bulls for progeny testing, genomic B / C / D = genomically tested bulls used respectively as sires of dams / as sires of sires / as sires of dams and sires of sires.

Sire use strategies: SU 5/5 = five bulls selected each year and used for five years, SU 1/5 = one bull selected each year and used for five years, SU 5/1 = five bulls selected each year and used for one year.

		Sire use strategies		
		SU 5/5	**SU 1/5**	**SU 5/1**
Sire selection strategy	**Conventional**	100%	131%	100%
	Genomic A	135%	146%	135%
	Genomic B	154%	150%	173%
	Genomic C	162%	162%	169%
	Genomic D	188%	215%	223%

Efficiency

Selection efficiencies are presented in Figure 1, which shows the evolution of the scenarios in terms of genetic gain and reductions in the genic standard deviation. In the SU 5/5 strategy all the genomic selection scenarios were between 14 and 38% more efficient than the conventional scenario. The most efficient were genomic scenarios A and C (38% increase) and the most comprehensive genomic selection scenario was not the most efficient one (14% increase). We attribute the differences in efficiency between genomic selection scenarios A, B and C versus D to greater accuracy of EBV with progeny testing and therefore greater spread of EBVs in the sires of sires. Sires of sires are under intense selection and have a major impact on the population. Thus, accurately estimating Mendelian sampling terms for this group of animals is important for maximizing conversion of standing genetic variation into genetic gain, which is in line with optimal contribution theory (Woolliams et al., 2015).

Although increasing the intensity of selection (SU 1/5) increased the genetic gain, it reduced the selection efficiencies by 35 – 50% compared to the corresponding scenarios in SU 5/5. We also see that the most comprehensive genomic scenario (Genomic D) became the least efficient one. The decrease in efficiencies in due to a larger loss in genetic variance caused by using only one sire. We can remedy this by selecting five bulls again in SU 5/1 strategy. The differences between the scenarios are smaller than in SU 1/5 and more similar to the ones in 5/5 strategy. However, the selection efficiencies are still decreased compared to the SU 5/5 strategy. This is due to a lower accuracies of the prediction in SU 5/1 strategy (0.1 unit drop in accuracy) attributable to a larger genetic distance between the training and testing population.

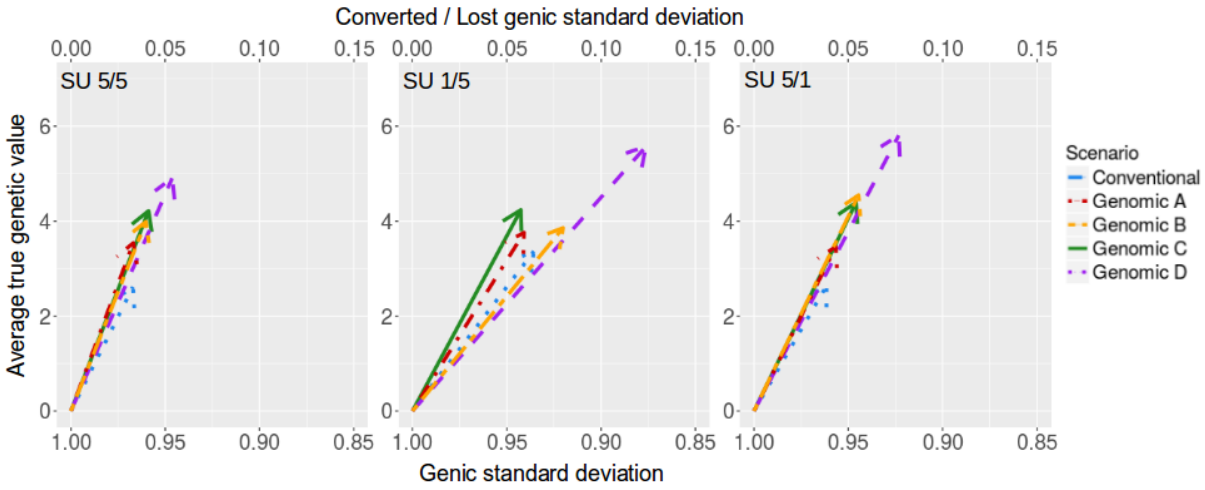

Figure 1: Change of genetic mean and genic standard deviation over the 20 years of selection by sire selection scenario within sire use strategies averaged across ten replicates.
Arrows point the direction of change. Sire selection strategies: conventional = only progeny tested bulls (no genomic information), genomic A = genomic information used for selection of young bulls for progeny testing, genomic B / C / D = genomically tested bulls used respectively as sires of dams / as sires of sires / as sires of dams and sires of sires.
Sire use strategies: SU 5/5 = five bulls selected each year and used for five years, SU 1/5 = one bull selected each year and used for five years, SU 5/1 = five bulls selected each year and used for one year.

Conclusion

To conclude, the developed simulator enables comparison of breeding scenarios using a model of a realistic cattle population. It can help breeding organisations to find the optimum strategy for using genomic information with the given resources.

List of References

Faux, A.-M., Gorjanc, G., Gaynor, R.C., Battagin, M., Edwards, S.M., Wilson, D.L., Hearne, S.J., Gonen, S., Hickey, J.M., 2016. AlphaSim: Software for Breeding Program Simulation. Plant Genome 9(3).

Legarra, A., Aguilar, I., Misztal, I., 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92(9): 4656–4663.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157(4): 1819–1829.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D.H., 2002. BLUPF90 and related programs (BGF90), in: Proc. 7th World Congress on Genetics Applied to Livestock Production. Presented at the WCGALP, Montpellier, France, pp. 1–2.

Rendel, J.M., Robertson, A., 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. J. Genet. 50(1): 1–8.

Thomasen, J.R., Egger-Danner, C., Willam, A., Guldbrandtsen, B., Lund, M.S., Sørensen, A.C., 2014. Genomic selection strategies in a small dairy cattle population evaluated for genetic gain and profit. J. Dairy Sci. 97(1): 458–470.

Woolliams, J.A., Berg, P., Dagnachew, B.S., Meuwissen, T.H.E., 2015. Genetic contributions and their optimization. J. Anim. Breed. Genet. 132(2): 89–99.