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

J. Obšteter¹, J. Jenko², J. M. Hickey² & G. Gorjanc²

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

Summary

Genomic selection offers the opportunity to increase genetic gain by reducing generation intervals and increasing accuracy of early selection decisions. However, the success of selection depends also on the extent and strategy of using genomic information in the breeding program. In small cattle populations, the use of genomic information is constrained due to limited resources and uncertainty about the technology. These limitations raise questions such as how many and which animals to genotype, in which selection paths to use the genomic information, and to what extent. Previous studies mainly explored the effect of parameters that influence the accuracy of genomic prediction on genetic gain by simulating generic cattle populations with discrete generations. Here we developed a simulator of realistic cattle populations with overlapping generations and selection. The simulated population reflected the Slovenian Brown Swiss population and the tested scenarios included conventional and four genomic selection scenarios. The genomic selection scenarios differed in the strategy and extent with which genomic information was used in the male selection paths. Scenarios were compared by genetic gain, generation interval, and efficiency of selection (defined as conversion of genetic variance into genetic gain). The scenario that used genomic information for pre-selection of young bulls for progeny testing resulted in 35% higher genetic gain per year than the conventional scenario. Selecting sires of dams and sires of sires based on genomic predictions increased genetic gain between 58 and 86% as compared to the conventional scenario. These increases are due to greater accuracies of early selection decisions and/or reduced generation intervals. The conventional scenario was the least efficient and the two less comprehensive genomic selection scenarios were 27% more efficient as the conventional scenario. While, the two most comprehensive genomic selection scenarios had the highest genetic gain, they were less efficient (between 16 and 24%) than the two less comprehensive genomic selection scenarios. The developed simulator will be used as a modelling tool for answering further questions regarding genomic selection in specific populations, such as the Slovenian Brown Swiss.

Keywords: genomic selection, simulation, breeding scenarios, small cattle population
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 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. An example is the InterGenomics project, which provides a joint genomic evaluation for Brown Swiss animals for participating countries. 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 software (Faux et al., 2016). In this study, we used this simulator to model the Slovenian Brown Swiss population. 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. In each generation, the following steps were performed: i) estimation of breeding values, ii) selection, and iii) mating of parents to simulate a new generation of individuals. Breeding values were estimated using blupf90 (Misztal et al., 2002) using conventional pedigree-based BLUP or single-step genomic BLUP (Legarra et al., 2009).

Scenarios

A conventional scenario and four genomic selection scenarios that differed in the use of genomic information were compared. We focused only on testing the use of genomics in the male selection paths (Table 1), while female selection paths were as in the conventional scenario. 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, which reflects the InterGenomics evaluation system for the breed. 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 criteria by scenario.

<table>
<thead>
<tr>
<th>Selection path</th>
<th>Conventional scenario</th>
<th>Genomic selection scenarios</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sires of sires</td>
<td>PT¹ bulls</td>
<td>PT*² bulls</td>
<td>PT¹ bulls</td>
<td>GT³ bulls</td>
<td>GT³ bulls</td>
<td></td>
</tr>
<tr>
<td>Sires of dams</td>
<td>PT¹ bulls</td>
<td>PT*² bulls</td>
<td>GT³ bulls</td>
<td>PT¹ bulls</td>
<td>GT³ bulls</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>¹PT = progeny tested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>²PT* = young bulls for progeny testing selected based on genomic breeding values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>³GT = genomically tested</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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), generation
interval, and efficiency. We defined 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 and Generation Intervals

Genetic gains and generation intervals for the tested scenarios are presented in Figure 1. The conventional scenario had the lowest genetic gain per year, i.e. 0.133 standard deviations (SD). Although the generation interval was not reduced in genomic selection scenario A compared to the conventional scenario, it resulted in a 35% (0.179 SD vs. 0.133 SD) higher genetic gain per year. The reason for this was greater accuracy of early selection decisions, since young bulls for progeny testing were selected based on genomic EBV instead of PA. The increase in genetic gain per year for genomic selection scenarios B, C, and D was proportional to the reduction in generation interval. Compared to the conventional scenario, the increase in genetic gain per year was 58% (0.210 SD vs. 0.133 SD) in genomic selection scenario B, 67% (0.222 SD vs. 0.133 SD) in genomic selection scenario C, and 86% (0.247 SD vs. 0.133 SD) in genomic selection scenario D. Since the reduction in generation interval in the sires of sires path was higher than the reduction in the sires of dams path, scenario C had higher genetic gain than scenario B. Scenario D resulted in the highest genetic gain since it used exclusively genomically tested bulls and had the largest reduction in generation intervals. All genomic selection scenarios also benefited from a higher accuracy of early selection decisions.
Figure 1. Genetic gain (a) and generation intervals (b) by scenario averaged across ten replicates.

Genomic selection scenarios: A = genomic information used for selection of young bulls for progeny testing, B / C / D = genomically tested bulls used respectively as sires of dams / as sires of sires / as sires of dams and sires of sires.

**Efficiency**

Genomic selection scenarios were more efficient than the conventional scenario. The most comprehensive genomic selection scenarios, which had the highest genetic gain (C and D), were less efficient than the less comprehensive genomic selection scenarios (A and B). This is presented in Figure 2, which shows the evolution of the scenarios in terms of genetic gain and reductions in the genic standard deviation. The conventional scenario had an efficiency of 18.9 units of genetic gain per unit loss of genic standard deviation. Genomic selection scenario C was 24% (23.5 vs. 18.9) more efficient than the conventional one and genomic selection scenario D 16% (21.9 vs. 18.9) higher efficiency than the conventional scenario. We attribute the differences in efficiency between genomic selection scenarios A and B versus C and 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).
Figure 2: Change of genetic mean and genic standard deviation over the 20 years of selection by scenario averaged across ten replicates. Arrows point the direction of change. Genomic selection scenarios: A = genomic information used for selection of young bulls for progeny testing, B / C / D = genomically tested bulls used respectively as sires of dams / as sires of sires / as sires of dams and sires of sires.

**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**