**Comparison of different genomic selection scenarios in a small cattle population by simulation**

*J. Obšteter\*, J. Jenko\*,†*, *J. M. Hickey†* & *G. Gorjanc ‡,†*

*\* Department of Animal Science, Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia*

*‡ Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia*

*† The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH259RG, United Kingdom*

*Jana Obšteter, Agricultural Institute of Slovenia, Department of Animal Science, Hacquetova ulica 17, 1000 Ljubljana, Slovenia*

*+386 1 280 51 34*

***jana.obsteter[@kis.si](mailto:Riberc@univit.com" \l "_blank) (Corresponding Author)***

# Introduction

This paper compares genetic gain, inbreeding, and efficiency of converting genetic variation into gain under different genomic selection scenarios in a small cattle population by simulation. Genomic selection has profoundly changed dairy breeding programs (Schaeffer, 2006; Garcia-Ruiz et al., 2016; Wiggans et al., 2017). Integration of genome-wide marker data into the traditional breeding programs has doubled the rate of genetic progress through a combination of decreased generation interval, increased selection accuracy for young animals, reduced costs of testing, and identification of recessive lethal alleles (Garcia-Ruiz et al., 2016; Wiggans et al., 2017). However, effective implementation of genomic selection requires a sizeable training population of genotyped and phenotyped animals, which is a limiting issue for small populations. TODO

[https://www.frontiersin.org/articles/10.3389/fgene.2018.00251/abstract](https://www.frontiersin.org/articles/10.3389/fgene.2018.00251/abstract" \l "_blank)

The latter diminishes the gain in accuracy of genomic prediction compared to parent average (PA) estimates what in turn decreases the advantage of using genomic information (the advantage of genomic over conventional selection) (Thomasen et al,. 2014).Alternatively small cattle populations can participate in international associations for a joint genomic prediction with international reference population or they can include cows in the reference population. However, an information of a cow in the reference is not equal to the information of a PT bull, therefore a larger number of cows has to be included to achieve desired accuracies (de Roos, 2011).

As mentioned, GS can also reduce the generation interval. The degree of reduction depends on the strategy of using the genomic information in a breeding program: in which selection paths to use the genomic information and to what extent. Previous studies mainly tested the scenarios in which the genomic information is used either for a pre-selection of young bulls for progeny testing (so called GS-PS) or for the selection of genomically tested young bulls directly (so called turbo scheme) against the conventional scenario with progeny tested sires without the use of genomic information (Pryce et al., 2010; Lillehammer et al., 2011; de Roos et al., 2011). The studies observed that we can increase the genetic gain up to 30% by using genomic information for a pre-selection step and up to 195% by using genomic information directly for selection of young bulls and dams as parents (M=100%!!!, realistic – 40% → up to 108%).

Further on, Thomasen et al., 2014, deterministically explored hybrid schemes that simultaneously use both progeny tested and young genomically tested bulls which changes the generation interval to an intermediate degree. The study varied the proportion of bull dams and cows mated with young / progeny tested sires and inspected the resulting annual monetary genetic gain (AMGG). They concluded that even in small populations, the genomic scenario was economically and genetically superior to the conventional scenario that uses progeny testes sires. However, with low reliabilities the hybrid scheme brought the higher AMGG whereas with high reliabilities the turbo scheme was the most superior (Thomasen et al., 2014).

In addition to maximizing the genetic gain, breeding schemes have to manage genetic variation to assure sustainability of selection. The aim is therefore to balance short- and long-term success of selection. While short-term success depends only on the genetic gain in the next (few) generation, long-term success enables increase of genetic gain after many generations of selection.

Some studies also focused onto how the use of genomic information affects inbreeding and obtained contradictory results. While some studies observed a decrease in inbreeding rate per year with the use of genomic information (Pryce et al., 2010; Lillehammer et al., 2011), the others observed and increase (de Roos et al., 2011). Although reducing the generation interval may increase annual inbreeding rate, managing inbreeding rate per generation is more important for avoiding inbreeding depression (Daetwyler et al., 2007). The reason for a lower rate of inbreeding with GS lies in a more accurate estimation of Mendelian sampling term – this results in lower co-selection of siblings that reduces inbreeding rate per generation.

However, a combine measure of genetic gain and genetic variability is needed to assure long-term competitiveness of selection. Efficiency of converting genetic variance into genetic gain measures how successful a specific scheme is in achieving both goals (Gorjanc et al., 2017).

Although GS is a well-established technology globally, small populations still struggle with the adoption. Contrary to other parameters of GS, less work was done exploring how the use of genomic information in a breeding program affects genetic gain. Although the existed studies explored the genetic gain and / or inbreeding, questions remain about the optimal use of genomic information in small realistic cattle populations with overlapping generation that allow for a long-term success of selection. (LONG TERM, DETERMIISTIC; DISCRETE GENERATION, LARGE POPULATION). Also, studies mainly explored the use of genomic information for a pre‑selection step or in a turbo scheme, but less work was done regarding the intermediate strategies. This study aimed to explore different strategies of the use of genomic information in small realistic cattle populations in order to determine the best one in terms of efficiency (the most efficient one).

# Material and Methods

First we developed a simulator of a realistic cattle population under selection which included all selection steps and allowed for the user to define all the selection parameters. We compared five sire selection scenarios in which we varied the criterion for the selection of sires for insemination of bull dams and cows. Additionally we tested three sire use strategies with varying number of bulls chosen each selection cycle and years kept in use. In contrast to these truncation selection scenarios we tested five OCS scenarios with varying target degrees between genetic gain and group coancestry. All tested scenarios were compared based on the achieved genetic gain, change (loss?) of genetic variance, inbreeding and efficiency of selection.

## Cattle breeding program simulator

We developed a simulator of a realistic cattle breeding program. The simulator was written as a Python wrapper around AlphaSim (Faux et al., 2016) and blupf90 (Misztal et al., 2002) software. The simulator enabled for the user to set all the selection parameters: the percentage of animals selected at each stage in each selection path, age at selection, selection criterion for each selection path, number of offspring per parent, years in use and number of selection cycles. The simulation was a continuous process resulting in overlapping generation. In each selection cycle three steps were performed: estimation of EBVs, selection of parents and mating. Breeding values were estimated as described below. Selection was performed based on single trait BVs.

## Simulated population

Genome included 108 base pairs arranged in ten chromosomes. We create two SNP‑chips, each including 20,000 distinct SNPs, from which one was used for selection (marker loci) and the other for the monitoring of inbreeding (neutral loci).From the segregating sites 10,000 were chosen at QTNs for a polygenic trait with heritability of 0.25. None of the QTNs were included on the SNP‑chips. The effects of the QTNs for each chromosome were drawn from a normal distribution. The simulated population mirrored Slovenian Brown-Swiss population participating in milk recording consisting of ~30,000 active individuals, ~10,000 of which cows. In total 60 generations were simulated with 8640 animals born each selection cycle. First 20 generations represented the burn-in population in which random mating was implemented. This was followed by 20 generations of conventional selection using PT bulls to achieve a population structure resembling a cattle population under selection. The last 20 generations implemented testing scenarios that differed in the use of genomic information in the male selection paths only.

## Selection of females

The selection of females was either random (heifers) or base on EBVs (cows). Out of newborn females 1% was removed in their first year reflecting stillbirths and deaths occurring soon after birth. Out of these female calves 90% were inseminated in the second year and became cows in the third year. In each subsequent lactation we removed 20% of the cows and after the fourth lactation all cows were culled. This totaled to 10,653 active cows that were all screened for the trait and had their EBV estimated. In each generation 8550 cows with highest EBVs were chosen for mating. After the second completed lactation and estimation of EBVs, 43 cows with highest EBVs were chosen as bull dams. Bull dams were kept in use for three lactations, completing five lactations in total. From 129 bull dams 90 were chosen each year for contracted mated with sires of sires to produce new generation of selection candidates.

## Selection of sires

Each year we obtained 45 male offspring from contracted matings. Sires were selected based on three overall strategies: a) progeny testing with pre-selection based on PAs (PT); b) progeny testing with pre-selection based on gEBVs (GS-PS); c) young genomically tested sires selected based on gEBVs (GS).

In the PT strategy, 8/27 were chosen for progeny testing based on their PAs in their second year. In the sixth year 4/8 bulls were selected after obtaining progeny testing results. In the GS-PS strategy all 45 male candidates were genomically tested in their first year and 8/45 were chosen for progeny testing in their second year based on their gEBVs. In the sixth year 4/8 bulls were selected after obtaining progeny testing results. In the GS strategy all 45 male candidates were genomically tested in their first year and in their second year 5/45 were chosen as young sires based on their gEBVs.

The chosen candidates entered either progeny testing or genomic testing. Young bulls for progeny testing were pre-selected either based on the performance test and EBVs (which equal PA since they have no information) (PT) or based on their gEBVs (GS-PS).

a) performance testing - progeny testing: 60% (n=27) of male calves from contracted mating were randomly chosen for performance testing. The random choice reflects the realistic situation where the availability of male calves depends also on unpredictable human factors (where not all the best male calves are available for performance testing). In the second year 8/27 bulls were chosen for progeny testing. The non-chosen bulls were used in natural service as cow sires. After five years 4/8 progeny tested bulls were chosen as proven bulls.

b) genomic testing - progeny testing: all male calves from contracted mating were genomically tested in their first year. In the second year 8/45 bulls were chosen for progeny testing. The non-chosen bulls were used in natural service as cow sires. After five years 4/8 progeny tested bulls were chosen as proven bulls.

c) genomic testing: all male calves from contracted mating were genomically tested in their first year. In the second year 5/45 bulls were chosen as young sires. The non-chosen bulls were used in natural service as cow sires.

## Tested scenarios

We created five testing scenarios of truncation selection by using different categories of sires (PT, GS-PS, GS) for insemination of cows and bull dams. Different combinations of bulls for cow and bull dams insemination reduces generation interval in the corresponding selection path. This allowed us to test the effect of varying degree of reduction in generation interval. The tested scenario included one conventional scenario without the use of genomic information and four genomic scenarios. i) PT scenario used PT sires exclusively for insemination of all cows and bull dams. ii) Similarly, GS-PS scenarios used GS-PS sires exclusively for insemination of all females. Next we created two hybrid scenarios that used GS-PS and GS sires simultaneously. iii) GS-C used young genomically tested sires for insemination of cow population and GS-PS sires for insemination of bull dams. iv) Contrary, in GS-BD we used GS sires for the insemination of bull dams and GS-PS for insemination of cows. v) GS scenarios, also referred to as “turbo”, used exclusively young GS sires for insemination of all females.

## Sire use strategy

All five truncation selection scenarios were tested within three sire use strategies. i) Every year we selected five sires and keep them in use for five years (SU 5/5). This strategy reflects the situation in the simulated Slovenian Brown Swiss population. ii) Every year we selected only one sire per year and kept him in use for five years (SU1/5). Hence this strategy had increased intensity of selection. iii) every year we selected five sires and replaced them all in the next year, i.e. kept them in use for only one year (SU5/1). This strategy reduced the generation interval even further.

## Optimum contribution selection

In addition to different settings of truncation selection, we created an optimum contribution selection (OCS) scenario. Contributions were optimized with heuristic evolutionary algorithm implemented in AlphaMate software (Gorjanc et al., 2018). In contrast to truncation selection scenarios, we did not specify the number of sires selected and years kept in use in the OCS scenario. Also, male contributions were not equalized nor limited. The selection and contributions of sires were set by the optimization, while dams were selected as described above with equalized contributions (one offspring per dam). The OCS scenario implemented genomic testing. All 45 male offspring from contracted matings were genotyped and added to the sire candidate pool along with already selected sires from the last five years. This sire candidate pool was subjected to AlphaMate for optimization along with selected dams for the next generation. The kinship was accounted for with H matrix constructed according to Legarra et al., 2009. The H relationship matrix was computed based on same set of animals as used for gBV estimation. The selected sires from the optimization were randomly mated to chosen dams according to optimized contributions. The goal of optimization were trigonometric degrees between genetic gain and group coancestry with smaller degrees prioritizing maximizing genetic gain over minimizing group coancestry. The simulation was ran in 20 replicates for each of the following target degrees: 15, 30, 45, 60 and 75.

## Estimation of breeding values

Breeding values were estimated using pedigree or single-step BLUP depending on whether the genomic information was available. In genomic scenarios we assumed a initial reference population of ~11,000 cows and 100 progeny tested sires for the genomic prediction. The reference population was updated each selection cycle by removing the oldest generation of cows and adding an equivalently large sample of cows from the currently active cow population to the reference. Each year all the male offspring of contracted mating were genotyped and added to the reference population.

## Rate of inbreeding

We computed the rate of inbreeding per year based on either pedigree or molecular information for all the tested scenario. The pedigree inbreeding was computed with AlphaRelate (AlphaSuite) following the procedure described in XXX. The pedigree rate of inbreeding (ΔFPed) was derived according to 1 - Ft = t \* (1 – ΔFPed)t from regression of log (1 – F) over generations of comparison (Pérez-Enciso, 1995). The corresponding effective population size (Neped) was computed as 1 / (2 \* ΔFPed). Molecular rate of inbreeding was computed based on a direct link of heterozygosity with inbreeding, i.e. Hett = Heto(1 – F) as described in Meuwissen et al., 2018. We computed heterozygosity separately for QTN, marker and neutral loci. Analogous to regression for Fped we regressed log(Het) over generations of comparison. Accordingly we then computed rate of inbreeding on QTN loci (ΔFQTN), marker loci (ΔFM) and neutral loci (ΔFN). Corresponding effective population sizes, namely NeQTN, NeM and NeN, were again computed as 1 / (2 \* ΔFQTN / ΔFM / ΔFN).

## Metrics of comparison

The scenarios were compared in terms of genetic gain, efficiency and inbreeding averaged across 20 replicates. Genetic gain was expressed with mean zero and in units of genetic standard deviation in the first generation of comparison. 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 genic variance because we believe it provides a better measure of genetic variability since it allows us to exclude the effect of linkage disequilibrium. 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 twenty replicates for each scenario on a per generation or cumulative basis.

# Results

Increasing the use of genomic information increased the genetic gain but also resulted in a higher loss of genic variance. Hybrid approaches of using genomic and progeny testing were the most efficient and had a higher effective population size than GS scenario making them suitable for a long-term selection success. Regarding sires use strategies, performing a faster turn-over of animals proved as a better way to increase genetic gain than increasing intensity, because it reached higher efficiency and Ne. [We can further increase the genetic gain and efficiency with the use of OCS].

Here we present the results for all three measures of comparison: genetic gain, inbreeding and efficiency. Within each of this measures we compare the effect of the use of genomic information, i.e. five truncation selection scenarios (vertical comparison); the effect of the sire use, i.e. three sire use strategies (horizontal comparison); and the effect of OCS on the specific metric.

## Genetic gain

~~Completely genomic scenario (GS) delivered the highest genetic gain. Overall, the genetic gain increased with the use of genomic information.~~

Genetic gain increased with earlier use of genomically tested young bulls and faster turnover of sires. This is shown in Table 2, which presents genetic gain by breeding program and sire selection-use strategy expressed as percentage change to the baseline scenario that had genetic gain of 2.5 genetic standard deviations. Compared to the baseline scenario introducing genomic selection of young bulls for progeny testing increased genetic gain by 37%. Genomic selection of young bulls for the insemination of cows or bull dams respectively increased genetic gain by 63% or 69%, and by 95% when used jointly for cows and bull dams. Reducing the use of 5 selected sires per year from 5 years to 1 year further increased genetic gain, between 11% and 144% compared to the baseline scenario. Reducing the number of selected sires per year from 5 to 1 and using that sire for 5 years also increased genetic gain, between 22% and 126% compared to the baseline scenario, but not compared to the strategy where use of 5 sires was reduced from 5 years to 1 year. The exception was genomic selection of bulls for insemination of cows (GS-C) where the extreme strategy of selecting 1 sire per year and using it for 5 years had hgiher genetic gain than the strategy of selecting 5 sires per year and using them for 1 year.

Scenarios with a higher reduction in generation interval delivered a higher genetic gain (Table 1, Supplementary). When we used GS sires for insemination of cow population (GS-C), we reduced the generation interval for sires of dams between 42 and 64% and increased the genetic gain between 63 and 94%. Similarly, when we used GS sires for insemination of bull dam population (GS-BD) – while cow population was inseminated with GS-PS bulls – we reduced generation interval for 58% in the sires of sires selection path and increased the genetic gain between 69 and 82%. However, the difference in genetic gain between GC‑C and GS‑BD was not significant in any of the sire use strategies. Lastly, the scenario that used GS sires exclusively resulted in the highest reduction in generation interval – between 44 and 68% in the sires of dams paths and between 53% and 74% in the sires of sires selection path – and in highest increase in genetic gain - between 95 and 144%.

Table 2: Genetic gain by breeding program and sire selection-use strategy expressed as percentage change to the baseline scenario that had genetic gain of 2.5 genetic standard deviations

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sire selection and use strategy | | |
| Breeding program | 5 sires/year, use 5 years | 5 sires/year, use 1 year | 1 sire/year, use 5 years |
| PT | / a, A | 119 a, AB | 2211 a, B |
| GS-PS | 3712 b, A | 5913 b, B | 5413 b, B |
| GS-C | 6314 c, A | 8719 c, B | 9421 c, B |
| GS-BD | 6915 c, A | 8418 c, B | 8220 c, B |
| GS | 9517 d, A | 14427 d, B | 12627 d, C |

PT = conventional progeny testing; GS-PS = genomic selection of young bulls for progeny testing, GS-C = genomic selection of young bulls for insemination of cows; GS-BD = genomic selection of young bulls for insemination of bull-dams; GS = genomic selection of young bulls for insemination of cows and bull dams. Lower-case letters denote statistically significant differences between breeding programs and upper-case letters between sire selection-use strategies.

Table S1: Generation interval by breeding program and sire selection-use strategy expressed relative to the conventional progeny testing (PT) with 5 sires selected each year that are used for 5 years.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sire selection and use | | | | | | | |
| Breeding program | 5 sires/year - use 5 years | | | 5 sires/year - use 1 year | | | 1 sire/year - use 5 years | |
|  | Sire of sires | Sire of dams | Sire of sires | | Sire of dams | Sire of sires | | Sire of dams |
| PT | / | / | -21 5 | | -18 4 | 2 5 | | 9 9 |
| GS-PS | 0 4 | 0 1 | -21 5 | | -18 5 | 2 5 | | 9 9 |
| GS-C | 0 4 | -42 18 | -21 5 | | -64 19 | 2 5 | | -42 20 |
| GS-BD | -58 14 | -1 1 | -58 14 | | -19 5 | -58 14 | | 8 9 |
| GS | -53 12 | -44 11 | -74 17 | | -68 16 | -53 13 | | -44 11 |

PT = conventional progeny testing; GS-PS = genomic selection of young bulls for progeny testing, GS-C = genomic selection of young bulls for insemination of cows; GS-BD = genomic selection of young bulls for insemination of bull-dams; GS = genomic selection of young bulls for insemination of cows and bull dams.

## OCS and genetic gain

Decreasing the target degrees in OCS scenarios increased the genetic gain (Table 2). This was expected since higher degrees prioritize increasing genetic gain over minimizing group coancestry. The OCS-15 scenario reached %% higher genetic gain than GS scenario in SU 5/5, XX% than GS in SU 1/5 and XX % higher than GS in SU 5/1 strategy.

## Effective population size

Ne computed based on pedigree information significantly differs from genomic Ne. However, the differences between Ne on neutral, marker and causal loci did not differ significantly. We are showing the results for pedigree Ne in Table 3 and for Ne on causal loci in Table 4. All the values for Ne in Tables 3 and 4 are relative to the PT scenario of the SU 5/5 strategy with values of 270 for pedigree Ne and 172 for Ne on causal loci. Pedigree Ne exceeded the Ne on causal loci for all tested scenarios up to 184%.

## Effect of the use of genomic information on effective population size

GS scenario had the lowest pedigree and genomic Ne in all sire use strategies (Table 3). We observed that the use of genomic information may increase pedigree Ne  in all three strategies. Pedigree Ne was increased when genomic information was used for pre‑selection of bull for insemination (GS-PS) in SU 5/5 and SU 5/1, for selection of bulls for insemination of cows (GS-C) in SU 5/5, and for selection of bulls for insemination of bull dams (GS-C) in SU 1/5. The completely genomic scenario (GS) had the lowest pedigree Ne in all three strategies. In contrast, the use of genomic information never increased genomic Ne (Table 4) and GS scenario has the lowest genomic Ne in all strategies. The genomic Nes are in concordance with the observed loss of genic variance that increases with the use of genomic information (Table 2, Supplementary).

## The effect of sire use strategy on effective population size

Faster turn-over of animals and increased intensity had different consequence on Ne of the population. Faster turn-over of animals managed to increase the pedigree Ne in PT scenario and to retain it in GS-PS, GS-C and GS‑BD scenario (compared to SU 5/5). Similar trend was observed for genomic Ne, where SU 5/1 scenarios retained the Ne of the SU 5/5 scenarios. The exception was the GS scenario, where we observed a decrease in pedigree and genomic Ne.

Increasing intensity of selection (SU 1/5) always decreased pedigree Ne compared to corresponding scenarios in SU 5/5 strategy. Further more, while the majority of SU 5/5 and SU 5/1 scenarios managed to increase pedigree Ne relative to the PT scenario of the SU 5/5 strategy (Table 3), all of the SU 1/5 scenarios failed to do so. Similarly, increasing intensity significantly decreased genomic Ne  on causal loci compared to all corresponding scenarios in SU 5/5 and SU 5/1 strategies.

Table 3: Effective population size by breeding program and sire selection-use strategy expressed relative to the conventional progeny testing (PT) with 5 sires selected each year that are used for 5 years that had pedigree Ne of 270 and genomic Ne of 172.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sire selection and use | | |
| Breeding program | 5 sires/year - use 5 years | 5 sires/year - use 1 year | 1 sire/year - use 5 years |
|  | Pedigree Ne | | |
| PT | / ab, A | 2018 a, B | -4911 ab, C |
| GS-PS | 4126 c, A | 4829 b, A | -3816 ac, B |
| GS-C | 2524 c, A | 2224 a, A | -5913 b, B |
| GS-BD | 122 a, A | 524 a, A | -2425 c, B |
| GS | -1619 b, A | -3217 c, B | -786 d, C |
|  | Pedigree Ne - AVERAGE | | |
| PT | / ab, A | 1817 a, B | -4512 ab, C |
| GS-PS | 3424 c, A | 3621b, A | -3415 ac, B |
| GS-C | 1920 c, A | 2124 a, A | -5512 b, B |
| GS-BD | -116 a, A | 418a, A | -2515 c, B |
| GS | -1616 b, A | -2713 c, B | -755 d, C |
|  | Genomic Ne at causal loci | | |
| PT | / a, A | 9 21 a, A | -42 12 a, B |
| GS-PS | -6 16 a, A | -13 21 b, A | -40 11 a, B |
| GS-C | -23 12 b, A | -26 17 c, A | -61 8 b, B |
| GS-BD | -29 12 b, A | -33 11 c, AB | -43 16 a, B |
| GS | -45 12 c, A | -56 10 d, B | -76 7 c, C |

PT = conventional progeny testing; GS-PS = genomic selection of young bulls for progeny testing, GS-C = genomic selection of young bulls for insemination of cows; GS-BD = genomic selection of young bulls for insemination of bull-dams; GS = genomic selection of young bulls for insemination of cows and bull dams. Lower-case letters denote statistical significance of the differences between the scenarios within a strategy and upper-case letters between strategies within a scenario.

~~Table 4: Genomic effective population at causal loci size by breeding program and sire selection-use strategy expressed relative to the conventional progeny testing (PT) with 5 sires selected each year that are used for 5 years with Ne = 172.~~

|  |  |  |  |
| --- | --- | --- | --- |
|  | ~~Sire selection and use~~ | | |
| ~~Breeding program~~ | ~~5 sires/year - use 5 years~~ | ~~5 sires/year - use 1 year~~ | ~~1 sire/year - use 5 years~~ |
| ~~PT~~ | ~~/~~ ~~a, A~~ | ~~9~~ ~~21~~ ~~a, A~~ | ~~-42~~ ~~12~~ ~~a, B~~ |
| ~~GS-PS~~ | ~~-6~~ ~~16~~ ~~a, A~~ | ~~-13~~ ~~21~~ ~~b, A~~ | ~~-40~~ ~~11~~ ~~a, B~~ |
| ~~GS-C~~ | ~~-23~~ ~~12~~ ~~b, A~~ | ~~-26~~ ~~17~~ ~~c, A~~ | ~~-61~~ ~~8~~ ~~b, B~~ |
| ~~GS-BD~~ | ~~-29~~ ~~12~~ ~~b, A~~ | ~~-33~~ ~~11~~ ~~c, AB~~ | ~~-43~~ ~~16~~ ~~a, B~~ |
| ~~GS~~ | ~~-45~~ ~~12~~ ~~c, A~~ | ~~-56~~ ~~10~~ ~~d, B~~ | ~~-76~~ ~~7~~ ~~c, C~~ |

~~PT = conventional progeny testing; GS-PS = genomic selection of young bulls for progeny testing, GS-C = genomic selection of young bulls for insemination of cows; GS-BD = genomic selection of young bulls for insemination of bull-dams; GS = genomic selection of young bulls for insemination of cows and bull dams. Lower-case letters denote statistical significance of the differences between the scenarios within a strategy and upper-case letters between strategies within a scenario.~~

Table X: Genetic variance by breeding program and sire selection-use strategy expressed relative to the conventional progeny testing (PT) with 5 sires selected each year that are used for 5 years with genetic variance 0.89 in generation 60.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sire selection and use | | |
| Breeding program | 5 sires/year - use 5 years | 5 sires/year - use 1 year | 1 sire/year - use 5 years |
| PT | / a, A | - 1.59.0 a, A | - 8.310.9 ab, B |
| GS-PS | - 6.45.0 ab, A | - 9.010.5 b, A | - 9.08.1 ab, A |
| GS-C | - 9.06.1 b, A | - 11.38.4 b, A | - 14.77.5 ac, A |
| GS-BD | 0.610.5 a, A | - 9.38.0 b, B | - 3.413.0 b, AB |
| GS | - 8.48.3 b, A | - 10.88.5 b, A | - 17.99.0 c, B |

PT = conventional progeny testing; GS-PS = genomic selection of young bulls for progeny testing, GS-C = genomic selection of young bulls for insemination of cows; GS-BD = genomic selection of young bulls for insemination of bull-dams; GS = genomic selection of young bulls for insemination of cows and bull dams. Lower-case letters denote statistical significance of the differences between the scenarios within a strategy and upper-case letters between strategies within a scenario.

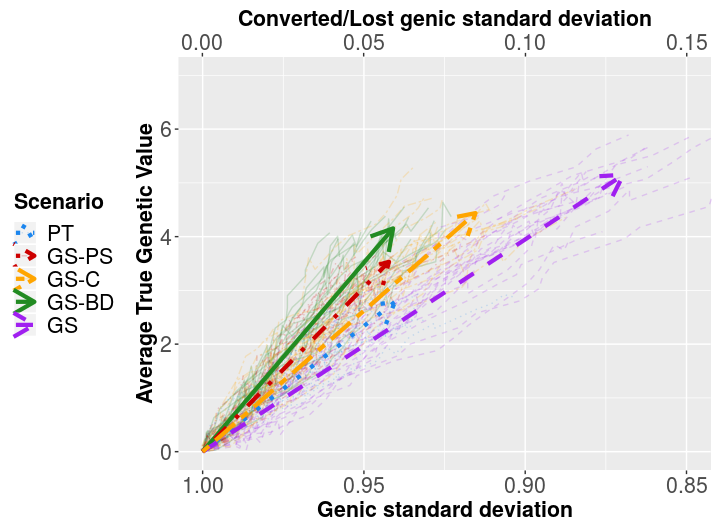
Table X: Genic variance by breeding program and sire selection-use strategy expressed relative to the conventional progeny testing (PT) with 5 sires selected each year that are used for 5 years with genic variance 0.94 in generation 60.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sire selection and use | | |
| Breeding program | 5 sires/year - use 5 years | 5 sires/year - use 1 year | 1 sire/year - use 5 years |
| PT | / a, A | - 1.81.6 a, B | - 5.91.7 a, C |
| GS-PS | - 0.51.2 ab, A | - 3.11.7 ab, B | - 5.52.0 a, C |
| GS-C | - 1.91.4 bc, A | - 5.11.7 c, B | - 10.82.2 b, C |
| GS-BD | - 2.61.1 c, A | - 4.31.5 bc, B | - 5.81.9 a, C |
| GS | - 5.01.9 d, A | - 9.81.5 d, B | - 19.52.8 c, C |

PT = conventional progeny testing; GS-PS = genomic selection of young bulls for progeny testing, GS-C = genomic selection of young bulls for insemination of cows; GS-BD = genomic selection of young bulls for insemination of bull-dams; GS = genomic selection of young bulls for insemination of cows and bull dams. Lower-case letters denote statistical significance of the differences between the scenarios within a strategy and upper-case letters between strategies within a scenario.

## Efficiency

Highest efficiency was achieved with simultaneous use of progeny and genomically tested sires. The results are shown in Table 4, which presents efficiency by breeding program and sire selection‑use strategy expressed as percentage change to the baseline scenario that has efficiency of 77. Efficiency of e.g. 77 means that when we burn all the genic variance we will reach genetic gain of 77 units. Compared to the baseline scenario, the use of genomic information increased efficiency. The highest increase, 33%, was achieved by using genomic information for selection of young sires for progeny testing. Efficiency increased by 30% when we introduced genomic selection of young sires for the insemination of cows and by 25% for the insemination of bull dams. Using young genomically tested sires exclusively increased efficiency by 12%, however, the increase was not significant. Scenarios – PT and GS - were the the least efficient ones. All three hybrid scenarios were significantly more efficient than PT scenario; GS‑PS and GS‑C scenario were significantly more efficient than both extreme scenarios – PT and GS (Table 4). Reducing the years of sire use from 5 to 1 still increased efficiency from 2 to 16% compared to the baseline scenario - with the exception of PT scenario, where we observed 18% decrease in efficiency. The hybrid scenarios again had the highest efficiency with GS-BD having the highest efficiency (16% increase). However, when compared to the original strategy of selecting 5 sires per year and using them for 5 years, the efficiencies of all scenarios were reduced. Reducing the number of sires selected per year to 1 and using it for 5 years decreased efficiencies of all scenarios compare do the baseline scenario as well as compared to the corresponding scenarios in the two other sire use strategies. Similarly, the two extreme scenarios - using no genomically tested young sires and using exclusively young sires – resulted in the lowest efficiency (35% decrease in PT scenario and 43% in GS scenario). The hybrid scenarios were the most efficient – genomic selection of young bulls for insemination of bulls dams was the most efficient with 2% decrease in efficiency compares to the baseline scenario.

Figure 1: Change of genetic mean and genic standard deviation over the 20 years of selection by breeding programs with 1 sire selected each year that is used for 5 years. Arrows point the direction of change. Sire selection strategies: conventional = only progeny tested bulls (no genomic information), PT = conventional scenario without use of genomic information; GS-PS = genomic information used for pre-selection of young bulls for progeny testing, GS-C / GS-BD / GS = genomically tested bulls used respectively for insemination of cows / bull dams / cows and bull dams.

## The effect of sire use strategies on selection efficiency

Despite larger genetic gain, increasing the intensity of selection and performing faster turn-over of animals decreased the efficiency. In SU 5/5 strategy the use of genomic information increases the efficiency from 12 to 33% (increases genetic gain from 37 to 95% and reduced genic variance from 0.5 to 5%) . In SU 5/1 all genomic scenarios had from 2 to 12% higher average efficiency relative to the reference scenario. However, compared to corresponding SU 5/5 scenarios, we observe lower average efficiencies of all scenarios - with the highest reduction observed for GS-PS scenario (21%) - although only efficiencies for GS-PS and GS-C were significantly reduced. We can link the decrease in efficiencies of the SU 5/1 scenario to the larger losses of genic variance in SU 5/1. We observed up to 5% loss of genic variance in SU 5/5 and up to 9.8% in SU 5/1 strategy. We attribute this to a 0.1 unit drop in accuracy compared to SU 5/5 strategy, which we believe is due to a larger genetic distance between the training and the testing population. Increasing the intensity of selection in the SU 1/5 strategy reduced efficiency compared to the reference scenario from 2 to 43%. We also observe the lowest efficiency for all the scenarios compared to corresponding scenarios in SU 5/5 and SU 5/1 strategies. The results are in concordance with results for genic variance, where scenarios in SU 1/5 strategy lost the most genic variance – up to 19.5%. The largest drop in efficiency compared to SU 5/5 strategy was observed for GS‑C scenario (58%).

Table 4: Efficiency of converting genetic variation into gain by breeding program and sire selection-use strategy expressed relative to the conventional progeny testing (PT) with 5 sires selected each year that are used for 5 years with efficiency of 77.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sire selection and use | | |
| Breeding program | 5 sires/year - use 5 years | 5 sires/year - use 1 year | 1 sire/year - use 5 years |
| PT | / a, A | -1812 a, B | -3512 ab, C |
| GS-PS | 3325 b, A | 1223 b, B | -1417 cd, C |
| GS-C | 3022 b, A | 920 b, B | -2811 ac, C |
| GS-BD | 2525 bc, A | 1616 b, A | -225 d, B |
| GS | 1220 ac, A | 222 b, A | -4312 b, B |

PT = conventional progeny testing; GS-PS = genomic selection of young bulls for progeny testing, GS-C = genomic selection of young bulls for insemination of cows; GS-BD = genomic selection of young bulls for insemination of bull-dams; GS = genomic selection of young bulls for insemination of cows and bull dams. Lower-case letters denote statistical significance of the differences between the scenarios within a strategy and upper-case letters between strategies within a scenario.

## OCS and efficiency

The efficiency of the OCS scenarios decreases with decreasing target degrees. The efficiencies ranged between XX and XX. Lower / higher on average than in SU 55 [čakam, da mi vsaj tri replike zvozijo, da vidim trend]. % higher.

The efficiencies of the \_\_\_ scenarios were reduced due to a larger loss in genic variance (% loss). The number of sires and their offspring were freely set by optimization. The average number of fathers in the OCS-75 scenarios was 3 (SD = XX), XX in OCS-45 (SD = ), XX in OCS-60 (SD = ) and 125 (SD = ) in OCS-15.

## Discussion for OCS efficiency

OCS scenarios \_\_\_\_ had lower efficiency due to a larger loss in genic variance. This is perhaps unexpected since OCS controls group coancestry and aims to maximize genetic gain for specifies value of group coancestry. The larger loss in genic variance is a result of a very low number of sires in OCS-75 scenario (mean = 3), since the sires were selected by optimization without any limitations for the number of sires and number of offspring per sire.. For example, the average number of sires in the truncation scenarios in SU 5/5 strategy was ~50 (proven, young, natural service bulls) with limited number of offspring. However, if we compare truncation selection and OCS scenarios with comparable genetic gain - for example SU 5/5 \_\_\_\_ and OCS-\_\_ - we see that the OCS scenarios loss less genic variance for the same genetic gain.

# Discussion

## TODO

## Genetic gain

As expected, the use of genomic information increased the genetic gain in all sire use strategies. This is due to higher accuracy for selection of young unphenotyped animals and reduction in generation interval. Due to a higher accuracy of the pre-selection step, we managed to substantially increase genetic gain without reducing generation interval (GS-PS scenario) - since gEBV (r = 0.79) were more accurate than PA (r = 0.30). In other genomic scenarios we used genomically tested sires directly and reduced the generation interval. This increased the genetic gain even further even though the accuracy of BVs for genomically tested sires at selection is lower than for progeny tested bulls in PT (and GS‑PS) scenario. The genetic gain increased with reduction in generation interval. This results suggest than we can alleviate the effect of the lower accuracy of genomic prediction with an earlier selection of sires, since the effect of the latter prevails. There was no significant difference in the genetic gain of GS-C and GS-BD scenarios with reduced generation interval in the sire of dams and sire of sires selection paths, respectively. This suggests that the amount of reduction in generation interval prevails over the path in which we use the genomic information (selection path is irrelevant and that genetic gain depends only on the amount of reduction).

The completely genomic scenario used GS sires for insemination of all cows – therefore all sires were selected at a lower accuracy. However, since generation interval was reduced in both sires of dams and sires of sires paths, the genetic gain still increased. These results are in concordance with previous studies that observed a higher genetic gain with the use of genomic information for a pre-selection for PT or for selection of young sires (“turbo” scheme) (Pryce et al., 2010; Lillehammer et al., 2011; de Roos et al., 2011; Thomasen et al., 2014).

Genetic gain in OCS scenarios (at which degrees) was increased due to two reasons: firstly, the optimization algorithm optimizes the selection of sires for a specified value of group coancestry; and secondly, the intensity of selection was increased since low degree OCS scenarios chose on average only three sires per year and replaced them all every subsequent year (faster turn over of sires). In general, OCS scenarios tended to replace the majority of bulls every year – therefore kept them in use only for one year (to moram preveriti, ne vem, če bo res). If we then compare truncation and OCS scenarios with comparable genetic gain we see that GS-PS corresponds to the OCS-75 scenario, that uses 125 bulls per year on average. This helps to explain the lower genetic gain of GS-PS scenario [other comparison – also GENERATION INTERVAL of the sires!!! → compare truncation and OCS with comparable generation interval for the sires!]. The average generation interval in the sires of sires psth in the OCS scenarios was XX – which is XX lower / higher than in Class / gen scenarios. This shows that OCS scenarios maximizes genetic gain by increasing the intensity of selection and decreasing the generation interval, therefore selecting only few top sires and replacing them all every year.

## Inbreeding | Effective population size

We observed a discordance in Ne computed based on pedigree or molecular information with the former exceeding the latter up to 166%. While pedigree information measures identity by descend, the identity on molecular markers measures identity by state (Meuwissen, 2018). This suggests that pedigree inbreeding is not capable of capturing the true inbreeding in the population. Our results are in concordance with previous findings that pedigree estimation tends to underestimated the true inbreeding in the population. On the other hand, the inbreeding estimated from QTN loci did not exceed inbreeding estimated from neutral or marker loci – there were no significant differences between Ne computed from neutral, selection and QTN loci. This is most likely a consequence of simulating a highly polygenic trait controlled by 1000 QTNs per chromosome. While the results from pedigree inbreeding suggest we could increase Ne by introducing genomic information in selection, the results from molecular markers suggest otherwise. According to the latter the Ne decreases with the use of genomic information which is in concordance with the observed loss of genic variance.

While increasing the intensity of selection severely decreased Ne compared to the SU 5/5 strategy (up to 74%), no such effect was observed for SU 5/1 strategy. These results are in concordance with previous studies reporting that a larger number of bulls counterbalances the effect of reduced effective number of bulls per generation caused by reduced generation interval (reviewed in Boichard et al., 2015). These results also suggest that we could reduce inbreeding in genomic scenarios by increasing the number of sires per generation. On the other hand, considering we do not observe a significant change in the Ne in SU 5/1 strategy suggests, a faster turn over of the animals - that could potentially benefit genetic diversity of the population - is counterbalanced with a reduction in generation interval.

In addition to different breeding strategies, inbreeding rate can be reduced by applying inbreeding control. In our simulation we did not perform any form of inbreeding control, except for limiting the number of offspring per bull to a realistic viable number of semen doses. In order to efficiently bound the rate of inbreeding, we should apply inbreeding control on the same basis as used for estimation of breeding values (Woolliams et al., 2015). Therefore, while for PT and GS-PS scenarios a pedigree based inbreeding control would suffice, the genomic scenarios would required control based on genomic relationship.

## Efficiency

We observed that efficiencies do not follow the pattern of genetic gain with the hybrid scenarios being the most efficient. This is due to differences in the loss of genic variance that increases with the genetic gain ( | the use of genomic information). GS and PT scenarios were the least efficient scenarios in all sire use strategies. The low efficiency of the GS scenario - despite its high genetic gain - is attributable to the relatively large loss of genic variance (up to 13% loss). The reason for this lies in the reduced accuracy of sire selection since genomic testing (mean r = 0.52) is less accurate than progeny testing (mean r = 0.75). In contrast, low efficiency of the PT scenario is due to its small genetic gain caused by long generation intervals although it retains the most genic variance (up to 6% loss).

Although not all differences were significant, the most efficient scenario in all three sires use strategies was one of the hybrid hybrid scenario – it was also significantly more efficient than PT scenario. Although the hybrid scenarios do not reach the genetic gain of the GS scenarios and do not maintain the genic variance of the PT scenario, they reach a balance between them that allow for a long-term genetic gain and sustainable selection. In the SU 5/5 strategy GS-PS was the most efficient scenario. A pre-selection step based on gEBVs allows us to identify the effect of Mendelian sampling term that is crucial for providing long-term genetic gain. In the SU 1/5 strategy we introduced an unrealistic condition of selecting only one bull per year. This results in a different pattern in the efficiencies of scenarios. While the GS remains the least efficient - due to a relatively large loss of genic variance (13%) - we observe a large drop in the efficiency of the GS-C scenario when compared to SU 5/5. This corresponds to a large increase in the loss of genic variance. In GS‑C scenarios the vast majority of the female population – with the exception of bull dams – is inseminated with genomically tested sire(s). [Tukaj ne znam pojasnit, zakaj GS-C izgubi veliko več genske variance v primerjavi z ostalimi scenariji kot pa V SU 5/5 in SU 5/1. Je res, da večino populacije (vse krave) osemenimo z enim genomsko testiranih bikom, katerega točnost je nižja (od PT) → pri čemer bi nižja točnost lahko pomenila večjo co-sib selection → vendar pa so ti biki genomsko testirani, kar pomeni ravno obratno – da najbrž izvajamo več izbire med družinami. Če bi bilo temu tako, bi lahko zapisali (to sem imela napisano prej, samo sem ugotovila, da najbrž ne drži): These results therefore suggest that in SU 1/5 strategy the effect of inseminating nearly entire population with a sire of lower accuracy is enhanced due to increased intensity. The effect of erroneously chosen sire is reflected throughout the population and results in a larger loss of genic variance. This effect is not as substantial in the GS-BD strategy since the bottle neck of selecting one sire for the majority of inseminations is made with larger accuracy (PT bulls)]. In SU 5/1 sire use strategies all three hybrid scenarios were significantly more efficient than both extreme scenarios. Again, the GS-PS scenario had the highest efficiency, although it did not significantly differ from efficiencies of GS-C and GS-BD scenarios. However, in SU5 5/1 the efficiencies were still reduced due to a faster turn-over of animals – this in in concordance with Gorjanc et al., 2017???

The results suggest that scenarios that use progeny tested and young genomically tested bulls simultaneously are the most efficient and therefore most appropriate for a long-term sustainability of selection (genetic gain?) – although they do not bring the highest genetic gain. We explain the success of these scenarios by a greater spread of the true BVs of the PT sires due to a higher accuracy of their EBVs. Hence hybrid schemes lose less genetic (genic) variance resulting in higher efficiencies. Regarding different sires use strategies, the results suggest that increasing the intensity of selection (SU 1/5) and reducing the generation interval even further (SU 5/1) have different consequences on short and long-term success of selection (genetic gain). Although both these strategies results in a larger genetic gain, increasing the intensity also causes a larger loss in genic variance and a severe drop in Ne. On the other hand, a faster turn-over of the animals resulted in the highest genetic gain, reduced efficiencies to a medium degree and retained the Ne. This suggests that a faster strategy could benefit breeding schemes and allow for a sustainable selection with high genetic gain.

Although the genomic scenarios lost more genic variance, the loss is still relatively small. The largest loss of genic variance was 13% and was observed for GS scenario in the SU 1/5 strategy. This could be explained by genomic selection increasing the frequency of rare alleles up to 0.5 will increase gen(et)ic variance in the population (Weller in sod., 2017).

Another point:

Thomasen et al., 2014, pointed out, that the benefit of genomic selection in small populations is undermined due to a small increase in the accuracy of gEBVs for young animals caused by a weak reference population. However, our results suggest that with a cow based reference population we can achieve an opposite result. Since in small populations the number of offspring in progeny testing is not as large, the accuracy of conventional EBVs for progeny tested bulls does not reach the accuracy of the large populations. On the other hand, with a cow reference population we can reach relatively high accuracies of genomic prediction comparable to the accuracies of the progeny tested bulls.

Invited review: A perspective on the future of genomic selection in dairy cattle

**[https://www.sciencedirect.com/science/article/pii/S0022030217307865](https://www.sciencedirect.com/science/article/pii/S0022030217307865" \l "_blank)**

 Genetic variance still quite large as we are increasing frequency of rare alleles

Sustainable dairy cattle selection in the genomic era

**[https://onlinelibrary.wiley.com/doi/pdf/10.1111/jbg.12150](https://onlinelibrary.wiley.com/doi/pdf/10.1111/jbg.12150" \l "_blank)**

 use more young bulls to maintain genetic diversity

Genomic selection requires genomic control of inbreeding.

[https://www.ncbi.nlm.nih.gov/pubmed/22898324](https://www.ncbi.nlm.nih.gov/pubmed/22898324" \l "_blank)

 must use G in OCS if we use gEBV

# Conclusion

TODO

# References

TODO