Genomic selection for any dairy breeding program via optimized investment in phenotyping and genotyping

Jana Obšteter[1*], Janez Jenko[2], Gregor Gorjanc[3,4]

[1]Department of Animal Science, Agricultural Institute of Slovenia, Ljubljana, Slovenia

[2]Geno Breeding and A.I. Association, Hamar, Norway

[3]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom

[4]Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

*** Correspondence:**
Jana Obšteter, Department of Animal Science, Agricultural Institute of Slovenia, Ljubljana, Slovenia
jana.obsteter@kis.si

Keywords: genomic selection, dairy breeding program, small populations, optimized investment, return on investment

Manuscript includes 6982 words, 1 table and 3 figures.

Abstract

This paper evaluates the potential of maximizing genetic gain in dairy cattle breeding by optimizing investment into phenotyping and genotyping. Conventional breeding focuses on phenotyping selection candidates or their close relatives to maximize selection accuracy for breeders and quality assurance for producers. Genomic selection decoupled phenotyping and selection and through this increased genetic gain per year compared to the conventional selection. Although genomic selection is established in well-resourced breeding programs, small populations and developing countries still struggle with the implementation. The main issues include the lack of training animals and lack of financial resources. To address this, we simulated a case-study of a small dairy population with a number of scenarios with equal resources yet varied use of resources for phenotyping and genotyping. The conventional progeny testing scenario had 11 phenotype records per lactation. In genomic scenarios, we reduced phenotyping to between 10 and 1 phenotype records per lactation and invested the saved resources into genotyping. We tested these scenarios at different relative prices of phenotyping to genotyping and with or without an initial training population for genomic selection. Reallocating a part of phenotyping resources for repeated milk records to genotyping increased genetic gain compared to the conventional scenario regardless of the amount and relative cost of phenotyping, and the availability of an initial training population. Genetic gain increased by increasing genotyping, despite reduced phenotyping. High-genotyping scenarios even saved resources. Genomic scenarios expectedly increased accuracy for young non-phenotyped male and female candidates, but also cows. This study shows that breeding programs should optimize investment into phenotyping and genotyping to maximise return on investment. Our results suggest that any dairy breeding program using conventional progeny testing with repeated milk records can implement genomic selection without increasing the level of investment.

1 Introduction

This paper evaluates the potential of maximizing genetic gain in dairy cattle breeding by optimizing investment into phenotyping and genotyping. Breeding programs strive to maximize genetic gain, which is a function of selection intensity, accuracy of selection, genetic variation, and generation interval. The conventional dairy breeding program uses an expensive and time-consuming progeny test. Genomic selection (Meuwissen et al., 2001; Schaeffer, 2006) achieves the same genetic progress faster and cheaper through a substantially reduced generation interval, increased accuracy of selection for young animals, and increased selection intensity of males (Schaeffer, 2006; Obšteter et al., 2019). Despite lower accuracy of sire selection compared to the conventional selection, genomic selection doubles the rate of genetic gain per year in dairy cattle (Wiggans et al., 2017).

All breeding programs operate with a limited resources allocated to breeding activities with the aim to maximize return on investment. Genomic selection is now a de-facto standard in leading breeding programs but is still challenging to implement in small national breeding programs or in developing countries. While leading breeding programs can service some small national breeding, developing countries require tailored breeding goals that will respond to the rapidly rising demand for a sustainable dairy production in local environment (Ducrocq et al., 2018; Marshall et al., 2019; Mrode et al., 2019). We hypothesize that these breeding programs need to evaluate priorities and could optimize the allocation of resources for phenotyping and genotyping to maximize return on investment. We base this hypothesis on the following simple examples (Table S1).

The accuracy of conventional (pedigree-based) estimates of breeding values increases with increasing heritability and increasing number of phenotype records per animal or its closest relatives (e.g. Mrode, 2005). For example, for a female-expressed trait with 0.25 heritability, the accuracy as a function of the number of repeated records per lactation (n) is 0.89 (n=10), 0.81 (n=5), 0.70 (n=2) and 0.62 (n=1). The corresponding accuracies for 100 sires tested on the total of 10,000 daughters are respectively 0.98 (n=10), 0.97 (n=5), 0.96 (n=2) and 0.93 (n=1).

The accuracy of genome-based estimates of breeding values similarly increases with increasing heritability and increasing number of phenotype records per genotyped animal. It also increases with increasing training population, decreasing genetic distance between training and prediction individuals, and decreasing number of effective genome segments (Daetwyler et al., 2008; Clark et al., 2011; Goddard et al., 2011). Following the previous example, assume 10,000 effective genome segments, 0.25 heritability, and a training population of 10,000 cows. The accuracy of genomic prediction for non-phenotyped animals as a function of the number of repeated records per lactation in the training population (n) is 0.76 (n=10), 0.71 (n=5), 0.63 (n=2), or 0.56 (n=1). These examples show diminishing returns with repeated phenotyping and a scope for optimizing return on investment in genomic breeding programs.

The accuracy of genomic prediction as a function of the number of genotyped and phenotyped cows (N) and the number of repeated records per lactation (n) is then 0.84 (N=20,000, n=5), 0.90 (N=50,000, n=2), or 0.93 (N=100,000, n=1). While these genomic prediction accuracies are lower

than with progeny testing, shorter generation interval enables larger genetic gain per unit of time (Schaeffer, 2006).

Previous studies also explored the value of adding females to the training population (Van Grevenhof et al., 2012; Gonzalez-Recio et al., 2014). They concluded that accuracy has diminishing returns with increasing the number of genotyped and phenotyped animals in the training population, hence additional females are most valuable when training population is small. However, real breeding programs involve overlapping generations, individuals with a mix of phenotype, pedigree, and genotype information, various selection intensities, and other dynamic components. Thus, evaluating the optimal allocation of resources into phenotyping and genotyping is beyond these simple examples.

The above examples suggest that repeated phenotyping could serve as an internal financial reserve to enable dairy breeding programs to implement genomic selection and maximize return on investment. In dairy breeding the most repeatedly and extensively recorded phenotypes are milk production traits. There are different milk recording methods that differ in the recording frequency (International Committee for Animal Recording, 2017). The recording interval ranges from daily recording to recording every nine weeks, which translates to between 310 and 5 records per lactation. The different recording methods have different costs, which vary considerably between recording systems, countries, and even their regions. For example, some organizations require payment of a participation fee plus the cost per sample, while others include the fee in the sample cost, or cover the costs in other ways. There is also a huge variance in the way dairy breeding programs are funded. Some are funded by farmers and breeding companies while in some countries data recording or breeding are subsidized by state or charities. There are also situations where farmers do not record phenotypes and all genetic progress is generated in "research" nucleus herds.

The aim of this study was to evaluate the potential of maximizing genetic gain by optimizing investment into phenotyping and genotyping in dairy breeding programs. Since milk production traits are example of repeated phenotypes with diminishing returns, we aimed to optimize investment into milk recording and genotyping. To this end we have compared a dairy breeding program with conventional progeny testing and several genomic breeding programs under equal financial resources. To implement genomic selection, we reduced the number of milk records per lactation and invested the saved resources into genotyping. We compared the breeding programs in a case-study of a small cattle population where implementing genomic selection is challenging. The results show that reallocating a part of phenotyping resources to genotyping increases genetic gain regardless of the cost and amount of genotyping, and the availability of an initial training population.

2 Methods

The study aimed to evaluate the effect of different investment into phenotyping and genotyping with a simulation of a case-study of a small dairy breeding program. The simulation mimicked a real dairy cattle population of ~30,000 animals analyzed in Obšteter et al. (2019). Here we evaluated 36 genomic scenarios against the conventional scenario, all with equal available resources, but varying extent of phenotyping and genotyping. The conventional scenario implemented progeny testing and

collected 11 phenotype records per lactation, while genomic scenarios reduced phenotyping and invested the saved resources into genotyping. The genomic scenarios differed in i) the number of phenotype records per lactation; ii) the relative cost of phenotyping and genotyping; and iii) the availability of an initial training population. All scenarios were compared on genetic gain and accuracy of selection.

2.1 Simulation of the base population, phenotype and historical breeding

The simulation mimicked a small dairy cattle breeding program of ~30,000 animals with ~10,500 cows. The introduction of effective genomic selection in such populations is challenging due to costs of assembling a training population and limited number of training animals. We use this population as a case-study to optimize investment into phenotyping and genotyping. The dairy breeding program aimed to improve production traits, which we simulated as a single polygenic trait. We used a coalescent process to simulate genome comprised of 10 cattle-like chromosomes, each with 10^8 base pairs, 1,000 randomly chosen causal loci, and 2,000 randomly chosen marker loci. We sampled the effects of causal loci from a normal distribution and use them to calculate animal's breeding value (a_i) for dairy performance (y_{ijkl}), which was affected also by a permanent environment (p_i), herd (h_j), herd-year (hy_{jk}), herd-test-day (htd_{jkl}), and residual environment (e_{ijkl}) effects: $y_{ijkl} = a_i + p_i + h_j + hy_{jk} + htd_{jkl} + e_{ijkl}$.

We sampled permanent environment effects from a normal distribution with zero mean and variance equal to the base population additive genetic variance (σ^2_a). We sampled herd, herd-year, and herd-test-day effects each from a normal distribution with zero mean and variance of 1/3 σ^2_a. Finally, we sampled residual environment effects from a normal distribution with zero mean and variance of σ^2_a. This sampling scheme gave a trait with 0.25 heritability and 0.50 repeatability. With the simulated genome and phenotype architecture we have initiated a dairy cattle breeding program and ran it for 20 years of conventional selection with progeny-testing based on 11 cow phenotype records per lactation. The detailed parameters of the simulation are described in Obšteter et al. (2019). In summary, in the breeding program we selected 3,849 out of 4,320 new-born females. We selected 139 bull dams out of cows in the second, third, and fourth lactation. We generated 45 male calves from matings of bull dams and progeny tested sires (elite matings). Out of these we chose 8 for progeny testing of which 4 were eventually selected as sires for widespread insemination of cows. We made all selection decisions based on pedigree-based estimates of breeding values. The 20 years represented historical breeding and provided a starting point for evaluating future breeding scenarios, which we ran for additional 20 years.

2.2 Scenarios

We evaluated 36 genomic scenarios with varying extent of phenotyping and genotyping against the conventional scenario, all with equal resources. The conventional scenario continued the breeding scheme from the historical breeding. It used progeny testing and 11 phenotype recordings per lactation (named C11), corresponding to the standard ICAR recording interval of 4 weeks (International Committee for Animal Recording, 2017). We assumed that this scenario represented

the total resources for generating data. We then created genomic scenarios by distributing resources between phenotyping and genotyping - we reduced phenotyping and invested the saved resources into genotyping. In genomic scenarios we selected females as in the conventional scenario and males on genomic estimates of breeding values. The number of genomically tested male candidates varied according to the genotyping resources in a specific scenario. We selected 5 males with the highest genomic estimates of breeding value as sires for widespread insemination of cows. We evaluated the genomic scenarios with a varying number of phenotype records per lactation, relative cost of phenotyping to genotyping, and the availability of an initial training population.

In genomic scenarios we reduced the number of phenotype records per lactation to between 10 and 1. Three of them followed ICAR standards of 9, 8, and 5 records per lactation, corresponding to recording intervals of 5, 6, and 9 weeks. Additionally, we tested the non-standard 10, 2, and 1 records per lactation. We named the scenarios as “GX” with X being the number of records per lactation. Genomic scenarios next varied the relative cost of phenotyping ($P) to genotyping ($G). We compared the cost of 11 phenotype records per lactation to the cost of one genotype. By increasing the number of records per lactation the cost per record decreased by 6%. The cost of genotyping remained constant. Based on a survey of several breeding programs, milk recording organizations, and genotyping providers we have considered three cost ratios of $P:$G: 2:1, 1:1, and 1:2. The reduction in phenotyping and the relative cost of phenotyping to genotyping dictated the number of genotyped animals (Table 1). For example, assume that the cost per milk record is €1.78 with 11 records per lactation (in total €19.55). Further assume that the cost per milk record is €1.89 with 10 records per lactation (in total €18.91). This second scenario saves €0.64 that can be used for genotyping. Applying this to our simulated populations with 10,852 active cows we could genotype 356 animals every year when $P:$G = 1:1.

We invested the saved resources into genotyping females and males in ratio 7:1 based on our previous work (Obšteter et al., 2019). We genotyped first parity cows. This maximized the accuracy of genomic prediction, since it reduced genetic distance between training and prediction population, prevented the loss of investment with culled heifers, and minimized the time to obtain a phenotype linked to a genotype. To maximize the genetic gain, we genotyped male calves from elite matings and other high parent average matings. In scenarios where the resources for genotyping females were larger than the cost of genotyping all first parity cows, we did not reallocate the excess of resources to genotyping males for consistency (we saved resources in those cases).

Lastly, we created scenarios with and without an initial training population for genomic prediction. When we assumed an initial training population was available, we genotyped all active cows (10,852) and progeny tested sires (100) before the first genomic selection of males. When an initial training population was not available, we yearly genotyped a designated number of first parity cows until the training population reached 2,000 cows. Once we reached this goal, we started to genotype both females and males as specified in Table 1. At that point we started genomic selection of males.

2.3 Estimation of breeding values

We selected animals based on their breeding values estimated from a pedigree or single-step genomic repeatability model with breeding value, permanent environment, and herd-year as random effects. We did not fit the herd-test-day effect as data structure of this small population did not enable its accurate estimation. We estimated breeding values once a year with blupf90 (Misztal et al., 2018) with default settings. In the estimation we included all available phenotype and pedigree records for all active, phenotyped, or genotyped animals, and additional three generations of their ancestors. We used at most 25,000 genotype records due to a limit in the academic software version. When we accumulated more than 25,000 genotyped animals, we removed genotypes of the oldest animals in favor of the latest genotyped cows and male selection candidates.

2.4 Analysis of scenarios

All scenarios had equal resources. We compared the scenarios based on their final genetic gain, which indicated return on investment, accuracy of selection, and selection intensity. We measured the genetic gain as an average true breeding value by year of birth and standardized it to have zero mean and unit standard genetic deviation in the first year of comparison. We measured the accuracy of breeding values as the correlation between true and estimated breeding values. We measured the accuracy separately for four groups of animals: i) male candidates (genotyped and non-phenotyped); ii) sires (currently used in artificial insemination); iii) female candidates (non-genotyped and non-phenotyped); and iv) cows (all active phenotyped cows and bull dams). We computed the intensities of two-stage selection (i) by integrating standard bivariate distribution $f(x,y,\rho)=\frac{1}{2\pi\sqrt{1-\rho}}e^{\frac{-1}{2(1-\rho)}(x^2+y^2-2\rho\cdot x\cdot y)}$ as $i=\frac{\int_{T_x}^{x}\int_{T_y}^{y} f(x,y,\rho)\cdot x\cdot dy\cdot dx}{p_2}$, where x is the parent average, y is the genomic breeding value, and ρ is the correlation between the variables (Young, 1964; Jopson et al., 2004). We computed the correlation ρ by dividing the scenario-specific accuracy of parent average by the accuracy of genomic prediction. T_x is the standardized cut-off for the proportion of all new-born male calves selected for genomic testing (p_1) based on parent average in the first selection stage. T_y is the standardized cut-off for the proportion of all newborn males selected as sires (p_2) based on genomic breeding values in the second selection stage. We repeated simulation of the base population and each scenario 10 times and summarized them with mean and standard deviation across the replicates. We used Tukey's multiple comparison test to test the significance of the difference between means.

3 Results

Genomic scenarios increased the genetic gain compared to the conventional scenario regardless of the number of phenotype records per lactation, relative cost of phenotyping to genotyping, and the availability of an initial training population. Genomic scenarios with an initial training population increased the genetic gain of the conventional scenario by up to 143%, despite reduced phenotyping. Genetic gain increased with increasing investment into genotyping. Genomic scenarios increased accuracy for non-phenotyped male and female candidates, and cows. Scenarios without an initial training population showed the same trends for genetic gain and accuracy. Although these scenarios

had a slightly smaller genetic gain due to delayed implementation of genomic selection, they still increased the genetic gain of the conventional scenario by up to 134%. We present these results in more details in the following sub-sections separately for settings with and without an initial training population available.

3.1 Genetic gain with an initial training population

Genomic scenarios with an initial training population increased the genetic gain of the conventional scenario using the same resources. The genetic gain increased with increasing investment in genotyping, despite reduced phenotyping (Figure 1 and Table S2). We show the corresponding intensities of sire selection in Table S3. In the $P:$G = 1:1 setting, the genomic scenarios increased the genetic gain of the conventional scenario between 79% and 143%. By reducing the number of phenotype records from 11 (C11) to 10 per lactation (G10), we saved resources for genotyping 355 animals per year (45 male candidates). This small change increased male selection intensity from 0.10 to 0.26 and coupled with a shorter generation interval increased the genetic gain by 79% (from 3.01 to 5.41). By reducing the number of phenotype records to nine or eight per lactation (G9 or G8), we respectively saved resources to genotype 800 or 1,345 animals per year (100 or 165 male candidates). This respectively increased male selection intensity to 0.39 or 0.49, and genetic gain by 109% or 120% (from 3.01 to 6.30 or 6.62). We achieved the highest genetic gain, between 135% and 143% of the conventional scenario (between 7.07 and 7.33), when we collected between five and one phenotype records per lactation. In these three scenarios we saved resources for genotyping between 3,230 and 3,850 (all) cows and between 465 and 1,125 male candidates per year, and achieved male selection intensity between 0.77 and 0.92.

Changing the relative cost of phenotyping to genotyping did not change the trend in genetic gain. In the G10 scenario of the $P:$G = 1:2 setting we yearly genotyped 182 animals (22 male candidates) and increased the genetic gain by 80% (from 3.01 to 5.43). In the G10 scenario of the $P:$G = 2:1 setting we yearly genotyped 710 animals (90 males candidates) and increased the genetic gain by 116% (from 3.01 to 6.50). When we maximized the investment into genotyping (G1), we genotyped between 565 and 2,245 male candidates and all females. This achieved a comparable genetic gain, between 136% and 143% of the conventional scenario, regardless of the relative cost of phenotyping to genotyping and male selection intensities.

The high-genotyping scenarios achieved the observed genetic gain without using all the resources (marked bold in Table S2). In these scenarios the resources designated to genotyping females exceeded the cost of genotyping all females. These savings could cover additional 11 phenotype records per lactation for between 169 and 5,950 animals, or between 85 and 11,900 additional genotypes.

In Figure 1 we also show the growth of the training population for genomic prediction. The training population started with ~10,000 individuals and grew until reaching 25,000 individuals. The increase was not linear through all generations, since the procedure for choosing the training animals changed when the training population exceed 25,000 (only latest females and male candidates included)

3.2 Accuracy with an initial training population

Compared to the conventional scenario, genomic scenarios increased accuracy for young non-phenotyped and genotyped male, non-phenotyped and non-genotyped female candidates, and cows, but decreased accuracy for sires (Figure 2 and Table S4). In the $P:$G=1:1 setting, the accuracy for young genomically tested male candidates ranged between 0.90 and 0.91 regardless of the amount of phenotyping and genotyping. This accuracy was between 0.53 and 0.54 higher compared to the pre-selection for progeny testing and between 0.03 and 0.04 lower compared to the sire selection in the conventional scenario. In contrast, the accuracy for already selected sires decreased with decreasing investment into phenotyping and was between 0.11 and 0.23 lower than in conventional scenario. We observed the lowest accuracy for sires (0.63) when we invested the most into genotyping (G1) and the highest (0.75) when we invested the most into phenotyping (G10).

The accuracy for non-genotyped female candidates and cows increased with increasing genotyping, despite reduced phenotyping. We observed the highest accuracy for female candidates (0.55 - 0.57) and cows (0.77 - 0.79) when we recorded between five and one phenotype record per lactation and invested the rest into genotyping. Compared to the conventional scenario, the genomic scenarios increased the accuracy between 0.03 and 0.11 for female candidates, and between 0.11 and 0.29 for cows.

Changing the relative cost of phenotyping to genotyping affected primarily the accuracy for female candidates and cows. In the majority of scenarios the accuracy increased with decreasing the relative cost of genotyping, which enabled more genotyping. We observed the largest difference of 0.06 for female candidates and 0.12 for cows when we changed the relative cost of phenotyping from half to twice the cost of genotyping. Changing the relative costs, however, did not change the accuracy trends.

3.3 Scenarios without an initial training population

3.3.1 Genetic gain

Genomic scenarios without an initial training population also increased the genetic gain of the conventional scenario using the same resources. The trends were in line with what we observed with an initial training population, that is, increasing genotyping increased genetic gain despite reduced phenotyping (Figure 3). However, all corresponding scenarios achieved between 2% and 28% smaller genetic gain than when an initial training population was available (Table S2).

In the $P:$G setting, genomic scenarios increased the genetic gain of the conventional scenario between 51% and 131%. Compared to when we had an initial training population, the corresponding scenarios achieved between 2% and 16% lower genetic gain. This difference was the largest when we invested the least into genotyping (G10). In this scenario we needed six years to build a training population of 2,000 cows and implement genomic selection, since we only genotyped 355 cows per year. We observed the smallest difference in the scenario that collected two phenotype records per lactations (G2) and implemented genomic selection in the first year.

Changing the relative cost of phenotyping to genotyping did not change the overall trend. In the $P:$G=1:2 setting, the genomic scenarios increased genetic gain of the conventional scenario between 31% and 126%. That was between 4% and 28% less than corresponding scenarios with an initial training population. In the $P:$G=2:1 setting, the genomic scenarios increased the genetic gain of the conventional scenario between 86% and 134%, which was between 3% and 14% less than corresponding scenarios with an initial training population.

3.3.2 Accuracy

Similar to the scenarios with an initial training population, genomic scenarios without it increased the accuracy for non-phenotyped male and female candidates, and cows (Figure 3 and Table S4). In the $P:$G=1:1 setting the accuracy for male candidates ranged between 0.84 and 0.91. In contrast to scenarios with an initial training population, the accuracy increased with increasing investment into genotyping. The accuracy for sires ranged between 0.64 and 0.74. Contrary to when we had an initial training population, we observed no clear trend of either increasing or decreasing accuracy with decreasing investment into phenotyping. For female candidates the accuracy ranged between 0.47 and 0.56, and for cows between 0.56 and 0.76. For both the accuracies followed the trends of when we had an initial training population - increasing genotyping increased the accuracy.

Changing the relative cost of phenotyping to genotyping affected the accuracy for non-genotyped female candidates, cows and male candidates. Decreasing the relative cost of genotyping to phenotyping increased the accuracy in the majority of the scenarios, particularly the low-genotyping ones.

4 Discussion

Our results show that any dairy breeding program using conventional progeny testing with repeated milk records can implement genomic selection without extra costs. While breeding programs have established funding for phenotyping, not all of them have established funding for genotyping. We show that by reallocating a part of phenotyping resources into genotyping, breeding programs can implement genomic selection and substantially increase genetic gain regardless of the amount and cost of genotyping, and availability of an initial training population. However, increasing investment in genotyping has diminishing returns, which suggests that breeding programs should optimize the investment into phenotyping and genotyping to maximize return on investment for selection and management. The results raise four discussion points: 1) how optimizing the investment in phenotyping and genotyping affects genetic gain; 2) how optimizing the investment in phenotyping and genotyping affects accuracy; 3) implications for dairy breeding programs; and 4) limitations of the study. We first discuss the results under equal cost of phenotyping and genotyping, and an initial training population available. We then discuss changes at different costs and no initial training population.

4.1 Genetic gain with an initial training population

4.1.1 Genomic vs. conventional selection

Implementing genomic selection by optimizing the investment in phenotyping and genotyping increased genetic gain compared to the conventional selection, mainly due to reduced generation interval in sire selection paths. This improvement is in agreement with previous theoretical studies (Schaeffer, 2006; Pryce et al., 2010; Obšteter et al., 2019). Empirical studies confirm this; in the US Holstein population the generation interval for the sires of sires and sires of dams paths recently decreased between 25% and 50% compared to the conventional selection (García-Ruiz et al., 2016). Van Grevenhof et al. (2012) also showed that when genomic selection halves the generation interval, a training population with ~2,000 individuals with own performance or ~3,500 individuals with ten progeny gives comparable response as conventional selection for a trait with intermediate heritability.

Another major advantage of the genomic scenarios was increased intensity of sire selection. A costly and lengthy progeny-testing limits the number of tested male candidates in conventional selection. Genomic selection significantly reduces the cost of testing (Schaeffer, 2006) and thus allows for testing more male candidates. In the US Holstein population, genomic selection improved the selection differential for all traits, particularly for traits with low heritability, such as health and fertility (García-Ruiz et al., 2016).

4.1.2 Increasing the investment into genotyping

Genetic gain increased with increased investment into genotyping. This was mainly due to higher intensity of sire selection, since more resources for genotyping allowed us to test more male candidates while selecting the same number. A larger investment into genotyping also increased the update and total size of the training population, which increased the accuracy of female selection (we discuss this in the next sub-section).

The genetic gain had diminishing relationship with investment into genotyping. This has important implications for dairy breeding programs, since they use phenotypes also for management, and we discuss this separately. The results showed that investing resources of more than six phenotype records into genotyping did not significantly improve the genetic gain. There are three reasons for this. First, increasing female training population has diminishing relationship with genetic gain (Van Grevenhof et al., 2012; Gonzalez-Recio et al., 2014). Since scenarios with an initial training population started with ~10,000 genotyped and phenotyped cows, enlarging the training population had a marginal effect. Consequently, the accuracy of sire selection in genomic scenario was high regardless of the amount of genotyping. Second, increasing investment into genotyping did not proportionally increase the size of the training population due to the limit of 25,000 animals of the training population. And third, the intensity of sire selection had diminishing return with increasing number of genotyped male candidates. This agrees with Reiner-Benaim et al. (Reiner-Benaim et al., 2017) that showed an increased genetic gain with increasing the number of tested male candidates, but with a diminishing return. While they achieved the maximum profit with four selected sires out of 1,721 tested candidates, they achieved 99% or 90% of the maximum profit with respectively 740 or 119 tested candidates. The same three reasons enabled comparable maximum genetic gain regardless of the relative price of phenotyping to genotyping. In general, selecting less than 2% of the

tested males and updating the training population with at least 35% of first-parity cows resulted in the maximum genetic gain.

While genetic gain increases with the number of cows in training population, it does not increase with the number of repeated records. The scenarios with the largest genetic gain therefore had a training population with many cows and few repeated records (Figure S1). However, since we used the single-step genomic prediction, the phenotypes of the non-genotyped animals contributed to the estimation as well. Effectively, all scenarios thus operated with the same number of phenotyped animals.

We should emphasize, that some of the high-genotyping scenarios achieved the observed genetic gain at a lower total cost, since they could not use all the saved resources for genotyping females in the studied population. The saved resources could be invested back into phenotyping females for milk production or novel traits, genotyping more male candidates, or other breeding actions.

4.2 Accuracy with an initial training population

Despite reduced phenotyping, genomic scenarios increased the accuracy for young non-phenotyped calves and cows. When the accuracy of parent average is already high, genomic prediction increases primarily the accuracy of the Mendelian sampling term. But when the accuracy of parent average is low, such as for the offspring of parents with little or no own or progeny information, genomic information increases accuracy both for the parent average and the Mendelian sampling term (Daetwyler et al., 2007; Wolc et al., 2011).

4.2.1 Accuracy for males

For male candidates, genomic prediction more than doubled the accuracy compared to the parent average used for pre-selection of male calves for progeny testing. This is in agreement with two-fold accuracy increase in dairy (Schaeffer, 2006) and layers (Wolc et al., 2011). Within the genomic scenarios, the accuracy for male candidates was high regardless of the amount of genotyping and phenotyping for two reasons. First, the accuracy of their parent average was high, since we tested offspring of elite matings. Second, starting with an initial 10,000 training population gave an adequate accuracy that was additionally boosted by using all available information jointly through the single-step genomic prediction. Using single-step genomic prediction also removed the bias due to pre-selection (Jibrila et al., 2020).

In contrast, reducing phenotyping decreased the accuracy of already selected sires. We believe this is due to two reasons. First, since sires are the very best animals, their breeding values are close together in the tail of the distribution. Due to small differences between the sires, each additional phenotypic record helps to differentiate them and thus increases the accuracy. Second, as we invested more into genotyping, the training population grew quicker and reached the limit of 25,000. At this point we removed sires' in favor of cows' genotypes, hence prediction for sires depended only on daughters' data and no longer on their own genotype. However, since this is the accuracy after the selection has already been made, it is not of great interest for breeding.

4.2.2 Accuracy for females

Genomic scenarios increased the accuracy for cows compared to the conventional scenario. Besides increasing the accuracy of Mendelian sampling term, using genomic information increases genetic connectedness between individuals from different management units (Yu et al., 2017; Powell et al., 2019). Increased connectedness in turn increases the accuracy of prediction regardless of the heritability and the number of causal loci or markers (Yu et al., 2018). This improvement is important because we selected bull dams for elite mating from cows.

The accuracy for cows increased with increasing investment into genotyping, despite reduced phenotyping due to three reasons. First, more cows had both genomic and phenotypic information, which increased the accuracy of their estimated breeding values. Second, more genotyped cows increased genetic connectedness (Yu et al., 2018). And third, investing more into genotyping translated into larger training population and its yearly update. As shown by previous studies (Van Grevenhof et al., 2012; Gonzalez-Recio et al., 2014), the accuracy of genomic prediction increases with increasing the size of a female training population. They showed that the accuracy of 0.70 is achieved with ~20,000 animals as in our study. Same studies shown that, as with genetic gain, accuracy had a diminishing return with the size of the training population. We observed plateau in accuracy when we invested more than six phenotype records (out of eleven) into genotyping.

Accuracy for female candidates followed the trend for cows, but at lower values. Female candidates were not genotyped nor phenotyped, hence their accuracy mainly reflected the accuracy of their parent average. Increasing genotyping increased the accuracy for cows and in turn increased the accuracy of female candidate`s parent average. The benefit of this increase was not large, since the intensity of female selection was low. However, there is potential for this benefit to be larger with sexed semen and embryo transfer.

4.3 Scenario without an initial training population

4.3.1 Genetic gain

We also considered that some populations do not have access to an initial training population and have to initialize one themselves. These genomic scenarios still increased genetic gain compared to the conventional scenario, but achieved lower genetic gain than corresponding scenarios with an initial training population available. Increasing the investment into genotyping compensated for starting without a training population in two ways. First, it shortened the time to obtain the targeted 2,000 genotypes required to implement genomic selection down to one year in high-genotyping scenarios. Second, it shortened the time to build a training population in which an additional record had negligible effect on accuracy (Gonzalez-Recio et al., 2014).

4.3.2 Accuracy

Accuracy in scenarios without an initial training population closely followed the trends of the corresponding scenarios with an initial training population available. We observed minor differences in the low genotyping scenarios that had reduced accuracy for male candidates and sires. We attribute this to a smaller training population. Buch et al. (2012) showed that for new traits and large-scale

recording, we can achieve 75% of the maximum genomic accuracy within first two to three years of recording. In our study we shortened this period even more by including the historical data through the single-step genomic prediction.

4.4 Implications

The results suggest that any dairy breeding program using conventional progeny testing with repeated milk records can implement genomic selection without extra costs by optimizing the investment of resources into breeding actions. Here we propose funding the genotyping with a part of resources for milk recording, since we can manipulate the number of repeated records. Breeding programs could reduce phenotyping for a different trait that they record repeatedly and is perhaps less crucial for management. They could also reallocate the funds from another breeding action.

Additionally, we could optimize which individuals to genotype and phenotype, as well as the computational costs . Selective phenotyping was shown to increase the accuracy of genomic prediction up to 20% with small sample sizes in plant breeding (Heslot & Feoktistov, 2017; Akdemir & Isidro-Sánchez, 2019). Similarly, selective genotyping of cows from the distribution tails has been shown to increase the accuracy of genomic prediction by 15% (Jenko et al., 2017). We expect this would further increase the return on investment, but increase the complexity of optimization. Regarding computing costs, the problem of a large number of genotypes can be alternatively solved by using methods with reduced computational costs. Examples of such methods are the algorithm for proven and young (Misztal et al., 2014) and singular value decomposition of the genotype matrix (Ødegård et al., 2018). Also, as shown in our study, we can achieve large genetic gain with a relatively small training population of recent genotypes.

The economic efficiency of breeding programs strongly depends on which stakeholders fund which breeding action. Different programs have different investment schemes, often complex. The scenarios presented in this paper are of little value for programs where funding for phenotyping and genotyping is disconnected. Similarly, optimizing the investment into phenotyping is not of interest for breeding programs with abundant use of automated milking systems where the cost of phenotyping does not depend on the number of records. However, in populations with small herds the use of automated milking systems is limited. Further on, genomic selection could benefit some settings more than others. For example, genomic information is especially important for generating genetic connectedness in systems with small herd sizes, geographically dispersed farms, and limited use of artificial insemination, often found in low- to mid-income countries (Powell et al., 2019). The same benefits are expected for small ruminant programs that do not actively exchange sires between herds (Kasap et al., 2018).

High cost of genotyping diminishes the benefit of the proposed solutions. The relative cost of phenotyping to genotyping at which genomic selection is not beneficial to conventional selection depends on a number of factors: i) the number of females in the recorded population, since it dictates the savings from reducing the phenotyping; ii) the number of phenotype records the breeders are willing to sacrifice; iii) the availability of an initial training population; and iv) the ratio of genotyped males and females. In our case, if genotyping was ten-times more expensive than phenotyping with

11 records, we would save resources to genotype between 36 (ten phenotypic records) and 900 (one phenotypic record) animals. While such numbers of genotyped animals could allow for efficient genomic selection if we had an initial training population and genotyped only males, it would probably not be viable if we had to build the training population ourselves.

We did not account for the benefits of genotyping besides genomic selection. Genomic information has additional value for i) parentage verification, parentage discovery, or correction of parentage errors; ii) management of monogenic diseases and traits, which can prevent economic losses caused by spreading lethal alleles or create economic gain by adding value to the products; iii) better monitoring and control of inbreeding (Sonesson et al., 2012) and optimization of matings (Obšteter et al., 2019); iv) determination of animals' breed composition, which serves to identify the most appropriate breed or cross-breed in a production system, or improve the structure of the training population (Marshall et al., 2019). Additional uses of genotypes increase the return on investment beyond what we measured in this study.

4.5 Limitations of the study

4.5.1 Reducing the number of phenotype records

Balancing phenotyping and genotyping can lead to conflicts between managing production (short-term goal) and achieving genetic gain (long-term goal). Producers use phenotype records to manage animals' health, reproduction and feed composition, which affect the success of dairy cattle production in the short-term. Besides managing production, milk recording is also important from an environmental perspective (Verbič et al., 2019), but so is genetic improvement. In general, about half of phenotypic improvement is due to management and half due to selection (Dekkers & Hospital, 2002).

The ownership of the data that drives dairy improvement is a matter of discussion in many breeding programs. When data collection is subsidized, the data are usually free for management and genetic improvement. Finding optimal frequency of recording for management and genetic improvement in such systems is crucial for optimal return on investment. When data collection is paid by producers, they can be available to breeding organizations for free or at a cost. Depending on the cost breeding organizations can save some resources by purchasing smaller number of repeated phenotypes per individual and genotype more selection candidates instead.

While genotype data is currently largely used for selection, the same data could also support management (predicting feed requirements, disease liability, etc.). Therefore, evaluating the value of phenotype and genotype data is complex and beyond the scope of this study. One possible way forward would be to compare variance between herd-test day effects and genetic variance to contrast the value of managing production and genetic gain in addition to comparing phenotypic and genetic trends (Dekkers & Hospital, 2002).

In practice, test day records are also used to compute the 305-day milk yield (International Committee for Animal Recording, 2020). The longest sampling interval tested in our study and still approved by ICAR was nine weeks, which yielded five records per lactation. Previous studies

showed that the correlation of predicting 305-day milk using five records per lactation with using either weekly or eleven records per lactation was between 0.98 and 0.99 (Pool and Meuwissen, 1999; Berry et al. 2005). However, some studies showed that prediction using less than eleven records can yield substantial bias (Gantner et al., 2008)

4.5.2 Single additive trait

We simulated milk yield as a single polygenic trait with additive genetic as well as herd, permanent environment, and residual environmental effects. We did not simulate nor account for non-additive genetic effects that also affect dairy performance (Fuerst & Sölkner, 1994; Ertl et al., 2014; Jiang et al., 2017). We note that we simulated permanent environment effects, which capture non-additive genetic effects and other individual specific environmental effects. We also simulated milk yield in different lactations as a single trait with constant heritability, whereas genetic correlation between different lactations and through the lactation is not unity (Meyer, 1984; Swalve & Vleck, 1987; Dong & van Vleck, 1989). If genetic correlation is less than one, the repeatability of the phenotype decreases and the value of a repeated record and its contribution to accuracy diminishes.

We simulated a trait with heritability of 0.25, since the majority of production traits recorded repeatedly show intermediate heritability. Previous studies also provide insights in how changing the heritability of the phenotype would affect the results. On one hand, at a lower heritability we would need more females in the training population until the contribution of additional female was negligible (Gonzalez-Recio et al., 2014). On the other hand, genomic selection is less affected by the heritability than conventional selection and hence more beneficial for traits with low heritability (Lillehammer et al., 2011; García-Ruiz et al., 2016).

4.5.3 Genomic selection of females

We did not use genomic selection for females, nor did we use reproductive technologies such as sexing semen or embryo transfer. This would further decrease the generation interval, increase selection intensity on female side, and in turn increase genetic gain of genomic scenarios even more (Pryce et al., 2010; García-Ruiz et al., 2016). Such an implementation of genomic selection requires only a minor modification of the design used in this study - genotyping heifers instead of first-parity cows. However, reproductive technologies require a larger modification and investment. Some of the scenarios saved resource and could invest into these technologies.

5 Author Contributions

JO designed the study, ran the simulation, analyzed the data, interpreted the results, and wrote the manuscript. JJ participated in study design, results interpretation, and manuscript revision. GG has initiated and supervised the work and contributed to all its stages.

6 Funding

JO acknowledges support from Agricultural Institute of Slovenia and core financing of Slovenian Research Agency (grant P4-0133). GG acknowledges support from the BBSRC to The Roslin Institute (BBS/E/D/30002275) and The University of Edinburgh’s Data-Driven Innovation Chancellor’s fellowship.

7 Acknowledgments

The authors acknowledge I. Pocrnić (The Roslin Institute, University of Edinburgh) for his help in interpreting the results and comments on the manuscript.

8 References

Akdemir, D., and J. Isidro-Sánchez. 2019. Design of training populations for selective phenotyping in genomic prediction. Sci Rep. 9:1446. https://doi.org/10.1038/s41598-018-38081-6

Berry, D. P., V. E. Olori, A. R. Cromie, R. F. Veerkamp, M. Rath, and P. Dillon. 2005. Accuracy of predicting milk yield from alternative milk recording schemes. Animal Science, 801:53–60. https://doi.org/10.1079/ASC34880053

Buch, L. H., M. Kargo, P. Berg, J. Lassen, and A. C. Sørensen. 2012. The value of cows in reference populations for genomic selection of new functional traits. Animal. 6:880–886. https://doi.org/10.1017/S1751731111002205

Clark, S. A., J. M. Hickey, and J. H. J. van der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. Genet. Sel. Evol. 43:18. https://doi.org/10.1186/1297-9686-43-18

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. Journal of Animal Breeding and Genetics, 124:369–376.

Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. PLoS ONE, 3:e3395. https://doi.org/10.1371/journal.pone.0003395

Dekkers, J. C. M. and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. Nat Rev Genet. 3:22–32. https://doi.org/10.1038/nrg701

Dong, M. C. and L. D. van Vleck. 1989. Correlations Among First and Second Lactation Milk Yield and Calving Interval. J. Dairy Sci. 72:1933–1936. https://doi.org/10.3168/jds.S0022-0302(89)79313-9

Ducrocq, V., D. Laloe, M. Swaminathan, X. Rognon, M. Tixier-Boichard, and T. Zerjal. 2018. Genomics for Ruminants in Developing Countries: From Principles to Practice. Front Genet. 9:251. https://doi.org/10.3389/fgene.2018.00251

Ertl, J., A. Legarra, Z. G. Vitezica, L. Varona, C. Edel, R. Emmerling, and K. U. Götz. 2014. Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. Genet Sel Evol. 46:40. https://doi.org/10.1186/1297-9686-46-40

Fuerst, C. and J. Sölkner. 1994. Additive and Nonadditive Genetic Variances for Milk Yield, Fertility, and Lifetime Performance Traits of Dairy Cattle. J Dairy Sci. 77:1114–1125. https://doi.org/10.3168/jds.S0022-0302(94)77047-8

Gantner, V., S. Jovanovac, N. Raguž, M. Klopčič and D. Solić. 2008. Prediction of lactation milk yield using various milk recording methods. Biotechnol Anim Husb. 24: 9–18.

García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggans, F. J. Ruiz-López, and C. P. V. Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. PNAS. 113:E3995–E4004. https://doi.org/10.1073/pnas.1519061113

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica. 136:245–257. https://doi.org/10.1007/s10709-008-9308-0

Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. J Anim Breed Genet. 128:409–421. https://doi.org/10.1111/j.1439-0388.2011.00964.x

Gonzalez-Recio, O., M. P. Coffey, and J. E. Pryce. 2014. On the value of the phenotypes in the genomic era. J Dairy Sci. 97:7905–7915. https://doi.org/10.3168/jds.2014-8125

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol. 42:5. https://doi.org/10.1186/1297-9686-42-5

Heslot, N. and V. Feoktistov. 2017. Optimization of selective phenotyping and population design for genomic prediction. BioRxiv. 172064. https://doi.org/10.1101/172064

ICAR DNA Working Group. 2017. ICAR Guidelines for Parentage Verification and Parentage Discovery Based on SNP Genotypes. The International Committee for Animal Recording.

International Committee for Animal Recording. 2017. Section 02—Cattle Milk Recording. Overview. In ICAR Guidelines. ICAR.

International Committee for Animal Recording. 2020. Section 02—Cattle Milk Recording. Procedure 2—Computing Lactation Yield. In ICAR Guidelines. ICAR.

Jenko, J., G. R. Wiggans, T. A. Cooper, S. A. E. Eaglen, W. G. de L. Luff, M. Bichard, R. Pong-Wong, and J. A. Woolliams. 2017. Cow genotyping strategies for genomic selection in a small dairy cattle population. J Dairy Sci. 100:439–452. https://doi.org/10.3168/jds.2016-11479

Jiang, J., B. Shen, J. R. O'Connell, P. M. VanRaden, J. B. Cole, and L. Ma. 2017. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. BMC Genomics. 18:425. https://doi.org/10.1186/s12864-017-3821-4

Jibrila, I., J. ten Napel, J. Vandenplas, R. F. Veerkamp, and M. P. L. Calus. 2020. Investigating the impact of preselection on subsequent single-step genomic BLUP evaluation of preselected animals. Genet Sel Evol. 52:42. https://doi.org/10.1186/s12711-020-00562-6

Jopson, N. B., P. R. Amer, and J. M. McEwan. 2004: Comparison of two-stage selection breeding programmes for terminal sire sheep. Pages 212-216 in Proceedings of the New Zealand Society of Animal Production 64.

Kasap, A., B. Mioc, J. M. Hickey, and G. Gorjanc. 2018. Genetic connectedness in the U.S. sheep industry. Page 590 in Book of Abstracts of the 69th Annual Meeting of the European Federation of Animal Science. Wageningen Academic Publishers, Wageningen, Netherlands. https://doi.org/10.3920/978-90-8686-871-1

Lidauer, M., E. A. Mäntysaari, and I. Strandén. 2003. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. Livest Prod Sci. 79: 73–86. https://doi.org/10.1016/S0301-6226(02)00142-2

Lillehammer, M., T. H. E.Meuwissen, and A. K. Sonesson. 2011. A comparison of dairy cattle breeding designs that use genomic selection. J Dairy Sci. 94:493–500. https://doi.org/10.3168/jds.2010-3518

Marshall, K., J. P. Gibson, O. Mwai, J. M. Mwacharo, A. Haile, T. Getachew, R. Mrode, and S. J. Kemp. 2019. Livestock Genomics for Developing Countries – African Examples in Practice. Front Genet, 10:297. https://doi.org/10.3389/fgene.2019.00297

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819–1829.

Meyer, K. 1984. Estimates of genetic parameters for milk and fat yield for the first three lactations in British Friesian cows. Animal Science. 38:313–322. https://doi.org/10.1017/S0003356100041519

Misztal, I., A. Legarra, and I. Aguilar. 2014. Using recursion to compute the inverse of the genomic relationship matrix. J Dairy Sci. 97:3943–3952. https://doi.org/10.3168/jds.2013-7752

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for BLUPF90 family programs. University of Georgia. http://nce.ads.uga.edu/wiki/doku.php?id=documentation

Mrode, R. A. 2005. Linear Models for the Prediction of Animal Breeding Values (Second edition). CAB International, Wallingford, UK.

Mrode, R., J. M. K. Ojango, A. M. Okeyo, and J. M. Mwacharo. 2019. Genomic Selection and Use of Molecular Tools in Breeding Programs for Indigenous and Crossbred Cattle in Developing Countries: Current Status and Future Prospects. Front Genet. 9:964. https://doi.org/10.3389/fgene.2018.00694

Obšteter, J., J. Jenko, J. M. Hickey, and G. Gorjanc. 2019. Efficient use of genomic information for sustainable genetic improvement in small cattle populations. J Dairy Sci. 102:9971–9982. https://doi.org/10.3168/jds.2019-16853

Ødegård, J., U. Indahl, I. Strandén, and T. H. E. Meuwissen. 2018. Large-scale genomic prediction using singular value decomposition of the genotype matrix. Genet Sel Evol. 50:6. https://doi.org/10.1186/s12711-018-0373-2

Pool, M. H. and T. H. E. Meuwissen. 1999. Prediction of Daily Milk Yields from a Limited Number of Test Days Using Test Day Models. J Dairy Sci. 82:1555–1564. https://doi.org/10.3168/jds.S0022-0302(99)75383-X

Powell, O., R. Mrode, C. R. Gaynor, M. Johnsson, G. Gorjanc, and J. M. Hickey. 2019. Genomic data enables genetic evaluation using data recorded on LMIC smallholder dairy farms. BioRxiv. 827956. https://doi.org/10.1101/827956

Pryce, J. E., M. E. Goddard, H. W. Raadsma, and B. J. Hayes. 2010. Deterministic models of breeding scheme designs that incorporate genomic selection. J Dairy Sci. 93:5455–5466. https://doi.org/10.3168/jds.2010-3256

Reiner-Benaim, A., E. Ezra, and J. I. Weller. 2017. Optimization of a genomic breeding program for a moderately sized dairy cattle population. J Dairy Sci. 100:2892–2904. https://doi.org/10.3168/jds.2016-11748

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet. 123:218–223. https://doi.org/10.1111/j.1439-0388.2006.00595.x

Sonesson, A. K., J. A. Woolliams, and T. H. E. Meuwissen. 2012. Genomic selection requires genomic control of inbreeding. Genet Sel Evol. 44:27. https://doi.org/10.1186/1297-9686-44-27

Swalve, H. and L. D. V. Vleck. 1987. Estimation of Genetic (Co) Variances for Milk Yield in First Three Lactations Using an Animal Model and Restricted Maximum Likelihood. J Dairy Sci. 70:842–849. https://doi.org/10.3168/jds.S0022-0302(87)80082-6

Van Grevenhof, E. M., J. A. M. Van Arendonk, and P. Bijma. 2012. Response to genomic selection: The Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. Gene Sel Evoul. 44:26. https://doi.org/10.1186/1297-9686-44-26

Verbič, J., J. Jenko, J. Jeretina, and D. Babnik. 2019. Milk urea concentration as a tool to reduce the nitrogen footprint of milk production in conditions of small scale farming. In Towards Precision Livestock Husbandry and Its Potential to Mitigate Ammonia and GHG Emissions: Abstracts' Leaflet. The 4th liveAGE meeting, Galilee, Israel.

Wiggans, G. R., J. B. Cole, S. M. Hubbard, and T. S. Sonstegard. 2017. Genomic Selection in Dairy Cattle: The USDA Experience. Annu Rev Anim Biosci. 5:309-327. https://doi.org/10.1146/annurev-animal-021815-111422

Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, and J. C. Dekkers. 2011. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. Genet Sel Evol. 43:23. https://doi.org/10.1186/1297-9686-43-23

Young, S. 1964. Multi-stage selection for genetic gain. Heredity. 19**:**131–145. https://doi.org/10.1038/hdy.1964.11

Yu, H., M. L. Spangler, R. M. Lewis, and G. Morota. 2017. Genomic Relatedness Strengthens Genetic Connectedness Across Management Units. G3 (Bethesda). 7:3543–3556. https://doi.org/10.1534/g3.117.300151

Yu, H., M. L. Spangler, R. M. Lewis, and G. Morota. 2018. Do stronger measures of genomic connectedness enhance prediction accuracies across management units? J Anim Sci. 96: 4490–4500. https://doi.org/10.1093/jas/sky316

Table 1: Number of genotyped animals per year by scenario and relative cost of phenotyping to genotyping ($P:$G). Scenarios are named "G" for genomic, followed by the number of phenotype records per lactation. For the $P:$G we compared the cost of 11 phenotype records per lactation to the cost of one genotype. The number of phenotype records and the $P:$G dictated the number of genotyped animals. We genotyped females (F) and males (M) in 7:1 ratio. We genotyped the females to update and increase the training population and males for selection.

	Scenario					
Relative cost	G10	G9	G8	G5	G2	G1
$P:$G = 1:2	160 F 22 M	350 F 50 M	590 F 85 M	1,610 F 235 M	3,230 F 465 M	3,850 F 565 M
$P:$G = 1:1	310 F 45 M	700 F 100 M	1,180 F 165 M	3,230 F 465 M	3,850 F 925 M	3,850 F 1,125 M
$P:$G = 2:1	620 F 90 M	1,400 F 200 M	2,360 F 335 M	3,850 F 925 M	3,850 F 1,845 M	3,850 F 2,245 M

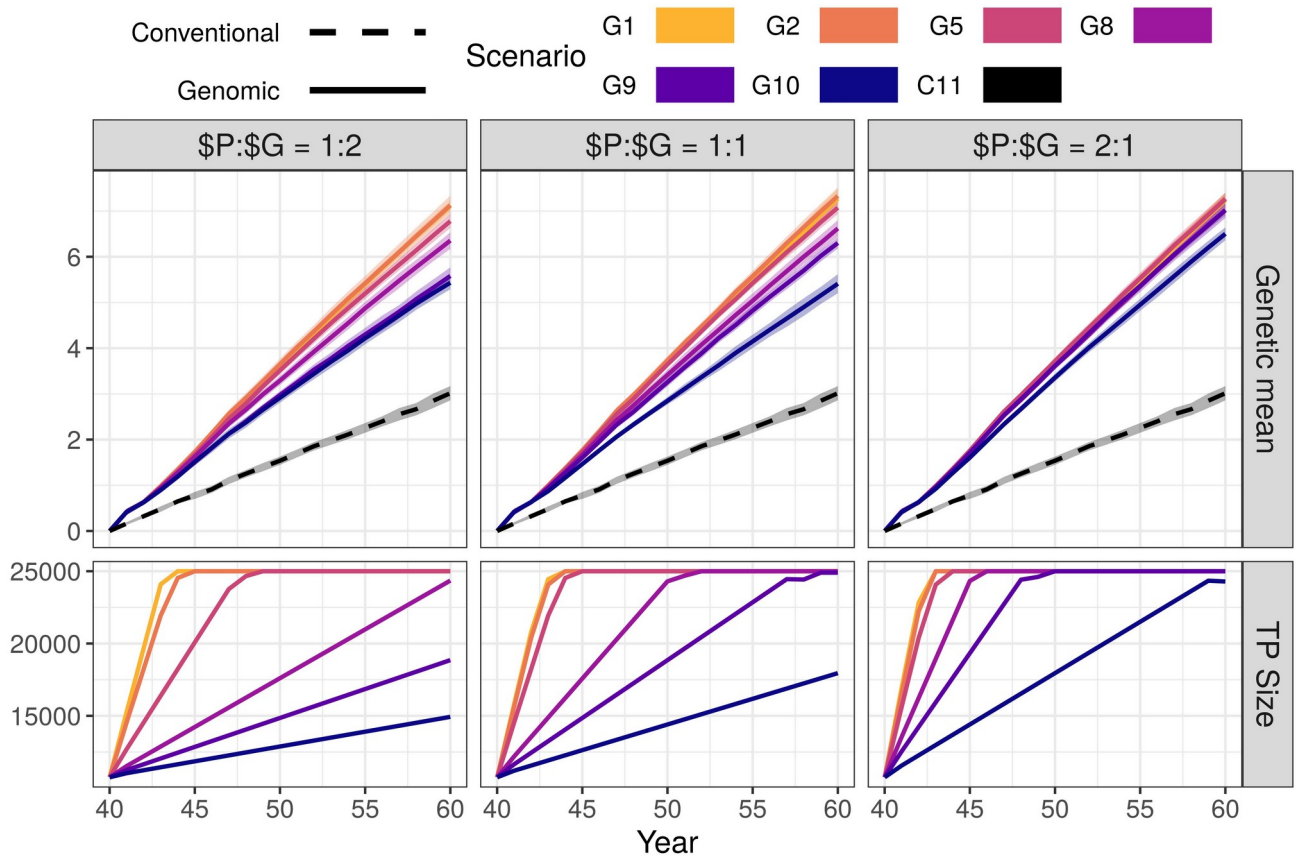

Figure 1. Genetic gain and training population size by scenario and relative cost of phenotyping to genotyping ($P:$G) with an initial training population (TP). The figure presents the means (lines) and 95% confidence intervals (polygons) across 10 replicates for the conventional (C) and genomic (G) scenarios, with numbers indicating the number of phenotype records per lactation. For the $P:$G we compared the cost of 11 phenotype records per lactation to the cost of one genotype.

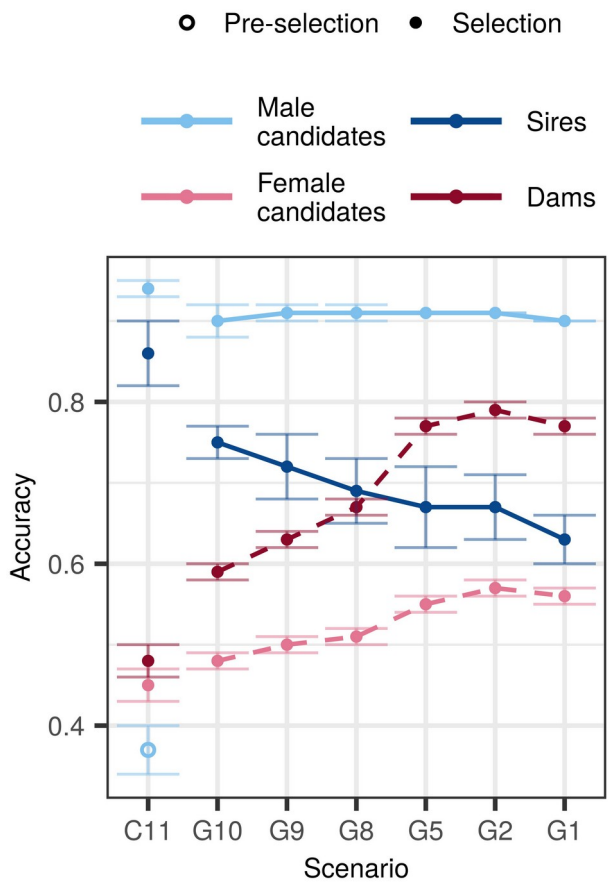

Figure 2. Accuracy by scenario with an initial training population and equal cost of phenotyping and genotyping. The figure presents the means (lines) and 95% confidence intervals (error bars) across 10 replicates for the conventional (C) and genomic (G) scenarios with numbers indicating the number of phenotype records per lactation. The cost of phenotyping was defined as the cost of 11 phenotype records per lactation, and the cost of genotyping as the cost of one genotype. Conventional selection implemented two-stage selection for males, hence we present the accuracy of pre-selection for progeny testing (empty point) and the accuracy of sire selection (solid point).

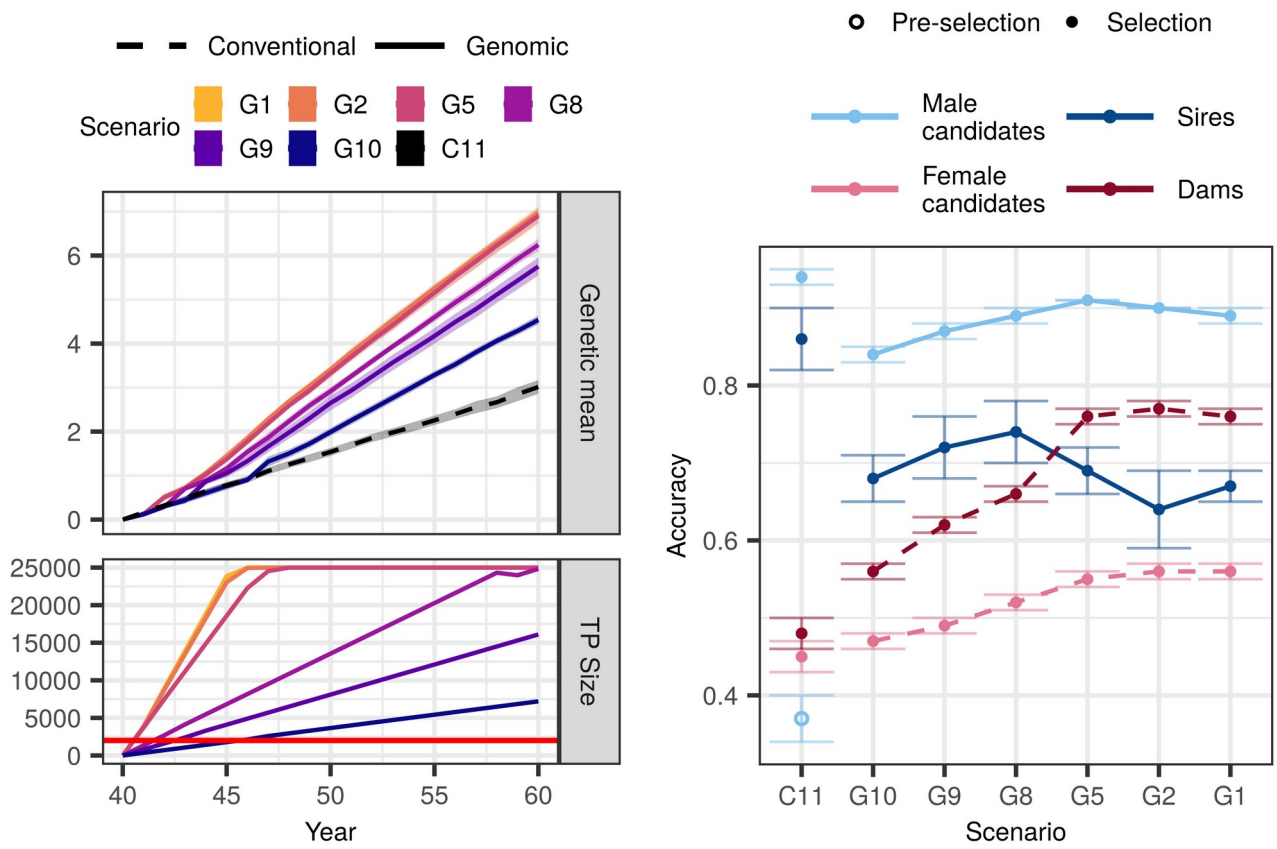

Figure 3. Genetic gain, training population size, and accuracy by scenario without an initial training population (TP) and equal cost of phenotyping and genotyping. The figure presents the means (lines or points) and 95% confidence intervals (polygons or error-bars) across 10 replicates for the conventional (C) and genomic (G) scenarios with numbers indicating the number of phenotype records per lactation. The cost of phenotyping was defined as the cost of 11 phenotype records per lactation, and the cost of genotyping as the cost of one genotype.The red line marks the condition of required 2,000 training animals to start genomic selection. Conventional selection implemented two-stage selection for males, hence we present the accuracy of the pre-selection for progeny testing (empty point) and the accuracy of sire selection (solid point).

Supplementary Material

1. Supplementary Figures and Tables

a. Supplementary Tables

Table S1 Accuracy of conventional and genomic selection with varying number of phenotypes and phenotyped animals. NoRec = Number of phenotypic records per lactation, NoDaughters = number or daughters per sire, r_{sire} = accuracy for sires, r_{cows} = accuracy for cows, $r_{non\text{-}pheno}$ = accuracy for non-phenotyped animals, NoPhenoCows = number of phenotyped cows, NoPhenoTotal = total number of phenotypes (number of phenotypes per lactation times the number of phenotyped cows).

NoPheno	NoDaughters	r_{sires}	r_{cows}	$r_{non\text{-}pheno}$	NoPhenoCows	NoPhenoTotal
Conventional selection, 100 sires						
Variable resources for phenotyping						
1	100	0.93	0.62	0.56	10,000	10,000
2	100	0.96	0.70	0.59	10,000	20,000
5	100	0.97	0.81	0.64	10,000	50,000
10	100	0.98	0.89	0.66	10,000	100,000
Fixed resources for phenotyping						
1	1000	0.99	0.63	0.59	100,000	100,000
2	500	0.99	0.71	0.61	50,000	100,000
5	200	0.99	0.82	0.64	20,000	100,000
10	100	0.98	0.89	0.66	10,000	100,000
Genomic selection						
Variable resources for phenotyping						
1	-	-	0.62	0.56	10,000	10,000
2	-	-	0.70	0.63	10,000	20,000
5	-	-	0.81	0.71	10,000	50,000
10	-	-	0.89	0.76	10,000	100,000
Fixed resources for phenotyping						
1	-	-	0.63	0.93	100,000	100,000
2	-	-	0.71	0.90	50,000	100,000
5	-	-	0.82	0.84	20,000	100,000
10	-	-	0.89	0.76	10,000	100,000

Table S2 Genetic gain by scenario, relative cost of phenotyping to genotyping ($P:$G), and availability of an initial training population. The table presents the means and standard deviations (subscript) across 10 replicates for the conventional (C) and genomic (G) scenarios, with numbers indicating the number of phenotype records per lactation. For the $P:$G we compared the cost of 11 phenotype records per lactation to the cost of one genotype. The scenarios in bold did not spend all the available resources. Lower-case letters denote statistically significant differences between scenarios within the same $P:$G and upper-case letters between different $P:$G within the same scenario.

		Relative cost of phenotyping (P) to genotyping (G)		
	Scenario	$P:$G = 1:2	$P:$G = 1:1	$P:$G = 2:1
	C11	$3.01_{0.22}^{a,A}$	$3.01_{0.22}^{a,A}$	$3.01_{0.22}^{a,A}$
With an initial training population	G10	$5.43_{0.20}^{b, A}$	$5.41_{0.29}^{b, A}$	$6.50_{0.20}^{b, B}$
	G9	$5.58_{0.26}^{b, A}$	$6.30_{0.17}^{c, B}$	$7.02_{0.24}^{c, C}$
	G8	$6.35_{0.25}^{c, A}$	$6.62_{0.25}^{d, B}$	$7.02_{0.17}^{c, C}$
	G5	$6.78_{0.21}^{d, A}$	$7.07_{0.20}^{e, B}$	$\mathbf{7.26_{0.19}^{c, B}}$
	G2	$7.13_{0.29}^{e, A}$	$\mathbf{7.33_{0.26}^{e, A}}$	$\mathbf{7.28_{0.17}^{c, A}}$
	G1	$\mathbf{7.11_{0.16}^{e,A}}$	$\mathbf{7.27_{0.28}^{e, A}}$	$\mathbf{7.24_{0.22}^{c,A}}$
Without an initial training population	G10	$3.93_{0.22}^{b, A}$	$4.54_{0.14}^{b, B}$	$5.61_{0.25}^{b, C}$
	G9	$4.64_{0.18}^{c, A}$	$5.75_{0.28}^{c, B}$	$6.52_{0.17}^{c, C}$
	G8	$5.61_{0.28}^{d, A}$	$6.24_{0.19}^{d B}$	$6.70_{0.25}^{cd, C}$
	G5	$6.43_{0.21}^{e, A}$	$6.90_{0.22}^{e, B}$	$\mathbf{7.05_{0.27}^{de, B}}$
	G2	$6.81_{0.28}^{f, A}$	$\mathbf{6.96_{0.17}^{e, A}}$	$\mathbf{7.00_{0.30}^{de, A}}$
	G1	$\mathbf{6.78_{0.29}^{f,A}}$	$\mathbf{6.92_{0.26}^{e, A}}$	$\mathbf{7.01_{0.23}^{e,A}}$

Table S3 Intensity of sire selection by scenario and relative cost of phenotyping to genotyping ($P:$G). PA = parent average, gEBV = genomic breeding value. We are presenting the bivariate intensities for a two-stage selection. The pre-selection step selects the animals with the highest parent average out of all available new born males to send into testing (progeny of genomical). The selection step selects the sires with the highest breeding values out of all tested males to use in artificial insemination. The scenarios are named C/G for conventional/genomic with numbers indicating the number of phenotype records per lactation. For the $P:$G we compared the cost of 11 phenotype records per lactation to the cost of one genotype.

		Relative cost of phenotyping (P) to genotyping (G)		
	Scenario	$P:$G = 1:2	$P:$G = 1:1	$P:$G = 2:1
	C11	0.10	0.10	0.10
With an initial training population	G10	0.17	0.27	0.37
	G9	0.28	0.39	0.55
	G8	0.38	0.49	0.68
	G5	0.59	0.77	0.91
	G2	0.77	0.91	0.98
	G1	0.83	0.92	0.99
Without an initial training population	G10	0.17	0.25	0.36
	G9	0.28	0.43	0.56
	G8	0.37	0.55	0.71
	G5	0.58	0.77	0.91
	G2	0.76	0.89	0.98
	G1	0.83	0.94	0.99

Table S4 Selection accuracy by scenario, relative cost of phenotyping to genotyping ($P:$G), and the availability of an initial training population. The table presents the means and standard deviations (subscript) across 10 replicates for the conventional (C) and genomic (G) scenarios, with numbers indicating the number of phenotype records per lactation. For the $P:$G we compared the cost of 11 phenotype records per lactation to the cost of one genotype. Conventional selection implemented two-stage selection for males, hence we present the accuracy of pre-selection for progeny testing (S1) and the accuracy sire selection (S2). In genomic scenarios the male candidates were genotyped and non-phenotyped. We also present the accuracy for sires currently used in artificial insemination (sires), for non-genotyped and non-phenotyped females (female candidates), and for all active phenotyped cows and bull dams (cows). Lower-case letters denote statistically significant differences between scenarios within the same $P:$G and upper-case letters between different $P:$G within the same scenario. Stars denote statistically significant difference between corresponding scenarios with and without an initial training population.

	With an initial training population			Without an initial training population		
Scenario	$P:$G=1:2	$P:$G=1:1	$P:$G=2:1	$P:$G=1:2	$P:$G=1:1	$P:$G=2:1
Male candidates						
C11, S1	$0.37_{0.04}^{a,A}$	$0.37_{0.04}^{a,A}$	$0.37_{0.04}^{a,A}$	$0.37_{0.04}^{a,A}$	$0.37_{0.04}^{a,A}$	$0.37_{0.04}^{a,A}$
C11, S2	$0.94_{0.01}^{b,A}$	$0.94_{0.01}^{b,A}$	$0.94_{0.01}^{b,A}$	$0.94_{0.01}^{b,A}$	$0.94_{0.01}^{b,A}$	$0.94_{0.01}^{b,A}$
G10	$0.89_{0.03}^{c,A}$	$0.90_{0.02}^{bc,AB}$	$0.91_{0.01}^{bc,B}$	$0.81_{0.03}^{b,A\,*}$	$0.84_{0.01}^{b,B\,*}$	$0.87_{0.01}^{b,C\,*}$
G9	$0.90_{0.03}^{bc,A}$	$0.91_{0.02}^{bc,A}$	$0.91_{0.01}^{bc,A}$	$0.85_{0.02}^{c,A\,*}$	$0.87_{0.01}^{bc,B\,*}$	$0.90_{0.01}^{bc,C\,*}$
G8	$0.91_{0.01}^{bc,A}$	$0.91_{0.01}^{bc,A}$	$0.91_{0.01}^{bc,A}$	$0.86_{0.01}^{cd,A\,*}$	$0.89_{0.01}^{c,B\,*}$	$0.90_{0.01}^{bc,B}$
G5	$0.91_{0.01}^{bc,A}$	$0.91_{0.00}^{bc,A}$	$0.91_{0.01}^{bc,A}$	$0.90_{0.01}^{d,A}$	$0.91_{0.01}^{c,A}$	$0.91_{0.01}^{c,A}$
G2	$0.91_{0.01}^{bc,A}$	$0.91_{0.00}^{bc,A}$	$0.90_{0.01}^{bc,A}$	$0.90_{0.01}^{d,A}$	$0.90_{0.01}^{c,A}$	$0.90_{0.01}^{bc,A}$
G1	$0.89_{0.01}^{c,A}$	$0.90_{0.01}^{c,A}$	$0.89_{0.01}^{c,A}$	$0.89_{0.01}^{cd,A}$	$0.89_{0.01}^{c,A}$	$0.89_{0.01}^{bc,A}$
Sires						
C11	$0.86_{0.05}^{a,A}$	$0.86_{0.05}^{a,A}$	$0.86_{0.05}^{a,A}$	$0.86_{0.05}^{a,A}$	$0.86_{0.05}^{a,A}$	$0.86_{0.05}^{a,A}$
G10	$0.75_{0.04}^{b,A}$	$0.75_{0.03}^{b,A}$	$0.73_{0.05}^{b,A}$	$0.67_{0.08}^{bc,A\,*}$	$0.68_{0.05}^{cde,A\,*}$	$0.67_{0.06}^{b,A\,*}$
G9	$0.76_{0.04}^{b,A}$	$0.72_{0.06}^{bc,AB}$	$0.69_{0.05}^{c,A}$	$0.70_{0.05}^{b,A\,*}$	$0.72_{0.05}^{bc,A}$	$0.71_{0.05}^{b,A}$
G8	$0.76_{0.03}^{b,A}$	$0.69_{0.05}^{cd,B}$	$0.68_{0.06}^{c,B}$	$0.71_{0.05}^{b,A\,*}$	$0.74_{0.05}^{b,A\,*}$	$0.70_{0.07}^{b,A}$
G5	$0.68_{0.07}^{c,A}$	$0.67_{0.08}^{de,A}$	$0.69_{0.04}^{c,A}$	$0.68_{0.05}^{bc,A}$	$0.69_{0.05}^{cd,A}$	$0.69_{0.03}^{b,A}$
G2	$0.67_{0.05}^{c,A}$	$0.67_{0.05}^{de,A}$	$0.67_{0.04}^{c,A}$	$0.65_{0.06}^{c,A}$	$0.64_{0.07}^{e,A}$	$0.69_{0.05}^{b,A}$
G1	$0.66_{0.06}^{c,A}$	$0.63_{0.05}^{e,A}$	$0.67_{0.04}^{c,A}$	$0.67_{0.04}^{bc,A}$	$0.67_{0.03}^{de,A}$	$0.69_{0.05}^{b,A}$
Female candidates						
C11	$0.45_{0.02}^{a,A}$	$0.45_{0.02}^{a,A}$	$0.45_{0.02}^{a,A}$	$0.45_{0.02}^{a,A}$	$0.45_{0.02}^{a,A}$	$0.45_{0.02}^{a,A}$
G10	$0.48_{0.01}^{ab,A}$	$0.48_{0.01}^{ab,A}$	$0.51_{0.01}^{b,B}$	$0.46_{0.02}^{ab,A\,*}$	$0.47_{0.02}^{ab,AB}$	$0.49_{0.01}^{b,B\,*}$
G9	$0.49_{0.02}^{b,A}$	$0.50_{0.01}^{b,B}$	$0.52_{0.01}^{b,C}$	$0.47_{0.02}^{ab,A\,*}$	$0.49_{0.02}^{bc,B}$	$0.52_{0.01}^{bc,C}$
G8	$0.51_{0.01}^{b,A}$	$0.51_{0.01}^{b,A}$	$0.54_{0.01}^{bc,B}$	$0.49_{0.02}^{bc,A\,*}$	$0.52_{0.01}^{cd,B}$	$0.53_{0.01}^{cd,C}$
G5	$0.51_{0.01}^{bc,A}$	$0.55_{0.01}^{c,B}$	$0.57_{0.01}^{c,C}$	$0.52_{0.01}^{cd,A}$	$0.55_{0.01}^{de,B}$	$0.57_{0.01}^{d,C}$
G2	$0.55_{0.01}^{cd,A}$	$0.57_{0.01}^{c,B}$	$0.57_{0.01}^{c,B}$	$0.55_{0.01}^{d,A}$	$0.56_{0.02}^{e,AB}$	$0.57_{0.01}^{d,B}$
G1	$0.56_{0.01}^{d,A}$	$0.56_{0.01}^{c,A}$	$0.56_{0.01}^{c,A}$	$0.55_{0.01}^{d,A}$	$0.56_{0.01}^{e,A}$	$0.56_{0.01}^{d,A}$
Cows						
C11	$0.48_{0.03}^{a,A}$	$0.48_{0.03}^{a,A}$	$0.48_{0.03}^{a,A}$	$0.48_{0.03}^{a,A}$	$0.48_{0.03}^{a,A}$	$0.48_{0.03}^{a,A}$
G10	$0.56_{0.02}^{b,A}$	$0.59_{0.02}^{b,B}$	$0.63_{0.01}^{b,C}$	$0.53_{0.01}^{b,A\,*}$	$0.56_{0.01}^{b,B\,*}$	$0.61_{0.01}^{b,C\,*}$
G9	$0.59_{0.03}^{bc,A}$	$0.63_{0.02}^{c,B}$	$0.70_{0.01}^{c,C}$	$0.57_{0.02}^{bc,A\,*}$	$0.62_{0.02}^{c,B}$	$0.68_{0.02}^{c,C\,*}$
G8	$0.62_{0.02}^{c,A}$	$0.67_{0.02}^{c,B}$	$0.74_{0.02}^{d,C}$	$0.60_{0.02}^{c,A\,*}$	$0.66_{0.01}^{d,B}$	$0.73_{0.02}^{d,C}$
G5	$0.70_{0.02}^{d,A}$	$0.77_{0.01}^{d,B}$	$0.79_{0.02}^{e,C}$	$0.69_{0.02}^{d,A}$	$0.76_{0.01}^{e,B}$	$0.78_{0.02}^{e,B}$
G2	$0.76_{0.02}^{e,A}$	$0.79_{0.02}^{d,B}$	$0.78_{0.01}^{e,AB}$	$0.76_{0.01}^{e,A}$	$0.77_{0.02}^{e,A\,*}$	$0.77_{0.01}^{de,A}$
G1	$0.77_{0.02}^{e,A}$	$0.77_{0.02}^{d,A}$	$0.77_{0.01}^{de,A}$	$0.76_{0.01}^{e,A}$	$0.76_{0.02}^{e,A}$	$0.76_{0.02}^{de,A}$

b. Supplementary Figures

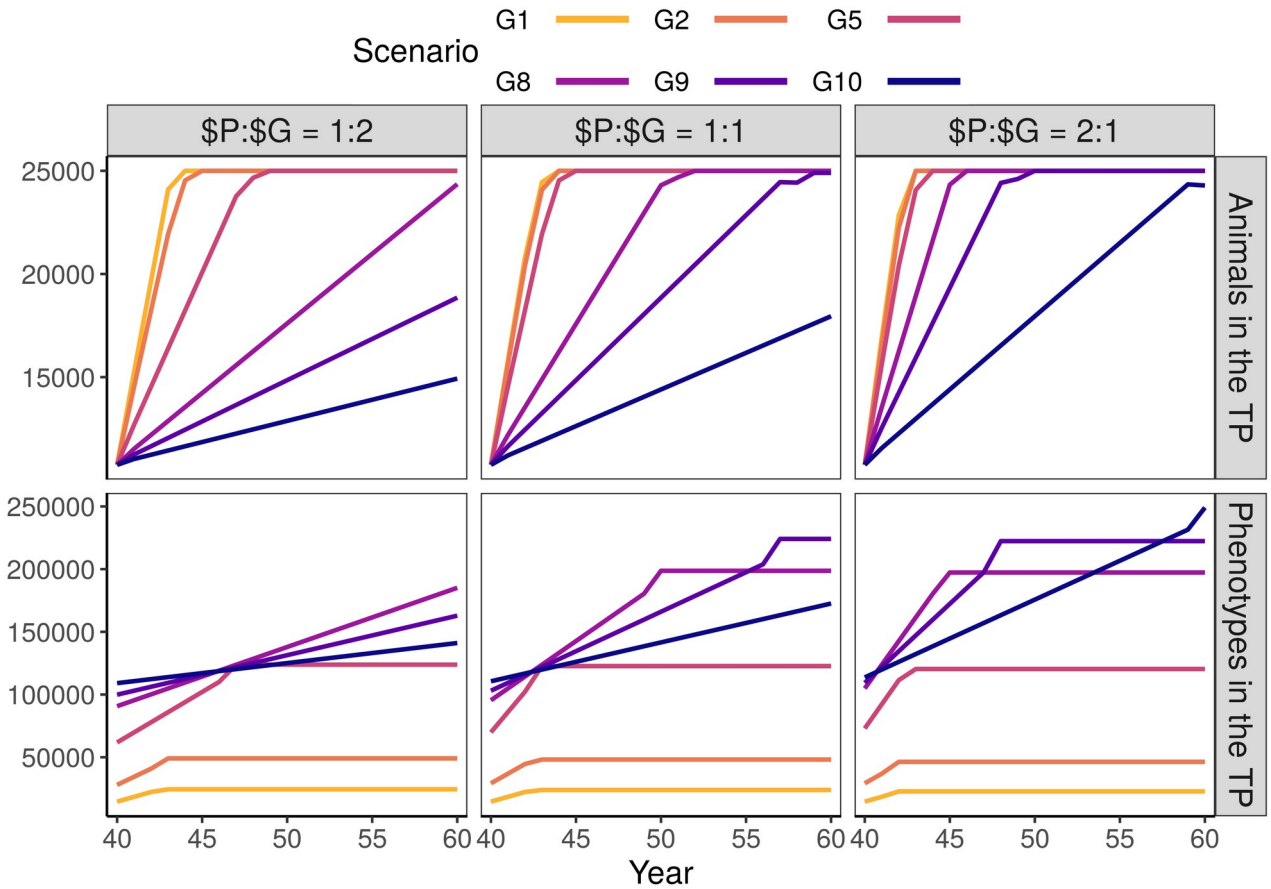

Figure S1 The number of animals and repeated phenotypes in the training population by scenario and relative cost of phenotyping to genotyping ($P:$G) with an initial training population (TP). The scenarios are named C/G for conventional/genomic with numbers indicating the number of phenotype records per lactation. For the $P:$G we compared the cost of 11 phenotype records per lactation to the cost of one genotype. In our simulation, scenarios traded repeated phenotype records for genotypes. Hence, the scenarios with the largest training population collected the least repeated records. These were also the scenarios that achieved the highest genetic gain.