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

Jana Obšteter1\*, Janez Jenko2, and Gregor Gorjanc3,4

1Department of Animal Science, Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia

2Geno Breeding and A.I. Association, Storhamargata 44, 2317 Hamar, Norway

3The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH259RG, United Kingdom

4Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia

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

E-mail addresses:

JO: [jana.obsteter@kis.si](mailto:jana.obsteter@kis.si)

JJ: [janez.jenko@geno.no](mailto:janez.jenko@geno.no)

[GG:](mailto:john.hickey@roslin.ed.ac.uk) [gregor.gorjanc@roslin.ed.ac.uk](mailto:gregor.gorjanc@roslin.ed.ac.uk)

# 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 achieve desired selection accuracy for breeders and quality for producers. Genomic selection decoupled phenotyping and selection and through this increased genetic gain per year compared to the conventional selection. However, genomic selection requires a large initial investment, which limits the adoption of genomic selection for some breeding programmes. We simulated a case-study of a small dairy population with a number of scenarios under equal available resources. 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 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 investment in genotyping, despite reduced phenotyping, and with high‑genotyping scenarios not even using all the available resources. Compared to the conventional scenario, genomic scenarios also increased accuracy for young non‑phenotyped male and female candidates, and cows. This study shows that breeding programmes should optimize investment into phenotyping and genotyping to maximise return on investment. Our results suggest that any dairy breeding programme using conventional progeny testing with repeated milk records can implement genomic selection without increasing the level of investment.

Keywords: genomic selection, dairy breeding programme, small populations, optimized investment

# introduction

This paper evaluates the potential of maximizing genetic gain in dairy cattle breeding by optimizing investment into phenotyping and genotyping. Breeding programmes strive to maximize genetic gain, which is a function of selection intensity, accuracy of selection, genetic variation, and generation interval. The conventional dairy breeding programme uses a lengthy (time consuming?) and expensive progeny test that also requires the infrastructure to rear and keep the males during the test. This limits the number of tested bulls and thus selection intensity, . The conventional programme allocates most of resources into phenotyping to achieve high accuracy of sire selection, since this is the main driver of genetic gain in conventional selection. Genomic selection (Meuwissen et al., 2001; Schaeffer, 2006), on the other hand, achieves genetic gain mainly through substantially reduced generation interval, increased accuracy of selection for young animals, and increased selection intensity of males (Obšteter et al., 2019; Schaeffer, 2006). The increased intensity stems from reduced cost and duration of genomic vs. progeny testing that allowed for a larger number of tested males. 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 programmes operate with a set amount of resources allocated to breeding activities with the aim to maximise return on investment. Genomic selection is now a de-facto standard in well-resourced breeding programmes, but is still challenging to implement in breeding programmes with limited resources. An example of such are breeding programmes of small populations that are additionally limited with number of animals for training population for genomic prediction and as selection candidates. Another example are developing countries, which have also not received the full benefit of livestock and genetic improvement technologies. Here, the problem is even more emphasized since the demand for animal source food is rapidly increasing and requires increased productivity and sustainability. The use of genomic technologies in developing countries is therefore an area of active research aiming to identify major problems and putative solutions (Ducrocq et al., 2018; Marshall et al., 2019; R. Mrode et al., 2019). Amongst these issues are the lack of investment related to the high cost of genotyping and lack of pedigree records that disable even conventional genetic evaluation. The major hurdle is the large initial investment in genotyping to establish a training population, though updating this population can also be challenging. We hypothesise that these breeding programs need to evaluate priorities and could optimize the allocation of resources for phenotyping and genotyping to maximise 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. R. A. Mrode, 2005). For example, for a female-expressed trait with 0.25 heritability, the accuracy for 10,000 cows 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 sire with 100 daughters each are 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 also increases with increasing heritability and increasing number of phenotype records per genotyped animal, but also with increasing training population, decreasing genetic distance between training and prediction individuals, and decreasing number of effective genome segments (Clark et al., 2011; Daetwyler et al., 2008; M. Goddard, 2009; M. E. Goddard et al., 2011; Habier et al., 2010). Following the previous example, assume 10,000 effective genome segments, 0.25 heritability, and a training population of 10,000 cows. The accuracy for sire selection as a function of the number of repeated records per lactation (n) is 0.76 (n=10), 0.71 (n=5), 0.63 (n=2), or 0.56 (n=1) These examples again shows diminishing returns with repeated phenotyping and a scope for optimizing return on investment in genomic breeding programmes.

We could invest the saved resources from reducing the number of phenotype records per daughter into genotyping. The accuracy of genomic prediction as a function of number of genotyped and phenotyped cows (N) and number of repeated records per lactation (n) is 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 (Gonzalez-Recio et al., 2014; Van Grevenhof et al., 2012). They concluded, that accuracy has diminishing returns with increasing the number of genotyped and phenotyped animals in the training population, hence additional female is most valuable when the training population is small.

The above examples suggest that repeated phenotyping could serve as an internal financial reserve to enable dairy breeding programmes to implement genomic selection. In dairy breeding the most repeatedly and extensively recorded phenotypes are milk production traits. There are different milk recording methods that differ in the party that carries out the recording, sampling scheme, recording and sampling frequency, and the number of milkings per day (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 also 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 programmes are funded. Some are funded by farmers and breeding companies. But, in some countries farmers get subsidies for milk testing. There are also situations where farmers do no testing at all and all genetic progress is generated in "research" nucleus herds. There are also cases where phenotyping is funded by philanthropic organisations.

The examples assumed simplified scenarios where cows form a training population and focused only on accuracy. Real breeding programmes involve overlapping generations, individuals with a mix of phenotype, pedigree, and genotype information, various selection intensities, and other dynamic components. While we can use the single-step genomic prediction to combine all phenotypic, pedigree, and genomic information (Gao et al., 2012; Gray et al., 2012; Lourenco et al., 2015) evaluating the optimal allocation of resources into phenotyping and genotyping is beyond simple examples.

The aim of this study was to evaluate the potential of maximizing genetic gain by optimizing investment into phenotyping and genotyping in dairy breeding programmes. 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 programme with conventional progeny testing and genomic testing under equal available resources. To implement genomic selection, we reduced the number of milk records per lactation and invested the saved resources into genotyping. We compared these strategies in a case-study of a small cattle breeding programme 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.

# 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 programme. The simulation mimicked a real dairy cattle population of ~30,000 animals analysed in our previous study (Obšteter et al., 2019). 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 saved resources to 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 tested scenarios were compared based on their genetic gain and accuracy of selection.

## Simulation of the base population, phenotype and historical breeding

The simulation mimicked a small dairy cattle breeding programme of ~30,000 animals with ~10,500 cows. The introduction of effective genomic selection in such populations is challenging due to a small number of animals available for selection and assembling a training population. We use this population as a case-study to optimize investment into phenotyping and genotyping. The breeding programme aimed to improve milk yield (milk trait / dairy traits / milk production?), which we simulated as a single polygenic trait. For this we used a coalescent process to simulate genome comprised of 10 cattle-like chromosomes, each with 108base 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 (*ai*) for dairy performance (*yijkl*), which was affected also by a permanent environment (*pi*), herd (*hj*), herd-year (*hyjk*), herd-test-day (*htdjkl*), and residual environment (*eijkl*) effects:

yijkl = ai + pi + hj + hyjk + htdjkl + eijkl.

We sampled permanent environment effects from a normal distribution with zero mean and variance equal to a base population additive genetic variance (*σ2A*). We sampled herd, herd-year, and herd-test-day effects each from a normal distribution with zero mean and variance of 1/3*σ2A*. Finally, we sampled residual environment effects from a normal distribution with zero mean and variance of *σ2A*. 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 programme 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 programme we selected 3,849 out of 4,320 new-born females. We selected 139 bull dams out of cows in second, third, and fourth lactation. We generated 45 male calves from matings of bull dams and progeny tested sires (elite matings) and out of these chose 8 for progeny testing of which 4 were eventually selected as sires for widespread use in artificial insemination. 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.

## Scenarios

We evaluated 36 genomic scenarios with varying the extent of phenotyping and genotyping against the conventional scenario. All scenarios had equal available resources. The conventional scenario continued the breeding scheme from the historical breeding. It used progeny testing and 11 phenotype records 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 available for generating the data. We then created genomic scenarios by distributing resources between phenotyping and genotyping - we reduced phenotyping and invested the saved resources into genotyping. In the genomic scenarios we selected females as in the conventional scenario and males on genomic prediction. The number of genomically tested male candidates varied according to the genotyping resources in a specific scenario. We selected the 5 males with the highest genomic breeding value as sires for widespread use in artificial insemination on genomic prediction. We evaluated the genomic scenarios with varying number of phenotype records per lactation, relative cost of phenotyping to genotyping, and the availability of an initial training population.

Genomic scenarios reduced the number of phenotype records per lactation to between 10 and 1. The scenarios followed ICAR standards of 9, 8, and 5 records per lactation, corresponding to recording intervals of 5, 6, and 9 weeks. Additionally, we created three non‑standard recording systems collecting 10, 2, and 1 record 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 one genotype to the cost of 11 phenotype records per lactation. Following the survey, we implemented quantity discount for increasing the number of milk recordings. By increasing the number of recordings the cost per recording decreased by 6%. The cost of genotyping remained constant. Based on a survey of several breeding programmes, 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 saved resources and the number of genotyped animals (Table 1). For example, assume that the cost per milk recording is 1.9€ when collecting 10 record per lactation and 1.8€ when collecting 11 records per lactation for a total of 19.6€ per lactation. Assume that this total equals the cost of genotyping one animal (1:1 setting). Reducing the number of records from 11 to 10 per lactation saves 6,959€ in our simulated population of 10,852 active cows, which suffices for genotyping of 356 animals.

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 maximise the genetic gain, we genotyped male calves from elite matings and other high parent average matings. In scenarios where the available 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.

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 evaluation. When 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.

## Estimation of breeding values

We selected the 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., 2002) 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 favour of the latest genotyped cows and male selection candidates.

## Analysis of scenarios

All scenarios had equal available resources. We compared the scenarios based on their final genetic gain, which indicated return on investment, and accuracy of selection. 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 repeated simulation of the base population and each scenario 10 times and summarised 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. We computed intensities of two-stage selection by integrating standard bivariate distribution:

as

where *x* is the parent average, *y* is the genomic breeding value, and *ρ* is the correlation between the variables computed by dividing the accuracy of parent average by the accuracy of genomic prediction. We assumed the accuracy of parent average of 0.5 and accuracy of genomic prediction of 0.8. Tx is the standardized cut-off for the proportion of all new born male calves selected for genomic testing (p1) based on parent average (first selection stage). Ty is the standardized cut‑off for the proportion of all new born males selected as sires (p2) based on genomic breeding values (second selection stage).

# 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.

**Genetic gain with an initial training population**

With the same available resources, genomic scenarios with an initial training population increased the genetic gain of the conventional scenario between 79% and 143%. 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 of those male candidates. This small change increased the male selection intensity from 0.16 to 0.38 and coupled with a shorter generation interval increased the genetic gain by 79% (from 3.01 to 5.41). By reducing the phenotype records to nine or eight per lactation (G9 or G8), we respectively saved resources to genotype 800 or 1,345 animals per year, of which 100 or 165 male candidates. This respectively increased the male selection intensity to 0.54 or 0.65, 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 fiveand 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 the male selection intensity between 0.85 and 0.96.

. Changing the relative cost of phenotyping to genotyping resulted in similar trend for genetic gain with the largest change in the scenario with the smallest amount of genotyping (G10). In this scenario, in the $P:$G=1:2 or $P:$G=2:1 setting, we respectively saved resources for genotyping 182 or 710 animals, of which 22 or 90 were males, and increased the genetic gain by 80% (from 3.01 to 5.43) or 116% (from 3.01 to 6.50). When we maximized the investment into genotyping (G1), we genotyped between 565 and 2,245 male candidates. We also genotyped all females and 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 available resources (marked bold in Table S2). In these scenarios the resources designated to genotyping females exceeded the cost of genotyping all females. The savings could cover between 42 and 23,800 additional phenotypes 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).

## Accuracy with an initial training population

Compared to the conventional scenario, genomic scenarios increased accuracy for young non‑phenotyped and genotyped male and 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 (between 0.55 and 0.57) and cows (between 0.77 and 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 trends.

## Genetic gain and accuracy without an initial training population

### Genetic gain

When an initial training population was not available, we increased the genetic gain of the conventional scenario between 31% and 134% by optimizing investment in phenotyping and genotyping (Figure 3 and Table S2). The trends were in line with what we observed with an initial training population, that is, increasing genotyping increased genetic gain despite reduced phenotyping. 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 already 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 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 initial training population.

### Accuracy

Similar to the scenarios with an initial training population, genomic scenarios without an initial training population 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 initial training population, the accuracy increased with increasing the 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 genotyping. 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, where 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.

# Discussion

1. Our results show that any dairy breeding programme using conventional progeny testing with repeated milk records can implement genomic selection without extra costs. While breeding programmes have established funding for phenotyping, not all of them have well established funding for genotyping. We show that by reallocating a part of phenotyping resources into genotyping, breeding programmes can implement genomic selection and substantially increase genetic gain regardless of the amount and cost of genotyping, and availability of an initial training population. 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 programmes; and 4) limitations of the study. In the following we first discuss the results under equal cost of phenotyping and genotyping, and initial training population available. We then discuss changes at different costs and no initial training population.

## Genetic gain with initial training population

### Genomic vs. conventional selection

1. 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 is in agreement with previous simulation studies. that showed increased genetic gain with genomic selection due to reduced generation interval compared to progeny test, despite reduced selection accuracy (Obšteter et al., 2019; Pryce et al., 2010; Schaeffer, 2006). Real data confirmed this and showed that in the US Holstein population the generation interval in 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 halfes 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. While the assumption of an available initial training population might not be realistic for some populations, it can be achieved through international collaboration (Jorjani, 2012).
2. Another major advantage of the genomic scenarios wasincreased 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 {Citation}.

### Increasing the investment into genotyping

1. 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)
2. The genetic gain had diminishing relationship with investment into genotyping. This has important implications for dairy breeding programmes, 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 (Gonzalez-Recio et al., 2014; Van Grevenhof et al., 2012). Since scenarios with 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 when there were at least 10,000 animals in the training population. Second, increasing investment into genotyping did not proportionally increase the size of the training population due to limit of 25,000 animals of the training population. And third, the intensity of sire selection had diminishing relationship with increasing genotyping. 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.
3. 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.
4. 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.

## Accuracy with initial training population

1. Despite reduced phenotyping, genomic scenarios increased the accuracy for young non‑phenotyped calves and cows.When the accuracy of parent average is high, genomic prediction increases the accuracy of the Mendelian sampling term. But when the accuracy of parent average is low, such as for animals with non‑phenotyped parents or parents with little 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).

### Accuracy for males

1. 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 in conventional scenario. 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).
2. In contrast, reducing phenotyping decreased the accuracy for selected sires. We believe this is due to two reasons. First, since sires are the very best animal, their breeding values are extreme and lie close together in the tail of the distribution. Due to small differences between the sires, each additional phenotypic record helps to distinguish between the 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 favour of cows’ genotypes, hence prediction for sires depended only on daughters data and no longer on own genotype. However, since this is the accuracy after the selection has already been made, it is not of great interest for breeding.

### Accuracy for females

1. 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 (Powell et al., 2019; Yu et al., 2017). This in turn increases the accuracy of prediction regardless of the heritability and the number of causal loci or markers (Yu et al., 2018). This had important implications, since we selected bull dams for elite mating from cows.
2. The accuracy for cows also 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 (Gonzalez-Recio et al., 2014; Van Grevenhof et al., 2012), 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 relationship with the size of the training population.We observed plateau in accuracy when we invested more than six phenotype records into genotyping.
3. Accuracy for female candidates followed the accuracy trend for the dams, 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 dams and in turn increased the accuracy of female candidate`s parent average. The benefit of this increase was not large, since the intensity of cow selection was low. However, there is potential for this benefit to be larger with sexed semen and embryo transfer.

## Scenario without an initial training population

### Genetic gain

1. We also considered that some small 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).
2. When implementing genomic selection with a delay due to building the training population, we ran a conventional selection with reduced phenotyping until we accumulated the targeted 2,000 genotypes. In this period, we did not observe decreased genetic gain compared to the conventional scenario with full phenotyping. This suggests that conventional programmes can reduce phenotyping until they accumulate genotypes to initiate genomic selection, without harming the genetic gain in the accumulation or transition period.

### Accuracy

1. 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.

## Implications

* 1. The results suggest that any dairy breeding programme 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 programmes 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.
     1. 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 (Akdemir & Isidro-Sánchez, 2019; Heslot & Feoktistov, 2017). 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, such as algorithm for proven and young (Misztal et al., 2014) or 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.
     2. The economic efficiency of breeding programmes strongly depends on which stakeholders fund which breeding action. Different programmes have different investment schemes, often intricate (complex?). The scenarios presented in this paper are of little value for programmes where funding for phenotyping and genotyping is disconnected. Similarly, optimizing the investment into phenotyping is not of interest for breeding programmes with abundant use of automated milking systems where the cost of phenotyping does not depend on the number of records. But in populations with small herds the use of automated system 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 programmes that do not actively exchange sires between herds (Kasap et al., 2018).
     3. High cost of genotyping diminishes the benefit of the proposed solutions The relative cost of phenotyping to genotyping at which genomic selection is not profitable (interesting) anymore depends on a number of factors: i) the number of females in the recorded population, since it dictates the savings from reducing the number phenotype records; ii) the number of phenotype records the breeders are willing to sacrifice; iii) the availability of an initial training population; iv) the ratio of genotyped males and females. In our case, if the genotyping was ten-times more expensive than phenotyping with 11 records, we would save the funds to genotype between 36 (10 records) and 900 (one record) animals. While this could suffice if we had an initial training population and if we genotyped only males, it would probably not be viable if we had to build the training population ourselves.

1. We did not account for the benefits of genotyping besides predicting genomic breeding values and selection. Genomic information has additional value for i) parentage verification, parentage discovery, or correction of parentage errors; ii)management of monogenic diseases and traits, that can prevent large 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. This could serve to identify the most appropriate breeding or cross-breed in a production system or to improve the structure of the training population (Marshall et al., 2019). Additional uses of genotypes increase the return on investment in genomic selection beyond what we measured in this study.

## Limitations of the study

### Reducing the number of phenotype records

* 1. 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 and feed composition, which affect milk yield and its composition. Besides managing production, milk recording is also important from an environmental perspective (Verbič et al., n.d.), but so is genetic improvement. In general, about half of phenotypic improvement is due to management and half due to selection (Dekkers & Hospital, 2002). Until recently, data that drives dairy selection has largely been collected for management and used for free in genetic evaluation. With the advent of genomic selection, the new genotype data is largely used for selection, but the same data could also serve in 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).

1. In practice, test day records are also used to compute the 305-day milk yield according to standard lactation curves using various regression methods (reviewed in ICAR Guidelines: Computing of Accumulated Lactation Yield, 2020). Previous studies explored how removing test day records affects the accuracy of predicting the lactation yield. Pool & Meuwissen, 1999, showed that the correlation of predicting 305-day yield based on weekly or 5-, 8- or 10-weekly records can respectively be as high as 0.99, 0.98, 0.97 or 0.96. Berry et al., 2005, similarly showed that the mean error of 305-day yield estimated from five test day records was 6.8kg with 0.99 correlation with 305-day yield estimated from 11 records. Kong et al., 2018, showed that while in the first lactation using six vs. three records increases accuracy between 0.01 and 0.31, in the second and third lactation the increase is marginal or even negative, depending on the breed. On the other hand, Gantner et al., 2008, showed that using eight instead of eleven records underestimates the 305-day milk yield by 500-1000 kg, although the correlation between predictions is 0.96. However, studies also showed that choice of the model affects the prediction outcome, hence the prediction could be optimized (Lidauer et al., 2003; Pool & Meuwissen, 1999). The longest sampling interval tested in our study and still approved by ICAR was nine weeks, which yielded five records per lactation. In most settings this sufficed to achieve the maximum genetic gain and selection accuracy.

### Single additive trait

1. 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 (Aliloo et al., 2016; Ertl et al., 2014; Fuerst & Sölkner, 1994; Jiang et al., 2017). We do note that permanent environment effects capture non-additive genetic effects and other individual specific environmental effects (Aliloo et al., 2016). We also simulated milk yield in different lactations as a single trait with constant heritability through the lactation, whereas genetic correlation between different lactations and through the lactation is not unity (Dong & van Vleck, 1989; Meyer, 1984; Swalve & Vleck, 1987). 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 even more.
2. We simulated a trait with heritability of 0.25, since the majority of recorded traits (traits in milk recording?) shows 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 more beneficial for lowly heritable traits, since it is less affected by the heritability as conventional selection (García-Ruiz et al., 2016; Lillehammer et al., 2011).

### Genomic selection of females

1. 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 (García-Ruiz et al., 2016; Pryce et al., 2010). 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.

# Conclusion

1. This study suggests that any dairy breeding programme using conventional progeny testing with repeated milk records can implement genomic selection with no extra costs by optimizing the investment into milk phenotyping and genotyping. We showed, that allocating some phenotyping resources into genotyping increased both genetic gain and selection accuracy for non-phenotyped candidates, despite reduced phenotyping. The increase was observed 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 programmes should optimize the investment into phenotyping and genotyping to maximise return on investment for selection and management.

# acknowledgment

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

# Appendix

1. **Table S1 Accuracy of conventional and genomic selection with varying number of phenotypes and phenotyped animals.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| NoPheno | NoDaughters | rsires | rcows | rnon-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 |

**NoRec = Number of phenotypic records per lactation, NoDaughters = number or daughters per sire, rsire = accuracy for sires, rcows = accuracy for cows, rnon-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).**

**Table S2 Genetic gain by scenario, relative cost of phenotyping to genotyping, and availability of an initial training population**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Relative cost of phenotyping (P) to genotyping (G) | | |
|  | Scenario | $P:$G = 1:2 | $P:$G = 1:1 | $P:$G = 2:1 |
|  | C11 | 3.010.22a,A | 3.010.22a,A | 3.010.22a,A |
| With initial  training population | G10 | 5.430.20b, A | 5.410.29b, A | 6.500.20b, B |
| G9 | 5.580.26b, A | 6.300.17c, B | 7.020.24c, C |
| G8 | 6.350.25c, A | 6.620.25d, B | 7.020.17c, C |
| G5 | 6.780.21d, A | 7.070.20e, B | **7.260.19c, B** |
| G2 | 7.130.29e, A | **7.330.26e, A** | **7.280.17c, A** |
| G1 | **7.110.16e,A** | **7.270.28e, A** | **7.240.22c,A** |
| Without initial training population | G10 | 3.930.22b, A | 4.540.14b, B | 5.610.25b, C |
| G9 | 4.640.18c, A | 5.750.28c, B | 6.520.17c, C |
| G8 | 5.610.28d, A | 6.240.19d B | 6.700.25cd, C |
| G5 | 6.430.21e, A | 6.900.22e, B | **7.050.27de, B** |
| G2 | 6.810.28f, A | **6.960.17e, A** | **7.000.30de, A** |
| G1 | **6.780.29f,A** | **6.920.26e, A** | **7.010.23e,A** |

\*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. 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.

**Table S3 Intensity of sire pre-selection and selection by scenario and relative cost of phenotyping to genotyping**

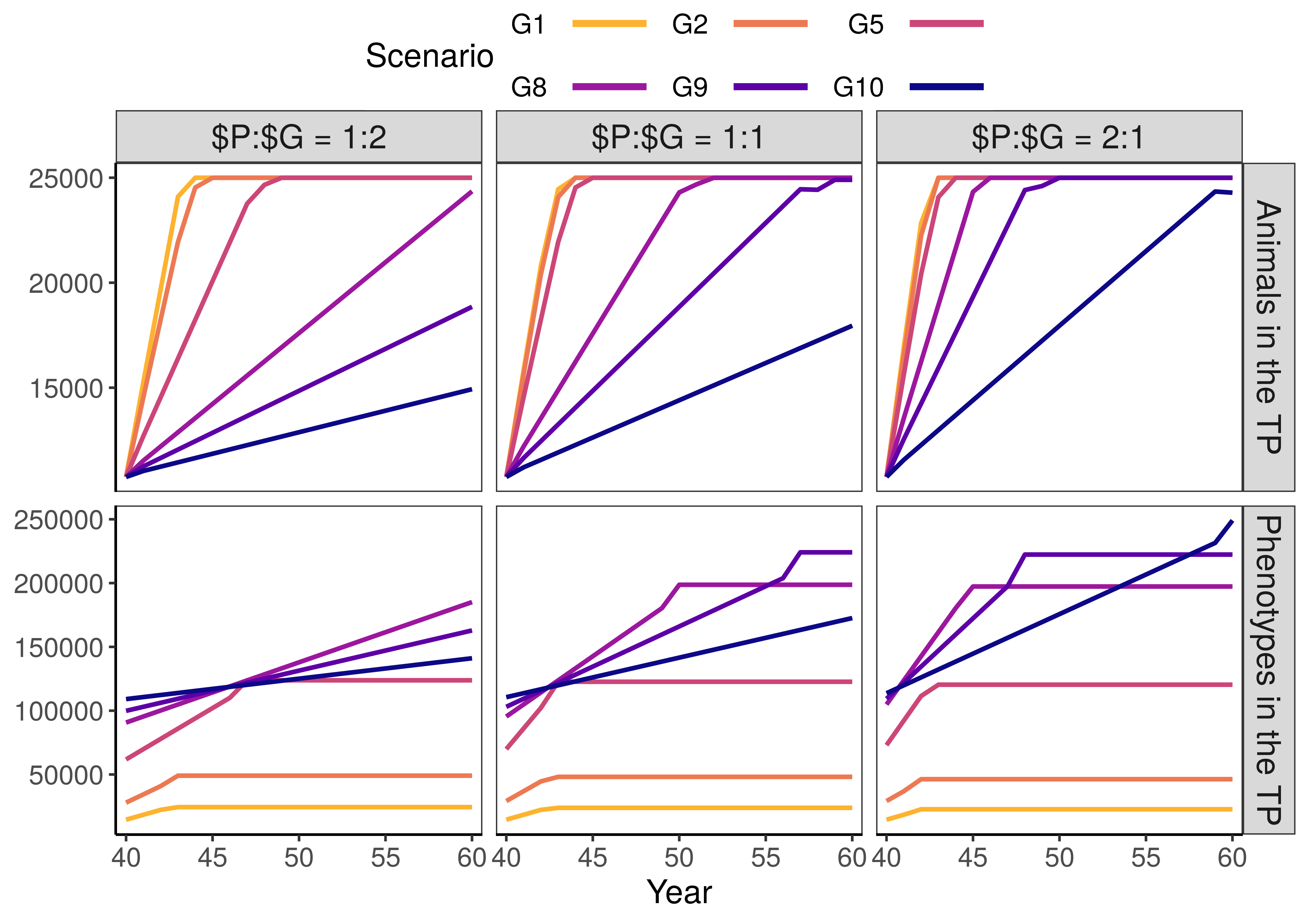
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Relative cost of phenotyping (P) to genotyping (G) | | | | | | |
|  | Pre-selection for testing based on PA | | | Selection of sires based on gEBVs | | | |
| Scenario | $P:$G = 1:2 | $P:$G = 1:1 | $P:$G = 2:1 | | $P:$G = 1:2 | $P:$G = 1:1 | $P:$G = 2:1 |
| C11 | 3.19 | 3.19 | 3.19 | | 0.80 | 0.80 | 0.80 |
| G10 | 2.89 | 2.65 | 2.41 | | 1.33 | 1.70 | 2.02 |
| G9 | 2.62 | 2.37 | 2.09 | | 1.75 | 2.06 | 2.34 |
| G8 | 2.43 | 2.17 | 1.87 | | 1.99 | 2.26 | 2.53 |
| G5 | 2.03 | 1.72 | 1.36 | | 2.40 | 2.64 | 2.87 |
| G2 | 1.72 | 1.36 | 0.92 | | 2.64 | 2.87 | 3.08 |
| G1 | 1.62 | 1.25 | 0.77 | | 2.71 | 2.93 | 3.14 |

$P:$G = relative cost of phenotyping ($P) to genotyping ($G), PA = parent average, gEBV = genomic breeding value. 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.

**Table S4 Selection accuracy by scenario, relative cost of phenotyping to genotyping ($P:$G), and the availability of an initial training population**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | With initial training population | | | Without 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 | 0.370.04a ,A | 0.370.04a,A | 0.370.04a,A | 0.370.04a,A | 0.370.04a,A | 0.370.04a,A |
| C11, S2 | 0.940.01b,A | 0.940.01b,A | 0.940.01b,A | 0.940.01b,A | 0.940.01b,A | 0.940.01b,A |
| G10 | 0.890.03c,A | 0.900.02bc,AB | 0.910.01bc,B | 0.810.03b,A \* | 0.840.01b,B \* | 0.870.01b,C \* |
| G9 | 0.900.03bc,A | 0.910.02bc,A | 0.910.01bc,A | 0.850.02c,A \* | 0.870.01bc,B \* | 0.900.01bc,C \* |
| G8 | 0.910.01bc,A | 0.910.01bc,A | 0.910.01bc,A | 0.860.01cd,A \* | 0.890.01c,B \* | 0.900.01bc,B |
| G5 | 0.910.01bc,A | 0.910.00bc,A | 0.910.01bc,A | 0.900.01d,A | 0.910.01c,A | 0.910.01c,A |
| G2 | 0.910.01bc,A | 0.910.00bc,A | 0.900.01bc,A | 0.900.01d,A | 0.900.01c,A | 0.900.01bc,A |
| G1 | 0.890.01c,A | 0.900.01c,A | 0.890.01c,A | 0.890.01cd,A | 0.890.01c,A | 0.890.01bc,A |
| Sires | | | | | | |
| C11 | 0.860.05a,A | 0.860.05a,A | 0.860.05a,A | 0.860.05a,A | 0.860.05a,A | 0.860.05a,A |
| G10 | 0.750.04b,A | 0.750.03b,A | 0.730.05b,A | 0.670.08bc,A \* | 0.680.05cde,A \* | 0.670.06b,A \* |
| G9 | 0.760.04b,A | 0.720.06bc,AB | 0.690.05c,A | 0.700.05b,A \* | 0.720.05bc,A | 0.710.05b,A |
| G8 | 0.760.03b,A | 0.690.05cd,B | 0.680.06c,B | 0.710.05b,A \* | 0.740.05b,A \* | 0.700.07b,A |
| G5 | 0.680.07c,A | 0.670.08de,A | 0.690.04c,A | 0.680.05bc,A | 0.690.05cd,A | 0.690.03b,A |
| G2 | 0.670.05c,A | 0.670.05de,A | 0.670.04c,A | 0.650.06c,A | 0.640.07e,A | 0.690.05b,A |
| G1 | 0.660.06c,A | 0.630.05e,A | 0.670.04c,A | 0.670.04bc,A | 0.670.03de,A | 0.690.05b,A |
| Female candidates | | | | | | |
| C11 | 0.450.02a,A | 0.450.02a,A | 0.450.02a,A | 0.450.02a,A | 0.450.02a,A | 0.450.02a,A |
| G10 | 0.480.01ab,A | 0.480.01ab,A | 0.510.01b,B | 0.460.02ab,A \* | 0.470.02ab,AB | 0.490.01b,B \* |
| G9 | 0.490.02b,A | 0.500.01b,B | 0.520.01b,C | 0.470.02ab,A \* | 0.490.02bc,B | 0.520.01bc,C |
| G8 | 0.510.01b,A | 0.510.01b,A | 0.540.01bc,B | 0.490.02bc,A \* | 0.520.01cd,B | 0.530.01cd,C |
| G5 | 0.510.01bc,A | 0.550.01c,B | 0.570.01c,C | 0.520.01cd,A | 0.550.01de,B | 0.570.01d,C |
| G2 | 0.550.01cd,A | 0.570.01c,B | 0.570.01c,B | 0.550.01d,A | 0.560.02e,AB | 0.570.01d,B |
| G1 | 0.560.01d,A | 0.560.01c,A | 0.560.01c,A | 0.550.01d,A | 0.560.01e,A | 0.560.01d,A |
| Cows | | | | | | |
| C11 | 0.480.03a,A | 0.480.03a,A | 0.480.03a,A | 0.480.03a,A | 0.480.03a,A | 0.480.03a,A |
| G10 | 0.560.02b,A | 0.590.02b,B | 0.630.01b,C | 0.530.01b,A \* | 0.560.01b,B \* | 0.610.01b,C \* |
| G9 | 0.590.03bc,A | 0.630.02c,B | 0.700.01c,C | 0.570.02bc,A \* | 0.620.02c,B | 0.680.02c,C \* |
| G8 | 0.620.02c,A | 0.670.02c,B | 0.740.02d,C | 0.600.02c,A \* | 0.660.01d,B | 0.730.02d,C |
| G5 | 0.700.02d,A | 0.770.01d,B | 0.790.02e,C | 0.690.02d,A | 0.760.01e,B | 0.780.02e,B |
| G2 | 0.760.02e,A | 0.790.02d,B | 0.780.01e,AB | 0.760.01e,A | 0.770.02e,A \* | 0.770.01de,A |
| G1 | 0.770.02e,A | 0.770.02d,A | 0.770.01de,A | 0.760.01e,A | 0.760.02e,A | 0.760.02de,A |

\*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. 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.



**Figure S1 The number of animals and repeated phenotypes in the training population**. The figure presents the results for three relative costs of phenotyping to genotyping ($P:$G). 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.

# References

Table 1 Number of genotyped animals per year by scenario and relative cost of phenotyping to genotyping.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Relative cost | Scenario | | | | | |
| 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 |

## Scenarios are named “G” for genomic, followed by the number of phenotype records per lactation. The number of phenotype records and the relative cost of phenotyping to genotyping ($P:$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.

Obšteter et al., 1

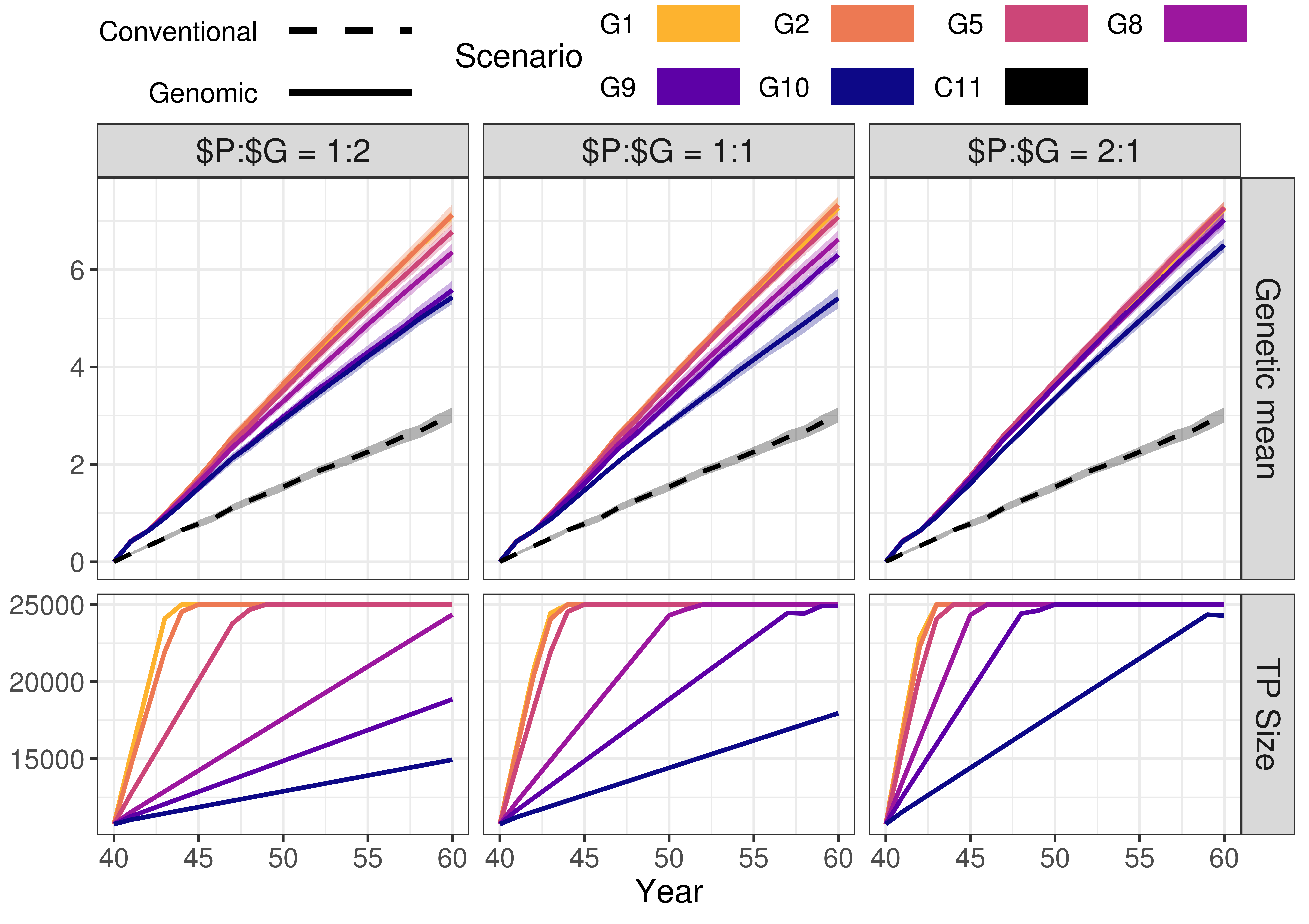


Figure 1 Genetic gain and training population size by scenario and relative cost of phenotyping to genotyping ($P:$G) with an initial training population (TP). The figure presents the means (lines) and 95% confidence intervals (polygons) across 10 replicates for the conventional (C) and genomic (G) scenarios,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

Obšteter et al., 2

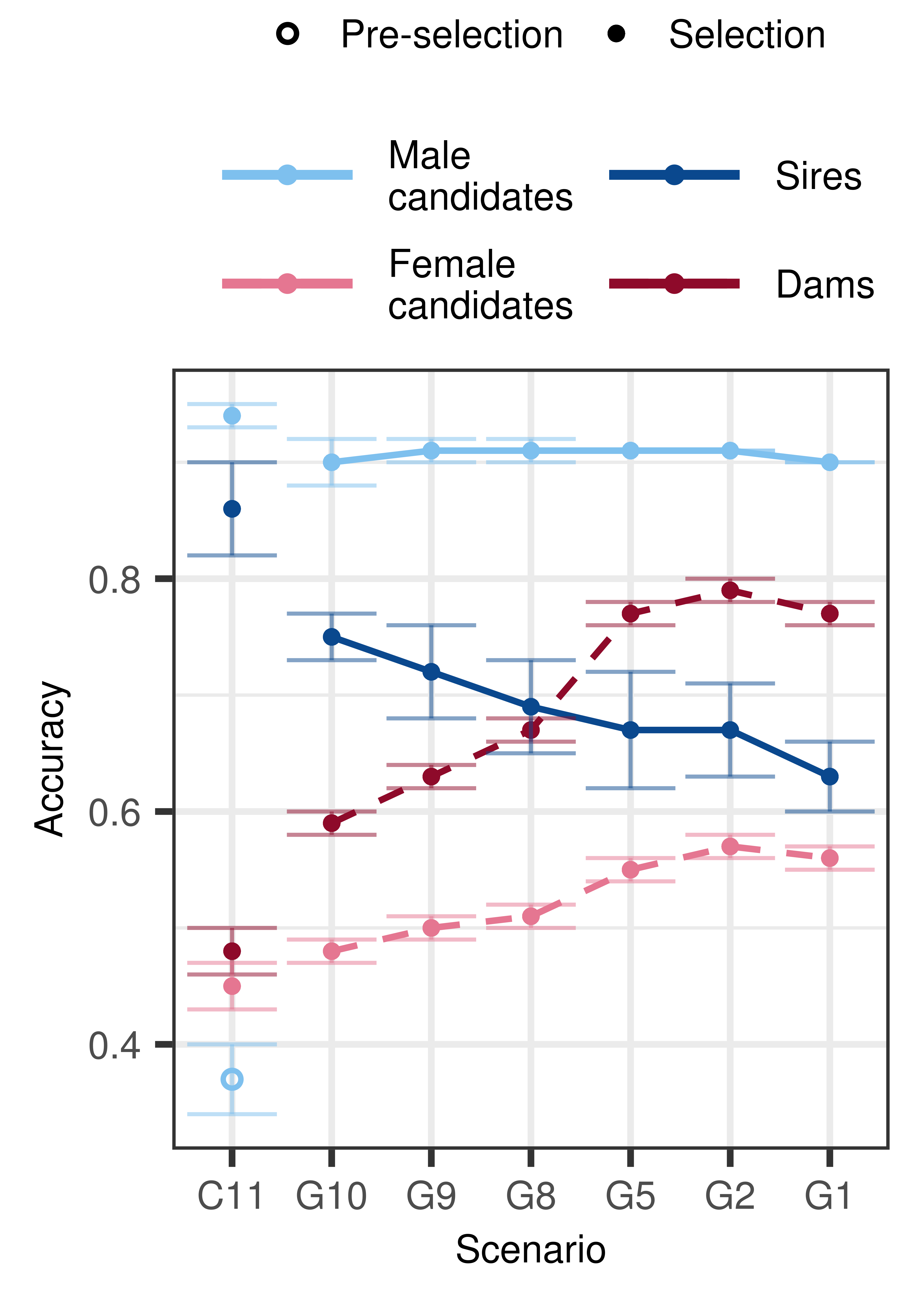


Figure 2 Accuracy by scenario with an initial training population and equal cost of phenotyping and genotyping. The figure presents the means (lines) and 95% confidence intervals (error bars) across 10 replicates for the conventional (C) and genomic (G) scenarios with numbers indicating the number of phenotype records per lactation. 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).

Obšteter et al., 3

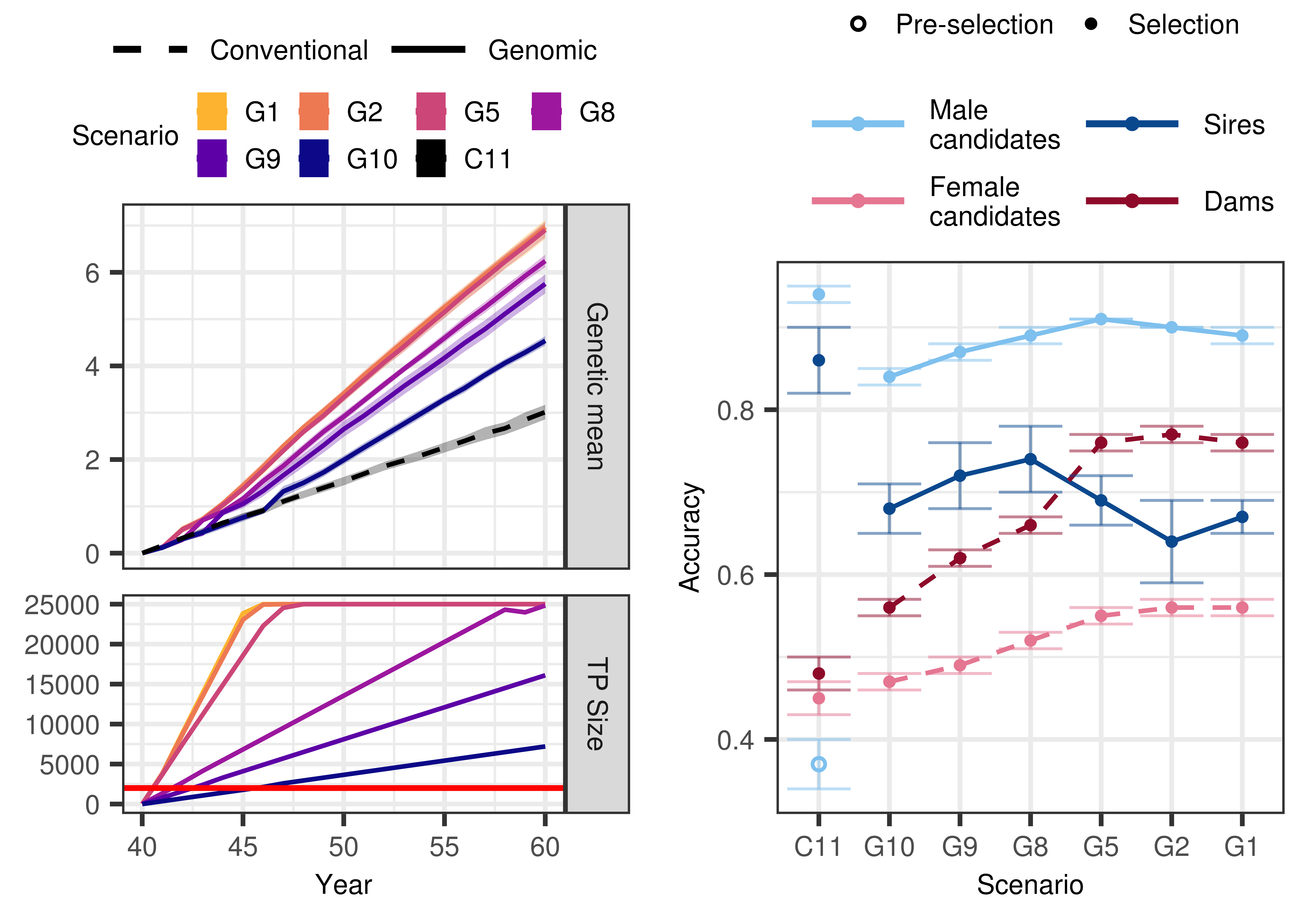


Figure 3 Genetic gain, training population size, and accuracy by scenario without an initial training population (TP) and equal cost of phenotyping and genotyping. The figure presents the means (lines or points) and 95% confidence intervals (polygons or error-bars) across 10 replicates for the conventional (C) and genomic (G) scenarios with numbers indicating the number of phenotype records per lactation. The 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).