Published in *J Anim Sci*. 2018 Nov; 96(11): 4532–4542.
doi: 10.1093/jas/sky330
PMCID: PMC6247857
PMID: 30107560

Received 2018 Jun 16; Accepted 2018 Aug 9; Published online 2018 Aug 11.

The impact of selective genotyping on the response to selection using single-step genomic best linear unbiased prediction

Jeremy T Howard,[1] Tom A Rathje,[2] Caitlyn E Bruns,[2] Danielle F Wilson-Wells,[2] Stephen D Kachman,[3] and Matthew L Spangler[1]

[1] Department of Animal Science, University of Nebraska-Lincoln, Lincoln, NE
[2] DNA Genetics, Columbus, NE
[3] Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE
Corresponding author: M. Spangler

Abstract

Across the majority livestock species, routinely collected genomic and pedigree information has been incorporated into evaluations using single-step methods. As a result, strategies that reduce genotyping costs without reducing the response to selection are important as they could have substantial economic impacts on breeding programs. Therefore, the objective of the current study was to investigate the impact of selectively genotyping selection candidates on the selection response using simulation. Populations were simulated to mimic the genome and population structure of a swine and cattle population undergoing selection on an index comprised of the estimated breeding values (**EBV**) for 2 genetically correlated quantitative traits. Ten generations were generated and genotyping began generation 7. Two phenotyping scenarios were simulated that assumed the first trait was recorded early in life on all individuals and the second trait was recorded on all versus a random subset of the individuals. The EBV

were generated from a bivariate animal model. Multiple genotyping scenarios were generated that ranged from not genotyping any selection candidates, a proportion of the selection candidates based on either their index value or chosen at random, and genotyping all selection candidates. An interim index value was utilized to decide who to genotype for the selective genotype strategy. The interim value assumed only the first trait was observed and the only genotypic information available was on animals in previous generations. Within each genotyping scenario 25 replicates were generated. Within each genotyping scenario the mean response per generation and the degree to which EBV were inflated/deflated was calculated. Across both species and phenotyping strategies, the plateau of diminishing returns was observed when 60% of the selection candidates with the largest index values were genotyped. When randomly genotyping selection candidates, either 80 or 100% of the selection candidates needed to be genotyped for there not to be a reduction in the index response. Across both populations, no differences in the degree that EBV were inflated/deflated for either trait 1 or 2 were observed between nongenotyped and genotyped animals. The current study has shown that animals can be selectively genotyped in order to optimize the response to selection as a function of the cost to conduct a breeding program using single-step genomic best linear unbiased prediction.

Keywords: genomic selection, genotype strategy, simulation

INTRODUCTION

Across the majority of livestock species, it has become a routine practice to genotype a proportion of the selection candidates in order to obtain a more accurate prediction of an animal's genetic merit (Berry et al., 2016; Knol et al., 2016). Furthermore, the incorporation of genomic information into routine genetic evaluations using multistep methods has, in general, been replaced with single-step methods. One of the issues with multiple-step methods is that they are more sensitive to biases when selective genotyping and phenotyping exists compared to single-step methods (Patry and Ducrocq, 2011; Masuda et al., 2017). One such single-step method, referred to as single-step genomic BLUP (**ssGBLUP**), utilizes a relationship matrix that blends full pedigree and genomic information to simultaneously evaluate genotyped and nongenotyped animals (Aguilar et al., 2010; Christensen and Lund, 2010). The ssGBLUP method does not rely on deregressed breeding values (Garrick et al., 2009), properly weights information from genotyped individuals and accounts for preselection bias of genomically selected parents without phenotypes (Legarra et al., 2014; Masuda et al., 2017).

Due to single-step methods being less sensitive to scenarios where selection candidates are selectively genotyped, strategies that minimize the cost of genotyping, while not reducing the response to selection, can be investigated. In general, across multiple livestock species, prior to having decided which animals to genotype some, albeit limited, phenotypes of economic importance are collected. For example, in swine, birth weight and average daily gain in the nursery can be collected along with birth weight and weaning weight in beef cattle prior to making selection decisions. As a result, information on early life traits can be utilized when deciding which animals to genotype in order to reduce the need to genotype animals with a low probability of being selected. The impact of selectively genotyping selection candidates over multiple generations on the long-term response to selection when estimating breeding values using ssGBLUP is currently unknown.

MATERIALS AND METHODS

No animal care approval was required because all data were simulated.

Simulated Data

To determine the impact of different genotyping strategies on the response to selection, a simulation and the generation of estimated breeding values (**EBV**) was conducted using the Geno-Diver software (Howard et al., 2017, V3). In order to understand if differences existed across species that have multiple offspring versus a single offspring, genomes and population structures that mimicked swine and cattle populations were generated.

Swine Genome and Population Structure

For the swine population, a genome with 5 chromosomes, each with a length of 136 Mb, was simulated. A length of 136 Mb was chosen based on the mean length of the swine autosomal chromosome. Within Geno-Diver, MaCS (Chen et al., 2009), a coalescence-based simulation program, was called to generate sequence data for 1,300 haplotypes within each chromosome. To generate levels of linkage disequilibrium (**LD**) in the sequence data that are similar to a swine population, the "Ne100_Scen2" option within Geno-Diver was utilized. The LD decay in the founder population is outlined in Supplementary Figure S1. After generating sequence information, 1,000 quantitative trait loci (**QTL**) and a marker panel consisting of 15,000 neutral markers were generated. The QTL and markers were spread equally across all 5 chromosomes resulting in 200 and 3,000 QTL and markers, respectively, within each chromosome. The number of markers per chromosome was chosen to resemble a medium density marker panel (e.g.,

Illumina PorcineSNP60K BeadChip; Illumina Inc.). In order for a QTL or marker to be chosen from the full set of base haplotypes, the minor allele frequency (**MAF**) had to be greater than 0.01 and 0.05, respectively.

The founder population consisted of 50 males and 400 females that were generated by randomly allocating base haplotypes, without replacement, to founder individuals across all chromosomes. Following the creation of the founder population, a forward-in-time simulation approach was utilized for a total of 10 generations. The population size for the forward-in-time portion was the same as the founder population and constant across generations. An animal was allowed to remain in the breeding population for a maximum of 8 generations. Male and female parents were replaced by selected offspring at a rate of 0.60 each generation. All parents were mated at random and each mating resulted in a total of 6 offspring. An offspring had an equal chance of being a male or female. Within a generation, a maximum of 2 selection candidates could be selected within each full-sib family.

Cattle Genome and Population Structure

For the cattle population, a genome with 5 chromosomes, each with a length of 87 Mb, was simulated. Similar to the swine population, a length of 87 Mb was chosen based on the mean length of the cattle autosomal chromosome. To generate levels of LD in the sequence data (n = 2,500 haplotypes) that is similar to a cattle population, the "Ne100_Scen1" within Geno-Diver was utilized when calling MaCS (Chen et al., 2009). Similar to the swine population, the LD decay in the founder population is outlined in Supplementary Figure S1. The "Ne100_Scen1" option generates lower levels of short-range LD compared to the "Ne100_Scen2" option that was utilized in the swine population. After generating sequence information, 1,000 QTL and a marker panel consisting of 8,750 neutral markers were generated and distributed equally across all 5 chromosomes. The number of markers per chromosome was chosen to resemble a medium density marker panel (e.g., Illumina BovineSNP50K BeadChip; Illumina Inc.).

The founder population consisted of 50 males and 1,000 females that were generated from the base haplotypes across all chromosomes. Similar to the simulated swine population, a forward-in-time simulation approach was utilized for 10 generations and the population size was constant across generations. The male and female parents were replaced by selected offspring at a rate of 0.40 and 0.20, respectively, each generation. An animal was allowed to remain in the breeding population for a maximum of 10 generations. All parents were mated at random and each mating resulted in 1 offspring that had an equal chance of being a male or female.

Genetic Architecture

Across both species, 2 genetically correlated quantitative traits were simulated. Two traits were generated in order to simulate an early life trait that was recorded prior to deciding whether to genotype a selection candidate and a second trait that was not recorded until after selection. Within each trait, additive effects were sampled from a gamma distribution and a correlation of 0.25 between the additive effects for trait 1 and 2 was generated following the method described in Hayashi and Iwata (2013). A range of correlations were initially investigated and no major differences were observed in terms of the proportion of genotyped animals that resulted in a diminishing rate of returns relative to genetic gain (data not shown). As a result, only the scenario with a correlation of 0.25 between the additive effects is described herein. For simplicity, it was assumed that all QTL had an impact on both traits. The marginal distribution to generate additive effects across both traits was assigned a scale and shape parameter of 0.4 and 1.66, respectively. A correlation between the additive effects was generated from 3 independent gamma distributions, $x1$, $x2$, $x3$, which were a Gamma (0.10,1.66), Gamma (0.30,1.66), and Gamma (0.30,1.66), respectively. Samples from $x1$, $x2$, and $x3$ had an equal chance of being positive or negative. The additive QTL effects for trait 1 and 2 were generated as $x1 + x2$ and $x1 + x3$, respectively. The phenotype for individual i and trait j (y_{ij}) was generated as:

$$y_{ij} = \mu_j + \sum_{q=1}^{n\text{QTL}} \gamma_{iq} a_{jq} + e_{ij},$$

where μ_j is the general mean for trait j, nQTL is the number of QTL, γ_{iq} is the genotype (i.e., 0 for the homozygote; 1 for the heterozygote; 2 for the alternative homozygote) for individual i at QTL q, a_{jq} is the additive effect for trait j at QTL q, and e_{ij} is a random residual ($e \sim N(0, \sigma^2_e)$) for individual i and trait j. The residuals were generated from independent normal distributions resulting in a residual covariance across traits being null. Across both traits, the additive effects were scaled to generate a trait with a heritability of 0.35. The phenotypic variance was set at 1.0; therefore, the residual variance was 0.65 across both traits. A range of heritability combinations were initially investigated and no major differences were observed in terms of the proportion of genotyped animals at the point of diminishing genetic gain (data not shown).

Selection and Phenotype Information

In order to build up the pedigree across both species, 2 generations of random selection and culling were utilized. For the remaining generations, animals were selected and culled based on an index comprised of the EBV for both traits. The index for individual i was constructed as outlined below:

$$\text{index}_i = \text{EBV}_{\text{Trait1}} \sigma_{\text{EBV}_{\text{Trait1}}} * 0.20 + \text{EBV}_{\text{Trait2}} \sigma_{\text{EBV}_{\text{Trait2}}} * 0.80,$$

where $\text{EBV}_{\text{Trait1}}$ is the EBV for individual i for trait 1, $\text{EBV}_{\text{Trait2}}$ is the EBV for individual i for trait 2, $\sigma_{EBV_{Trait1}}$ is the standard deviation of EBV for trait 1 on animals born in generation 2 and $\sigma_{EBV_{Trait2}}$ is the standard deviation of EBV for trait 2 on animals born in generation 2. The standard deviation across both EBV was calculated in generation 2 because it was the generation when selection began. The EBV that were used to generate the index were estimated based on a bivariate animal model as outlined below:

$$y = \mathbf{X}b + \mathbf{Z}u + e,$$

where y is a vector of phenotypic observations, b is a vector of fixed effects, u is a vector of random additive genetic effects, e is a vector of random residuals, and **X** and **Z** are incidence matrices relating observations to the fixed and random additive genetic effects, respectively. The only fixed effect was the intercept. It was assumed that the $\text{var}(u) = K \otimes G$, $\text{var}(e) = I \otimes R$, and the $\text{cov}(a,e) = 0$, where **G** and **R** are 2 × 2 matrices of variance and covariance components for random animal and residual effects and **K** is a relationship kernel. Starting at generation 7 and continuing through all remaining generations, an animal had the potential to be genotyped. As a result, EBV from generations 3 to 6 were estimated using a relationship kernel based on pedigree information (**A**; $u \sim N(0, \sigma_{2u}\mathbf{A})$). For the remaining generations, EBV were estimated using a relationship kernel that is a blend of pedigree and genomic information (Aguilar et al., 2010; Christensen and Lund, 2010) referred to as **H** ($u \sim N(0, \sigma_{2u}\mathbf{H})$). When EBV were estimated using the **A** or **H** matrix, the method will be referred to as pBLUP and ssGBLUP, respectively. When constructing the inverse of **H**, an initial genomic relationship matrix ($\mathbf{G_{raw}}$) was constructed as

$$\mathbf{G_{raw}} = \mathbf{MM}' 2 \sum p_j (1 - p_j),$$

where **M** is a genotype incidence matrix that has been centered based on allele frequencies (VanRaden, 2008) and p is the allele frequency of the second allele at the jth SNP. The allele frequencies were estimated from all genotyped animals that were utilized when estimating breeding values. As outlined in Vitezica et al. (2011), $\mathbf{A_{22}}$ and $\mathbf{G_{raw}}$ need to be compatible. The $\mathbf{A_{22}}$ matrix refers to the pedigree-based relationship for genotyped animals and was constructed as outlined in Colleau (2002). Therefore, $\mathbf{G_{raw}}$ was adjusted to make the mean diagonal and mean of all elements equal the mean diagonals and mean of all elements of $\mathbf{A_{22}}$ as outlined in Christensen et al. (2012). A weighted genomic relationship ($\mathbf{G_w}$; $0.95\mathbf{G_{raw}} + 0.05\mathbf{A_{22}}$) was utilized when blending

genomic and pedigree information. Lastly, when constructing the inverse of H ($\mathbf{H^{-1}}$), the τ and Ω values for scaling the inverse of $\mathbf{G_w}$ and $\mathbf{A_{22}}$ were both set at 1.0.

Across both species, 2 types of phenotyping scenarios were investigated in order to understand the impact of different genotyping strategies on a dense (i.e., growth rate) versus sparsely recorded trait (i.e., feed intake). Within both scenarios, the first trait was observed on all selection candidates and resembled an early life trait, but the second trait was either observed on all selection candidates (dense_dense) or only a random proportion of the selection candidates (dense_sparse). For trait 2 in the dense_sparse scenario, phenotypes were allocated randomly across all selection candidates prior to selection. The second trait was observed after an animal was selected and therefore selection candidates lacked phenotypic information for the second trait at the time of selection in the dense_sparse scenario. Given breeding values were estimated from a bivariate animal model, information on the second trait was generated based on the genetic correlation between trait 1 and 2. As a result, the EBV for the first and second trait was not the average EBV of the 2 parents. Within each sex, 20% and 40% of the selection candidates for the swine and cattle scenario, respectively, had phenotypes recorded for the second trait in the dense_sparse scenario.

Genotyping Scenarios

Starting at generation 7, ten different genotyping scenarios were generated that ranged from not genotyping any selection candidates, a proportion of the selection candidates based on either their interim index breeding value or chosen at random, and genotyping all selection candidates. These scenarios are outlined in Table 1. Within each phenotype scenario all genotyping scenarios were investigated. For the genotyping scenario where a proportion of the animals with the highest index breeding value were genotyped, an interim index value was calculated prior to a genotyping decision being made. An interim value was generated that assumed the first trait was observed while the second trait was not observed and the only genotypic information available was on animals in previous generations. It should be noted the interim value was only utilized to decide who to genotype and an updated index value that included genotypic information, if it was available, on the selection candidates was calculated prior to selection.

Table 1.

Summary of genotyping scenarios and total number of animal genotypes across all generations by species

Genotyping scenario	Summary	Mean number genotyped[1]	
		Swine	Beef
pblup	No parents and selection candidates are genotyped.	0	0
random20	All selected parents and 20% of the selection candidates genotyped at random.	3,605	2,735
index20	All selected parents and 20% of the selection candidates with the highest index breeding value.	3,293	2,498
random40	All selected parents and 40% of the selection candidates genotyped at random.	5,737	3,550
index40	All selected parents and 40% of the selection candidates with the highest index breeding value.	5,257	3,216
random60	All selected parents and 60% of the selection candidates genotyped at random.	7,928	4,371
index60	All selected parents and 60% of the selection candidates with the highest index breeding value.	7,650	4,050
random80	All selected parents and 80% of the selection candidates genotyped at random.	10,165	5,206
index80	All selected parents and 80% of the selection candidates with the highest index breeding value.	10,050	5,050
all	All the parents when genotyping was started and all selection candidates for the remaining generations.	12,450	6,050

[1]Within a genotype scenario, the mean number of genotyped animals across all generations was averaged across the 2 phenotyping scenarios.

Evaluation of Scenarios

Within each genotyping scenario, a total of 25 replicates were generated. Within each replicate, the mean true breeding values (**TBV**) for trait 1 and 2 along with the mean true index value within each generation were utilized to calculate the mean response per generation. The mean response was calculated as the difference in the associated value for all animals born in generation 10 and all animals born in generation 2. Furthermore, the correlation between TBV and EBV for trait 1 and 2 on the selection candidates within each generation, referred to as accuracy, was calculated. Lastly, the degree to which EBV were inflated/deflated across different genotype scenarios for the selection candidates was quantified by the coefficient of regression of TBV on EBV. The expected coefficient of regression is a value of 1.0, which implies the EBV are not inflated/deflated. For each metric, the 95% confidence interval was calculated across all replicates based on a randomized complete block design with replicates (i.e., block) and genotype scenario considered fixed.

RESULTS

The mean index selection response per generation for the swine and cattle population across different genotyping and phenotyping scenarios is outlined in Fig. 1. Each of the scenarios displayed a plateauing pattern in the index response as the proportion of genotyped animals approached 100%. For the scenario where only a portion of animals were genotyped at random, increases in the index response were slower as the proportion of genotyped animals increased. Alternatively, when selection candidates were genotyped based on having higher index values, selection response was quicker suggesting that genotyping more individuals provided minimal improvement in the selection response. For example, across both species and phenotyping strategies genotyping the top 60% of the selection candidates based on their index value within each sex did not result in a statistical significant (P-value > 0.05) change in the selection response compared to genotyping all selection candidates. Alternatively, when genotyping individuals at random, 80% to 100% of the selection candidates needed to be genotyped to avoid a statistically significant reduction in the index response compared to genotyping all selection candidates. The mean response to selection for trait 1, trait 2, and the index across different genotyping and phenotyping scenarios for the swine and cattle population is outlined in Supplementary Tables S1 and S2, respectively. Across both species and phenotyping scenarios, the results for trait 2 in terms of the proportion genotyped without a significant reduction in the response to selection were similar to the index response results. Lastly, no major differences in the

mean selection response for trait 1 existed across genotyping or phenotyping strategies for both species. This result is not unexpected given the phenotypes for trait 1 were observed and utilized when predicting the interim value and therefore contained more information prior to genotyping compared to trait 2 that was not observed on selection candidates. However, in general, across species and phenotyping strategies, a decrease in the response for trait 1 and an increase in the response for trait 2 were observed when EBV were estimated utilizing genomic and pedigree information (i.e., ssGBLUP across genotyping scenarios) compared to pedigree information only (i.e., pBLUP).

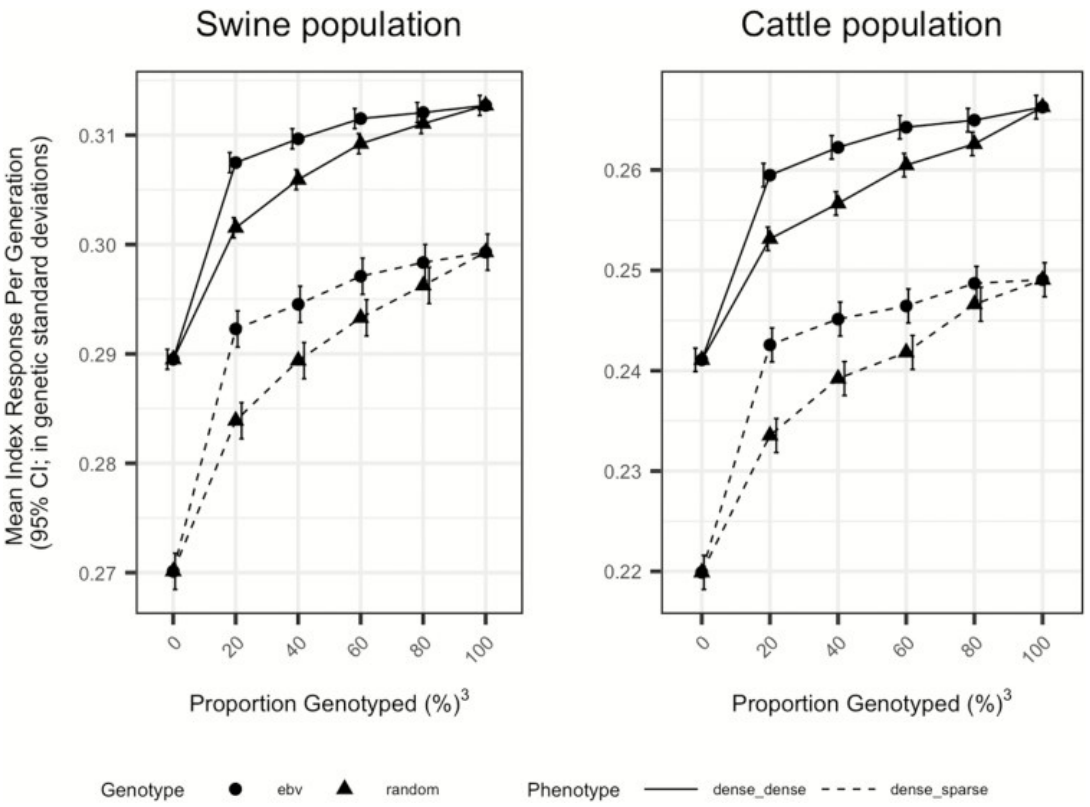

Figure 1.

Mean index true breeding value response per generation across different genotyping[1] and phenotyping[2] scenarios for simulated swine and cattle populations. [1]The genotype strategy refers to, when applicable (i.e., 20% to 80% genotyped), the criteria used to determine who to genotype. For the EBV strategy, individuals in the top index true breeding value percentile were genotyped, whereas for the random strategy individuals were genotyped at random. [2]The dense_dense phenotype strategy refers to all individuals obtaining a phenotype for both of the traits that are in the index. The dense_sparse phenotype strategy refers to all individuals obtaining a phenotype for the first trait while only a fraction (20% in swine and 40% in cattle) of the individuals obtained a phenotype for the second trait. [3]The proportion genotyped refers to the

proportion of selection candidates genotyped within each generation. The 0% refers to no animals genotyped (i.e., traditional pedigree-based selection), 20% to 80% refers to the proportion genotyped based on the genotyping scenario and 100 refers to all selection candidates being genotyped.

Outlined in Table 2 is the mean accuracy of the EBV for nongenotyped and genotyped animals across different genotyping scenarios for the swine and cattle populations. In general, across both species and phenotype scenarios, the gain in accuracy for genotyped animals was negligible for trait 1 given phenotypic information was available at the time of selection. On the other hand, for trait 2 the EBV accuracy increased for genotyped animals as the proportion of animals increased and the degree at which the accuracy changed depended on the genotyping strategy. When animals were chosen to be genotyped based on their index value, accuracy increased to a greater extent as compared to genotyping the same proportion of individuals at random. The increase in accuracy was even larger for the phenotyping scenario where trait 2 was sparsely recorded (i.e., dense_sparse) compared to the scenario where both traits were densely recorded (i.e., dense_dense). However, across both phenotyping scenarios and both species, the accuracy when selectively genotyping was numerically lower compared to randomly genotyping selection candidates at the same proportion. Although, across both species and phenotyping strategies, genotyping 60% of the selection candidates based on their index value resulted in a negligible (P-value > 0.05) reduction in the index response compared to the scenario when all animals were genotyped. Lastly, minor changes in the accuracy for nongenotyped animals were observed for trait 1 and under random genotyping of individuals for trait 2. When selectively genotyping individuals, the accuracy for nongenotyped animals for trait 2 decreased as a greater proportion of the selection candidates with a high index value were genotyped. The selection candidates that were in the nongenotyped group had genotyped parents that were older and therefore this likely resulted in compatibility issues between $\mathbf{G}$ and $\mathbf{A}_{22}$. Although the error surrounding the accuracy for nongenotyped animals for trait 2 was larger compared to the accuracy for nongenotyped animals for trait 1.

Table 2. Mean accuracy for selection candidates born after generation 6 across different genotyping[1] and phenotyping scenarios[2] for nongenotyped (NG) and genotyped (G) animals in the swine and cattle population

		Phenotype scenario 1				Phenotype scenario 2			
		Trait 1		Trait 2		Trait 1		Trait 2	
Population	**Genotyping scenario**	**NG**	**G**	**NG**	**G**	**NG**	**G**	**NG**	**G**
Swine	pblup	0.71	–	0.45	–	0.70	–	0.35	–
	random20	0.73	0.86	0.51	0.74	0.72	0.86	0.43	0.58
	index20	0.71	0.85	0.44	0.65	0.71	0.84	0.37	0.47
	random40	0.73	0.88	0.52	0.76	0.72	0.87	0.44	0.60
	index40	0.71	0.87	0.39	0.68	0.70	0.86	0.33	0.50
	random60	0.73	0.88	0.52	0.78	0.72	0.88	0.43	0.61
	index60	0.71	0.88	0.36	0.71	0.70	0.88	0.31	0.53
	random80	0.73	0.89	0.52	0.78	0.72	0.89	0.44	0.62
	index80	0.70	0.89	0.33	0.74	0.70	0.88	0.26	0.57
	all	–	0.90	–	0.79	–	0.90	–	0.63
Mean 95% confidence interval range		0.05	0.04	0.09	0.15	0.05	0.05	0.11	0.16
Cattle	pblup	0.70	–	0.47	–	0.70	–	0.40	–
	random20	0.73	0.86	0.52	0.77	0.72	0.86	0.46	0.66
	index20	0.72	0.84	0.42	0.68	0.71	0.84	0.37	0.56
	random40	0.73	0.87	0.52	0.77	0.73	0.86	0.48	0.68
	index40	0.71	0.85	0.37	0.71	0.71	0.85	0.33	0.59
	random60	0.73	0.88	0.53	0.78	0.73	0.87	0.48	0.68
	index60	0.71	0.87	0.32	0.73	0.71	0.86	0.30	0.61
	random80	0.74	0.88	0.53	0.79	0.73	0.88	0.47	0.69
	index80	0.70	0.87	0.27	0.76	0.72	0.87	0.25	0.65
	all	–	0.88	–	0.79	–	0.88	–	0.69
Mean 95% confidence interval range		0.05	0.04	0.10	0.10	0.05	0.04	0.11	0.12

[1]See Table 1 for a description of the genotyping scenarios.

[2]The dense_dense phenotype strategy refers to all individuals obtaining a phenotype for both of the traits that are in the index. The dense_sparse phenotype strategy refers to all individuals obtaining a phenotype for the first trait while only a fraction (20% in swine and 40% in cattle) of the individuals obtained a phenotype for the second trait.

Outlined in Table 3 is the mean regression of TBV on the EBV for nongenotyped and genotyped animals across different genotyping scenarios for the swine and cattle population. Across both cattle and swine populations, no differences in the regression coefficient between nongenotyped and genotyped animals for either trait 1 or 2 were observed within a given genotyping and phenotyping scenario. Therefore, the EBV for genotyped animals for a given trait are not inflated/deflated to a greater degree compared to nongenotyped animals for the same trait within the same scenario. For the swine population, the degree of inflation/deflation in the EBV was minimal and the 95% CI contained the value of 1.0 across all genotyping and phenotyping strategies. Alternatively, for multiple genotyping scenarios and across both phenotyping scenarios the EBV in the cattle population were slightly deflated based on the 95% CI not containing 1.0. Deflated EBV occurred more often for trait 2 that was not observed on selection candidates. It should be noted that even though some genotyping scenarios resulted in deflated EBV, the regression coefficient between nongenotyped and genotyped animals was not statistically different (P-value > 0.05).

Table 3. Mean inflation[1] of breeding values in the selection candidates born after generation 6 across different genotyping[2] and phenotyping scenarios[3] for nongenotyped (NG) and genotyped (G) animals in the swine and cattle population

		Phenotype scenario 1				Phenotype scenario 2			
		Trait 1		Trait 2		Trait 1		Trait 2	
Population	**Genotyping scenario**	**NG**	**G**	**NG**	**G**	**NG**	**G**	**NG**	**G**
Swine	pblup	1.00	–	0.95		0.99	–	0.95	–
	random20	1.00	1.00	1.00	1.02	0.99	0.99	1.00	1.01
	index20	0.99	1.00	1.03	1.05	0.99	0.99	1.07	1.04
	random40	1.00	1.00	1.00	1.02	0.99	1.00	1.01	1.02
	index40	1.00	1.01	1.03	1.05	0.99	1.01	1.07	1.03
	random60	1.00	1.01	1.00	1.03	1.00	1.01	1.00	1.01
	index60	1.00	1.02	1.01	1.04	0.99	1.01	1.06	1.03
	random80	1.00	1.01	1.00	1.02	0.99	1.00	1.00	1.01
	index80	1.00	1.02	1.04	1.03	1.00	1.02	1.00	1.03
	all	–	1.01	–	1.03	–	1.01	–	1.02
Mean 95% confidence interval range		0.07	0.05	0.21	0.13	0.08	0.05	0.32	0.19
Cattle	pblup	1.00	–	0.99	–	1.00	–	1.00	–
	random20	1.01	1.02	1.08	1.13*	1.01	1.05*	1.08	1.12*
	index20	1.01	1.01	1.09	1.14*	1.01	1.03	1.12	1.12*
	random40	1.01	1.02	1.08	1.13*	1.01	1.04*	1.10	1.13*
	index40	1.01	1.01	1.08	1.13*	1.02	1.03	1.12	1.14*
	random60	1.01	1.03	1.08	1.12*	1.01	1.04*	1.10	1.13*
	index60	1.00	1.02	1.04	1.13*	1.02	1.04*	1.11	1.13*
	random80	1.02	1.03	1.08	1.12*	1.00	1.04*	1.07	1.12*
	index80	1.01	1.03	1.00	1.12*	1.03	1.04*	1.04	1.13*
	all	–	1.03	–	1.11*	–	1.03*	–	1.12*
Mean 95% confidence interval range		0.10	0.07	0.26	0.15	0.10	0.07	0.34	0.18

[1]Inflation is coefficient of regression of TBV on EBV and values with an * have a 95% confidence interval that does not contain 1.0.

[2]See Table 1 for a description of the genotyping scenarios.

[3]The dense_dense phenotype strategy refers to all individuals obtaining a phenotype for both of the traits that are in the index. The dense_sparse phenotype strategy refers to all individuals obtaining a phenotype for the first trait while only a fraction (20% in swine and 40% in cattle) of the individuals obtained a phenotype for the second trait.

DISCUSSION

Using simulation, this study has provided evidence that animals can be selectively genotyped as a means to reduce the cost of genotyping without any reduction in the long-term genetic gain when breeding values are estimated using ssGBLUP. The use of genomic selection across the majority of livestock species has resulted in a large number of animals that are routinely genotyped. Therefore, methods that strategically select animals within a breeding program to genotype that reduce routine genotyping costs, without any reduction in the response to selection, are important to optimize the response to selection as a function of the cost to conduct a breeding program. Previous research has been conducted on the impact of different genotyping strategies (Lillehammer et al., 2011; Buch et al., 2012; Tribout et al., 2012; Lillehammer et al., 2013), although the impact of different genotyping strategies within the context of ssGBLUP has not been investigated. The use of ssGBLUP in routine evaluations is attractive because it is less sensitive to scenarios where animals are selectively genotyped and/or genomic preselection exists compared to multistep methods (Patry and Ducrocq, 2011; Masuda et al., 2017). As illustrated by Masuda et al. (2017), when incorporating genomic information into traditional pedigree-based EBV using multistep methods in dairy cattle, genomic preselection for genotyped sires and cows resulted in biased genetic trends across time. Furthermore, the authors found that the bias was reduced when EBV were estimated using ssGBLUP (Masuda et al., 2017).

A plateau in the index response to selection as a greater proportion of the selection candidates were genotyped was observed when choosing animals to genotype with the highest index value compared to a nearly linear increase in the selection response as more animals were genotyped at random. A similar trend was observed for the response to selection for trait 2 when selectively versus randomly genotyping selection candidates. Across both species and phenotyping strategies, the plateau of diminishing returns was observed when only 60% of the selection candidates with the largest index values were genotyped. A similar result was observed by Tribout et al. (2012), such that genotyping a limited number of preselected candidates significantly reduced financial costs, while preserving most of the benefits in terms of genetic trends. As a result, the cost of genotyping can be reduced by not genotyping selection candidates that have a low probability of being selected. Phenotypic information from the first trait along with parent average information on the second trait was included when generating the interim index value, which was utilized to determine whether an animal was genotyped. Therefore, to some degree, information on the Mendelian sampling term for the second trait is generated through the genetic correlation with the first (observed) trait, although

genotype information provides a more precise estimate of the Mendelian sampling term. When genotyping a proportion of the selection candidates at random, information on the parent average and Mendelian sampling values are not utilized when deciding who to genotype, both of which provide information on the probability of an animal being selected to serve as a parent. As a result, a greater proportion of animals needed to be genotyped to ensure all animals that have a high probably of being selected to serve as parents are genotyped, which is what was observed. For example, 80% to 100% of the selection candidates needed to be genotyped when genotyping was done at random in order for there not to be a reduction in the index response. The genotype proportion with diminishing returns is likely to be population specific and depends on the proportion of the selection candidates that are selected within a given generation and the mating design. For example, assortative mating plans result in a subset of the families with a high probability of generating selection candidates compared to random mating which was utilized in the simulation. As a result, the genotype proportion with diminishing returns needs to be taken in the context of a population breeding design. Lastly, when EBV were estimated using ssGBLUP instead of pBLUP, the selection response for trait 1 was reduced and increased for trait 2. When EBV were estimated with pBLUP, EBV for trait 2 had a lower accuracy and the resulting EBV were regressed more toward zero resulting in a lower EBV standard deviation compared to trait 1. As a result, under pBLUP the EBV for the second trait contributed less to the overall index compared to the EBV estimated using ssGBLUP.

Genotyping an animal resulted in a large increase in the accuracy and an even larger increase was observed when a selection candidate did not have phenotypic information on the trait. For strategies that genotyped a certain proportion of the selection candidates, the increase in accuracy as a greater number of animals were genotyped was negligible for trait 1 as a result of phenotypic information being available at the time of selection. For strategies that genotyped a certain percentage of the selection candidates, the accuracy of genotyped animals for trait 2 increased as more selection candidates were genotyped and the increase in accuracy was dependent on the genotyping strategy. For the selective genotyping strategy, the accuracy increased to a greater extent as more selection candidates were genotyped compared to the random genotyping strategy. Across both species and both phenotyping strategies, when genotyping the same proportion of animals, the accuracy was numerically larger under the random scenario compared to the selective genotype strategy. The accuracy when selectively genotyping at a given percentage is, in part, lower than randomly genotyping at the same percentage due to only having a portion of the full-sib and/or half-sib families genotyped. As a result, additive genetic variation explained by the markers is not being fully captured, which is verified by a smaller numerical difference in the accuracy of selective versus random genotyping as a greater proportion of the animals

are genotyped. It should be noted that the accuracy in this context is population-wide and does not reflect the standard error associated with an individual animal's EBV. As a result, selective genotyping allows for one to obtain a more precise EBV prediction (i.e., individual animal accuracy) for animals which have a high probability of being parents without any significant reduction (P-value > 0.05) in the population-wide accuracy. For random selection, a EBV prediction was more accurate, but an animal with a low and high probability of being selected has an equal chance of getting genotyped. For example, when genotyping the same proportion of animals, the numerically largest difference in accuracy for selective genotyping versus random was observed at 20%, although the selection response was larger for the selective genotyping scenario versus the random genotyping scenario. This highlights the importance of genotyping selection candidates in order to obtain an estimate of the Mendelian sampling term. For the nongenotyped animals, minor changes in the accuracy were observed for trait 1 and when randomly genotyping selection candidates for trait 2. Lastly, the accuracy for the nongenotyped animals for trait 2 decreased as a greater proportion of the high index value selection candidates were genotyped, although the error surrounding the accuracy estimate was much larger for trait 2 compared to trait 1. In the nongenotyped group for trait 2, as more individuals were genotyped the nongenotyped group was comprised of selections candidates whose parents were older compared to the genotyped group. As a result, selection candidates with older genotyped parents along with changes in allele frequencies and the additive genetic variance across time likely resulted in compatibility issues between $\mathbf{G}$ and $\mathbf{A_{22}}$. In a real population, these issues are not likely to arise due to multiple traits being selected for simultaneously and as a result less change is expected for each trait. In order to verify that the decrease was partially explained by older genotyped parents with nongenotyped offspring, a simulation similar to the swine scenario, but with discrete generations (i.e., parents are only allowed to serve as parents for 1 generation) was generated (results not shown). With discrete generations, the accuracy for nongenotyped individuals on trait 2 no longer decreased as a greater proportion of the selection candidates were selectively genotyped.

Across both species and phenotyping strategies and within each genotyping scenario for trait 1 and 2, the degree of inflation/deflation in EBV was similar across nongenotyped versus genotyped selection candidates. This is of primary importance in order to alleviate issues when comparing the EBV for animals that are not genotyped versus have genotyped information. Furthermore, across all genotype scenarios in the swine population, the 95% CI contained 1.0, although for some genotype scenarios in the cattle population the 95% CI did not contain 1.0. As outlined in Koivula et al. (2015) and more recently in Martini et al. (2017), different scaling values for $\mathbf{G}$ and $\mathbf{A_{22}}$ when setting up ssGBLUP will impact the degree that EBV are inflated/deflated. As a result, the

choice of the blending factors can be optimized, although outside the scope of the current manuscript, in order to minimize the amount EBV are inflated/deflated.

Across both phenotyping strategies the same plateau was observed in terms of the genotype proportion, but the response was lower in the dense_sparse scenario compared to the dense_dense across both populations. Therefore, optimizing the number of phenotypes and genotypes simultaneously needs to be investigated in order to further optimize the response to selection as a function of the cost to run a breeding program. Furthermore, under the dense_sparse scenario, it was assumed that within each sex, 20% and 40% of the selection candidates in the swine and cattle scenario obtained phenotypes for trait 2. A simplistic scenario was generated herein. Admittedly, in cases where traits are sex-limited, when the density of phenotypic information varies across sexes for other reasons, or when having phenotypic information on certain traits necessitates genotyping, the proportion of selection candidates that need to be genotyped could be impacted.

CONCLUSIONS

When simulating 2 phenotyping scenarios, the current study has shown that animals can be selectively genotyped in order to reduce the cost of genotyping animals, with minimal reduction in the response to selection. Using a simulated swine and cattle population, the plateau of diminishing returns was observed when only 60% of the selection candidates with the largest index values were genotyped. Therefore, selective genotype can be utilized to optimize the response to selection as a function of the cost to conduct a breeding program. Further research investigating the optimization of genotyping and phenotyping strategies is needed.

Supplementary Material

Supplementary Materials

Click here for additional data file.(3.1M, docx)

ACKNOWLEDGMENTS — This project is based on research that was partially supported by the Nebraska Agricultural Experiment Station with funding from the Hatch Act (accession number 1011203) through the USDA National Institute of Food and Agriculture.

LITERATURE CITED

1. Aguilar I., Misztal I., Johnson D. L., Legarra A., Tsuruta S., and Lawlor T. J. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. doi:10.3168/jds.2009-2730 [PubMed] [CrossRef] [Google Scholar]
2. Berry D. P., Garcia J. F., and Garrick D. J., 2016. Development and implementation of genomic predictions in beef cattle. Anim. Front. 6:32–38. doi:10.2527/af.2016-0005 [CrossRef] [Google Scholar]
3. Buch L. H., Sørensen M. K., Berg P., Pedersen L. D., and Sørensen A. C. 2012. Genomic selection strategies in dairy cattle: strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. J. Anim. Breed. Genet. 129:138–151. doi:10.1111/j.1439-0388.2011.00947.x [PubMed] [CrossRef] [Google Scholar]
4. Chen G. K., Marjoram P., and Wall J. D. 2009. Fast and flexible simulation of DNA sequence data. Genome Res. 19:136–142. doi:10.1101/gr.083634.108 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
5. Christensen O. F., and Lund M. S. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. doi:10.1186/1297-9686-42-2 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
6. Christensen O. F., , Madsen P., , Nielsen B., , Ostersen T., , and Su G. Single-step methods for genomic evaluation in pigs. Animal 6:1565–1571. doi:10.1017/S1751731112000742 2012 [PubMed] [Google Scholar]
7. Colleau J. J. 2002. An indirect approach to the extensive calculation of relationship coefficients. Genet. Sel. Evol. 34:409–421. doi:10.1051/gse:2002015 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
8. Garrick D. J., Taylor J. F., and Fernando R. L. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55. doi:10.1186/1297-9686-41-55 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
9. Hayashi T., and Iwata H. 2013. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics 14:34. doi:10.1186/1471-2105-14-34 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
10. Howard J. T., Tiezzi F., Pryce J. E., and Maltecca C. 2017. Geno-Diver: a combined coalescence and forward-in-time simulator for populations undergoing selection for

complex traits. J. Anim. Breed. Genet. 134:553–563. doi:10.1111/jbg.12277 [PubMed] [CrossRef] [Google Scholar]
11. Knol E. F., Nielsen B. and Knap P. W. 2016. Genomic selection in commercial pig breeding. Anim. Front. 6:15–22. doi:10.2527/af.2016-0003 [CrossRef] [Google Scholar]
12. Koivula M., Strandén I., Pösö J., Aamand G. P., and Mäntysaari E. A. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775–2784. doi:10.3168/jds.2014-8975 [PubMed] [CrossRef] [Google Scholar]
13. Legarra A., Chistensen O. F., Aguilar I., and Misztal I. 2014. Single step, a general approach for genomic selection. Livest. Prod. Sci. 166:54–65. doi:10.1016/j.livsci.2014.04.029 [CrossRef] [Google Scholar]
14. Lillehammer M., Meuwissen T. H., and Sonesson A. K. 2011. Genomic selection for maternal traits in pigs. J. Anim. Sci. 89:3908–3916. doi:10.2527/jas.2011-4044 [PubMed] [CrossRef] [Google Scholar]
15. Lillehammer M., Meuwissen T. H., and Sonesson A. K. 2013. Genomic selection for two traits in a maternal pig breeding scheme. J. Anim. Sci. 91:3079–3087. doi:10.2527/jas.2012-5113 [PubMed] [CrossRef] [Google Scholar]
16. Martini J. W. R., Schrauf M. F., Garcia-Baccino C. A., Pimentel E. C. G., Munilla S., Rogberg-Muñoz A., Cantet R. J. C., Reimer C., Gao N., Wimmer V., et al. 2017. The effect of the H^{-1} scaling factors τ and ω on the structure of H in the single-step procedure. Genet. Sel. Evol. 50:16. doi:10.1186/s12711-018-0386-x [PMC free article] [PubMed] [CrossRef] [Google Scholar]
17. Masuda Y., VanRaden P. M., Misztal I., and Lawlor T. J. 2017. Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. J. Dairy Sci. 101:5194–5206. doi:10.3168/jds.2009-2730 [PubMed] [CrossRef] [Google Scholar]
18. Patry C., and Ducrocq V. 2011. Accounting for genomic pre-selection in national BLUP evaluations in dairy cattle. Genet. Sel. Evol. 43:30. doi:10.1186/1297-9686-43-30 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
19. Tribout T., Larzul C., and Phocas F. 2012. Efficiency of genomic selection in a purebred pig male line. J. Anim. Sci. 90:4164–4176. doi:10.2527/jas.2012-5107 [PubMed] [CrossRef] [Google Scholar]
20. VanRaden P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. doi:10.3168/jds.2007-0980 [PubMed] [CrossRef] [Google Scholar]
21. Vitezica Z. G., Aguilar I., Misztal I., and Legarra A. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb). 93:357–366. doi:10.1017/S001667231100022X [PubMed] [CrossRef] [Google Scholar]