CHANGES IN GENETIC PARAMETERS AND ACCURACY OF GENOMIC SELECTION

by

JORGE ANGEL HIDALGO MORENO

(Under the Direction of Ignacy Misztal)

ABSTRACT

The use of genomic information increases the accuracy and persistence of predictions, and decreases the generation interval, speeding up genetic changes in populations. Intensive changes caused by selection can reduce genetic variation and can strengthen undesirable genetic correlations. Stability of genomic predictions is high for proven animals or animals without new data when ample amount of pedigree, phenotypic and genomic data is available to estimate accurately the value of nearly all the independent chromosome segments segregating in the population. The objective of this dissertation was to investigate changes in genetic parameters, stability, and decay of genomic predictions over time in populations under genomic selection. Datasets from pigs, beef cattle and broilers were used. Genetic parameters for fitness and growth traits were estimated by Gibbs sampling in a pig population. Over 10 years, heritabilities for fitness and growth traits decreased ~25 and ~50%, respectively. Genetic correlations between fitness and growth traits that were initially positive remained stable, while those that were initially negative became more negative. Stability of predictions was evaluated during 1 year in a beef cattle population contrasting monthly evaluations of estimated breeding values (EBV) and genomic EBV for genotyped animals with or without their own phenotypes or progeny phenotypes. Average absolute changes for EBV were about two times smaller than for GEBV, except for animals with

new progeny phenotypes (≤0.12 and ≤0.11 additive genetic standard deviations (SDa) for EBV and GEBV). The maximum absolute changes for EBV (≤2.95 SDa) were greater than for GEBV (≤1.59 SDa). Decay of genomic predictions was assessed in a broiler population across 7 years in the progeny, grand progeny, and great grand progeny. The use of genomic data increased accuracy and persistence of genomic predictions about two-fold compared to traditional evaluation. Accuracy of genomic predictions declined over generations, on average, 40% from progeny to grand progeny and 4% from grand progeny to great grand progeny. Strong selection reduced heritabilities and emphasized the antagonistic genetic relationships between fitness and growth traits. Genomic predictions are as stable as traditional evaluations for animals with new phenotypic data and with less extreme changes because of increased accuracy.

INDEX WORDS: accuracy, genetic correlation, genetic variance, genomic information, independent chromosome segments

CHANGES IN GENETIC PARAMETERS AND ACCURACY OF GENOMIC SELECTION

by

JORGE ANGEL HIDALGO MORENO

B.S., Universidad Autónoma Chapingo, Mexico, 2009

M.S., Universidad Autónoma Chapingo, Mexico, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

CHANGES IN GENETIC PARAMETERS AND ACCURACY OF GENOMIC SELECTION

by

JORGE ANGEL HIDALGO MORENO

Major Professor:	Ignacy Misztal
Committee:	Daniela Lourenco
	Romdhane Rekaya

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2021

DEDICATION

To Ise, Giorgio, Chuyo and Sofana.

ACKNOWLEDGEMENTS

Dr. Misztal, this was an invaluable opportunity and I have no way to thank you enough. Thank you for accepting me as your student, and for all the guidance not only in the academic but also in the personal life. Dr. Lourenco, thank you for having always your door open to me and for the support in all ways. Dr. Rekaya, thank you for all the encouragement and your advice during my studies. Dr. Tsuruta and Dr. Masuda, special thanks for the guidance with all my questions about the research projects. Thank you everybody in the group for allowing me to be part of this family.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Genetic variation is crucial in animal breeding programs because it determines the response to selection. Directional and stabilizing selection, which are the most used selection methods in animal breeding programs, generate a negative correlation between pairs of loci because of linkage disequilibrium. Therefore, genetic variation in the generation after selection will be reduced. Accuracy, persistence, and stability of genomic predictions depend on the amount of data and genetic parameters. When the amount of information is enough to accurately estimate the value of nearly all independent chromosome segments segregating in a population, accuracy, persistence, and stability of genomic predictions will be high. If genetic parameters change over time because of selection, these changes must be accounted for in the estimation of accuracy, persistence, and stability of predictions to have realistic estimates.

The objective of this dissertation was to investigate changes in genetic parameters, stability, and decay of genomic predictions over time in populations under genomic selection. In Chapter 2, a literature review is presented. Chapter 3 presents the investigation about changes in genetic parameters over ten years in a pig population undergoing selection. In chapter 4, an assessment of the stability of genomic predictions over one year in a large, genotyped beef cattle population is presented. In chapter 5, the decay of genomic predictions in the progeny, grand progeny, and great grand progeny in a broiler population is investigated across seven years of traditional and genomic selection. The general conclusions of this dissertation are presented in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

In 1970, Crow and Kimura proved that under the infinitesimal model assumptions (Fisher, 1918), selection can change the phenotypic mean with negligible impact on the genetic variation in the population unless it is small, or selection is very intensive. Bulmer (1971) considered the effect of one generation of selection on the genetic variance in the next generation. Assuming additive and dominance components, Bulmer (1971) proved that the genetic variance (σ_u^2) under selection is the sum of two components; the equilibrium genetic variance ($\sum_{i=1}^{N} \sigma_{g_i}^2$; N is the total number of loci), and the linkage disequilibrium contribution ($\sum_{i \neq j} \sigma_{g_i,g_j}$). Therefore, the genetic variance under selection can be expressed as:

$$\sigma_u^2 = \sum_{i=1}^{N} \sigma_{g_i}^2 + 2 \sum_{i \neq j} \sigma_{g_i,g_j}$$

where g_i is the genotypic value at the ith locus, and σ_{g_i,g_j} is the covariance between the ith and jth loci.

The equilibrium genetic variance term (first term on the right-hand side) represents the value that the genetic variance would have in the absence of linkage disequilibrium, whereas the linkage disequilibrium term (second term on the right-hand side) originates from the covariance between pairs of loci introduced by selection. In the absence of selection, this covariance (linkage disequilibrium contribution) is zero. However, under directional or stabilizing selection, the covariance will be negative, reducing the genetic variance. Conversely, under disruptive selection, the covariance will be positive, increasing the genetic variance.

In the short term, the effects of linkage disequilibrium in selected populations are more important than the changes in gene frequencies (Bulmer, 1976). Therefore, the changes in the genetic variance are mainly explained by the covariance between pairs of loci. Recombination between unlinked loci removes linkage disequilibrium, and its contribution is not fully passed to the next generation. Indeed, linkage disequilibrium is roughly halved every generation, allowing a partial recovery of the genetic variance lost. After repeated generations of selection, an equilibrium is eventually achieved, and genetic variance will be stable (Villanueva and Kennedy, 1990).

Over longer time, and under selection, changes in gene frequencies will be more important. Any initial genetic variation is removed by selection and genetic drift, and further selection response will depend on the creation of new genetic variation, mainly by mutation (Walsh and Lynch, 2018). The dynamic of the changes in the genetic variance is not obvious, and usually is difficult to explain the coexistence of enough genetic variance and steady selection response in most of the traits across populations (Walsh and Blows, 2009). However, it is true that the genetic variance is not a static parameter, and it can change over time. As stated initially by Fisher (1930), selection changes the genetic variance; the magnitude and direction of the changes depend on gene frequencies, dominance, epistasis, linkage disequilibrium and mating system.

In multivariate selection, the changes in the genetic variance of a selected trait will cause changes in the genetic variance of correlated traits. No matter the sign of the correlation, there is a reduction on the genetic variance of correlated traits (Walsh and Lynch, 2018). Changes in genetic covariances are more difficult to predict than changes in genetic variances. Two different genetic mechanisms contribute to genetic covariances, linkage disequilibrium, and pleiotropy, and both can change over time. As stated before, selection creates linkage disequilibrium, and alleles at different loci affecting single traits are co-inherited, creating a correlation between pairs of loci.

With pleiotropy, 1 allele influences 2 or more traits. Complexity arises because to predict medium to long-term selection responses in covariances, the distribution of allelic effects across all loci is critical, and this information is never available.

Itoh (1991) theoretically demonstrated that multi-trait selection changes the genetic correlation always in an undesirable direction. Thus, the remaining genetic variation upon which future selection must act for response will face more constrains. Animal breeding programs should evaluate not only trends for the mean of the population but also for the genetic parameters to evaluate the success of the breeding program.

The stability and accuracy of genomic predictions depend on the amount of data and population parameters. Genomic predictions have increased accuracy compared to traditional genetic evaluations because genomic information yields more accurate estimates of the Mendelian sampling effects (Hayes et al., 2009; Cole and VanRaden, 2011). Accuracy of genomic predictions is a function of the proportion of the genetic variance captured by the single nucleotide polymorphisms (SNP) and the accuracy of the SNP effect estimates, which depends on the amount of phenotypic data available, the heritability of the trait, and the statistical method of choice (Dekkers, 2007; Goddard, 2009).

Accuracy and persistence of genomic predictions will be high in a population with a large number of phenotypes if all the genetic variation in the population is captured. Genomic information has limited dimensionality, meaning there is a threshold for the amount of data that is not redundant and can increase the accuracy of genomic predictions. This information is equivalent to the number of independent chromosome segments segregating in a finite population and can be estimated as $4N_eL$ (Stam, 1980), where N_e is the effective population size and L is the length of the genome in Morgans. Equivalently, the number of independent chromosome segments can be

estimated as the number of the largest eigenvalues explaining 98% of the variation in the genomic relationship matrix (Pocrnic et al., 2016a).

Pocrnic et al. (2016b) estimated the number of independent chromosome segments to be ~5K in pigs, ~15K in Angus beef cattle, and ~4K in broilers. These numbers correspond to the number of core animals in the algorithm for proven and young animals (APY; Misztal et al., 2014). Pocrnic et al. (2019) reported that accuracies were marginally smaller when using 25% instead of 100% of the optimal number of core animals in APY, suggesting that genomic selection acts on clusters of independent chromosome segments rather than on individual independent chromosome segments.

The APY is used to exploit the limited dimensionality of the genomic information (Misztal et al., 2021) in single-step genomic best linear unbiased prediction (ssGBLUP; Aguilar et al., 2001). The ssGBLUP method requires the inversion of the genomic relationship matrix ($\mathbf{G}$), which can be difficult to obtain when the number of genotyped animals is large (i.e., more than 100,000 genotyped animals) due to computing bottlenecks (memory, time, etc.). Thus, the use of direct inversion of $\mathbf{G}$ in ssGBLUP limits the number of genotyped animas that can be included in the evaluation system. In the APY approach, based on the theory of the limited dimensionality of the genomic information (Misztal et al., 2016), it is assumed that N core animals contain the information to estimate the value of all independent chromosome segments segregating in the population. Therefore, a generalized inverse of the genomic relationship matrix can be constructed by recursions on N core animals. Then, breeding values of noncore animals are obtained based on recursion of breeding values (u) of noncore (n) on core (c) animals as follows:

$$\mathbf{u}_{\mathrm{n}} = \mathbf{P}_{\mathrm{nc}}\mathbf{u}_{\mathrm{c}} + \boldsymbol{\varepsilon},$$

where $\mathbf{P}$ relates breeding values of noncore to core animals and $\boldsymbol{\varepsilon}$ is an estimation error. In matrix notation:

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_c \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{u}_c \\ \boldsymbol{\varepsilon} \end{bmatrix}$$

The genomic relationship matrix in APY ($\mathbf{G}_{APY}$) is:

$$\mathbf{G}_{APY} = \begin{bmatrix} \mathbf{G}_{cc} & \mathbf{G}_{cn} \\ \mathbf{G}_{nc} & \mathbf{G}_{nn} \end{bmatrix}$$

$$\text{var}(\mathbf{u}) = \mathbf{G}_{APY}\sigma_u^2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{P} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \text{var}(\mathbf{u}_c) \\ \text{var}(\boldsymbol{\varepsilon}) \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{P} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

where, σ_u^2 is the additive genetic variance.

The main advantage of APY is that only $\mathbf{G}_{cc}$, the core set of $\mathbf{G}_{APY}$, is explicitly inverted, whereas the coefficients in the inverse for all the remaining animals are calculated as linear functions of the inverse of $\mathbf{G}_{cc}$. The inverse of $\mathbf{G}_{APY}$ is:

$$\mathbf{G}_{APY}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{P}' \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \text{var}(\mathbf{u}_c)^{-1} & \mathbf{0} \\ \mathbf{0} & \text{var}(\boldsymbol{\varepsilon})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{P} & \mathbf{I} \end{bmatrix} \sigma_u^{-2}$$

$\mathbf{P} = \mathbf{G}_{nc}\mathbf{G}_{cc}^{-1}$ and $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{M}\sigma_u^2 = \text{diag}(g_{ii} - g_{i,c}\mathbf{G}_{cc}^{-1}g_{c,i})\sigma_u^2$ for individual i in the noncore group. The final formula is:

$$\mathbf{G}_{APY}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{cc}^{-1}\mathbf{G}_{nc} \\ \mathbf{I} \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} -\mathbf{G}_{nc}\mathbf{G}_{cc}^{-1} & \mathbf{I} \end{bmatrix}$$

The $\mathbf{M}$ matrix is not usually diagonal but ignoring its off-diagonal elements may result in an approximation that can lead to a reduction in the accuracy of ssGBLUP. However, Bradford et al. (2017) demonstrated that when the number of core animals is equal or greater than the number of the largest eigenvalues explaining 98% of the variation in $\mathbf{G}$, the accuracy no longer increases, indicating that the information in the off-diagonal elements of $\mathbf{M}$ is redundant and can be ignored. The same authors studied the optimal definition of the core set in APY and concluded that the

choice of animals did not affect prediction accuracy and that the random choice is better for computational reasons.

The APY reduces computational cost, and it was successfully used to construct the inverse of the genomic relationship matrix for up to 2.3 million genotyped animals (Masuda et al., 2019). An important point related to the use of APY in ssGBLUP evaluations is that the core set have to represent the genetic variation in the population, and as stated before, this requirement is fulfilled when the number of core animals is equal to the number of the largest eigenvalues representing 98% of the variation in **G**, assuming that the remaining 2% is noise (Pocrnic et al., 2016a). Over time, the number of genotyped individuals in livestock populations is increasing, and the core set must be updated. As the noise term (e.g., error term) is different for every core set, changing the core set creates fluctuations in the estimated breeding values (Misztal et al., 2021).

The study of the core-dependent fluctuations in APY is important to evaluate the magnitude of the changes in estimated breeding values and to understand when to sample a new core set. Changes in estimated breeding values will be present also when new information is added to the evaluation system. The amount of data used in the evaluation system is related to the standard error, accuracy, and possible changes of predictions (Van Vleck, 2016). The larger the amount of data available for an animal, the more stable its prediction is, and the lower the changes that may occur when new information is added.

For an animal without new own or progeny phenotypes, the estimated breeding value (EBV) from the traditional best linear unbiased prediction (BLUP) is very stable even with moderate accuracy. In contrast, genomic EBV (GEBV) with moderate or low accuracy can change in the absence of new own or progeny phenotypes added in the evaluation system. The reason for

those larger changes in GEBV than in EBV is a higher number of links between animals through genomic than pedigree relationships, and decay of genomic information across generations.

The EBV can be decomposed into a parental average, a yield deviation, and a progeny contribution (VanRaden and Wiggans, 1991). Additionally, the decomposition of GEBV also includes a direct genomic value and a pedigree prediction, and the last one is needed to avoid double counting of relationships (Aguilar et al., 2010). Without new phenotypic information for an animal, the EBV based on parental average is expected to be stable (assuming that the parents had high EBV accuracy). However, the addition of new progeny phenotypes can result in large changes in EBV for an animal because, in the absence of large progeny groups (i.e., few or no progeny), the additional information is sizeable respective to the parental average.

In the case of GEBV, every genotyped animal with phenotypic information influences the direct genomic value of all genotyped animals. Consequently, the GEBV of animals with no additional phenotypic information of their own or their relatives could change. However, if the reference population is large, the accuracy of the direct genomic value will be high (Lourenco et al., 2015); thus, additional phenotypic records would have a lower impact on GEBV because they would contribute with less information than the direct genomic value.

Genomic information can be considered an extra source of "information" in the evaluation system. For milk yield in Holsteins, the information from the sire and the dam, both with 99% reliability, is equivalent to having 14 daughters with phenotypic records (VanRaden and Wiggans, 1991). The genotype of an animal provides information equivalent to 37.5 daughters for milk yield, 240.6 daughters for daughter pregnancy rate, and 780.2 daughters for heifer conception rate (https://queries.uscdcb.com/eval/summary/comparexml_menu.cfm?R_menu=v_2004.v_Young_Bulls.v_Holstein_wddx#StartBody). For weaning weight in Angus beef cattle, having the

genotype of an unproven bull is equivalent to having 27 calves with weaning weight records (https://www.angus.org/AGI/GenomicEnhancedEPDs.pdf). The study of the changes in GEBV when new information is added is an important topic as it involves genomic relationships instead of pedigree relationships in comparison with the classical EBV.

Accuracy of genomic predictions is an important parameter in a breeding program because of its direct relationship with the selection response. The decay of accuracy over time in initial genomic selection studies using stochastic simulations was small. Meuwissen et al. (2001) found that the accuracy for a trait with major genes, in the absence of artificial selection, decreased from 0.84 to 0.72 after 5 generations without phenotyping the genotyped animals. Muir (2007), also using a stochastic simulation concluded that accuracy of genomic selection in breeding programs decays faster for traits under selection.

Using simulated data from layers, Wolc et al. (2015) reported that after 3 years of selection the accuracy remained almost stable, decaying from 0.77 to 0.73 when the breeding program included new animals with genotypes and phenotypes every generation. On the contrary, the accuracy declined from 0.77 to 0.34 when no new animals with phenotypes were included, and the selection response was ~30% smaller. In the same study, the results observed using real data for approximately 2,700 genotyped animals were consistent with those of the simulation; however, the accuracy was lower.

In the absence of inbreeding, the relatedness and potential contributions of ancestors of an animal, based on pedigree relationships, decline 50% for each generation traced back in the pedigree. Therefore, very distant ancestors have small or even negative effect on the accuracy of predictions of the youngest animals (Lourenco et al., 2014). The decline of relatedness and potential contributions based on genomic relationships will depend on the method used to compute

G and whether it is based on identical by state or identical by descent relationships. Thus, it is of interest to study the contribution of genotypes, pedigree, and phenotypes from distant generations to the accuracy of genomic predictions in the selection candidates.

One of the most used methods to estimate the accuracy of genomic predictions is a cross-validation approach called predictive ability, in which the correlation between genomic predictions and phenotypes adjusted for fixed effects is computed (Legarra et al., 2008). The statistic of this method is sensitive to incorrect heritabilities or precorrection of phenotypes and may yield biased results if those are incorrect (Legarra and Reverter 2018). Legarra and Reverter (2018) proposed a semi-parametric method based on linear regression (i.e., the **LR** method) that relies on the comparison of successive genomic evaluations based on partial and whole data. The statistics of this method do not require the pre-correction of phenotypes and might better estimate the accuracy of genomic predictions.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 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(2):743-752. https://doi.org/10.3168/jds.2009-2730

Bulmer, M. G. 1971. The effect of selection on genetic variability. Am. Nat. 105:201–211. https://doi.org/10.1086/282718

Bulmer, M. G. 1976. The effects of selection on genetic variability: a simulation study. Genet. Res. 28(2):101-117. https://doi.org/10.1017/S0016672300016797

Cole, J. B., and P. M. VanRaden. 2011. Use of haplotypes to estimate Mendelian sampling effects and selection limits. J. Anim. Breed. Genet. 128:446–455. https://doi.org/10.1111/j.1439-0388.2011.00922.x

Dekkers, J. C. M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331–341. doi: https://doi.org/10.1111/j.1439-0388.2007.00701.x

Fisher, R. A. 1 930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh. 52(2):399-433. https://doi.org/10.1017/S0080456800012163

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

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. (Camb). 91:47–60. https://doi.org/10.1017/S0016672308009981

Itoh, Y. 1991. Changes in genetic correlations by index selection. Genet. Sel. Evol. 23:301. https://doi.org/10.1186/1297-9686-23-4-301

Legarra A., Reverter A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet Sel Evol. 50:53. https://doi.org/10.1186/s12711-019-0511-5

Legarra, A., C. Robert-Granié, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611– 618. https://doi.org/10.1534/genetics.108.088575

Lourenco, D. A. L., I. Misztal, S. Tsuruta, I. Aguilar, T. J. Lawlor, S. Forni, and J. I. Weller. 2014. Are evaluations of young animals benefiting from the past generations? J. Dairy Sci. 97:3930–3942. https://doi.org/10.3168/jds.2013-7769

Lourenco, D. A. L., S. Tsuruta, B. O.Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, J. K. Bertrand, T. S. Amen, L. Wang, D. W. Moser, et al. 2015. Genetic evaluation using single-step genomic BLUP in American Angus. J. Anim. Sci. 93:2653–2662. https://doi.org/10.2527/jas.2014-8836

Masuda, Y., S. Tsuruta, E. Nicolazzi, and I. Misztal. 2019. Singlestep GBLUP including more than 2 million genotypes with missing pedigrees for production traits in US Holstein. In: Interbull Meeting; Cincinnati (OH); June 22 to 24, 2019; Available from https://interbull.org/static/web/10_30_Masuda_final.pdf.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819-1829. https://doi.org/10.1534/genetics.110.116590

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

Misztal, I. 2016. Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. Genetics 202(2):401–409. https://doi.org/10.1534/genetics.115.182089

Misztal, I., S. Tsuruta, I. Pocrnic, and D. Lourenco. 2021. Core-dependent changes in genomic predictions using the algorithm for proven and young animals in single-step genomic best linear unbiased prediction. J. Anim. Sci. 98(12):1-8. https://doi.org/10.1093/jas/skaa374

Muir, W. M. 2007. Comparison of genomic and traditional BLUP- estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342–355. https://doi.org/10.1111/j.1439-0388.2007.00700.x

Muir, W. M. 2007. Comparison of genomic and traditional BLUP- estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342–355. https://doi.org/10.1111/j.1439-0388.2007.00700.x

Pocrnic, I., D. A. L. Lourenco, K. Y. Masuda, and I. Misztal. 2019. Accuracy of genomic BLUP when considering a genomic relationship matrix based on the number of the largest eigenvalues: a simulation study. Genet. Sel. Evol. 51:75. https://doi.org/10.1186/s12711-019-0516-0

Pocrnic, I., D. A. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016a. The dimensionality of genomic information and its effect on genomic prediction. Genetics 203:573–581. https://doi.org/10.1534/genetics.116.187013

Pocrnic, I., D. A. Lourenco, Y. Masuda, and I. Misztal. 2016b. Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species. Genet. Sel. Evol. 48:82. https://doi.org/10.1186/s12711-016-0261-6

Stam, P. 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. Genet. Res. 35:131–155. https://doi.org/10.1017/S0016672300014002

Van Vleck, L. D. 2016. EPDs and risk. Faculty Papers and Publications in Animals Science. 898:107–114. www.BeefImprovement.org. Accessed May 4, 2020.

VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737–2746. https://doi.org/10.3168/jds.S0022-0302(91)78453-

Villanueva, B., and B. W. Kennedy. 1990. Effect of selection on genetic parameters of correlated traits. Theor. Appl. Genet. 80:746–752. https://doi.org/10.1007/BF00224187

Walsh, B. and M. Lynch. 2018. Evolution and selection of quantitative traits. Oxford, United Kingdom.

Walsh, B. and M. W. Blows. 2009. Abundant genetic variation + strong selection = multivariate genetic constrains: a geometry view of adaptation. Annu. Rev. Ecol. Evol. Syst. 40:41-59. https://doi.org/10.1146/annurev.ecolsys.110308.120232

Wolc, A., H. H. Zhao, J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, C. Stricker, D. Habier, R. L. Fernando, D. J. Garrick, S. J. Lamont, and J. C. M. Dekkers. 2015. Response and inbreeding from a genomic selection experiment in layer chickens. Gen. Sel. Evol. 47:59. https://doi.org/10.1186/s12711-015-0133-5

CHAPTER 3

CHANGES IN GENETIC PARAMETERS FOR FITNESS AND GROWTH TRAITS IN PIGS UNDER GENOMIC SELECTION[1]

[1] Jorge Hidalgo, Shogo Tsuruta, Daniela Lourenco, Yutaka Masuda, Yijian Huang, Kent A. Gray, and Ignacy Misztal. 2020. *Journal of Animal Science*. 98(2):1-12. Reprinted here with permission of the publisher.

ABSTRACT

Genomic selection increases accuracy and decreases generation interval, speeding up genetic changes in the populations. However, intensive changes caused by selection can reduce the genetic variation and can strengthen undesirable genetic correlations. The purpose of this study was to investigate changes in genetic parameters for fitness traits related with prolificacy (FT1) and litter survival (FT2 and FT3), and for growth (GT1 and GT2) traits in pigs over time. The data set contained 21,269 (FT1), 23,246 (FT2), 23,246 (FT3), 150,492 (GT1) and 150,493 (GT2) phenotypic records obtained from 2009 to 2018. The pedigree file included 369,776 animals born between 2001 and 2018, of which 39,103 were genotyped. Genetic parameters were estimated with bivariate models (FT1-GT1, FT1-GT2, FT2-GT1, FT2-GT2, FT3-GT1 and FT3-GT2) using 3-year sliding subsets. Computations were performed with (GEN) or without genotypes (PED) with a Bayesian implementation using the GIBBS3F90 program. For GEN (PED), the changes in heritability from the first to the last year interval, i.e., from 2009-11 to 2015-18 were 8.6 to 5.6 (7.9 to 8.8) for FT1, 7.8 to 7.2 (7.7 to 10.8) for FT2, 11.4 to 7.6 (10.1 to 7.5) for FT3, 35.1 to 16.5 (32.5 to 23.7) for GT1, and 35.9 to 16.5 (32.6 to 24.1) for GT2. Differences were also observed for genetic correlations as they changed from -0.31 to -0.58 (-0.28 to -0.73) for FT1-GT1, -0.32 to -0.50 (-0.29 to -0.74) for FT1-GT2, -0.27 to -0.45 (-0.30 to -0.65) for FT2-GT1, -0.28 to -0.45 (-0.32 to -0.66) for FT2-GT2, 0.14 to 0.17 (0.11 to 0.04) for FT3-GT1, and 0.14 to 0.18 (0.11 to 0.05) for FT3-GT2. Strong selection in pigs reduced heritabilities and emphasized the antagonistic genetic relationships between fitness and growth traits. With genotypes considered, heritability estimates were smaller and genetic correlations were greater than estimates with only pedigree and phenotypes. When selection is based on genomic information, genetic parameters estimated without this information can be biased because pre-selection is not accounted for by the model.

INTRODUCTION

The main purpose of genomic selection is to accelerate the genetic progress. This is accomplished by increasing the accuracy of selection and decreasing the generation interval. However, the breeder's equation includes the genetic variance, and accuracies in multi-trait selection depend on genetic correlations (Walsh and Lynch, 2018). If genetic parameters change as a result of strong selection, the genetic gain as predicted using old parameters may not be realized.

In general, heritability and genetic correlations are expected to change under selection, with the amount of change dependent on the intensity of selection and initial genetic variance (Falconer and Mackay, 1996; Walsh and Lynch, 2018). Under directional selection, negative linkage disequilibrium (LD) is created introducing a negative correlation between pairs of loci, which decreases genetic variance, heritability, and selection response. Conversely, under disruptive selection, positive LD is generated introducing a positive correlation between pairs of loci (Bulmer, 1971; Walsh and Lynch, 2018). Two different genetic mechanisms contribute to genetic covariances, LD and pleiotropy, and both can change over time. Selection creates LD, and alleles at different loci affecting single traits are co-inherited, creating a correlation between pairs of loci. With pleiotropy, one allele influences two or more traits.

The distribution of allelic effects is critical to predict medium to long-term selection response in covariances. Changes in genetic covariances are likely to be more unpredictable than changes in genetic variances (Walsh and Lynch, 2018). Under directional selection, ignoring the reduction in genetic variance leads to overestimation of the accuracy of selection (Bijma, 2012; Gorjanc et al., 2015). In the multi-trait selection, the genetic covariances also play an important role and appropriate estimates have to be used to calculate correct accuracies. Additionally, genetic

covariances can evolve away from the direction favored by selection as a result of the introduced correlation between pairs of loci, making it harder to realize genetic changes in the desirable direction for each trait.

Estimation of genetic parameters over time is complex, especially under genomic selection. A general method would be to use a random regression model over time (Tsuruta et al., 2004). However, such a model is computationally expensive, especially with a large number of genotyped animals, and its ability to model complex changes is limited by the type and order of the regression functions. Another option is to use time intervals (i.e., slices) so that only a fraction of the data is utilized in each analysis. However, changes inside the intervals are averaged, and intervals need to be large enough to avoid biases due to earlier selection (Cesarani et al., 2019).

The objective of this study was to investigate changes in genetic parameters for fitness and growth traits in pigs under genomic selection using data in time intervals. We defined three year intervals to have enough data and the subsequent intervals overlapped two years, i.e., the first interval included data from 2009 to 2011, the next included data from 2010 to 2012 and so on, and the last interval included four years (2015 to 2018) because the last year contained few data points.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not needed as data were obtained from preexisting databases.

Data

Data for fitness traits related to prolificacy (**FT1**) and litter survival (**FT2** and **FT3**), and for classical growth (**GT1** and **GT2**) traits, recorded from 2009 to 2018, were provided by Smithfield Premium Genetics (Roanoke Rapids, North Carolina, USA). Initial pedigree consisted

of 369,776 animals from one line born between 2001 and 2018, of which 39,103 were genotyped. The objective of the breeding program is to increase growth traits, FT1 and FT2, and decrease FT3. Numbers of animals with genotypes, phenotypes, and in the pedigree per year and interval are shown in Table 3.1 and Table 3.2, respectively.

Analyses and Computations

Variance components were estimated using a Bayesian approach via the Gibbs sampling algorithm as implemented in GIBBS3F90 program (Misztal et al., 2014) with (**GEN**) or without genotypes (**PED**). The analyses were separately performed in each interval. A single Gibbs chain of total length of 100,000 rounds was initially generated. After discarding the initial 10,000 samples as burn-in, one in every 10 samples was stored to compute means and standard deviations of the posterior distributions. The means were used as estimates of the (co)variance components, and their posterior standard deviations were considered to be a measurement of their estimation errors.

The bivariate model could be expressed in matrix notation as:

$$\begin{bmatrix} y_f \\ y_g \end{bmatrix} = \begin{bmatrix} X_f & 0 \\ 0 & X_g \end{bmatrix} \begin{bmatrix} b_f \\ b_g \end{bmatrix} + \begin{bmatrix} Z_f & 0 \\ 0 & Z_g \end{bmatrix} \begin{bmatrix} a_f \\ a_g \end{bmatrix} + \begin{bmatrix} W_f & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} pe_f \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & C_g \end{bmatrix} \begin{bmatrix} 0 \\ cl_g \end{bmatrix} + \begin{bmatrix} e_f \\ e_g \end{bmatrix}$$

where f and g stand for fitness (FT1, FT2, and FT3) and growth traits (GT1 and GT2), respectively; y is the vector of observations; b is a vector of systematic effects (as stated below); a is the vector of direct additive genetic effects; pe is the vector of permanent environment effects; cl is the vector for common litter environment effects; e is the vector for random residual effects; X, Z, W, and C are incidence matrices relating the elements of y to elements of b, a, pe, and cl, respectively. The covariance matrix was assumed to be:

$$Var\begin{bmatrix} a_f \\ a_g \\ pe_f \\ cl_g \\ e_f \\ e_g \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}\sigma^2_{af} & \boldsymbol{T}\sigma_{af,ag} & 0 & 0 & 0 & 0 \\ \boldsymbol{T}\sigma_{ag,af} & \boldsymbol{T}\sigma^2_{ag} & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{I}\sigma^2_{pef} & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{I}\sigma^2_{cl_g} & 0 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{I}\sigma^2_{ef} & 0 \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{I}\sigma^2_{eg} \end{bmatrix}$$

where σ^2_{af}, σ^2_{ag}, and $\sigma_{af,ag}$ are variances for direct additive genetic effects for fitness traits, direct additive genetic effects for growth traits, and their covariances, respectively; σ^2_{pef} is the variance for permanent environment effects for fitness traits; $\sigma^2_{cl_g}$ is the variance for common litter environment effects for growth traits; and σ^2_{ef} and σ^2_{eg} are variances of residual effects for fitness and growth traits, respectively; **I** is the identity matrix; **T** is equal to **A** when only pedigree information is used as the covariance structure for the direct additive genetic effects, or **H** when genomic and pedigree information are jointly used to compute relationships. According to Aguilar et al. (2010), the inverse of **H** is:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where $\mathbf{A}^{-1}$ is the inverse of a pedigree-based relationship matrix for all animals included in the analysis; $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree-based relationship matrix for genotyped animals only, and $\mathbf{G}^{-1}$ is the inverse of a genomic relationship matrix. The systematic effects included in *b* were farm, year of farrowing, month of farrowing, age at measurement, and sex (included only for growth traits).

RESULTS AND DISCUSSION

Heritability

The changes in heritability estimated by GEN (PED) from the first to the last year interval, i.e., from 2009-11 to 2015-18 were 8.6 to 5.6 (7.9 to 8.8) for FT1, 7.8 to 7.2 (7.7 to 10.8) for FT2, 11.4 to 7.6 (10.1 to 7.5) for FT3, 35.1 to 16.5 (32.5 to 23.7) for GT1, and 35.9 to 16.5 (32.6 to 24.1) for GT2. The posterior means and standard deviations for heritabilities estimated by GEN and PED are shown in Fig. 3.1. The evolution of the heritabilities over time was similar with and without genomic information. Heritabilities for FT1 and FT2 were nearly stable, whereas heritabilities for FT3 and growth traits decreased over time. According to Walsh and Lynch (2018), the stronger the intensity of selection and the greater the heritability, the larger the LD and stronger the reduction in the genetic variance.

In this population the reduction in heritabilities was greater for growth traits, explained by their greater heritability and the stronger intensity of selection on these traits (evidenced by genetic trends; Fig. 3.7). The reduction in the heritability for FT3 was observed in the first three intervals, accompanied by an undesirable slight increase in the genetic trend (Fig. 3.7), suggesting that an unfavorable correlated response was the main cause.

Heritability for FT3 showed minor changes since 2012-14, possibly as a result of weak selection (Fig. 3.7). Another possible explanation is that this trait reached the equilibrium. Assuming an infinitesimal model and infinite population size with repeated cycles of selection, Bulmer (1971) showed that an equilibrium is eventually achieved in which the genetic variance lost by selection is regenerated by recombination. Villanueva and Kennedy (1990) studied the effect of selection in two traits. When selecting for one trait, variances of the directly and indirectly selected traits were reduced, equilibrium values were reached in approximately four rounds of

directional selection. Holm et al. (2004) reported constant heritability values for number born alive at first and second parity (0.10±0.01) in a pig population under selection.

In a research related to survival in dairy cattle, van Pelt et al. (2016) found that the mean survival increased over time, whereas genetic and residual variances and heritability decreased. The heritability changed from 0.06±0.07 (1989-93) to 0.01±0.06 (2009-13). In a related study, the heritabilities for productive life in dairy cattle (Tsuruta et al., 2004) were constant over time, most likely because of the lack of directional selection in this trait, which was evidenced by genetic trends. Haile-Mariam and Pryce (2015) reported a reduction in heritabilities for survival from 0.07 (1993-94) to 0.03 (2007-08) and for calving interval from 0.06 (1993-94) to 0.03 (2007-08). They stated that the observed decrease was related to a reduction in the genetic variance and an increase in the residual variance for calving interval.

In this research we found similar results, where the reduction in heritabilities was associated with a decrease in genetic variance and an increase in residual variance. Posterior means and standard deviations for additive genetic variance, environmental variance (permanent environment for fitness traits and common litter environment for growth traits) and residual variance are shown in Fig. 3.2, Fig. 3.3, and Fig. 3.4, respectively. Genetic variance was nearly flat for FT1 and FT2, whereas for FT3 and growth traits it decreased. Environmental variance was stable for all the traits, except for FT3 where a reduction was observed. The residual variance increased over time for FT1, FT2, and growth traits, but was stable for FT3. The increase in the residual variance could be due to scaling effect. As the means for FT1, FT2 and growth traits increased (improved), those residual variances increased, but mean for FT3 showed minor changes, tending to decrease (improve) as the residual variance.

In summary, the reduction in heritability for FT3 was associated with a reduction in the additive genetic variance, the reduction in heritabilities of growth traits was due to the combination of a decrease in additive genetic variance and an increase in residual variance. However, it is important to highlight that the reduction in additive genetic variance played a key role in the observed changes, explaining an important part of it. Thus, breeding programs should take this into consideration.

Genetic Correlations

The changes in genetic correlations estimated by GEN (PED) from the first to the last year interval, i.e., from 2009-11 to 2015-18 were -0.31 to -0.58 (-0.28 to -0.73) for FT1-GT1, -0.32 to -0.50 (-0.29 to -0.74) for FT1-GT2, -0.27 to -0.45 (-0.30 to -0.65) for FT2-GT1, -0.28 to -0.45 (-0.32 to -0.66) for FT2-GT2, 0.14 to 0.17 (0.11 to 0.04) for FT3-GT1, and 0.14 to 0.18 (0.11 to 0.05) for FT3-GT2. Posterior means and standard deviations for genetic correlations computed by GEN and PED are shown in Fig. 3.5. The genetic correlations between FT3 and growth traits were roughly stable over time, whereas the genetic correlations of FT1 and FT2 with growth traits decreased. A possible explanation for the constant correlation between FT3 and growth traits is that the LD between these two traits has stabilized.

McMillan et al. (1995) studied the effects of simultaneous selection on the genetic correlation. They used selection index and concluded that positive and negative genetic correlations tended to decline. However, unequal heritabilities and unequal relative economic weights reduced the rate of change with the greatest imbalance tending to hold the genetic correlation constant or move it toward zero. In a study about selection in a single trait, Villanueva and Kennedy (1990) found that the genetic correlation between the trait under direct selection and the trait indirectly selected always decrease in absolute value, whereas genetic correlations

between two traits indirectly selected can either decrease or increase in absolute value, depending not only on the signs but also on the magnitudes of the parameters involved.

Strandén et al. (1993) studied how the genetic correlation changes under selection either on single or both traits in a dairy breeding program using simulations. They simulated directional selection by truncation to increase two traits and reported that the absolute value of the genetic correlation usually decreased with single trait selection. However, when the initial genetic correlation was low and the residual correlation had the same sign and was high, the genetic correlation increased. With selection on both traits, the change in genetic correlation was always negative, i.e., the traits became less positively correlated or more negatively correlated after selection, in agreement with the results found by McMillan et al. (1995).

Itoh (1991) theoretically demonstrated that multi-trait selection changes the genetic correlation always in an undesirable direction. These results agree with the ones found in the present study involving multi-trait selection. In fact, the genetic correlations of FT1 and FT2 with growth traits became more antagonistic and the remaining genetic variation upon which future selection must act for response will face a more undesirable genetic covariance. It is important to consider these changes in the breeding program, and genetic parameters need to be updated regularly. It is important to investigate the trends for the genetic variances and covariances and also for the mean breeding values to detect directional selection.

Holm et al. (2004) studied the effect of selection for production traits in pigs. They estimated genetic correlations for number born alive at first and second parity with adjusted age at 100 kg live weight (0.60 and 0.42), individual feed consumption from 25 to 100 kg (0.23 and 0.20), and percentage of lean meat content (-0.12 and -0.24). The authors concluded that directional

selection for production traits resulted in more unfavorable genetic correlations in the second parity compared to the first parity.

In a study of dairy cattle, Haile-Mariam and Pryce (2015) reported that genetic correlations of survival with milk yield declined from 0.45 at the beginning of the study (1993-94) to -0.15 at the end (2009-10), whereas the genetic correlation between calving interval and milk yield became more unfavorable and increased from 0.31 to 0.50 over the same period. The genetic correlation between survival and calving interval also became more antagonistic, declining from -0.67 to -0.87. Similar results were found by Lawlor et al. (2002), the genetic correlation between milk yield and productive life decreased from 0.26 (1981) to -0.08 (1996). Another study reported changes in genetic correlations between somatic cell scores and milk yield from positive (0.25) to negative (-0.15) from first to later lactations (Banos and Shook, 1990). It is important to state that in dairy cattle, changes in genetic parameters could be affected by management practices.

According to McMillan et al. (1995), changes in the genetic parameters over time can affect selection decisions. The genetic covariance changes more rapidly than genetic variances. The use of initial genetic parameter estimates without considering the changes and unchanging selection weights for the traits in the selection index could have potentially negative effects in the overall genetic gain. In an example provided by Cheverud (1984), selection to increase two traits, with a genetic correlation of -0.80, was five times slower than if the genetic correlation was 0.0.

The posterior means and standard deviations estimated for additive genetic covariances are shown in Fig. 3.6. The additive genetic covariances among FT3 and growth traits were stable. The additive genetic covariances of FT1 and FT2 with growth traits were stable with genomic information but decreased if only pedigree and phenotypes were used. The genetic correlations of FT1 and FT2 with growth traits decreased both with and without genomic information. Therefore,

the changes observed in genetic correlations in our study are mainly due to the fluctuations in the additive genetic variances as they were stronger than in the genetic covariances.

Genomic versus pedigree-based analyses

Heritability estimates were similar for fitness traits with GEN and PED, whereas for both growth traits, the heritabilities calculated with GEN were initially larger but declined faster (Fig. 3.1). This greater initial heritabilities could be because genomic relationships in the limited interval were possibly more informative than the pedigree ones, and because the genomic selection effectively started around 2014, which was evidenced by genetic trends (Fig. 3.7). The steeper decline past 2014 is likely because the analyses with the genomic information account for genomic pre-selection, which avoids estimation bias (Patry and Ducrocq, 2011b).

Using genomic data before the implementation of genomic selection, Forni et al. (2011) studying a pig population and Veerkamp et al. (2011) studying a dairy cattle population found similar genetic parameter estimates between genomic and pedigree-based analyses. These results are in agreement with our estimates of heritabilities for fitness traits.

Raidan et al. (2018) reported greater heritabilities in genomic analyses for adaptative and growth traits in beef cattle. Momen et al. (2017), in broiler chickens, found that heritabilities using genomic information were greater for body weight at 35 days of age and ultrasound area of breast meat. However, their results were mixed for hen-house egg production. In our study, the heritabilities for growth traits were larger at the beginning but lower at the end, suggesting that the use of time intervals is appropriate to study changes in genetic parameters over time.

Genetic correlations with GEN, in general were greater than the estimates with only pedigree and phenotypes (Fig. 3.5). Momen et al. (2017) found similar results for genetic correlations in broiler chickens; genetic correlations between body weight at 35 days of age and

hen-house egg production were -0.192 and -0.020 with pedigree-based and genomic analyses, respectively, and genetic correlations for ultrasound area of breast meat and hen-house egg production were -0.206 and -0.154 with pedigree-based and genomic analyses, respectively. The latter authors also reported that genetic correlations between body weight at 35 days of age and ultrasound area of breast meat were similar with pedigree-based (0.484) and with genomic analyses (0.497).

For populations undergone genomic selection, variance components or breeding values estimated using only pedigree and phenotypes are biased because an important piece upon selection is not used in the model. According to Patry and Ducrocq (2011), this bias is due to pre-selection or because of the assumption that the mean of the mendelian sampling is 0. The pre-selection is clear when genetic trends under genomic and non-genomic (i.e., BLUP) analyses are compared. Masuda et al. (2018) showed that the trend for protein yield in US Holsteins under BLUP leveled off, whereas the trend for single-step GBLUP showed greater genetic gain, which agreed with phenotypic trends. In our study the bias was clear for the heritabilities estimated for growth traits.

Genetic trends estimated by GEN and PED are showed in Fig. 3.7. According to Masuda et al. (2018), if there is a real downward bias using PED because an underestimation of breeding values, the trend by GEN should be greater than by PED for recent animals (for traits in which an increase is the goal). Our results agree with this statement as genetic trends by GEN or genomic estimated breeding values were greater than those by PED or estimated breeding values in the last three years (2016-2018) for FT1, FT2, and growth traits. Over time, the genetic trends were similar for FT3, mainly due to the weak selection on this trait. However, for FT1, FT2, and growth traits,

the difference between genetic trends by PED and GEN was nearly constant until 2015, but the two trends started to diverge in 2016.

In populations under selection with selective genotyping or when the genotyped population does not represent well the pedigree population (i.e., genotyped animals from only a few recent generations), a method that accounts for all available pedigree, genotypes, and phenotypes−single-step genomic restricted maximum likelihood (ssGREML)−produces the most accurate variance components compared with REML or genomic REML if only phenotypes of genotyped animals are considered (Cesarani et al., 2019).

The selection response is dictated by the breeder's equation. With reduction of the genetic variance and assuming other parts of the breeder's equation are constant (generation interval, intensity and accuracy of selection), the selection response should be reduced. With increased accuracy due to increasing number of genotyped animals (Table 3.1), again assuming other parts of the breeder's equation constant, the selection response should accelerate. In our study, the selection response was not reduced (Fig. 3.7), possibly because the reduction of the genetic variance was compensated by increased accuracy of selection.

CONCLUSIONS

Under genomic selection, the heritabilities of FT3 and growth traits decrease, and the negative genetic correlations become more negative. Subsequently, genetic gains may not be as high as computed using initial genetic parameters although they could keep constant if the accuracy of selection is increasing due to more genotyped animals. In populations undergone genomic selection, variance components estimated without genomic information are possibly

biased. When the number of genotyped animals is large, variance components over time can be estimated using a model with time intervals.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 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.

Banos, G. and Shook, G. E. 1990. Genotype by Environment Interaction and Genetic Correlations Among Parities for Somatic Cell Count and Milk Yield. J. Dairy Sci. 73:2563–2573. doi:10.3168/jds.S0022-0302(90)78942-4.

Bijma, P. 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. J. Anim. Breed. Genet. 129:345–358. doi:10.1111/j.1439-0388.2012.00991.x.

Bulmer, M. G. 1971. The Effect of Selection on Genetic Variability. Am. Nat. 105:201–211. doi:10.1086/282718.

Cesarani, A., I. Pocrnic, N. P.P. Macciota, B. O, Fragomeni, I. Misztal, and D. A. L. Lourenco. 2019. Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. J. Anim. Breed. Genet. 136:40–50. doi.org/10.1111/jbg.12367.

Cheverud, J. M. 1984. Quantitative genetics and developmental constraints. J. Theor. Biol. 110:155–171. doi.org/10.1016/S0022-5193(84)80050-8.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics. United Kingdom: Pearson Education Ltd.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet. Sel. Evol. 43:1. doi:10.1186/1297-9686-43-1.

Gorjanc, G., P. Bijma, and J. M. Hickey. 2015. Reliability of pedigree-based and genomic evaluations in selected populations. Genet. Sel. Evol. 47:65. doi:10.1186/s12711-015-0145-1.

Haile-Mariam, M., and J. E. Pryce. 2015. Variances and correlations of milk production, fertility, longevity, and type traits over time in Australian Holstein cattle. J. Dairy Sci. 98:7364–7379. doi:10.3168/jds.2015-9537.

Holm, B., M. Bakken, G. Klemetsdal, and O. Vangen. 2004. Genetic correlations between reproduction and production traits in swine. J. Anim. Sci. 82:3458–3464. doi.org/10.2527/2004.82123458x.

Itoh, Y. 1991. Changes in genetic correlations by index selection. Genet. Sel. Evol. 23:301. doi:10.1186/1297-9686-23-4-301.

Lawlor, T. J., S. Tsuruta, L. Klei, and I. Misztal. 2002. Use of a random regression model to investigate changes in genetic parameters over time. Proc. 7th WCGALP, Montpellier, France. 17:06.

Masuda, Y., P.M. VanRaden, I. Misztal, and T. J. Lawlor. Differing genetic trend estimates from traditional and genomic evaluations of genotyped animals as evidence of preselection bias in US Holsteins. 2018. J. Dairy Sci. 101:5194–5206. doi.org/10.3168/jds.2017-13310.

McMillan, I, R. W. Fairfull, G. W. Friars, and M. Quinton. 1995. The effect of simultaneous selection on the genetic correlation. Theor. Appl. Genet. 91:776-779. doi: 10.1007/BF00220958.

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf. (Accessed January 6, 2019).

Momen, M., A. A. Mehrgardi, A. Sheikhy, A. Esmailizadeh, M. A. Fozi, A. Kranis, B. D. Valente, G. J. M. Rosa, and D. Gianola. 2017. A predictive assessment of genetic correlations between traits in chickens using markers. Genet. Sel. Evol. 49:16. doi:10.1186/s12711-017-0290-9.

Patry, C., and V. Ducroq. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. 2011. J. Dairy Sci. 94:1011–1020. doi:10.3168/jds.2010-3804.

Patry, C., and V. Ducroq. Accounting for genomic pre-selection in national BLUP evaluations in dairy cattle. 2011b. Genet. Sel. Evol. 43:30. doi:10.1186/1297-9686-43-30.

Raidan, F. S. S., L. R. Porto-Neto, Y. Li, S. A. Lehnert, and A. Reverter. 2018. Weighting genomic and genealogical information for genetic parameter estimation and breeding value prediction in tropical beef cattle. J. Anim. Sci. 96:612–617. doi:10.1093/jas/skx027.

Strándem, I., E. A. Mäntysaari, and A. Mäki-Tanila. 1993. Change in genetic correlation due to selection using animal model evaluation. J. Anim. Breed. Genet. 110:412–422. doi.org/10.1111/j.1439-0388.1993.tb00754.x.

Tsuruta, S., I. Misztal, and T. J. Lawlor. 2004. Genetic Correlations Among Production, Body Size, Udder, and Productive Life Traits Over Time in Holsteins. J. Dairy Sci. 87:1457–1468. doi:10.3168/jds.s0022-0302(04)73297-x.

van Pelt, M. L., V. Ducrocq, G. de Jong, M. P. L. Calus, and R. F. Veerkamp. 2016. Genetic changes of survival traits over the past 25 yr in Dutch dairy cattle. J. Dairy Sci. 99:9810–9819. doi:10.3168/jds.2016-11249.

Veerkamp, R. F., H. A. Mulder, R. Thompson, and M. P. L. Calus. 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. J. Dairy Sci. 94:4189–4197. doi:10.3168/jds.2011-4223.

Villanueva, B. and B. W. Kennedy. 1990. Effect of selection on genetic parameters of correlated traits. Theor. Appl. Genet. 80:746-752. doi.org/10.1007/BF00224187.

Walsh, B. and M. Lynch. 2018. Evolution and selection of quantitative traits. United Kingdom: Oxford.

TABLES

Table 3.1. Number of animals with genotypes, phenotypes, and in the pedigree for fitness (FT1, FT2 and FT3) and growth traits (GT1 and GT2) per year

Year	Pedigree	Genotypes	Fitness traits			Growth traits	
			FT1	FT2	FT3	GT1	GT2
2009	12,154	65	1,183	2,016	2,016	5,775	5,775
2010	16,474	923	2,182	2,237	2,237	7,858	7,858
2011	17,669	1,654	3,301	3,377	3,377	10,417	10,417
2012	26,177	1,455	2,506	2,579	2,579	14,656	14,656
2013	29,917	1,492	2,692	2,856	2,856	19,331	19,331
2014	27,947	1,957	3,438	3,620	3,620	18,391	18,392
2015	34,313	8,496	3,339	3,463	3,463	22,393	22,393
2016	39,536	10,881	2,018	2,203	2,203	24,783	24,783
2017	32,722	9,338	610	895	895	23,382	23,382
2018	13,309	2,842	-	-	-	3,506	3,506
Total	250,218	39,103	21,269	23,246	23,246	150,492	150,493

Table 3.2. Number of animals with genotypes, phenotypes, and in the pedigree for fitness traits (FT1, FT2, and FT3) and growth traits (GT1 and GT2) per interval

Interval	Pedigree[1]	Genotypes	Fitness traits			Growth traits	
			FT1[2]	FT2[2]	FT3[2]	GT1[2]	GT2[2]
2009-11	46,297 (24,479)	2,642	6,666 (2,151)	7,630 (2,195)	7,630 (2,195)	24,050 (2,582)	24,050 (2,582)
2010-12	60,320 (35,131)	4,032	7,989 (3,334)	8,193 (3,400)	8,193 (3,400)	32,931 (3,971)	32,931 (3,971)
2011-13	73,763 (47,075)	4,601	8,499 (3,821)	8,812 (3,915)	8,812 (3,915)	44,404 (4,597)	44,404 (4,597)
2012-14	80,041 (56,168)	4,904	8,636 (3,943)	9,055 (4,055)	9,055 (4,055)	52,378 (4,894)	52,379 (4,894)
2013-15	92,177 (65,098)	11,945	9,469 (4,238)	9,939 (4,351)	9,939 (4,351)	60,115 (10,851)	60,116 (10,851)
2014-16	101,796 (72,249)	21,334	8,795 (4,087)	9,286 (4,177)	9,286 (4,177)	65,567 (4,177)	65,568 (4,177)
2015-18	119,880 (84,171)	28,715	5,967 (3,112)	6,561 (3,358)	6,561 (3,358)	70,558 (25,892)	70,558 (25,892)

[1]Total number of animals in the interval (number of animals used in the analyses after tracing back all animals with phenotypes or genotypes up to 3 generations of their ancestors).
[2]Number of animals with phenotypes (number of genotyped animals with phenotypes).

FIGURES

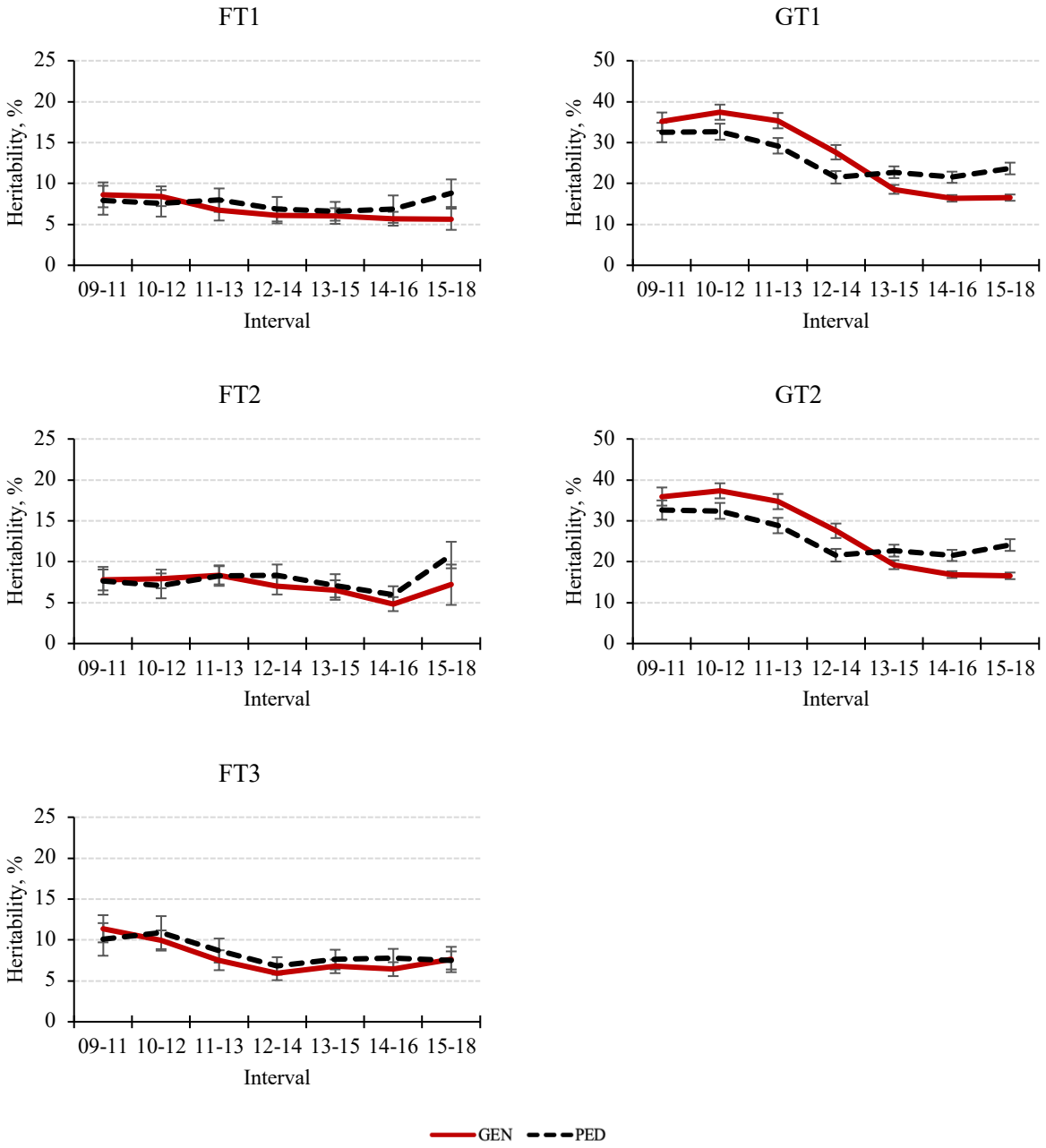

Figure 3.1. Posterior means and standard deviations for heritabilities of fitness (FT1, FT2 and FT3) and growth traits (GT1 and GT2) estimated with (GEN) or without (PED) genotypes. Heritabilities for FT1 and FT2 were nearly stable, while heritabilities for FT3 and growth traits decreased over time, estimates for FT2 increased in the last interval (2015-18), and estimates for FT3 showed minor changes since 2012-14.

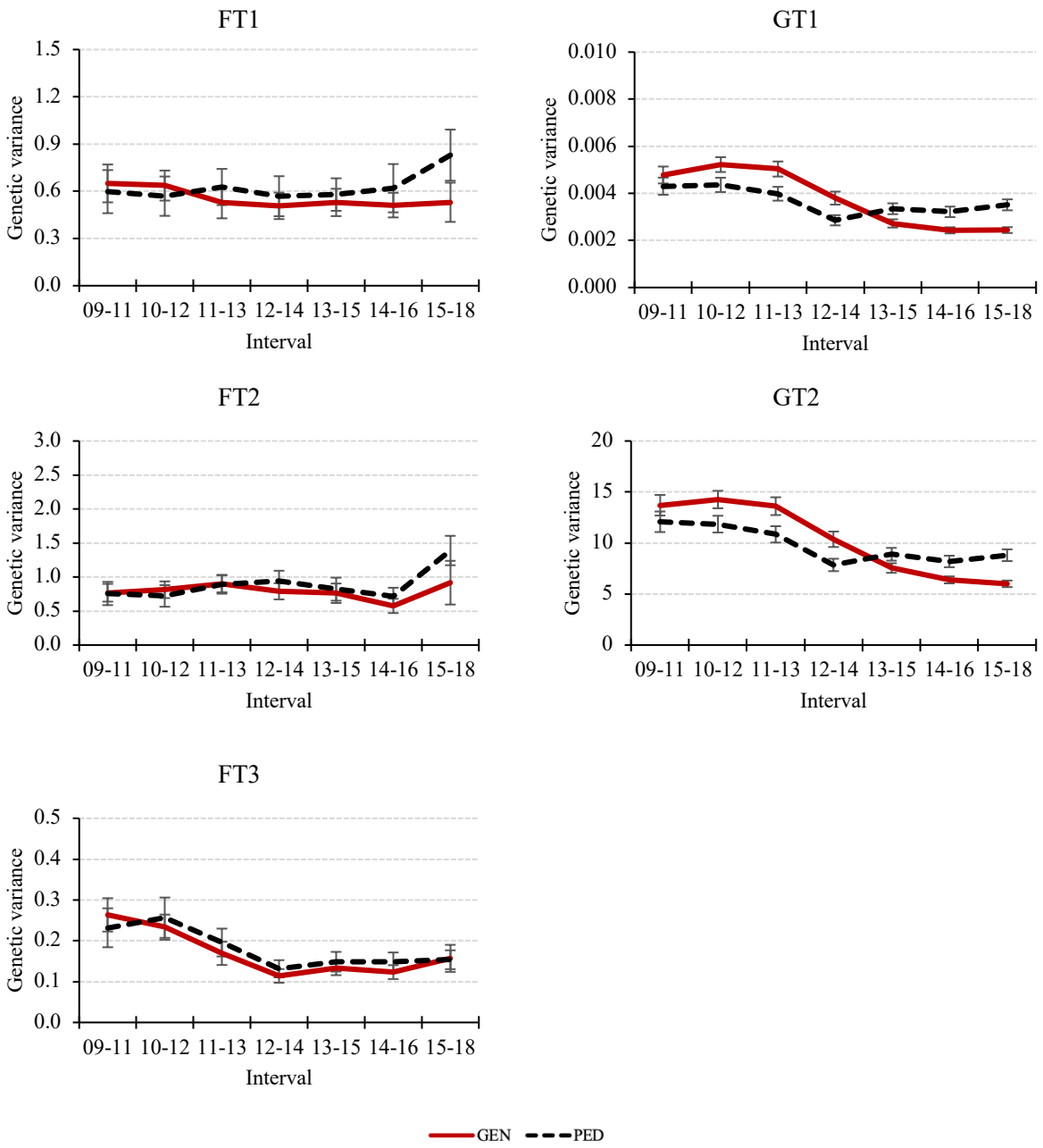

Figure 3.2. Posterior means and standard deviations for additive genetic variances of fitness (FT1, FT2 and FT3) and growth traits (GT1 and GT2) estimated with (GEN) or without (PED) genotypes. Genetic variance was nearly flat for FT1 and FT2, while for FT3, and growth traits it decreased.

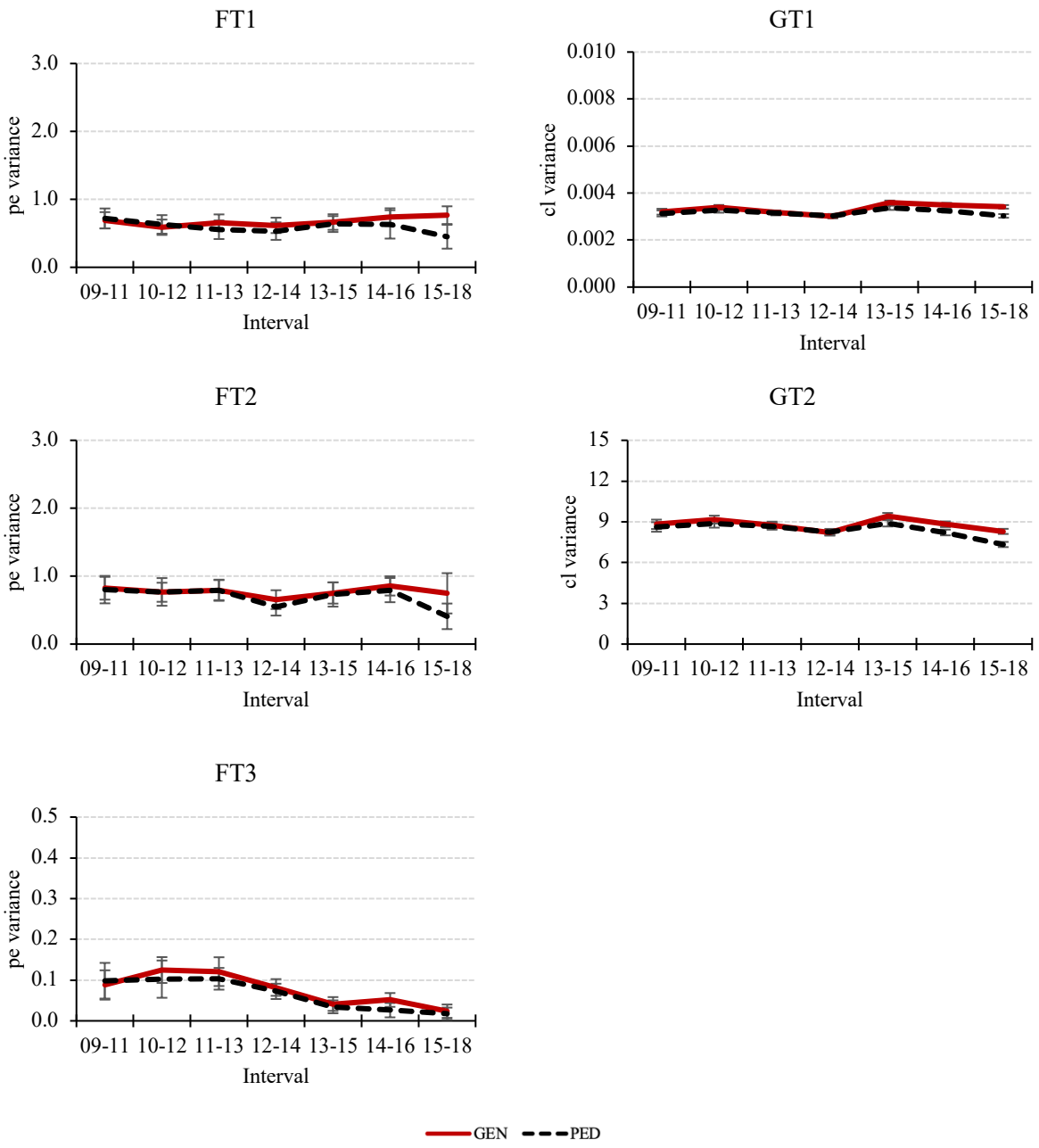

Figure 3.3. Posterior means and standard deviations for permanent environment (pe) variances of fitness traits (FT1, FT2 and FT3) and for common litter (cl) environment variances of growth traits (GT1 and GT2) estimated with (GEN) or without (PED) genotypes. Environmental variance was stable for all the traits, with exception for FT3 showing a reduction.

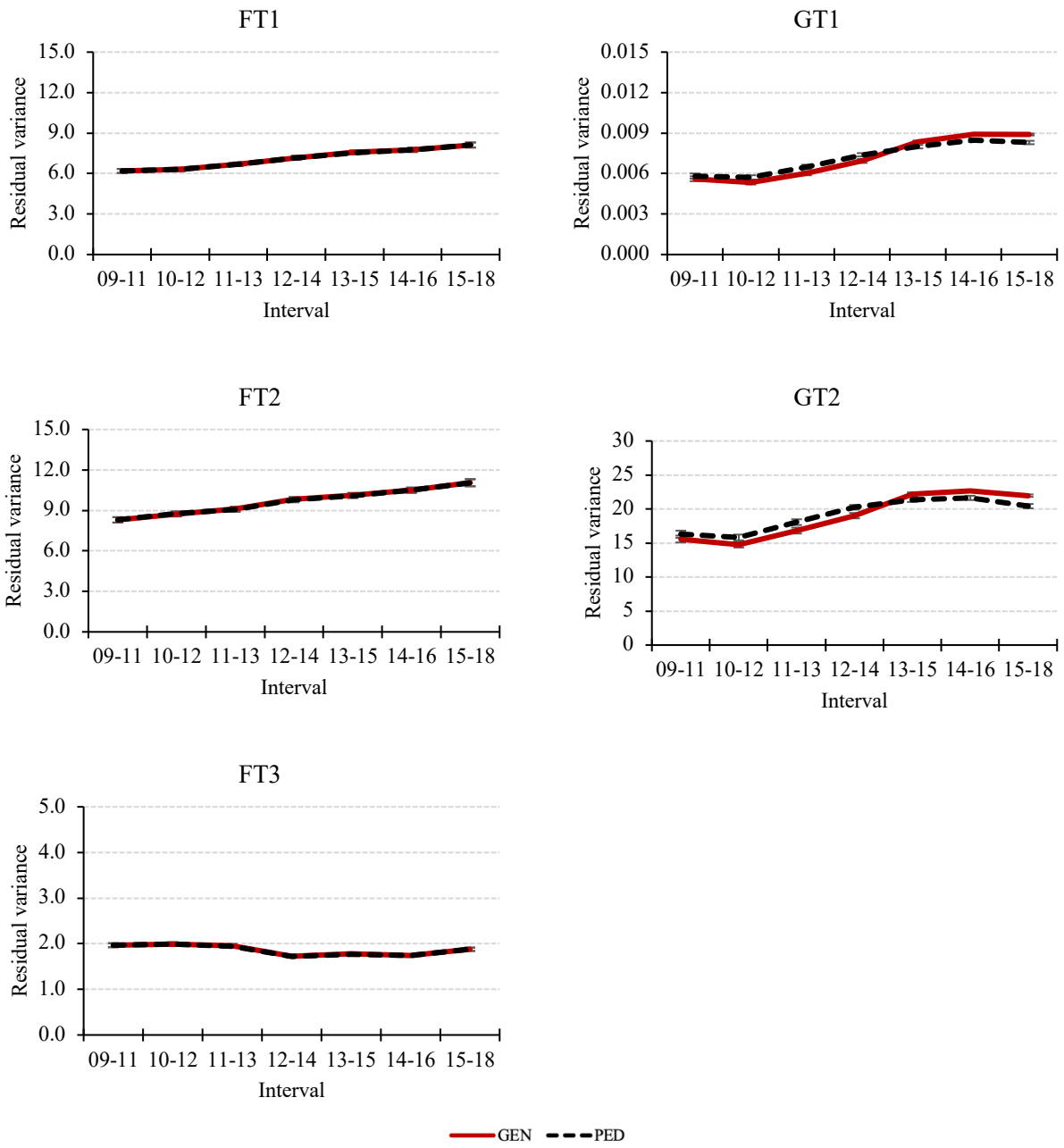

Figure 3.4. Posterior means and standard deviations for residual variances of fitness (FT1, FT2 and FT3) and growth traits (GT1 and GT2) estimated with (GEN) or without (PED) genotypes. Residual variance increased over time for all the traits, with exception for FT3 showing a stable value.

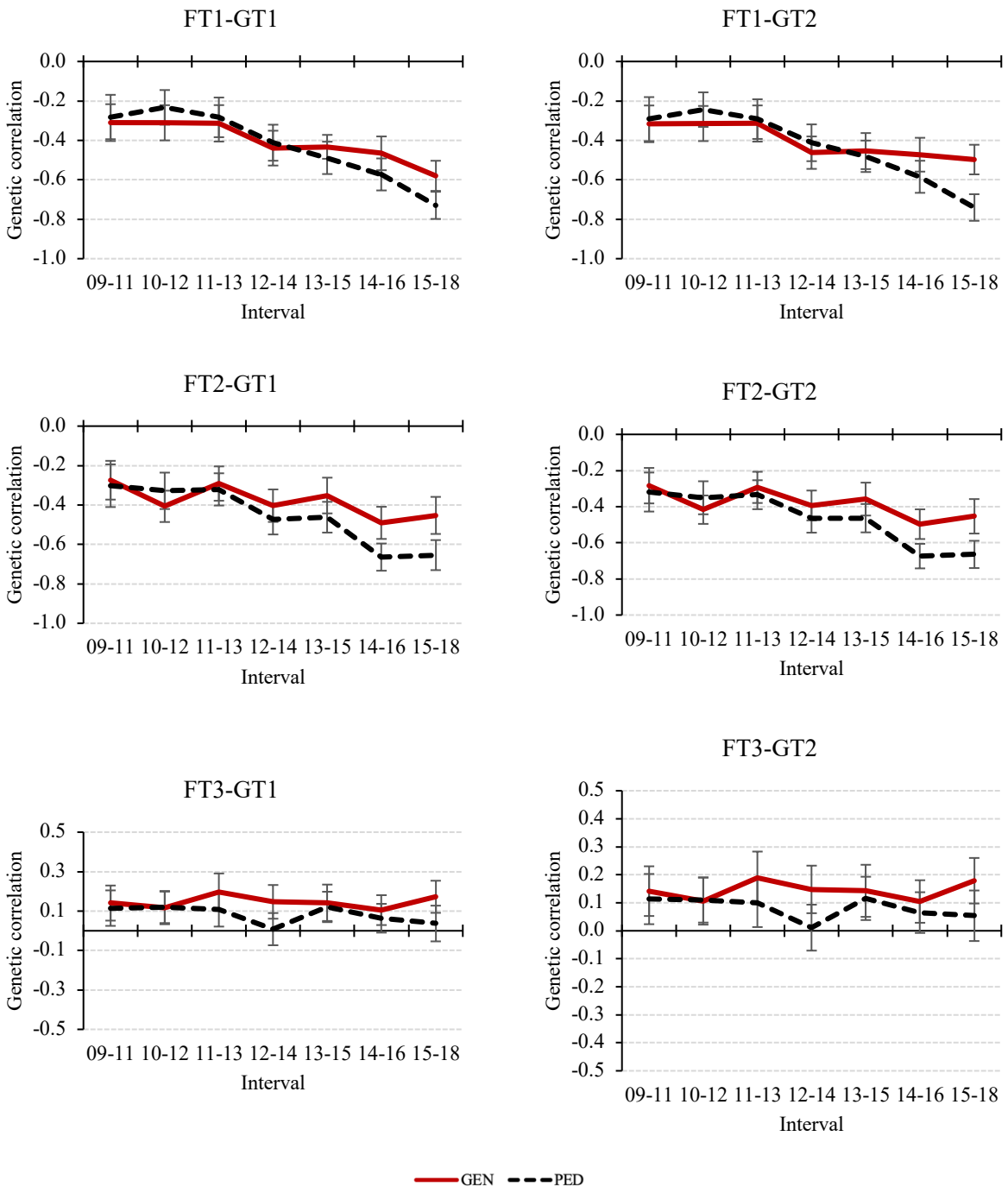

Figure 3.5. Posterior means and standard deviations for genetic correlations among fitness (FT1, FT2 and FT3) and growth traits (GT1 and GT2) estimated with (GEN) or without (PED) genotypes. The genetic correlations between FT3 and growth traits were roughly stable over time, whereas the genetic correlations of FT1 and FT2 with growth traits decreased.

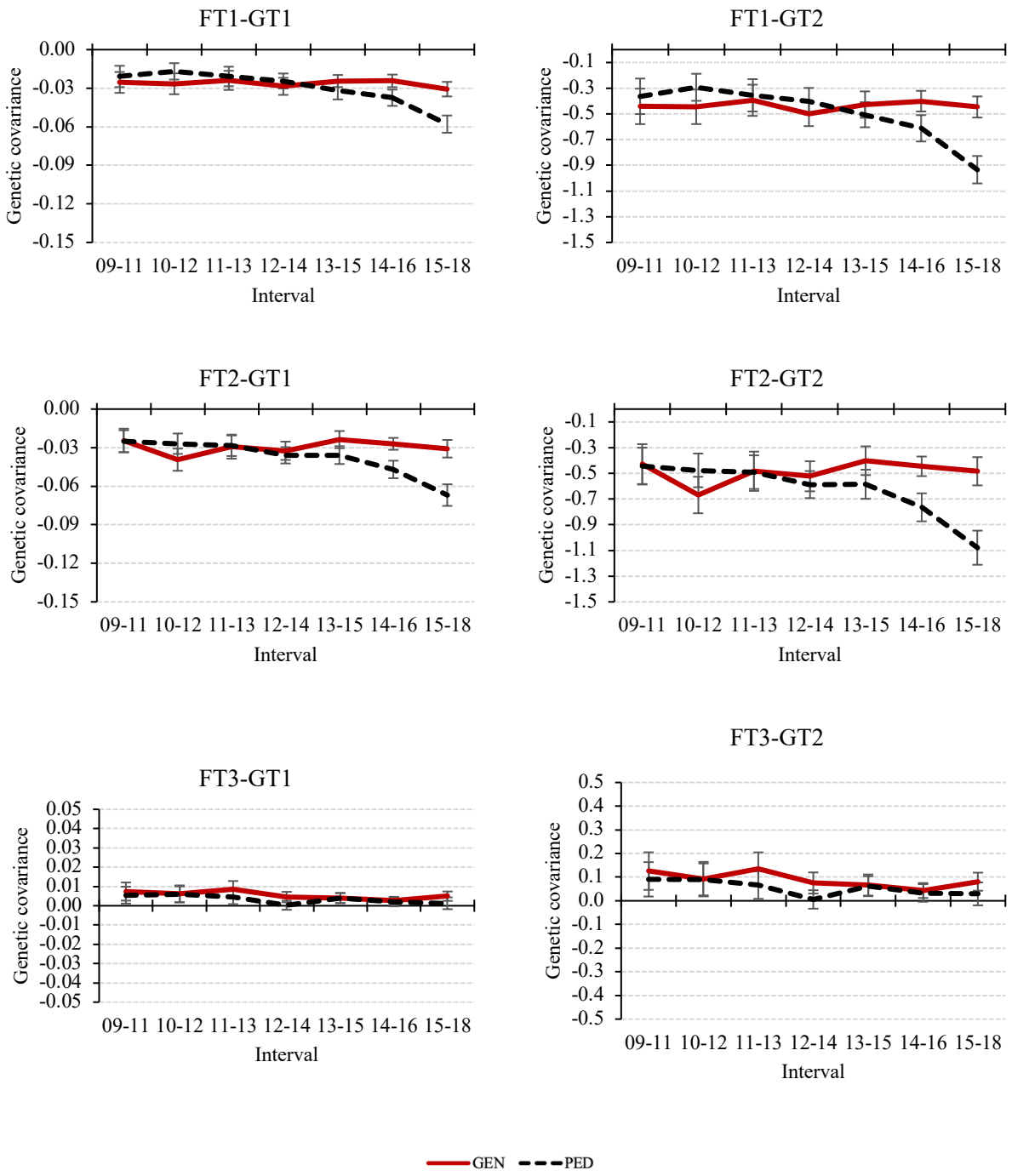

Figure 3.6. Posterior means and standard deviations for additive genetic covariances among fitness (FT1, FT2 and FT3) and growth traits (GT1 and GT2) estimated with (GEN) or without (PED) genotypes. The additive genetic covariances of FT1 and FT2 with growth traits were stable with GEN, whereas they decreased with PED.

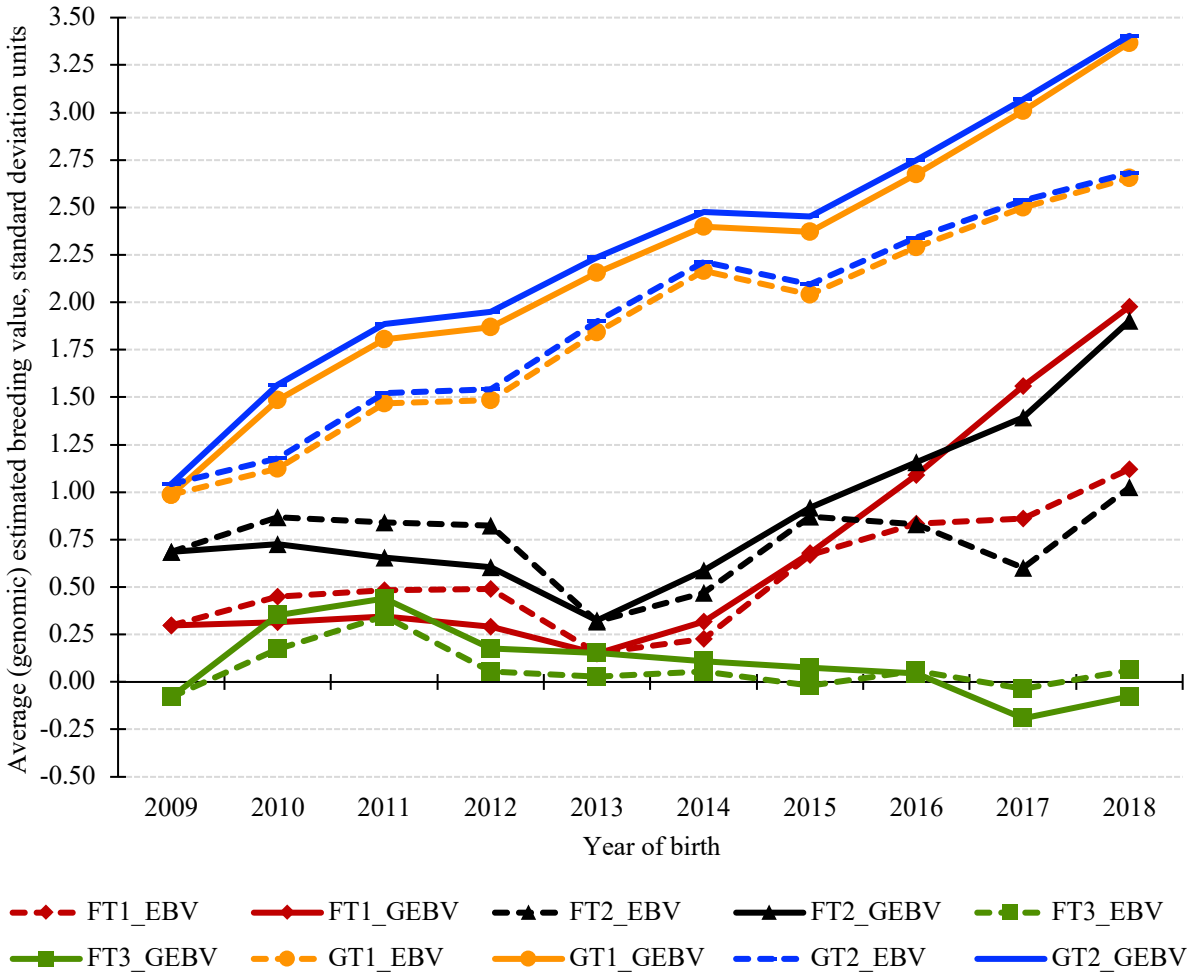

Figure 3.7. Genetic trends in standard deviation units of estimated breeding values (EBV) and genomic estimated breeding values (GEBV) for fitness (FT1, FT2 and FT3) and growth traits (GT1 and GT2). Genetic trends were similar for FT3. For FT1, FT2 and growth traits, the difference between genetic trends was nearly constant until 2015, but the two trends started to diverge in 2016.

CHAPTER 4

CHANGES IN GENOMIC PREDICTIONS WHEN NEW INFORMATION IS ADDED[2]

[2] Jorge Hidalgo, Daniela Lourenco, Shogo Tsuruta, Yutaka Masuda, Stephen Miller, Matias Bermann, Andre L. S. Garcia, and Ignacy Misztal. 2021. *Journal of Animal Science*. 99(2):1-10. Reprinted here with permission of the publisher.

ABSTRACT

Stability of genomic evaluations depends on amount of data and population parameters. When the dataset is large enough to estimate the value of nearly all independent chromosome segments (~10K in American Angus cattle), the accuracy and persistency of breeding values will be high. The objective of this study was to investigate changes in estimated breeding values (**EBV**) and genomic EBV (**GEBV**) across monthly evaluations for one year in a large genotyped population of beef cattle. The American Angus data used included 8.2 million records for birth weight, 8.9 for weaning weight, and 4.4 for post-weaning gain. A total of 10.1 million animals born until December 2017 had pedigree information, 484,074 were genotyped. A truncated dataset included animals born until December 2016. To mimic a scenario with monthly evaluations, 2017 data were added one month at a time to estimate EBV using BLUP, and GEBV using single-step genomic BLUP with the algorithm for proven and young (**APY**) with core group fixed for one year or updated monthly. Predictions from monthly evaluations in 2017 were contrasted with the predictions of the evaluation in December 2016 or the previous month for all genotyped animals born until December 2016 with or without own phenotypes or progeny phenotypes. Changes in EBV and GEBV were similar across traits, only results for weaning weight are presented. Correlations between evaluations from December 2016 and the 12 consecutive evaluations were ≥ 0.97 for EBV and ≥ 0.99 for GEBV. Average absolute changes for EBV were about two times smaller than for GEBV, except for animals with new progeny phenotypes (≤ 0.12 and ≤ 0.11 additive genetic SD (**SDa**) for EBV and GEBV). The maximum absolute changes for EBV (≤ 2.95 SDa) were greater than for GEBV (≤ 1.59 SDa). The average(maximum) absolute GEBV changes for young animals from December 2016 to January and December 2017 ranged from 0.05(0.25) to 0.10(0.53) SDa. Corresponding ranges for animals with new progeny phenotypes were from

0.05(0.88) to 0.11(1.59) SDa for GEBV changes. Average absolute change in EBV(GEBV) from December 2016 to December 2017 for sires with ≤ 50 progeny phenotypes was 0.26(0.14) and for sires with >50 progeny phenotypes was 0.25(0.16) SDa. Updating the core group in APY without adding data created an average absolute change of 0.07 SDa in GEBV. Genomic evaluations in large genotyped populations are stable and persistent as the traditional genetic evaluations, with less extreme changes.

INTRODUCTION

The stability of genomic evaluations depends on amount of data and population parameters, provided that no changes are made to the model. When the data are large enough to estimate the value of nearly all independent chromosome segments, the accuracy of genomic predictions will be high (Pocrnic et al., 2019), hence their persistency and stability will also be high. Independent chromosome segments are DNA regions that are inherited together in linked blocks (Stam, 1980) that are also called linkage disequilibrium blocks (Muir, 2007). The number of blocks that are inherited together depends on the past demographic history, the recombination landscape, the length of the genome ($\boldsymbol{L}$) and the mating structure of the population (Stam, 1980; Slatkin, 2008). The mating structure in finite populations under random mating is represented by the effective population size ($\boldsymbol{N_e}$), thus the independent chromosome segments can be quantified as $4N_eL$ (Stam, 1980). In Angus beef cattle populations, the number of independent chromosome segments ranges from 10k to 15k (Pocrnic et al., 2016a). Given two populations with the same number of genotyped animals and phenotypes, the one with smaller N_e would yield genomic predictions of greater accuracy because a lower number of independent chromosome segments would need to be estimated (Misztal et al., 2020a).

Pocrnic et al. (2016b) showed in a simulated dataset that the accuracy of genomic selection was maximized when the genomic estimated breeding values (**GEBV**) of genotyped animals were conditioned on GEBV from a number of genotyped animals equal to the number of independent chromosome segments ($4\ N_e L$), or equivalently, the number of the largest eigenvalues explaining 98% of the variation in the genomic relationship matrix (**G**). Using data from the American Angus beef cattle population, Pocrnic et al. (2016a) found that the accuracy of genomic predictions was marginally smaller using the number of the largest eigenvalues explaining 98% instead of 99 or 100% of the variation in **G**. The number of independent chromosome segments corresponds to the number of core animals in the algorithm for proven and young (**APY**; Misztal et al., 2014a, Pocrnic et al., 2016b). Pocrnic et al. (2019) reported that accuracies were marginally smaller when using 25% instead of 100% of the optimal number of core animals in APY, suggesting that genomic selection acts on clusters of independent chromosome segments rather than on individual independent chromosome segments. In the same study, the authors also showed that a small amount of phenotypic data allowed only the estimation of the largest clusters (i.e., eigenvalues), and that the four largest clusters explained 10% of the variation in **G**. This would yield only moderate genomic prediction accuracies; thus, many more phenotypes would be required for additional improvements in accuracy.

Accuracy and possible changes in predictions when more data are added to the evaluation system are both based on the standard error of prediction (Van Vleck, 2016). The larger the amount of data available for an animal, the more stable its prediction is, and the lower the size of changes that may occur. Estimated breeding values (**EBV**) from traditional best linear unbiased prediction (**BLUP**) are very stable even for animals with moderate accuracies. This outcome generated a high level of confidence in this method. Stability of predictions when new data are included is a

desirable feature of genomic evaluations for proven animals or animals without new data. However, changes may occur due to limited accuracies, higher number of links between animals through genomic than pedigree relationships, and decay of genomic information across generations. Considering the large amount of phenotypic and genomic data available and the limited dimensionality of the genomic information in the American Angus population (independent chromosome segments = 10,605; N_e = 113), we hypothesized that all independent chromosome segments could be estimated accurately, and the stability of genomic predictions would be high for proven animals or animals without new data.

Genomic evaluations for large genotyped populations based on the single-step genomic BLUP (**ssGBLUP**) require special algorithms like APY for computational feasibility and efficiency. Recently, Misztal et al. (2019) showed that GEBV can change when the core group is updated, even if no extra data is added, and those changes can be as high as one additive genetic SD (**SDa)**. Such changes happen because the APY relies on recursions of breeding values of noncore on core animals (Misztal et al., 2014). The recursion formula for noncore animals in APY assumes that a fraction of variation in **G** (~98%) is due to information and the rest is noise (~2%). The noise is modeled in the formula by an error term, which varies with different random samples of core animals, leading to small changes in GEBV (Misztal et al., 2020b). Keeping the same core animals for a given time is a good strategy to minimize changes in GEBV.

The main objective of this study was to investigate changes in genomic predictions across monthly evaluations for one year to assess the stability of predictions when new animals with phenotypic and genomic information are included in the evaluations. A second objective was to compare the changes in GEBV from adding new data and updating the core group in APY.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not needed because information was obtained from pre-existing databases.

Data

The dataset was provided by the American Angus Association and contained phenotypes for birth weight, weaning weight (N = 8,881,124), and post-weaning gain. The pedigree file consisted of 10,129,980 animals born between 1955 and 2017. A total of 484,074 animals had genotypes for 39,774 SNPs after quality control. Numbers of animals with genotypes, phenotypes for weaning weight, and pedigree records from December of 2016 to December of 2017 are shown in Table 4.1. Most of the data were added between January and April of 2017 because the vast majority of calves are born in the spring. Therefore, the 2017 data were added based on birth date of the animals because this information was more abundant than processing dates. It is important to highlight that the dynamics of the data inclusion in the official database is based on processing dates, and the amount of data added to the database could have a different pattern. The number of animals added to the database from December of 2016 to December of 2017 was 329,858, of which 91,075 had genomic information and 250,897 had phenotypes for weaning weight.

Analyses and Computations

The analyses were carried out with the BLUP90IOD2OMP1 program (Misztal et al., 2014b) utilizing a three-trait model used by Angus Genetics Inc. (St. Joseph, MO) for routine genetic evaluations (Lourenco et al., 2015). To compute traditional pedigree-based predictions, the analyses included pedigree and phenotypes, and to compute genomic predictions the analyses included pedigrees, phenotypes and genotypes. When genotypes were included, GEBV were computed by ssGBLUP using APY. The **G** matrix was constructed based on VanRaden (2008),

blended with 10% of the pedigree relationship matrix for genotyped animals ($\mathbf{A}_{22}$) to avoid singularity problems, and then rescaled to have the same means of diagonals and off-diagonals as $\mathbf{A}_{22}$ (Vitezica et al., 2011).

To mimic a system with monthly evaluations, phenotypes, genotypes, and pedigree information were added every month, from January to December of 2017. We computed traditional BLUP and ssGBLUP evaluations monthly and compared the resulting EBV and GEBV for genotyped animals born until December of 2016 with the values obtained in the evaluation from December of 2016 or with the values obtained in the evaluation of the previous month. For example, we compared the evaluations obtained with data until May of 2017 with the evaluations obtained with data until December of 2016 and also with the evaluations obtained with data until April of 2017. We computed correlations between EBV in December of 2016 and EBV in the 12 monthly evaluations in 2017 as well as correlations between GEBV in December of 2016 and GEBV from the 12 monthly evaluations in 2017. In addition, we obtained the distribution of changes for EBV and GEBV when comparing monthly evaluations in 2017 with evaluations in December of 2016, and computed average, top 1%, and maximum absolute changes in EBV and GEBV expressed as SDa and kg. Correlations and changes were evaluated separately for six groups of animals: all genotyped animals born until December of 2016 (n = 392,999), genotyped animals born until December of 2016 with (n = 387,743) and without (n = 5,256) phenotypes, genotyped animals born until December of 2016 with (n = 45,848) and without (n = 347,151) new progeny phenotypes after December of 2016, and for young genotyped animals born in 2016 with neither phenotypes nor progeny (n = 1,444).

To implement ssGBLUP with APY, two scenarios were used. In the first scenario, 20k genotyped animals from the December of 2016 dataset were randomly selected to be the core

group, and this group remained unchanged in all the genomic evaluations of 2017. In the second scenario, updated core groups were utilized from December of 2016 to December of 2017, meaning that every month the core was updated considering also the genotyped animals newly added to the database, which means 20k animals were again randomly selected from all the genotyped animals available at that point in time. This second scenario was useful to investigate the impact of adding new data versus the impact of changing the core group. The comparison between the two evaluations in December of 2016 using different core groups was termed contrast zero. The evaluations obtained with fixed and updated core are hereafter referred as GEBV and GEBV_UC, respectively.

RESULTS AND DISCUSSION

Only results for weaning weight are presented here because monthly changes in EBV and GEBV were similar among traits. Correlations between evaluations in December of 2016 and the 12 subsequent monthly evaluations in 2017 were greater than or equal to 0.97 for EBV and 0.99 for GEBV. Figure 1a shows the distribution of cumulative changes in EBV for weaning weight between the evaluations in December of 2016 and the 12 monthly evaluations in 2017 for the set of all genotyped animals born until December of 2016. Similarly, Fig. 4.1b presents the corresponding distribution of cumulative changes in GEBV between these 12 contrasts. The main cumulative changes occurred in the first 4 months of 2017 because approximately 90% of the phenotypes and genotypes were added during this period (spring calving season; Table 4.1). An increasing dispersion was observed from January to April (Fig. 4.1 and Table 4.3). Changes from May to December of 2017 were negligible in comparison with the previous months (Fig. 4.1 and

Table 4.4) because substantially less pedigree, phenotypic, and genotypic information was added to the dataset (Table 4.1).

Figure 2 presents average, top 1%, and maximum absolute changes in EBV and GEBV for weaning weight across evaluations for all genotyped animals born until December of 2016 in the American Angus population. Table 4.2 shows average and maximum (in parenthesis) absolute changes between evaluations in December of 2016 and evaluations in January of 2017 and between evaluations in December of 2016 and evaluations in December of 2017 for the six comparison groups of genotyped animals born until December of 2016. For all the genotyped animals, the average, top 1%, and maximum absolute changes between December of 2016 and January of 2017 were 0.02, 0.19, and 2.20 SDa for EBV and 0.05, 0.15, and 0.88 SDa for GEBV, respectively (Fig. 4.2). The corresponding average, top 1%, and maximum absolute changes between December of 2016 and December of 2017 were 0.06, 0.37, and 2.95 SDa for EBV and 0.10, 0.35, and 1.59 SDa for GEBV (Fig. 4.2). As expected, EBV and GEBV absolute changes were larger over a 12-month period than over a one-month period.

The average absolute changes for EBV were approximately two times smaller than the average absolute changes for GEBV for all genotyped animals, genotyped animals with and without own records, genotyped animals without new progeny records after December of 2016, and for genotyped young animals (Table 4.2). Genotyped animals with new progeny records after December of 2016 had similar average absolute changes for EBV and GEBV (Table 4.2).

The average absolute changes for EBV were greater for the group of animals with new progeny phenotypes added after December of 2016, about two times compared to the other groups. Therefore, in this case the changes were driven by the new phenotypes. The average absolute changes for GEBV were similar across the six comparison groups, thus in this case, not just the

new phenotypes, but also the new genotypes were the main forces driving the changes (Tables 4.2, 4.3, and 4.4). In general, the group of animals with new progeny phenotypes added after December of 2016 presented the greater average absolute changes, but even in this case, and after one year of adding new data, the average absolute changes were small (≤ 0.12 SDa for EBV and ≤ 0.11 for GEBV). The maximum absolute changes were always greater for EBV than for GEBV, and in line with the theory of outliers in the normal distribution (≤ 2.59 SDa for EBV and ≤ 1.95 SDa for GEBV). Lastly, the maximum absolute changes were smaller for young animals, a desirable feature because this group represents selection candidates. The changes in (G)EBV are bounded by the accuracy or precision of the estimates; therefore, smaller maximum changes for GEBV, when compared to EBV, are likely because the genomic information is expected to reduce the prediction error variance (Misztal et al. et al., 2020b). The reason for this is the more accurate estimation of mendelian sampling effects with genomics (Hayes et al., 2009; Cole and VanRaden, 2011).

Figure 4.3 presents the distribution of the changes in EBV(a), GEBV_UC(b), and GEBV (c) when contrasting the predictions from December of 2016 with those from December of 2017 for genotyped sires (n = 7,299) with own weaning weight phenotype and progeny phenotypes. Average absolute changes for sires with 50 or less progeny were 0.26, 0.15, and 0.14 SDa for EBV, GEBV_UC, and GEBV, respectively. Average absolute changes for sires with more than 50 progeny were 0.25, 0.17, and 0.16 SDa, in the same order. Maximum absolute changes in EBV, GEBV_UC, and GEBV were 2.95, 1.62, and 1.59 SDa for sires with 50 or less progeny, and 2.11, 1.19, and 1.17 SDa for sires with more than 50 progeny. In general, changes were greater for sires with no or few progeny before December of 2016. Proven sires with more than 50 progeny had similar average absolute changes, but smaller maximum absolute changes than sires with 50 or

less progeny. For this group of sires with own phenotype and progeny phenotypes, the average and maximum absolute changes in GEBV were smaller than in EBV. Additionally, updating the core group in APY generated marginal increase in the absolute average and maximum changes.

Looking at the decomposition of EBV and GEBV might help to understand the changes. The EBV can be decomposed into a parental average, a yield deviation, and a progeny contribution (VanRaden and Wiggans, 1991). Additionally, the decomposition of GEBV also includes a direct genomic value and a pedigree prediction, the last one is needed to avoid double counting of relationships (Aguilar et al., 2010; Lourenco et al., 2015). Without new information for an animal (i.e., new progeny phenotypes), the EBV based on parental average is expected to be stable (assuming that the parents had high accuracy EBV). However, the addition of new progeny phenotypes can result in large changes in EBV for an animal because, in the absence of large progeny groups (i.e., few or no progeny), the additional information is sizeable respective to the parental average. In this case, the magnitude of the change in EBV has a direct relationship with the phenotypic information that is added.

In the case of GEBV, every genotyped animal with phenotypic information influences the direct genomic value of all genotyped animals. Consequently, the GEBV of animals with no additional phenotypic information of their own or their relatives could change. However, if the reference population is large, the accuracy of the direct genomic value will be high (Lourenco et al., 2015), thus additional phenotypic records would have a lower impact on GEBV because they would contribute with less information than the direct genomic value. In our study, average absolute changes for GEBV in the group of genotyped animals with new progeny records after December of 2016 were similar compared with the remaining groups, indicating that the accuracy is high, which was expected because of the large reference population.

For milk yield in Holsteins, the information from the sire and the dam, both with 99% reliability, is equivalent to having 14 daughters with phenotypic records (VanRaden and Wiggans, 1991). With genomic information, the genotype of an animal provides information equivalent to 37.5 daughters for milk yield, 240.6 daughters for daughter pregnancy rate, and 780.2 daughters for heifer conception rate (https://queries.uscdcb.com/eval/summary/comparexml_menu.cfm?R_menu=v_2004.v_Young_Bulls.v_Holstein_wddx#StartBody). For weaning weight in Angus beef cattle, having the genotype of an unproven bull is equivalent to having 27 calves with weaning weight records (https://www.angus.org/AGI/GenomicEnhancedEPDs.pdf).

Genotypes create stronger relationship ties among animals, and at the same time reduce the prediction error variance of GEBV as demonstrated by Misztal et al. (2020b) in a research study carried out to investigate the magnitude of changes in relation to accuracy. In the preceding research study, the authors compared two genomic evaluations with the same amount of data but different core groups in APY and concluded that the largest differences were for animals with accuracy lower than 0.7; the animals with greater accuracy (smaller prediction error variance) had considerably smaller changes. In our study, the stronger relationship ties and the reduced prediction error variance resulted in greater average absolute changes in GEBV, but lower maximum absolute changes than those from EBV. In the group of animals with new progeny phenotypes after December of 2016 the average absolute changes in GEBV were similar to the average absolute changes in EBV, and the maximum absolute changes were smaller for GEBV than those for EBV. Additionally, genotyped sires of progeny with weaning weight phenotypes had smaller changes for GEBV than for EBV, indicating that genomic evaluations were as stable

as traditional evaluations and with less extreme changes because of their greater accuracy when animals had new phenotypic information.

An implicit assumption in this study was that the amount of information was sufficient to estimate the values of nearly all independent chromosome segments and their clusters with high accuracy. The predictivity of GEBV (Legarra et al., 2008), calculated as the correlation between phenotypes adjusted for fixed effects (using the whole dataset) and GEBV (estimated using a partial dataset where the phenotypes for validation animals were removed), for animals with weaning weight records and born in 2017 (n = 250,897) was 0.44. Assuming a heritability of weaning weight (0.20), the population accuracy calculated as predictivity/$\sqrt{\text{heritability}}$ would be very high (0.98). This formula depends on the heritability of the trait in the validation population and the adjustments to phenotypes. Although the computed accuracy seems overly high, the actual accuracy was likely high, thus explaining the high persistency of GEBV in this study. Hidalgo et al. (2020) found that the heritability for traits under strong genomic selection change over time, which could also lead to changes in GEBV accuracies. Thus, validation methods that do not explicitly depend on heritabilities (e.g., LR validation from Legarra and Reverter, 2018) may provide better estimates of GEBV accuracies. The LR method assumes that the additive genetic variance for the subset of validation animals is known. This additive genetic variance can be computed and is smaller than the population-based estimate in populations undergoing selection; however, its computation is intricate and requires a special method based on Gibbs Sampling (Sorensen et al., 2001). The implementation of this method is out of the scope of this research study and deserves future research.

Table 4.3 shows average and maximum absolute changes in EBV, GEBV, and GEBV_UC (i.e., updated core) for all genotyped animals born until December of 2016, those that had new

progeny records added after December of 2016, and the young ones. All changes are shown in SDa and in kg, but discussions are in SDa for simplicity. The contrast zero represents absolute changes between initial evaluations in December of 2016 and evaluations in December of 2016 with updated core. The contrasts one to twelve represent absolute changes among evaluations in December of 2016 and evaluations from January to December of 2017 adding data, and either with fixed or updated core.

The average absolute changes updating the core group for APY and adding new pedigree, genotypic, and phenotypic information (contrasts one to twelve) were about 0.02 SDa greater than with fixed core group. The maximum absolute changes were similar for the groups of all genotyped animals born until December of 2016 and those that had new progeny records added after December of 2016, but about 0.12 SDa greater for young animals when compared to using a fixed core group. The average absolute changes for contrast zero were 0.07 for all groups of animals, which was similar to those changes in contrast two with fixed core group. This means that the impact of changing the core animals in the estimation of GEBV is the same as the impact of adding two months of data but keeping the same core, it is important to highlight that we refer to those months with the majority of data. Therefore, the changes in GEBV because of core updating may have more impact when no data or a small amount of data is added from one evaluation to the next. To avoid larger changes, the core group can be fixed for a period of time, i.e., one year, and updates should be done when large amounts of data are added to the evaluation system.

Table 4.4 contains average and maximum absolute changes in EBV, GEBV, and GEBV_UC for all genotyped animals born until December of 2016, those that had new progeny records added after December of 2016, and the young ones. All changes are shown in SDa and in kg, but discussions are in SDa for simplicity. The contrasts one to twelve represent absolute

changes among evaluations of 2017 in the current and the precedent month. The largest average absolute changes in both EBV and GEBV were observed in the first four contrasts (from January to April of 2017) and were marginal afterward. The maximum absolute changes were larger in the first four contrasts and decreased afterward, as expected, because less information was added from May to December of 2017. Updating the core group in APY caused an average absolute change in GEBV of ~0.7 SDa. The maximum absolute changes for GEBV updating the core group in APY ranged from 0.61 to 1.17 SDa for all the genotyped animals and those with new progeny phenotypes. Maximum absolute changes were smaller for the group of young animals (~0.4 SDa).

CONCLUSIONS

When new pedigrees, genomic, and phenotypic data were included in the genomic evaluations of animals in the American Angus population, average absolute changes in GEBV were about twice as large as those in EBV for most genotyped animals, but for the ones with new progeny phenotypic records. The most extreme maximum absolute changes were observed for EBV. The deep pedigree and ample phenotypic and genomic data from the American Angus population yielded accurate estimates for all independent chromosome segments and their clusters, contributing to high persistency of genomic evaluations and leading to stable genomic evaluations during the year of the study. Genomic evaluations were as stable as traditional evaluations for animals with new progeny phenotypic records. If the algorithm for proven and young is used, changes in GEBV due to changes in the core group can be avoided by keeping the same core group for one year and updating it when a large amount of data is added.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 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(2):743-752. doi.org/10.3168/jds.2009-2730

Cole, J. B., and P. M. VanRaden. 2011. Use of haplotypes to estimate Mendelian sampling effects and selection limits. J. Anim. Breed. Genet. 128: 446-455. https://doi.org/10.1111/j.1439-0388.2011.00922.x

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. (Camb.). 91:47-60. doi: 10.1017/S0016672308009981.

Hidalgo, J., S. Tsuruta, D. Lourenco, Y. Masuda, Y. Huang, K. A. Gray, and I. Misztal. 2020. Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. J. Anim. Sci. 98: 1-12. doi:10.1093/jas/skaa032

Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet. Sel. Evol. 50:53. https://doi.org/10.1186/s12711-018-0426-6

Legarra, A., C. Robert-Granié, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611– 618. doi:10.1534/genetics.108.088575

Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, J. K. Bertrand, T. S. Amen, L. Wang, D. W. Moser, and I. Misztal. 2015. Genetic evaluation using single-step genomic BLUP in American Angus. J. Anim. Sci. 93:2653-2662. doi.org/10.2527/jas.2014-8836

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

Misztal, I., D. Lourenco, and A. Legarra, 2020a. Current status of genomic evaluation. J. Anim. Sci. 98:4648–4655. doi:10.1093/jas/skaa101

Misztal, I., S. Tsuruta, D.A.L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014b. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all1.pdf. (Accessed May 4th, 2020.)

Misztal, I., S. Tsuruta, I. Pocrnic, and D. Lourenco. 2019. *Changes in predictions when using different core animals in the APY algorithm*. Proceedings of the 70th Annual Meeting EAAP; August 26 to 30, 2019; Ghent, Belgium; p. 593.

Misztal, I., S. Tsuruta, I. Pocrnic, and D. Lourenco. 2020b. Core-dependent changes in genomic predictions using the algorithm for proven and young in single-step genomic best linear unbiased prediction. J. Anim. Sci. https://doi.org/10.1093/jas/skaa374

Muir, W. M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342–355. doi:10.1111/j.1439-0388.2007.00700.x

Pocrnic, I., D. A. L. Lourenco, K. Y. Masuda, and I. Misztal. 2019. Accuracy of genomic BLUP when considering a genomic relationship matrix based on the number of the largest eigenvalues: a simulation study. Genet. Sel. Evol. 51:75. doi: 10.1186/s12711-019-0516-0

Pocrnic, I., D. A. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016a. The dimensionality of genomic information and its effect on genomic prediction. Genetics 203:573–581. doi:10.1534/genetics.116.187013

Pocrnic, I., D. A. Lourenco, Y. Masuda, and I. Misztal. 2016b. Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species. Genet. Sel. Evol. 48:82. doi:10.1186/s12711-016-0261-6

Slatking, M. 2008. Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9: 477-485. https://doi.org/10.1038/nrg2361

Sorensen, D., R. Fernando, and D. Gianola. 2001. Inferring the trajectory of genetic variance in the course of artificial selection. Genet. Res. 77:83-94. https://doi.org/10.1017/S0016672300004845

Stam, P. 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. Genet. Res. 35:131–155.

Van Vleck, L. Dale. 2016. EPDs and risk. Faculty Papers and Publications in Animals Science. 898.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423. doi.org/10.3168/jds.2007-0980

VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737-2746. doi: 10.3168/jds.S0022-0302(91)78453-1.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (*Camb*). 93:357–366. doi:10.1017/S001667231100022X.

TABLES

Table 4.1. Number of animals in the pedigree, animals with genotypes, and animals with weaning weight in the American Angus population[1]

Period	Pedigree	Genotypes	Weaning weight
12/2016	9.800	0.393	8.630
01/2017	9.876	0.416	8.688
02/2017	9.965	0.443	8.763
03/2017	10.046	0.462	8.831
04/2017	10.081	0.467	8.860
05/2017	10.092	0.469	8.868
06/2017	10.095	0.469	8.869
07/2017	10.097	0.470	8.871
08/2017	10.104	0.474	8.874
09/2017	10.117	0.479	8.879
10/2017	10.125	0.482	8.881
11/2017	10.128	0.483	8.881
12/2017	10.130	0.484	8.881

[1]In millions.

Table 4.2. Average (maximum) absolute changes in EBV and genomic EBV for weaning weight between December of 2016 and January of 2017 and between December of 2016 and December of 2017 for genotyped animals born until 2016 in the American Angus population[1]

Contrast[2]	Genotyped animals					
	All	With records	Without records	With new records after 2016	Without new records after 2016	Young[3]
C1_EBV	0.02 (2.20)	0.02 (2.20)	0.01 (0.70)	0.04 (2.20)	0.02 (1.15)	0.02 (0.70)
C1_GEBV	0.05 (0.88)	0.05 (0.88)	0.04 (0.25)	0.05 (0.88)	0.05 (0.45)	0.05 (0.25)
C12_EBV	0.06 (2.95)	0.06 (2.95)	0.04 (1.44)	0.12 (2.95)	0.05 (1.39)	0.05 (0.69)
C12_GEBV	0.10 (1.59)	0.10 (1.59)	0.10 (0.53)	0.11 (1.59)	0.10 (0.75)	0.10 (0.53)

[1]Absolute changes in additive genetic standard deviation units.
[2]C1 = contrast between December of 2016 and January of 2017; C12 = contrast between December of 2016 and December of 2017.
[3]Born in 2016 with neither phenotypes nor progeny.

Table 4.3. Average and maximum absolute changes in EBV and genomic EBV for weaning weight among evaluations in December of 2016 and evaluations from January to December of 2017 for genotyped animals born until December of 2016 in the American Angus population[1]

Group	Contrast[2]	Average						Maximum					
		EBV		GEBV		GEBV_UC[3]		EBV		GEBV		GEBV_UC	
All	C0	-	-	-	-	0.07	(0.71)	-	-	-	-	0.76	(7.97)
	C1	0.02	(0.22)	0.05	(0.48)	0.08	(0.86)	2.20	(23.12)	0.88	(9.21)	1.08	(11.35)
	C2	0.04	(0.38)	0.07	(0.77)	0.10	(1.05)	2.81	(29.54)	1.53	(16.08)	1.48	(15.56)
	C3	0.05	(0.49)	0.09	(0.94)	0.11	(1.19)	2.71	(28.49)	1.54	(16.19)	1.51	(15.87)
	C4	0.05	(0.54)	0.10	(1.00)	0.12	(1.23)	2.84	(29.85)	1.60	(16.82)	1.67	(17.55)
	C5	0.05	(0.55)	0.10	(1.02)	0.12	(1.25)	2.84	(29.85)	1.60	(16.82)	1.51	(15.87)
	C6	0.05	(0.55)	0.10	(1.02)	0.12	(1.25)	2.95	(31.01)	1.60	(16.82)	1.55	(16.29)
	C7	0.05	(0.56)	0.10	(1.02)	0.12	(1.25)	2.95	(31.01)	1.60	(16.82)	1.54	(16.19)
	C8	0.05	(0.57)	0.10	(1.03)	0.12	(1.26)	2.96	(31.11)	1.59	(16.71)	1.62	(17.03)
	C9	0.05	(0.57)	0.10	(1.05)	0.12	(1.26)	2.95	(31.01)	1.59	(16.71)	1.58	(16.61)
	C10	0.06	(0.58)	0.10	(1.05)	0.12	(1.27)	2.95	(31.01)	1.59	(16.71)	1.62	(17.03)
	C11	0.06	(0.58)	0.10	(1.05)	0.12	(1.27)	2.95	(31.01)	1.59	(16.71)	1.61	(16.92)
	C12	0.06	(0.58)	0.10	(1.05)	0.12	(1.27)	2.95	(31.01)	1.59	(16.71)	1.62	(17.03)
New progeny records	C0	-	-	-	-	0.07	(0.70)	-	-	-	-	0.69	(7.23)
	C1	0.04	(0.41)	0.05	(0.48)	0.08	(0.87)	2.20	(23.12)	0.88	(9.21)	0.88	(9.22)
	C2	0.07	(0.78)	0.08	(0.80)	0.10	(1.07)	2.81	(29.54)	1.53	(16.08)	1.48	(15.56)
	C3	0.10	(1.05)	0.10	(1.01)	0.12	(1.24)	2.71	(28.49)	1.54	(16.19)	1.51	(15.87)
	C4	0.11	(1.16)	0.10	(1.08)	0.12	(1.29)	2.84	(29.85)	1.60	(16.82)	1.67	(17.55)
	C5	0.11	(1.19)	0.11	(1.10)	0.13	(1.31)	2.84	(29.85)	1.60	(16.82)	1.51	(15.87)
	C6	0.11	(1.19)	0.11	(1.10)	0.13	(1.31)	2.95	(31.01)	1.60	(16.82)	1.55	(16.29)
	C7	0.11	(1.20)	0.11	(1.10)	0.13	(1.32)	2.95	(31.01)	1.60	(16.82)	1.54	(16.19)
	C8	0.12	(1.23)	0.11	(1.11)	0.13	(1.33)	2.96	(31.11)	1.59	(16.71)	1.62	(17.03)
	C9	0.12	(1.26)	0.11	(1.14)	0.13	(1.33)	2.95	(31.01)	1.59	(16.71)	1.58	(16.61)
	C10	0.12	(1.28)	0.11	(1.15)	0.13	(1.35)	2.95	(31.01)	1.59	(16.71)	1.62	(17.03)
	C11	0.12	(1.29)	0.11	(1.15)	0.13	(1.36)	2.95	(31.01)	1.59	(16.71)	1.61	(16.92)
	C12	0.12	(1.29)	0.11	(1.15)	0.13	(1.36)	2.95	(31.01)	1.59	(16.71)	1.62	(17.03)
Young[4]	C0	-	-	-	-	0.07	(0.73)	-	-	-	-	0.62	(6.54)
	C1	0.02	(0.21)	0.05	(0.50)	0.08	(0.88)	0.70	(7.31)	0.25	(2.59)	0.37	(3.87)
	C2	0.04	(0.37)	0.08	(0.81)	0.10	(1.09)	0.71	(7.44)	0.36	(3.73)	0.47	(4.97)
	C3	0.05	(0.48)	0.09	(0.99)	0.12	(1.23)	0.69	(7.23)	0.52	(5.41)	0.69	(7.25)
	C4	0.05	(0.53)	0.10	(1.03)	0.12	(1.25)	0.69	(7.25)	0.55	(5.77)	0.75	(7.90)
	C5	0.05	(0.53)	0.10	(1.04)	0.12	(1.24)	0.69	(7.25)	0.54	(5.63)	0.73	(7.63)
	C6	0.05	(0.53)	0.10	(1.04)	0.12	(1.25)	0.69	(7.25)	0.53	(5.54)	0.57	(5.95)
	C7	0.05	(0.53)	0.10	(1.05)	0.12	(1.25)	0.69	(7.27)	0.52	(5.51)	0.60	(6.32)
	C8	0.05	(0.53)	0.10	(1.05)	0.12	(1.26)	0.69	(7.28)	0.52	(5.41)	0.59	(6.16)
	C9	0.05	(0.54)	0.10	(1.07)	0.12	(1.25)	0.69	(7.28)	0.53	(5.53)	0.64	(6.67)
	C10	0.05	(0.54)	0.10	(1.08)	0.12	(1.26)	0.69	(7.22)	0.53	(5.56)	0.55	(5.77)
	C11	0.05	(0.54)	0.10	(1.07)	0.12	(1.28)	0.69	(7.22)	0.53	(5.58)	0.62	(6.50)
	C12	0.05	(0.54)	0.10	(1.08)	0.12	(1.29)	0.69	(7.22)	0.53	(5.55)	0.77	(8.06)

[1]Absolute changes in additive genetic standard deviation units (in kg inside the parenthesis).
[2]C0 = contrast between two evaluations in December of 2016 with different core; C*x* = contrast between December of 2016 and month *x* of 2017; with *x* varying from 1 (January) to 12 (December).
[3]GEBV_UC = GEBV from a scenario where the core set is updated every month.
[4]Born in 2016 with neither phenotypes nor progeny.

Table 4.4. Average and maximum absolute changes in EBV and genomic EBV for weaning weight among monthly evaluations in 2017 and evaluations in the previous month for genotyped animals born until December of 2016 in the American Angus population[1]

Group	Contrast[2]	Average						Maximum					
		EBV		GEBV		GEBV_UC[3]		EBV		GEBV		GEBV_UC	
All	C1	0.02	(0.22)	0.05	(0.48)	0.08	(0.86)	2.20	(23.12)	0.88	(9.21)	1.08	(11.35)
	C2	0.02	(0.25)	0.05	(0.55)	0.09	(0.91)	2.06	(21.65)	1.15	(12.09)	1.10	(11.56)
	C3	0.02	(0.22)	0.04	(0.46)	0.08	(0.85)	1.91	(20.08)	1.07	(11.25)	0.97	(10.21)
	C4	0.01	(0.11)	0.02	(0.26)	0.07	(0.76)	1.63	(17.13)	0.66	(6.98)	0.85	(8.97)
	C5	0.00	(0.04)	0.01	(0.13)	0.07	(0.73)	1.33	(13.98)	0.45	(4.69)	0.96	(10.12)
	C6	0.00	(0.01)	0.01	(0.07)	0.07	(0.73)	1.68	(17.66)	1.09	(11.46)	1.17	(12.30)
	C7	0.00	(0.02)	0.01	(0.06)	0.07	(0.73)	1.55	(16.29)	0.72	(7.54)	0.95	(10.00)
	C8	0.00	(0.04)	0.01	(0.12)	0.07	(0.73)	1.08	(11.35)	0.44	(4.60)	0.84	(8.81)
	C9	0.01	(0.06)	0.01	(0.15)	0.07	(0.73)	1.29	(13.56)	0.70	(7.35)	0.88	(9.26)
	C10	0.00	(0.03)	0.01	(0.11)	0.07	(0.72)	1.00	(10.51)	0.44	(4.61)	0.94	(9.88)
	C11	0.00	(0.01)	0.00	(0.05)	0.07	(0.72)	0.43	(4.53)	0.18	(1.85)	0.84	(8.81)
	C12	0.00	(0.00)	0.00	(0.02)	0.07	(0.72)	0.27	(2.80)	0.11	(1.10)	0.87	(9.09)
New progeny records	C1	0.04	(0.41)	0.05	(0.48)	0.08	(0.87)	2.20	(23.12)	0.88	(9.21)	0.88	(9.22)
	C2	0.05	(0.51)	0.05	(0.58)	0.09	(0.93)	2.06	(21.65)	1.15	(12.09)	1.10	(11.56)
	C3	0.05	(0.48)	0.05	(0.51)	0.08	(0.88)	1.91	(20.08)	1.07	(11.25)	0.83	(8.72)
	C4	0.02	(0.23)	0.03	(0.28)	0.07	(0.77)	1.63	(17.13)	0.66	(6.98)	0.70	(7.32)
	C5	0.01	(0.09)	0.01	(0.14)	0.07	(0.73)	1.33	(13.98)	0.45	(4.69)	0.72	(7.53)
	C6	0.00	(0.03)	0.01	(0.07)	0.07	(0.73)	1.68	(17.66)	1.09	(11.46)	1.17	(12.30)
	C7	0.00	(0.03)	0.01	(0.06)	0.07	(0.73)	1.55	(16.29)	0.72	(7.54)	0.95	(10.00)
	C8	0.01	(0.07)	0.01	(0.12)	0.07	(0.73)	1.08	(11.35)	0.44	(4.60)	0.84	(8.81)
	C9	0.01	(0.10)	0.01	(0.15)	0.07	(0.73)	1.29	(13.56)	0.70	(7.35)	0.72	(7.58)
	C10	0.01	(0.06)	0.01	(0.11)	0.07	(0.73)	1.00	(10.51)	0.44	(4.61)	0.64	(6.73)
	C11	0.00	(0.02)	0.00	(0.05)	0.07	(0.72)	0.43	(4.53)	0.18	(1.85)	0.61	(6.42)
	C12	0.00	(0.01)	0.00	(0.02)	0.07	(0.72)	0.27	(2.80)	0.11	(1.10)	0.65	(6.84)
Young[4]	C1	0.02	(0.21)	0.05	(0.50)	0.08	(0.88)	0.70	(7.31)	0.25	(2.59)	0.37	(3.87)
	C2	0.02	(0.23)	0.05	(0.57)	0.09	(0.97)	0.58	(6.09)	0.25	(2.62)	0.42	(4.46)
	C3	0.02	(0.21)	0.04	(0.46)	0.08	(0.87)	0.56	(5.87)	0.23	(2.43)	0.61	(6.45)
	C4	0.01	(0.11)	0.03	(0.27)	0.07	(0.77)	0.22	(2.34)	0.14	(1.51)	0.53	(5.58)
	C5	0.00	(0.04)	0.01	(0.13)	0.07	(0.74)	0.12	(1.23)	0.06	(0.58)	0.40	(4.20)
	C6	0.00	(0.01)	0.01	(0.07)	0.07	(0.75)	0.04	(0.41)	0.03	(0.29)	0.34	(3.62)
	C7	0.00	(0.01)	0.01	(0.06)	0.07	(0.74)	0.14	(1.46)	0.05	(0.56)	0.37	(3.93)
	C8	0.00	(0.03)	0.01	(0.13)	0.07	(0.75)	0.07	(0.73)	0.08	(0.82)	0.40	(4.15)
	C9	0.00	(0.05)	0.01	(0.15)	0.07	(0.74)	0.12	(1.24)	0.07	(0.74)	0.46	(4.79)
	C10	0.00	(0.04)	0.01	(0.11)	0.07	(0.72)	0.12	(1.25)	0.05	(0.49)	0.39	(4.08)
	C11	0.00	(0.01)	0.01	(0.05)	0.07	(0.70)	0.05	(0.49)	0.02	(0.24)	0.41	(4.32)
	C12	0.00	(0.00)	0.00	(0.02)	0.07	(0.70)	0.02	(0.22)	0.01	(0.12)	0.47	(4.97)

[1]Absolute changes in additive genetic standard deviation units (in kg inside the parenthesis).
[2]C*x* = contrast between month *x* of 2017 and the precedent month; with *x* varying from 1 (January) to 12 (December).
[3]GEBV_UC = GEBV from a scenario where the core set is updated every month.
[4]Born in 2016 with neither phenotypes nor progeny.

FIGURES

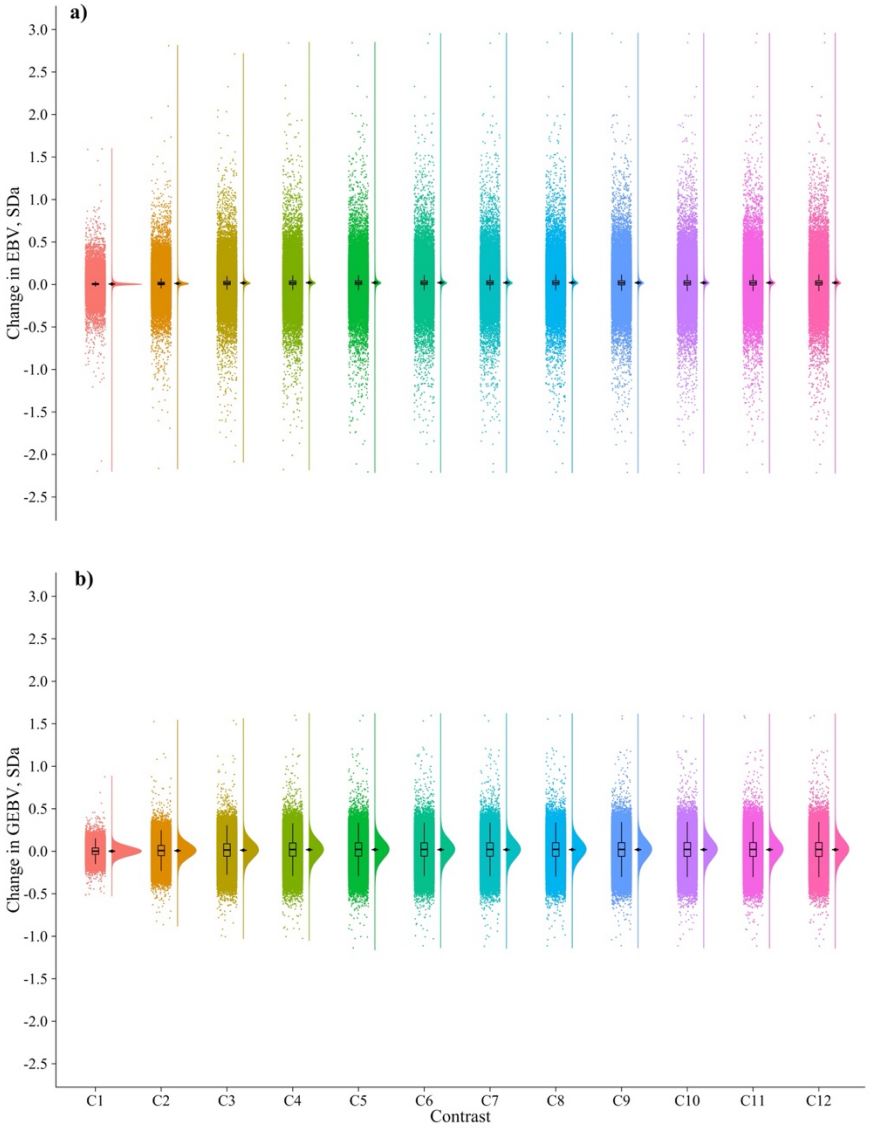

Figure 4.1. Distribution of cumulative changes in EBV (a) and GEBV (b) contrasting the evaluation for weaning weight in December of 2016 with subsequent monthly evaluations in 2017 for all genotyped animals [C*x* = contrast between December of 2016 and month *x* of 2017; with *x* varying from 1 (January) to 12 (December)]. All changes and statistics are presented in additive genetic standard deviation (SDa) units. The major changes ocurred in the first four evaluations due to a considerable increase in the number of phenotypes and genotypes from January to April of 2017.

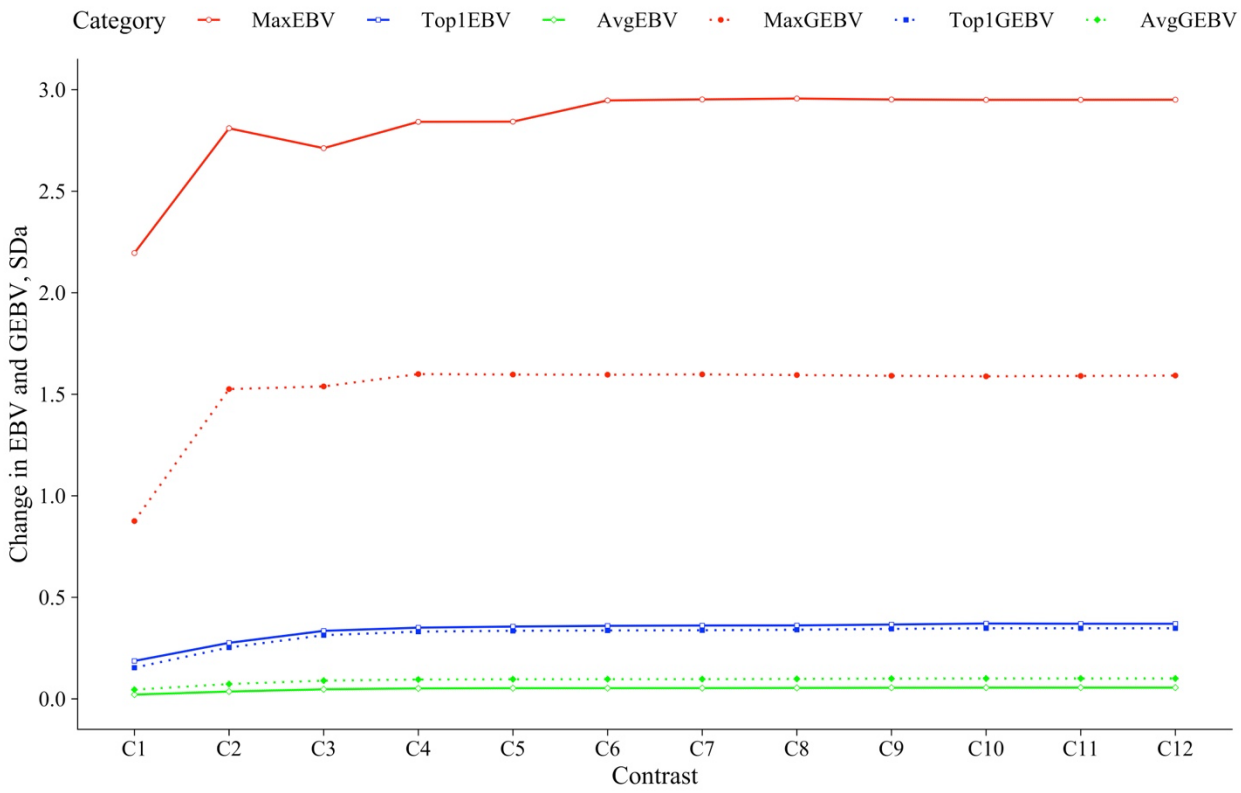

Figure 4.2. Maximum (Max), top 1% (Top1) and average (Avg) absolute changes in EBV and GEBV contrasting the evaluation for weaning weight in December of 2016 with subsequent monthly evaluations in 2017 for all genotyped animals [C*x* = contrast between December of 2016 and month *x* of 2017; with *x* varying from 1 (January) to 12 (December)]. All changes and statistics are presented in additive genetic standard deviation (SDa) units. Average absolute changes in GEBV were approximately two times greater than those in EBV, but maximum and top 1% absolute changes were larger in EBV.

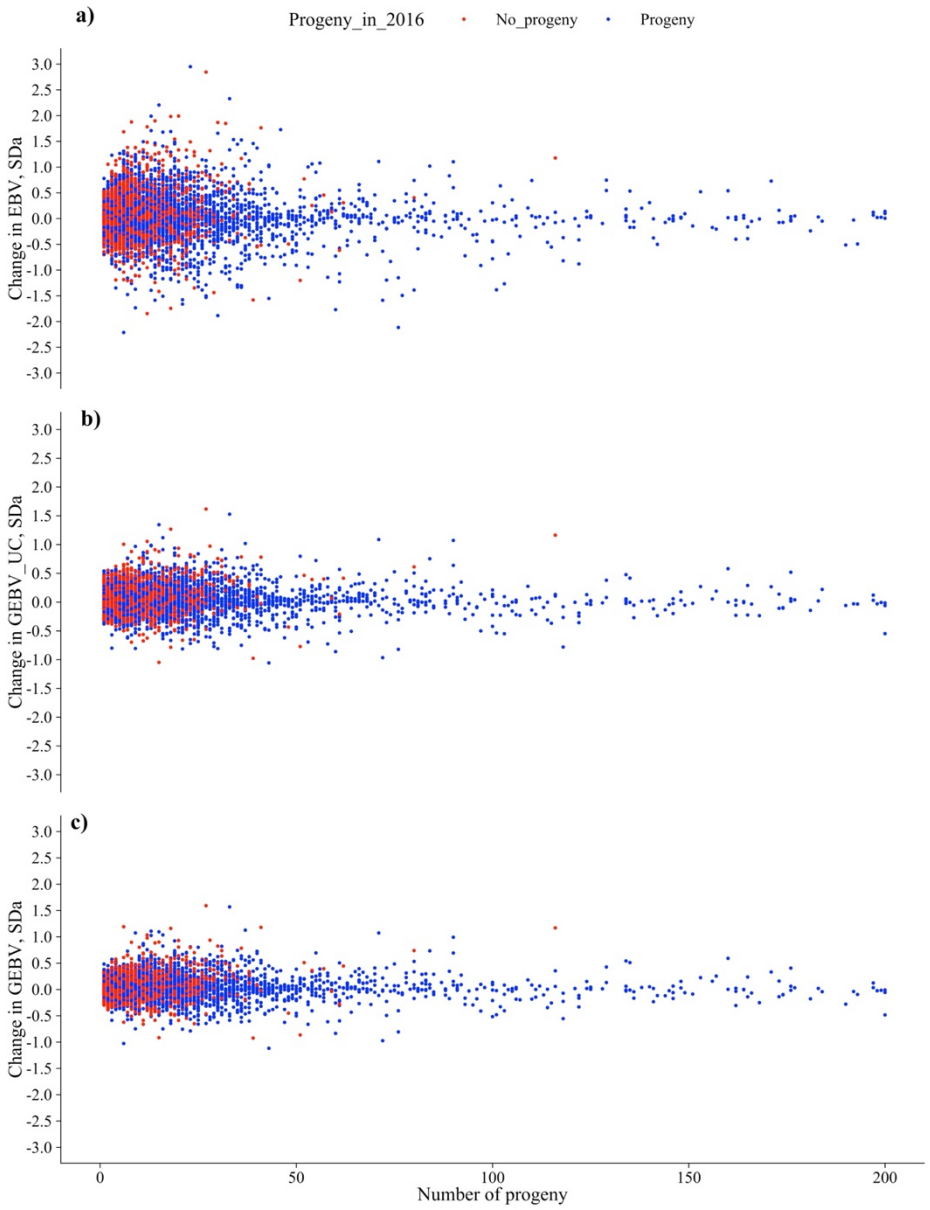

Figure 4.3. Distribution of changes in EBV (a), GEBV_UC (b), and GEBV (c) as a function of the number of progeny for genotyped sires with own weaning weight phenotype and progeny with weaning weight phenotypes, when contrasting the evaluations for weaning weight in December of 2016 with the evaluations in December of 2017. Changes are presented in additive genetic standard deviation (SDa) units. Average and maximum absolute changes in GEBV were smaller than those changes in EBV; updating the core group in APY generated marginal increase in the absolute average and maximum changes.

CHAPTER 5

DECAY OF ACCURACY OF GENOMIC PREDICTIONS OVER TIME IN BROILERS[3]

[3] Jorge Hidalgo, Daniela Lourenco, Shogo Tsuruta, Yutaka Masuda, Vivian Breen, Rachel Hawken, Matias Bermann, and Ignacy Misztal. To be submitted to *Journal of Animal Science*.

ABSTRACT

Accuracy of genomic selection is an important component of the selection response. The objectives of this research were to: 1) investigate the trends for accuracy of predictions over time in a broiler population accumulating phenotypes, genotypes, and pedigrees; 2) test if the data from distant generations is useful to maintain the accuracy of predictions in selection candidates; 3) compare accuracy estimates obtained with predictive ability and linear regression (**LR**) methods. The data used contained 820K phenotypes for a growth trait (**GT**), 200K for two feed efficiency traits (**FE1 and FE2**), and 42K for a carcass yield trait (**CY**). The pedigree included 1,252,619 animals born across 7 years, of which 154,318 from the last 4 years were genotyped. Before genotypes became available for the training populations, the accuracy was nearly stable even with the accumulation of phenotypes and pedigrees. The inclusion of genotypes in the training populations lead to an increase in accuracy of 88, 106, 166 and 53% for GT, FE1, CY, and FE2, respectively. The decay of accuracy over time was evaluated in predictions either of animals that were born in the 3 following generations or in the years after the training populations. The accuracy in both cases declined faster without than with genomic information in the training populations. When genotypes were not available, the average decline in accuracy across traits was 51% from the first to the second generation of validation, and 26% from the second to the third generation of validation. When genotypes were available, the average decline across traits was 36% from the first to the second generation of validation, and it was 4% from the second to the third generation of validation. The accuracy in the last 3 generations was the same when the training population included 5 or 2 years of data, and a marginal decrease was observed when the training population included just 1 year of data. In this broiler population, training sets including genomic information provide an increased accuracy and persistence of genomic predictions about two-fold compared to

training sets without genomic data. The two most recent years of pedigree, phenotypic and genomic data are enough to maintain the accuracy of predictions in selection candidates. Accuracy estimates obtained by predictive ability and the LR methods have the same trend, with differences in their magnitude.

INTRODUCTION

Accuracy of genomic selection is an important parameter in animal breeding programs because of its direct relationship with the selection response. This parameter is a function of the proportion of the genetic variance captured by the single nucleotide polymorphisms (**SNP**) and the accuracy of the SNP effect estimates, which depends on the amount of phenotypic data available, the genetic architecture and the heritability of the trait, and the statistical method used (Dekkers, 2007; Goddard, 2009).

The decay of accuracy of predictions over time in initial genomic selection studies using stochastic simulations was small. Meuwissen et al. (2001) found that the accuracy for a trait with major genes, in the absence of artificial selection, decreased from 0.84 to 0.72 after 5 generations without phenotyping the genotyped animals. Muir (2007), also using a stochastic simulation concluded that accuracy of genomic selection in breeding programs decays faster for traits under selection.

Using simulated data from layers, Wolc et al. (2015) reported that after 3 years of selection the accuracy remained almost stable, decaying from 0.77 to 0.73 when the breeding program included new animals with genotypes and phenotypes every generation. On the contrary, the accuracy declined from 0.77 to 0.34 when no new animals with phenotypes were included, and the

attained selection response was ~30% smaller. In the same study, the results observed using real data for approximately 2,700 genotyped animals were consistent with those of the simulation; however, the accuracy was lower.

Genomic selection acts on independent chromosome segments (VanRaden, 2008; Goddard, 2009) and on its clusters (Pocrnic et al., 2019). The number of independent chromosome segments segregating in a finite population can be estimated as $4N_eL$ (Stam, 1980), where N_e is the effective population size and L is the length of the genome in Morgans. Equivalently, the number of independent chromosome segments can be estimated as the number of the largest eigenvalues explaining 98% of the variation in the genomic relationship matrix (**G**; Pocrnic et al., 2016a). For a broiler population, assuming L = 30 Morgans, the number of independent chromosome segments was ~4K and N_e was 44 (Pocrnic et al., 2016b).

Based on the findings of Pocrnic et al. (2016b), we hypothesized that the accuracy of genomic predictions in a broiler population under selection would be high, with a small decay over time if predictions are based on at least 4K genotyped animals with high individual accuracy. The persistence of accuracy of genomic predictions should be high under the assumption of the additive model even with strong selection on the traits of interest in the breeding program.

In the absence of inbreeding, the relatedness and potential contributions of ancestors of an animal, based on pedigree relationships, decline 50% for each generation traced back in the pedigree. Therefore, very distant ancestors have small or even negative effect on the accuracy of predictions of the youngest animals (Lourenco et al., 2014). The decline of relatedness and potential contributions based on genomic relationships will depend on the method used to compute **G** and whether it is based on identical by state or identical by descent relationships. Thus, it is of

interest to study the contribution of genotypes, pedigree, and phenotypes from distant generations to the accuracy of genomic predictions in the selection candidates.

One of the most used methods to estimate the accuracy of genomic predictions is a cross-validation test called predictive ability, in which the correlation between genomic predictions and phenotypes adjusted for fixed and random effects other than the additive genetic and the residual effects is computed (Legarra et al., 2008). According to Legarra and Reverter 2018, the statistic of this method is sensitive to incorrect heritability or precorrection of phenotypes and may yield biased results if those are incorrect. Legarra and Reverter (2018) proposed a semi-parametric method based on linear regression (i.e., the **LR** method) that relies on the comparison of successive evaluations based on partial and whole data. The statistics of this method do not require the pre-correction of phenotypes and might better estimate the accuracy of predictions.

The objectives of this research were to: 1) investigate the trends for accuracy of predictions over time in a broiler population accumulating phenotypes, genotypes, and pedigrees; 2) test if the data from distant generations is useful to maintain the accuracy of predictions in selection candidates; 3) compare accuracy estimates obtained by predictive ability and the LR methods.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not needed as data were obtained from preexisting databases.

Data and variance components

The dataset used in this research study was provided by Cobb-Vantress Inc. (Siloam Springs, AR). The pedigree included 1,252,619 purebred broilers born across 7 years. A total of 154,318 animals were genotyped for 60K SNP panel. After quality control 44,448 markers were

kept for the analyses. All the genotyped animals belonged to the four most recent years in the dataset. The whole dataset contained 820,110 phenotypes for a growth trait (**GT**), 200,093 phenotypes for a feed efficiency trait (**FE1**), 42,895 phenotypes for a carcass yield trait (**CY**), and 203,060 phenotypes for a second feed efficiency trait (**FE2**). The GT and CY phenotypes were recorded on birds at 35 d of age, FE1 and FE2 phenotypes were measured during a 1-week period after 35 d of age.

Variance components were estimated using the average information restricted maximum likelihood algorithm as implemented in the AIREMLF90 software (Misztal et al., 2014a). The analysis to estimate variance components was performed using the four-trait animal model described in Lourenco et al. (2015) and included all pedigree and phenotypic data available (i.e., the analysis was based on a traditional pedigree-based model). The heritabilities for the 4 traits ranged from 0.20 to 0.55, the genetic correlations ranged from -0.15 to 0.31, and the phenotypic correlations ranged from -0.01 to 0.43.

Analyses and computations

Genomic estimated breeding values were obtained with the four-trait model using single-step genomic best linear unbiased prediction (ssGBLUP; Aguilar et al., 2010) and the algorithm for proven and young (APY; Misztal et al., 2014b); BLUP90IOD2OMP1 was the software of choice (Misztal et al., 2014a). The construction of **G** was based on VanRaden (2008), and this matrix was blended with 5% of the block of **A** corresponding to the genotyped animals ($\mathbf{A}_{22}$) to avoid singularity problems. The rescaling of **G** to match $\mathbf{A}_{22}$ was by means of diagonals and off-diagonals (Chen et al., 2011). To implement APY, the eigenvalue decomposition of **G** was done to determine the number of the largest eigenvalues explaining 98% of the variation. Based on that, a core set of 5,173 genotyped animals was randomly selected.

Training and validation populations

The effect of increasing or decreasing the size of the training population on accuracies of predictions was evaluated. Data across the 7 years were split according to animals' year of birth. In the first scenario the objective was to investigate the evolution of accuracy over time accumulating phenotypes, genotypes, and pedigrees. Thus, the training population was increased progressively by adding 1 year of data. For example, the first training population included all the animals with phenotypes born in the first year, the second training population included animals with phenotypes born in the first two years, and so on until the sixth training population, which included 6 years of data.

The first 3 years of data did not include genotyped animals; therefore, only the last 4 years of data contributed with genotyped animals to the training populations. This allowed us to evaluate the impact of including genomic information in the training population on the accuracy of predictions of the validation animals. When the training population included up to the first 3 years of data (i.e., no genomic information), validation was done on animals with phenotypes, whereas for the training populations including also genotyped animals, validation was done on animals with both phenotypes and genotypes in the whole data.

The validation populations were defined following two approaches: 1) the validation populations were the animals born in the 3 generations after the training population, that is, the progeny, grand progeny, and great grand progeny of animals in the training population. 2) the validation populations were defined as the animals born in the years after the training population. For example, if the training population included the first 3 years of data, the validation populations were animals born in years 4 to 7. In the second approach, each validation population included

animals born only in 1 year. The phenotypes of animals and their contemporaries in the validation populations were removed from the analyses.

In the second scenario the objective was to test if the data from distant generations help to avoid a decrease in accuracy of predictions in selection candidates. The largest training population in this scenario that included 5 years of data was progressively reduced removing the oldest animals. For example, when the training population was reduced from 5 to 4 years of data, we removed data from the first year. The second year of data was removed resulting in a training population with 3 years of data (years 3 to 5 of data), and so on until we kept only the last year of data in the training population (fifth year of data). The validation populations in this case were always the animals born in the 3 generations after the last year of the training population (i.e., after the fifth year of data). The number of animals in the training populations is presented in Table 1. The number of animals in the validation populations by generation and by year is presented in Table 2 and Table 3, respectively.

In the first scenario, where data accumulated over the years, the core set in APY remained unchanged. In contrast, in the analyses removing old data, the core set was updated after removing the fourth year of data because animals with genotypes that were born in that year were excluded from the analysis. Therefore, a new core set of 5,173 genotyped animals was randomly selected within the genotyped animals available at that point in time.

Accuracy of genomic predictions

Validation of genomic estimated breeding values was done by cross-validation using the predictive ability method (Legarra et al., 2008). Predictive ability was defined as the correlation between genomic estimated breeding value and phenotype adjusted for fixed and random effects

other than the additive genetic and residual effects ($r_{GEBV,\hat{Y}}$). Accuracy of estimated breeding values ($\widehat{Acc}$) for birds in the validation populations was calculated as follows:

$$\widehat{Acc} = \frac{r_{GEBV,\hat{Y}}}{\sqrt{h^2}}$$

where, h^2 is the heritability of the trait.

Adjusted phenotypes were calculated using PREDICTF90 software (Misztal et al., 2014a) with all the pedigree, phenotypic, and genomic information available. The formula by Legarra et al. (2008) is dependent on the heritability for the validation population and on the unbiased estimates of the effects used to adjust phenotypes. The method LR (from linear regression) may provide better estimates of accuracy because it does not dependent on the adjustments present in the predictive ability formula (Legarra and Reverter, 2018). In the LR method, statistics are computed for the validation individuals by comparing estimated breeding values obtained using the whole data available ($\hat{u}_w$) with estimated breeding values obtained using partial data ($\hat{u}_p$). In the latter, phenotypes of validation individuals are removed from the analyses. Thus, both methods yield estimates of the correlation between the estimated and the "true" breeding value. However, the LR method measures the accuracy of predicting future breeding value instead of measuring the accuracy of predicting future performance, as in the predictive ability method. Accuracies using the method LR ($\widehat{Acc_{LR}}$) were calculated as follows:

$$\widehat{Acc_{LR}} = \sqrt{\frac{cov(\hat{u}_w, \hat{u}_p)}{(1-\bar{F})\sigma_a^2}}$$

where, $cov(\hat{u}_w, \hat{u}_p)$ is the covariance between breeding values obtained with the whole dataset and breeding values obtained with the partial dataset, $\bar{F}$ is the average inbreeding coefficient in the validation population, and σ_a^2 is the additive genetic variance of the population.

RESULTS AND DISCUSSION

Effect of increasing the size of training population

The trends for the accuracy (based on predictive ability) in the validation populations per generation and per year are shown in Figure 1 and Figure 2, respectively. The accumulation of phenotypes and genotypes over time increased the accuracy. Traits with lower heritability required more phenotypes to achieve a good level of accuracy. For instance, CY was twice as heritable as FE2, the number of phenotypes was five times larger for FE2, and the accuracy attained was similar for both traits.

When genotypes were not available for animals in the training populations (three first years of data), the accuracy in the validation populations was nearly stable or the gain was small (~15%, on average across traits) even accumulating pedigrees and phenotypes (Figure 1 and Figure 2). As an illustration, the accuracy for GT in the progeny was 0.28, 0.24, and 0.33 when the training population included, one, two, and 3 years of data, respectively. In contrast, when genotyped birds were available in the training populations, on average, the accuracy in validation populations was approximately two-fold in all the traits but in FE2. The accuracy in the progeny increased 88% (0.33 vs. 0.62) for GT, 106% (0.35 vs. 0.72) for FE1, 166% (0.24 vs. 0.64) for CY, and 53% (0.43 vs. 0.66) for FE2, when the training population included 4 years of data compared to when it included 3 years of data (Figure 1).

The same trend was observed in the scenario of validation per year. The accuracy in the validation animals born one year after the training population increased 118% (0.27 vs. 0.58) for GT, 79% (0.31 vs. 0.56) for FE1, 100% (0.24 vs. 0.47) for CY, and 30% (0.41 vs. 0.54) for FE2 when the training population included 4 years of data compared to when it included 3 years of data (Figure 2).

The benefits of including genomic information on accuracy of predictions were also reported across several species and traits. Wolc et al. (2011) reported an average increase ~17% for predictions of 16 traits in layers. In broilers, Lourenco et al. (2015) estimated an increase of ~50% in accuracy of predictions for growth and efficiency related traits. Vallejo et al. (2017) reported that genomic information doubled the accuracy of predictions in a study related to disease resistance in rainbow trout. Garcia et al. (2018) found that genomic information increased accuracy up to 36% for predictions of residual carcass weight in channel catfish. In dual purpose cattle, Cesarani et al. (2020) found an increase of 37% in accuracy for predictions of milkability. Macedo et al. (2020a) reported an increase in accuracy of ~33% for predictions of milk production in a sheep population. Bermann et al. (2020) reported a gain of 15% in accuracy of predictions studying mortality in broilers.

The gain in accuracy is attributed to a more precise estimation of the mendelian sampling terms (Hayes et al., 2009; Cole and VanRaden, 2011). The magnitude of the gain in accuracy of genomic predictions depends on the number, distribution, and contribution of genotypes and phenotypes, as well as on the selection intensity on traits (Lourenco et al., 2015). In our study the gain in accuracy by the inclusion of genotypes was greater than in most of the previous studies, which can be explained by the large number of phenotypes and genotypes in this broiler population.

Across the validation populations, the accuracy obtained before genotyped birds were included in the training populations (three first years of data) declined faster than with genotyped birds, suggesting that the effects of the independent chromosome segments were estimated with more precision in the latter case. According to Bastiaansen et al. (2012), pedigree relationships are not able to predict the random segregation of independent chromosome segments to the next

generations, whereas this segregation can be traced by markers. Therefore, the proportion of variance explained by linkage disequilibrium with markers decays less than the proportion of variance explained by family structure.

When no genotyped birds were included in the training populations, the average decline across traits and training populations was 51% from progeny to grand progeny, 77% from progeny to great grand progeny, and 26% from grand progeny to great grand progeny. In contrast, accuracy decayed 36% from progeny to grand progeny, 40% from progeny to great grand progeny, and only 4% from grand progeny to great grand progeny when training populations included genotyped birds (Figure 1). Habier et al. (2007) reported similar results comparing several methods to estimate breeding values. In their study the persistence was also always greater in genomic-based methods than in pedigree-based methods.

Our results agree with the expectation of a 50% decay in pedigree relationships every generation. Genomic relationships are expected to decay at a slower rate (Wolc et al., 2011), which was confirmed by the increased persistence of genomic predictions. Within our research study, the loss of accuracy across traits when the training populations included genotyped birds was important from progeny to grand progeny, and the loss from grand progeny to great grand progeny was marginal. Thus, the advantage of using genomic information became more important as breeding values from more distant validation generations were predicted. Wolc et al. (2011) also reported more persistence in the genomic evaluations than in genetic evaluations, as well as an important decline in the accuracy of the first generation, with smaller losses in the successive generations.

The decay of accuracy over generations was similar when the training population included 1 or 2 years of genomic information (Figure 1). Therefore, although the decay in accuracy was

smaller when genomic information was used in the genomic evaluation, it is important to highlight the need for continuing phenotyping animals to minimize the decay in accuracy (Sonesson and Mewuissen, 2009).

The persistence of genomic predictions had a similar trend when validation was done by years. When no genotyped birds were included in the training populations (three first years of data), the average decline in accuracy across traits and training populations was 62% from the first to the second year of validation, 87% from the first to the third year of validation, and 25% from the second to the third year of validation. Also, the persistence of genomic predictions was greater when training populations included genotyped birds. In the latter case, on average, the accuracy decreased 13% from the first to the second year of validation and 20% from the first to the third year of validation (Figure 2).

Trends for the accuracy in the next year after the training populations are shown in Figure 3. Although the accumulation of data along the years resulted in an increase in accuracy, the addition of the last year of data (sixth year of data) to the training population did not increase accuracy for CY and had a negative impact on the accuracy for GT and FE2.

As the formula to obtain the accuracy by predictive ability is a ratio, possible explanations for the decline in the accuracy are a decrease in the correlation between the genomic estimated breeding values and the adjusted phenotypes, which is unlikely because it is expected that the accumulation of recent data produce at least the same correlation; another possibility is a decline in the heritability of the trait that was not accounted for by the formula (Legarra and Reverter, 2018). Selection reduces the genetic variance and heritabilities, this was proved analytically (Bulmer, 1971) and by simulations (Bijma, 2012; Gorjanc et al., 2015). Using real data, recent

studies reported reduction in genetic variances and heritabilities as a result of selection (Bulmer effect) and drift (Hidalgo et al., 2020; Macedo et al., 2021; Tsuruta et al., 2021).

It would be useful to have theoretical formulas for accuracies for the next and successive generations given different parameters such as population size, effective population size, number of animals with genotypes and phenotypes, breeding structure, genetic parameters, etc. In theory, one would expect to have similar theoretical accuracies for traits with similar parameters, the accuracies would increase over time with more data, persistence would be higher with more data, and there would be a consistent decay of accuracy from progeny to grand progeny and successive generations without recording more phenotypes. However, this study would defy such expectations as accuracy for GT was the same in the grand progeny, and it declined in the great grand progeny regardless of more data in the training population including 5 years of data than in the training population including 4 years of data (Figure 1). In FE2 the accuracy was the same for the grand progeny and for the great grand progeny contrasting the same training populations (5 vs. 4 years of data; Figure 1).

In addition, when the training population included 5 years of data, the accuracy was the same for the grand progeny and it was greater for the great grand progeny in a trait with less data and lower heritability (i.e., GT vs. FE1; Figure 1). Also, the accuracy was the same in the grand progeny and in the great grand progeny for several cases (i.e., training populations) in all the traits. In other words, the expected decay in accuracy was not consistent.

Such phenomena can be explained by the selection intensity and by changes in genetic parameters over time. Lourenco et al. (2015) looked at accuracies of genomic predictions obtained with genotypes of males, females, or both in a broiler population. The accuracies were strongly dependent on selection intensity per sex, which means the accuracy for strongly selected traits on

males were depressed when the training population included genotypes of both sexes. Based on the accuracy, it was possible to understand the selection practices. Such findings were theoretically addressed by Bijma (2012). Hidalgo et al. (2020) found that the genetic parameters based on pedigree and phenotypes differ from the ones computed with the inclusion of genomic information. Likewise, the trends over time are contrasting. If heritability decline over time, the accuracy computed from predictive ability using a previous estimate of heritability is underestimated. Also, if the genetic correlations change, the correlated responses will change accordingly. Therefore, a realistic theory for the accuracy and persistence of genomic predictions would have to include the effects of selection and changes in genetic parameters. An extra complication would be accounting for epistatic changes if they occur faster with the genomic selection.

Effect of decreasing the size of training population

Figure 4 shows the trend for the accuracy of genomic estimated breeding values in the last 3 generations of validation when data from distant generations were removed. The accuracy of genomic predictions in the last 3 generations was the same when the training population included 5 or 2 years of data. When the training population included just the nearest year of data (i.e., the fifth year of data) to the 3 generations of validation, a marginal decrease in accuracy of genomic predictions was observed. For instance, in the progeny the accuracy decreased from 0.96 to 0.93 for GT, from 0.81 to 0.78 for FE1, from 0.83 to 0.81 for CY, and it remained at 0.91 for FE2.

Differences were greater in the next generations. In the great grand progeny, the accuracy decreased from 0.43 to 0.39 for GT, from 0.57 to 0.53 for FE1, from 0.46 to 0.38 for CY, and from 0.42 to 0.37 for FE2. Our results suggest that in this population, 2 years of data (the closest to the selection candidates), including pedigree, phenotypic, and genomic information are enough to

maintain the accuracy. However, future research is needed to assess the decay in accuracy with the elimination of more years of information, including genotypes.

Similar results were reported by Lourenco et al. (2014) where no decrease was observed in the accuracy of genomic evaluations for final score in US Holsteins after removing 2 or 3 generations (12 – 17 years) of phenotypes and pedigrees. This was also true for evaluations of reproductive traits in pigs after removing up to five generations (15 years) of phenotypes and pedigrees. The conclusions of Lourenco et al. (2014) agree with our findings that retaining two or three generations of phenotypes are enough to maintain the accuracy; decreased computing cost is an advantage of the reduced data size.

Bastiaansen et al. (2012) showed by simulation that very shallow training populations (1 generation of data) in comparison with multi-generational training populations (5 generations of data) lead to a larger reduction in accuracy in the first generation; however, differences were marginal from the second to the tenth generation. Their results were not in agreement with ours because when present, the differences were larger in the more distant generations.

Predictive ability vs. LR method

Figure 5 shows accuracies of genomic predictions estimated by predictive ability and LR methods in the 3 generations after the training populations. Overall, the trends for accuracies were similar with both methods and the general conclusions remained. The correlation between accuracy by predictive ability and the accuracy by LR was 0.98 for GT, 0.91 for FE1, 0.94 for CY, and 0.93 for FE2 illustrating good agreement between both methods. The agreement (correlation) was greater for traits with larger amount of data. The reason for comparing the two methods to compute accuracy using correlations was to show that the trends may be similar independent of

the method. Although correlations do not imply that both methods yield similar estimates in magnitude, that was verified with the actual values (Figure 5).

The estimates of accuracy in both methods were similar; however, still some differences were observed. The observed increase in accuracy due to including genomic information was smaller (~0.08 percentage points across traits) in the estimates by LR method; the accuracy in the progeny increased 56% for GT, 77% for FE1, 111% for CY, and 39% for FE2 (Figure 5).

The observed decline in accuracy over time was also smaller in the LR estimates. Before genomic information was included in the training populations the accuracy decreased 41% from progeny to grand progeny, 60% from progeny to great grand progeny, and 19% from grand progeny to great grand progeny. When genomic information was included in the training populations, the accuracy decreased 14% from progeny to grand progeny 16% from progeny to great grand progeny, and 2% from grand progeny to great grand progeny (Figure 5).

According to Legarra and Reverter (2018), the accuracy estimated using the predictive ability method is sensitive to how well the model is correcting for effects, without implying that the effects are appropriate to the model; therefore, bias in the estimation of the fixed effects can create an upward bias in the estimated accuracy. The latter fact can explain the greater accuracy estimated in some cases by predictive ability, such as the accuracies in progeny validation populations especially when training populations included genomic information (Figure 5). However, there was not clear conclusion about this discrepancy since accuracies estimated by predictive ability in the grand progeny and great grand progeny were smaller than those estimated by LR in most of the cases (Figure 5).

A concern related to the accuracies estimated by predictive ability method is that estimates can be negative. For instance, for FE1 estimates in the grand progeny and great grand progeny

were negative when the training population included 1 and 2 years of data (Figure 5). In fact, it was not possible to estimate accuracies using the LR method in the same validation populations because the $cov(\hat{u}_w, \hat{u}_p)$ was negative. Therefore, this issue remained, and it just reflects the fact that the more distant the training population is from the validation population, the lower the predictive power.

Macedo et al. (2020b) evaluated by simulation the performance of the statistics of the LR method when the evaluation model used incorrect heritabilities (over and underestimated) and when the evaluation model does not account for an environmental trend present in the data. The accuracy was well estimated in all cases and was more precise as the amount of information increased. This suggests that accuracy by the LR method could be a more realistic metric for the accuracy of genomic predictions in this broiler population, yet both methods led to the same general conclusions, possibly because of the large amount of information in this population.

CONCLUSIONS

In this broiler population, training populations including genomic information provide an increase in accuracy and persistence of genomic predictions about twice as large as training populations without genomic data. There is a general decline in accuracy when predicting the performance of relatives distant from the training populations like grand progeny and great grand progeny, and the decay is larger for grand progeny than for great grand progeny. Accuracies are greater when most recent data are accumulated. The most recent two years of pedigree, phenotypic, and genomic information provide persistent accuracy of predictions for selection candidates in the next three generations beyond the training population. Accuracy estimates

obtained by predictive ability and the LR methods have similar trends, with differences in magnitude.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 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(2):743–752. https://doi.org/10.3168/jds.2009-2730

Bastiaansen, J. W. M., A. Coster, M. P. L. Calus, J. A. M. van Arendonk, and H. Bovenhuis. 2012. Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. Gen. Sel. Evol. 24:44. https://doi.org/10.1186/1297-9686-44-3

Bermann, M., A. Legarra, M. K. Hollifield, Y. Masuda, D. Lourenco, and I. Misztal. 2020. Validation of genomic and pedigree predictions from threshold models using the linear regression (LR) method: an application in chicken mortality. J. Anim. Breed. Genet. 138:4–13. https://doi.org/10.1111/jbg.12507

Bijma, P. 2012. Accuracies of estimated breeding values from ordinary genetic evaluations do not reflect the correlation between true and estimated breeding values in selected populations. J. Anim. Breed. Genet. 129:345–358. https://doi.org/10.1111/j.1439-0388.2012.00991.x

Bulmer, M. G. 1971. The effect of selection on genetic variability. Am. Nat. 105:201–211. https://doi.org/10.1086/282718

Cesarani, A., A. Garcia, J. Hidalgo, L. Degano, D. Vicario, N. P. P. Macciotta, and D. Lourenco. 2020. Genomic information allows for more accurate breeding values for milkability in

dual purpose Italian Simmental cattle. J. Dairy Sci. 104(5):5719–5727. https://doi.org/10.3168/jds.2020-19838

Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2011. Effect of different genomic relationship matrices on accuracy and scale1. J. Anim. Sci. 98:2673–2679. https://doi.org/10.2527/jas.2010-3555

Cole, J. B., and P. M. VanRaden. 2011. Use of haplotypes to estimate Mendelian sampling effects and selection limits. J. Anim. Breed. Genet. 128:446–455. https://doi.org/10.1111/j.1439-0388.2011.00922.x

Dekkers, J. C. M. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331–341. doi: https://doi.org/10.1111/j.1439-0388.2007.00701.x

Garcia, A. L. S., B. Bosworth, G. Waldbieser, I. Misztal, S. Tsuruta, and D. A. L. Lourenco. 2018. Development of genomic predictions for harvest and carcass weight in channel catfish. Gen. Sel. Evol. 50:66. https://doi.org/10.1186/s12711-018-0435-5

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

Gorjanc, G., P. Bijma, and J. M. Hickey. 2015. Reliability of pedigree based and genomic evaluations in selected populations. Genet. Sel. Evol. 47:65. https://doi.org/10.1186/s12711-015-0145-1

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. Genetics. 177:1389–2397. https://doi.org/10.1534/genetics.107.081190

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. (Camb.). 91:47–60. https://doi:10.1017/S0016672308009981

Hidalgo, J., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, Y. Huang, K. A. Gray, and I. Misztal. 2020. Changes in genetic parameters for fitness and growth traits in pigs under genomic selection. J. Anim. Sci. 98(2):1–12. https://doi.org/10.1093/jas/skaa032

Legarra A., Reverter A. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet Sel Evol. 50:53. https://doi.org/10.1186/s12711-019-0511-5

Legarra, A., C. Robert-Granié, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611–618. https://doi.org/10.1534/genetics.108.088575

Lourenco, D. A. L., B. O. Fragomeni, S. Tsuruta, I. Aguilar, B. Zumbach, R. J. Hawken, A. Legarra, and I. Misztal. 2015. Accuracy of estimated breeding values with genomic information on male, females, or both: an example on broiler chicken. Genet Sel Evol. 47:56. https://doi.org/10.1186/s12711-015-0137-1

Lourenco, D. A. L., I. Misztal, S. Tsuruta, I. Aguilar, T. J. Lawlor, S. Forni, and J. I. Weller. 2014. Are evaluations of young animals benefiting from the past generations? J. Dairy Sci. 97(6):3930–3942. https://doi.org/10.3168/jds.2013-7769

Macedo, F. L., O. F. Christensen, and A. Legarra. 2021. Selection and drift reduce genetic variation for milk yield in Manech Tete Rousse dairy sheep. J. Dairy Sci. Communications. 2(1):31–34. https://doi.org/10.3168/jdsc.2020-0010

Macedo, F., L. A. Reverter, and A. Legarra. 2020b. Behavior of the Linear Regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. J. Dairy Sci. 103(1):529–544. https://doi.org/10.3168/jds.2019-16603

Macedo, F., O. F. Christensen, J. M. Astruc, I. Aguilar, Y. Masuda, and A. Legarra. 2020a. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. Genet. Sel. Evol. 52:47 https://doi.org/10.1186/s12711-020-00567-1

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829. https://doi.org/10.1534/genetics.110.116590

Misztal, I., A. Legarra, and I. Aguilar. 2014b. Using recursion to compute the inverse of the genomic relationship matrix. J. Dairy Sci. 97(6):3943–3952. https://doi.org/10.3168/jds.2013-7752

Misztal, I., S. Tsuruta, D.A.L. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014a. Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf (Accessed May 4th, 2020.)

Muir, W. M. 2007. Comparison of genomic and traditional BLUP- estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342–355. https://doi.org/10.1111/j.1439-0388.2007.00700.x

Pocrnic, I., D. A. L. Lourenco, K. Y. Masuda, and I. Misztal. 2019. Accuracy of genomic BLUP when considering a genomic relationship matrix based on the number of the largest

eigenvalues: a simulation study. Genet. Sel. Evol. 51:75. https://doi.org/10.1186/s12711-019-0516-0

Pocrnic, I., D. A. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016a. The dimensionality of genomic information and its effect on genomic prediction. Genetics 203:573–581. https://doi.org/10.1534/genetics.116.187013

Pocrnic, I., D. A. Lourenco, Y. Masuda, and I. Misztal. 2016b. Dimensionality of genomic information and performance of the Algorithm for Proven and Young for different livestock species. Genet. Sel. Evol. 48:82. https://doi.org/10.1186/s12711-016-0261-6

Sonesson A. K., J., and T. H. Meuwissen. 209. Testing strategies for genomic selection in aquaculture breeding programs. Gen. Sel. Evol. 41:37. https://doi.org/10.1186/1297-9686-41-37

Stam, P. 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. Genet. Res. 35:131–155. https://doi.org/10.1017/S0016672300014002

Tsuruta S., T. J. Lawlor, D. A. L. Lourenco, and I. Misztal. 2021. Bias in genomic predictions by mating practices for linear type traits in a large-scale genomic evaluation. J. Dairy Sci. 104(1):662–667. https://doi.org/10.3168/jds.2020-18668

Vallejo, R. L., T. D. Leeds, G. Gao, J. E. Parson, K. E. Martin, J. P. Evenhuis, B. O. Fragomeni, G. D. Wines, and Y. Palti. 2017. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. Gen. Sel. Evol. 49:17. https://doi.org/10.1186/s12711-017-0293-6

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11):4414–4423. doi.org/10.3168/jds.2007-0980

Wolc, A., H. H. Zhao, J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, C. Stricker, D. Habier, R. L. Fernando, D. J. Garrick, S. J. Lamont, and J. C. M. Dekkers. 2015. Response and inbreeding from a genomic selection experiment in layer chickens. Gen. Sel. Evol. 47:59. https://doi.org/10.1186/s12711-015-0133-5

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

TABLES

Table 5.1. Number of animals with phenotypic records in the training populations

Training population Years	Trait[1]			
Accumulating data	GT	FE1	CY	FE2
1	104,993	24,059	5,830	24,753
1-2	224,193	52,590	12,613	53,861
1-3	342,282	82,214	18,214	83,940
1-4	464,088	111,521	23,915	113,487
1-5	596,956	141,521	30,381	143,823
1-6	711,252	171,232	36,749	173,905
Removing old data				
1-5	596,956	141,521	30,381	143,823
2-5	491,963	117,462	24,551	119,070
3-5	372,763	88,931	17,768	89,962
4-5	254,674	59,307	12,167	59,883
5	132,868	30,000	6,466	30,336

[1]GT = growth trait; FE1 = feed efficiency trait one; CY = carcass yield trait; FE2 = feed efficiency trait two.

Table 5.2. Number of animals in the validation populations by generations

Training population Years	Validation population[1]	Trait[2]			
		GT	FE1	CY	FE2
1	P	90,673	21,341	4,927	21,797
	GP	101,843	25,416	5,134	25,853
	GGP	111,953	26,824	5,008	27,081
1-2	P	95,485	23,730	4,150	24,134
	GP	106,400	25,024	4,893	25,245
	GGP	102,146	25,264	5,150	25,511
1-3	P	98,337	22,715	4,289	22,909
	GP	99,476	23,776	4,857	24,021
	GGP	103,909	24,028	5,161	24,298
1-4	P	30,023	21,969	4,884	22,226
	GP	30,473	22,888	4,760	23,157
	GGP	31,607	22,086	4,768	22,346
1-5	P	31,795	23,730	4,859	24,027
	GP	31,594	22,021	4,509	22,280
	GGP	17,489	12,651	2,921	12,759

[1]P = progeny; GP = grand progeny; GGP = great grand progeny.
[2]GT = growth trait; FE1 = feed efficiency trait one; CY = carcass yield trait; FE2 = feed efficiency trait two.

Table 5.3. Number of animals in the validation populations by year[1]

Year of validation	Trait[2]			
	GT	FE1	CY	FE2
2	119,200	28,531	6,783	29,108
3	118,089	29,624	5,601	30,079
4	33,343	24,856	5,652	25,066
5	39,068	28,695	6,442	29,022
6	39,859	29,633	6,341	30,003
7	41,019	28,769	5,948	29,063

[1]The number of animals presented corresponds to the smallest training population including one year of data, when the training population included two years of data, the first year of validation was the year 3 in the table. The same logic applies for the successive and larger training populations.
[2]GT = growth trait; FE1 = feed efficiency trait one; CY = carcass yield trait; FE2 = feed efficiency trait two.

FIGURES

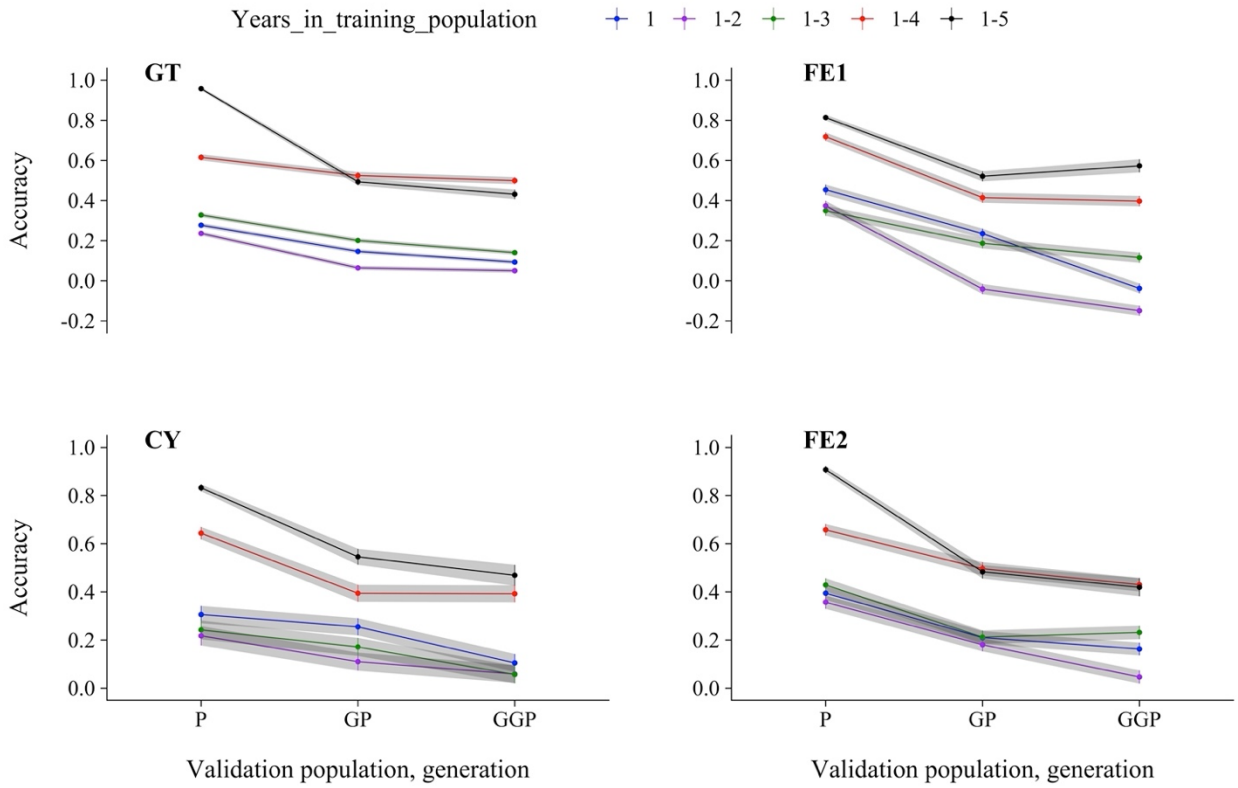

Figure 5.1. Accuracy (calculated by predictive ability) and 95% confidence interval of estimated breeding values in progeny (P), grand progeny (GP) and great grand progeny (GGP) of different training populations acummulating pedigree, phenotypic and genomic information for growth trait (GT), feed efficiency trait one (FE1), carcass yield trait (CY) and feed efficiency trait two (FE2) in broilers. Accuracies were nearly stable with the accumulation of pedigree and phenotypes. The addition of genomic information increased the accuracies approximately two-fold.

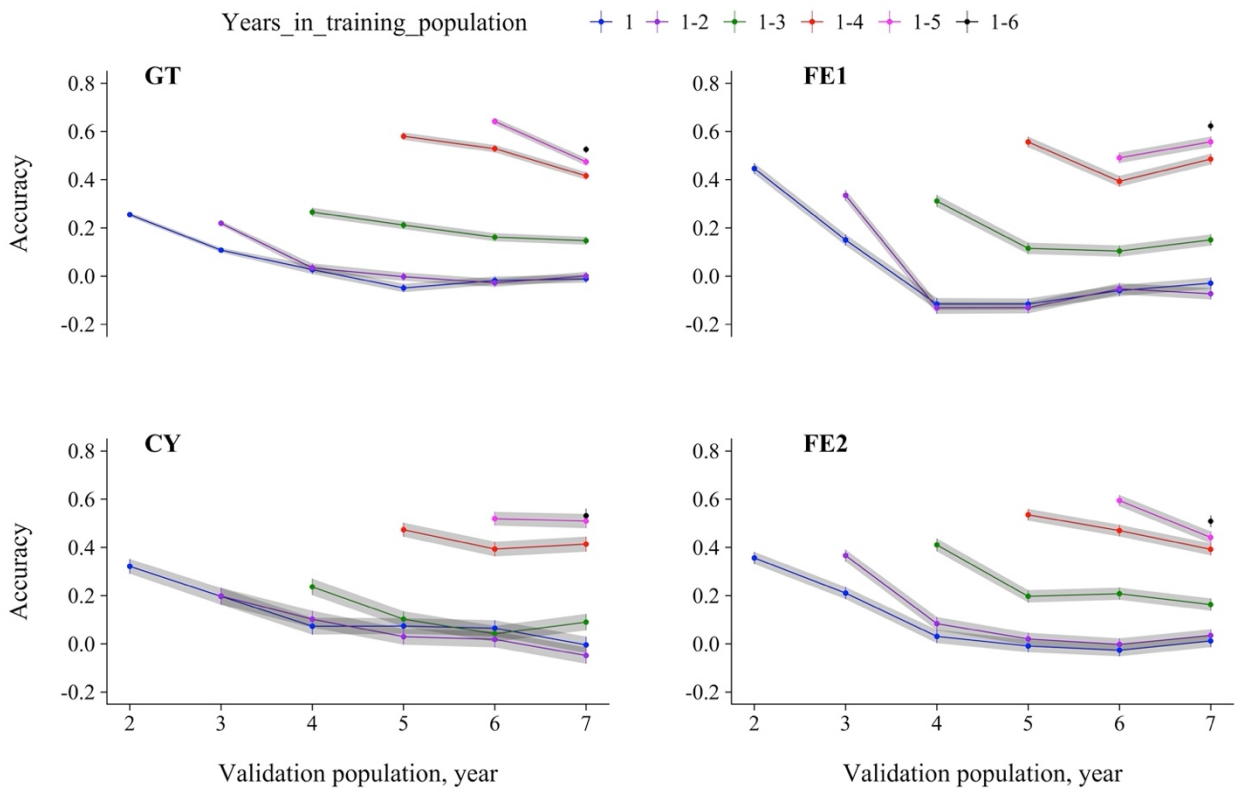

Figure 5.2. Accuracy (calculated by predictive ability) and 95% confidence interval of estimated breeding values in animals born during the succesive years after different training populations acummulating pedigree, phenotypic and genomic information for growth trait (GT), feed efficiency trait one (FE1), carcass yield trait (CY) and feed efficiency trait two (FE2) in broilers. Accuracies were nearly stable with the accumulation of pedigree and phenotypes. The addition of genomic information increased the accuracies approximately two-fold.

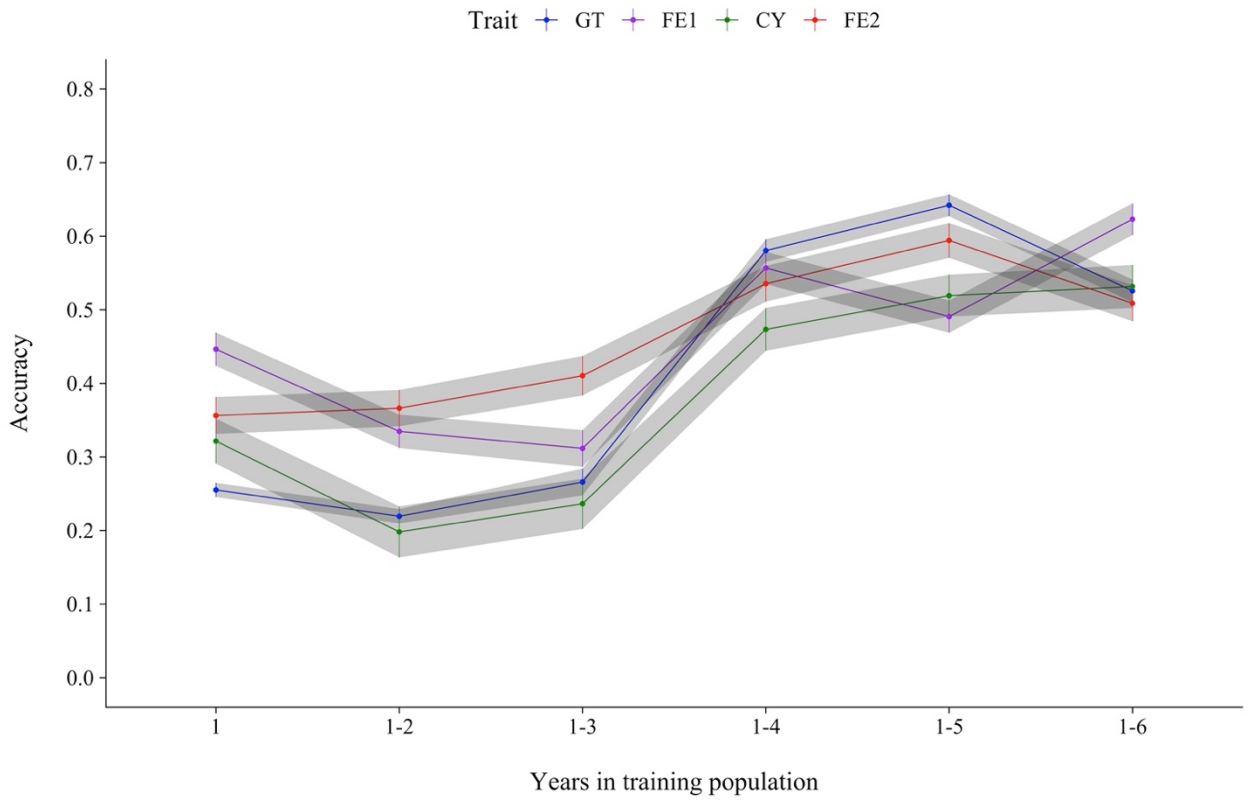

Figure 5.3. Accuracy (calculated by predictive ability) and 95% confidence interval of estimated breeding values in animals born during the next year after different training populations acummulating pedigree, phenotypic and genomic information for growth trait (GT), feed efficiency trait one (FE1), carcass yield trait (CY) and feed efficiency trait two (FE2) in broilers. Overall, the accumulation of data increased the accuracy although the increase was no steady for FE1 and CY. After the accumulation of 5 years of data, adding one more year of data did not increased the accuracy in all the traits; the accuracy remained the same for CY, while for GT and FE2 it dropped.

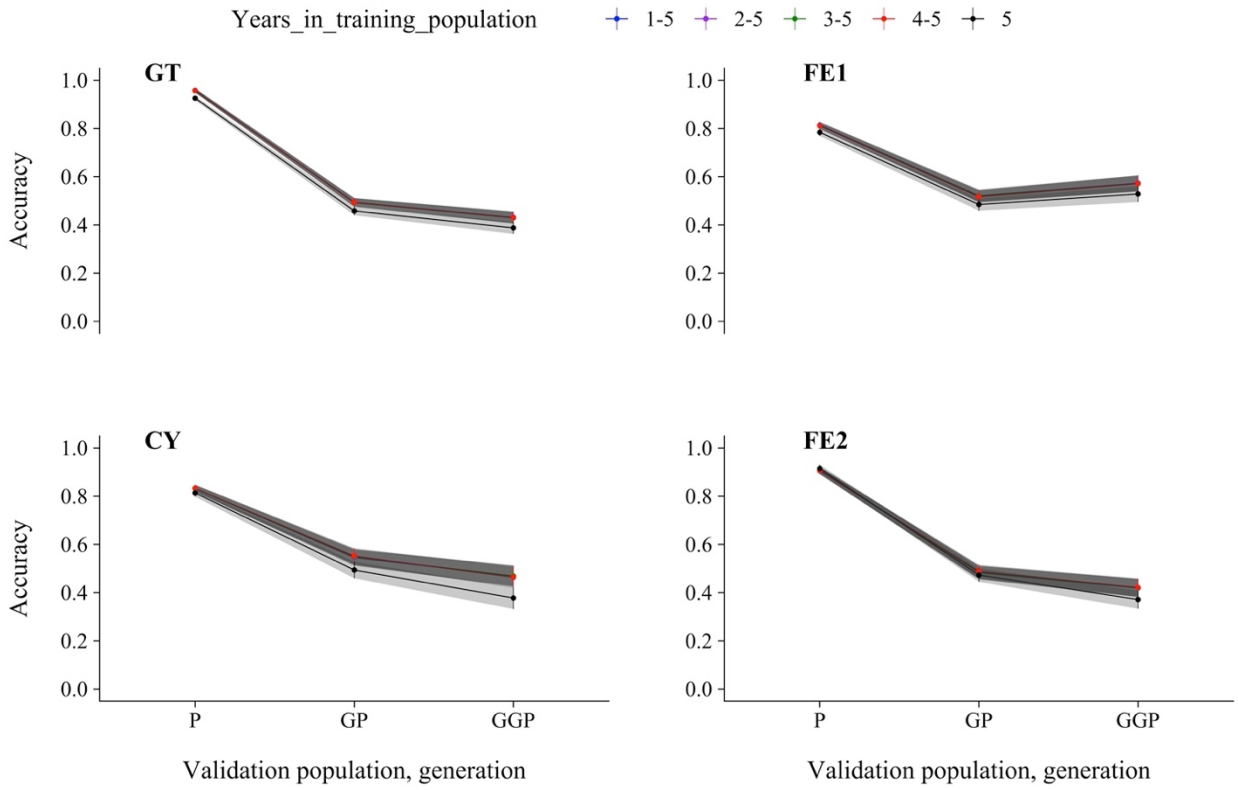

Figure 5.4. Accuracy (calculated by predictive ability) and 95% confidence interval of estimated breeding values in progeny (P), grand progeny (GP) and great grand progeny (GGP) of different training populations reducing pedigree, phenotypic and genomic information for growth trait (GT), feed efficiency trait one (FE1), carcass yield trait (CY) and Feed efficiency trait two (FE2) in broilers. Validation populations were always the 3 generations after the fifth year of data, accuracies were the same including 5 or 2 years of data in the training population. A marginal decay was observed in the accuracies when the training population included only the last year of data.

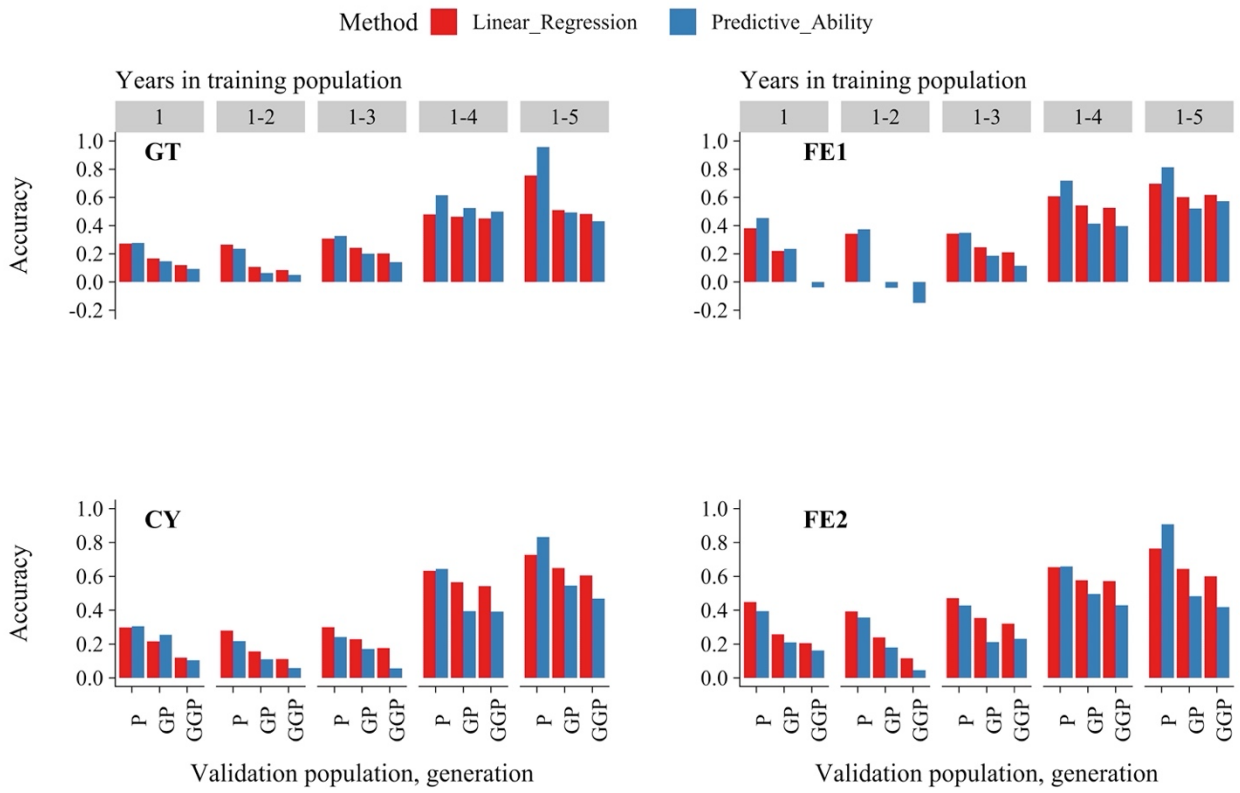

Figure 5.5. Accuracy calculated by predictive ability and linear regression methods of estimated breeding values in progeny (P), grand progeny (GP) and great grand progeny (GGP) of different training populations accumulating pedigree, phenotypic and genomic information for growth trait (GT), feed efficiency trait one (FE1), carcass yield trait (CY) and feed efficiency trait two (FE2) in broilers. The trends for accuracies were similar with both methods, differences were observed in the magnitude of the estimates.

CHAPTER 6

CONCLUSIONS

Directional selection reduces the genetic variance and emphasizes the antagonistic genetic relationships between fitness and growth traits. The reduction in the genetic variance is stronger for traits with greater heritability and under more intensive selection. Negative genetic correlations become more negative, whereas positive and weak genetic correlations remain stable. In populations undergoing genomic selection, variance components estimated without genomic information are possibly biased. Genomic estimated breeding values are as stable as traditional EBV for animals with new phenotypic data. For animals without new phenotypic data, average changes in GEBV are two times larger than average changes in EBV because of stronger links through genomic than pedigree relationships. For proven sires, changes in GEBV are smaller than changes in EBV. Genomic information increases the accuracy and persistence of GEBV leading to smaller maximum changes compared to maximum changes in EBV. The accuracy of genomic predictions declines when predicting the performance of relatives in subsequent generations without phenotypic data. The accumulation of phenotypes, genotypes, and pedigrees increases the accuracy. The most recent two years of data provide persistent accuracy of predictions for selection candidates three generations beyond the training population. Accuracy estimates obtained by predictive ability and LR methods have similar trends, with slight differences in magnitude. The changes in genetic parameters over time should be considered in the calculation of accuracy, persistence and selection response in animal breeding programs to have realistic estimates.