Running head: Starting point of genomic selection

Detecting effective starting point of genomic selection by divergent trends from BLUP and ssGBLUP in pigs, beef cattle, and broilers

Rostam Abdollahi-Arpanahi*[1], Daniela Lourenco*, Ignacy Misztal*

*Department of Animal and Dairy Science, University of Georgia, Athens, GA, 30602, USA

[1]Corresponding author: rostam.abdollahi@uga.edu ; Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA 30602

Abstract

Genomic selection has been adopted nationally and internationally in different livestock and plant species. However, understanding whether genomic selection has been effective or not is an essential question for both industry and academia. Once genomic evaluation started being used, estimation of breeding values with pedigree BLUP became biased because this method does not consider selection using genomic information. Hence, the effective start point of genomic selection can be detected in two possible ways including the divergence of genetic trends and Realized Mendelian sampling (RMS) trends obtained with BLUP and Single-step genomic BLUP (ssGBLUP). This study aimed to find the start date of genomic selection for a set of economically important traits in three livestock species by comparing trends obtained using BLUP and ssGBLUP. For this purpose, three datasets comprised a pig dataset with 117k genotypes and 1.3M animals in pedigree, Angus cattle dataset consisted of ~842k genotypes and 11.5M animals in pedigree, and a purebred broiler chicken dataset included ~154k genotypes and 1.3M birds in pedigree were used. The genetic trends for pigs diverged for the genotyped animals born in 2014 for average daily gain and backfat. In beef cattle, the trends started diverging in 2009 for weaning weight and in 2016 for postweaning gain, with little diverging for birth weight. In broiler chickens, the genetic trends estimated by ssGBLUP and BLUP diverged at breeding cycle 6 for two out of three production traits. The RMS trends for the genotyped pigs diverged for animals born in 2014, more for average daily gain than for backfat. In beef cattle, the RMS trends started diverging in 2009 for weaning weight and in 2016 for postweaning gain, with a trivial trend for birth weight. In broiler chickens, the RMS trends from ssGBLUP and BLUP diverged strongly for two production traits at breeding cycle 6, with a slight divergence for another trait. Divergence of the genetic trends from ssGBLUP and BLUP indicates onset of the genomic selection. Presence of

trends for RMS indicates selective genotyping, with or without the genomic selection. The onset of genomic selection and genotyping strategies agree with industry practices across the 3 species. In summary, the effective start of genomic selection can be detected by the divergence between genetic and RMS trends from BLUP and ssGBLUP.

Keywords: breeding values, genetic gain, genomic preselection, Mendelian sampling, single-step GBLUP

List of Abbreviations

BLUP: Best Linear Unbiased Prediction

ssGBLUP: single step Genomic Best Linear Unbiased Prediction

EBV: Estimated Breeding Value(s)

GEBV: Genomic Estimated Breeding Value(s)

SNP: Single Nucleotide Polymorphism

RMS: Realized Mendelian Sampling

APY: Algorithm for Proven and Young

ADG: Average Daily Gain

BF: Backfat

BTW: Birth Weight

WW: Weaning Weight

PWG: Post Weaning Gain

PA: Parent Average

PC: Progeny Contribution

YD: Yield Deviation

GI: Genomic Information

Introduction

Genomic selection has been widely recognized as a successful tool for genetic improvement, as evident by the extensive genotyping in various livestock and plant species (Misztal et al., 2020; VanRaden, 2020). Genomic selection allows to preselect young animals and also parents with higher accuracy than with BLUP (Patry and Ducrocq, 2011a; Tyrisevä et al., 2018b). However, the actual gains with genomic selection depend on a number of factors, aside from the genetic parameters. These include the choice of animals for genotyping, quality of methods for genomic prediction, and fraction of genotyped animals used for breed improvement. Genotyping is not effective if only parents with large number of progenies are genotyped because their BLUP evaluations are already accurate. A genomic selection scheme using simple single-trait models, possibly with few phenotypes, may be less accurate than BLUP selection with more complete data and models (Muir, 2007). Finally, if genotyping is used only for marketing, e.g., young bull sales to commercial farms, such genotyping has no effect on the genetic improvement of the breeding population.

With a large investment in genomic selection, it is of interest to find out the onset of the genomic selection and whether it is successful over the long run. There are several possible ways to find out the start date of genomic selection. One way to investigate the onset of genomic selection is by analyzing differences in genetic trends by BLUP and single-step genomic BLUP (ssGBLUP). Under genomic selection, BLUP cannot account for the fact that animals are being selected based on genomic information before having their phenotypes recorded (i.e., genomic preselection) and is therefore biased (Party and Ducrocq, 2009; Patry and Ducrocq, 2011b). On the contrary, ssGBLUP accounts for all sources of information jointly and is expected to be less affected by preselection bias (Legarra et al., 2009; VanRaden and Wright, 2013; Legarra et al., 2014). Superior

79 genetic trends by ssGBLUP compared to BLUP have been reported in several cases. Masuda et al.
80 (2018) presented trends for milk yield in Holsteins by BLUP and ssGBLUP. While the trend by
81 ssGBLUP increased at the expected beginning of the genomic selection, the trend by BLUP leveled
82 off. Koivula et al. (2018) reported that including the genotypes of culled bull calves in the
83 ssGBLUP analysis leads to higher genetic trends for milk production traits of Nordic Red Dairy
84 Cattle compared to the situation where genomic information of the culled bull calves is ignored.

85 Another way to investigate the onset of genomic selection is by analyzing genetic and phenotypic
86 trends, expecting accelerating trends under genomic selection (Misztal et al., 2020). However, both
87 trends are affected by changes in selection policies and incur some lag time. Additionally, changes
88 in genetic parameters over time (Hidalgo et al., 2020) may cause fluctuations in the genetic trend.
89 The third way is by analyzing realized Mendelian sampling (RMS) trends derived by genomic and
90 traditional evaluations (Tyrisevä et al., 2018a; Tyrisevä et al., 2018b). Genetic selection works by
91 selecting animals with superior Mendelian sampling. The selection is based on phenotypes and
92 progeny records in BLUP, and additionally on genomic information with genomic methods
93 (Lourenco et al., 2020). When some animals are selected for superior Mendelian sampling, the
94 average Mendelian sampling for all the animals is still zero, but for the selected animals is different
95 than zero. Therefore, RMS for genotyped animals is likely to be different than zero with selective
96 genotyping based on performance for both BLUP and ssGBLUP. Additionally, RMS is likely to
97 be zero for both methods when genotyping involves all young animals or is random. However, the
98 magnitude of RMS by ssGBLUP will be bigger because of the higher accuracy of genomic EBV
99 (GEBV). Not only the accuracy is higher, but the average GEBV is usually greater than the average
100 EBV, which translates into superior genetic trends. This study aimed to find the onset of genomic

101 selection by comparing the genetic and Mendelian sampling trends derived by ssGBLUP versus
102 BLUP in pigs, Angus cattle, and broiler chickens.

Materials and Methods

Pig data

The pig data consisted of 934,148 records for average daily gain (ADG) and 856,546 for Backfat (BF) collected until 2019, and 1,310,240 animals in pedigree, of which 117,091 were genotyped for 43,910 SNP markers after quality control. This dataset was provided by Genus PIC (Hendersonville, TN). The descriptive statistics of studied traits can be seen in Table 1.

American Angus data

Genotypes, pedigree, and phenotypes for three traits including birth weight (BTW, N=9,003,125), weaning weight (WW, N=9,506,570) and post weaning gain (PWG, N=4,671,702) of Angus beef cattle were provided by the American Angus Association (St. Joseph, MO). The pedigree consisted of 11,573,108 animals, of which 842,199 were genotyped for 39,766 SNP markers. The quality control of genotypes was conducted as in Lourenco et al. (2015b). The descriptive statistics of studied traits in American Angus can be seen in Table 2.

Broiler chicken data

The broiler chicken data were provided by Cobb-Vantress Inc. (Siloam Springs, AR). The dataset comprised phenotypes records on a purebred broiler chickens across 32 breeding cycles for three production traits referred as T1, T2 and T3. Each eight breeding cycles comprise one generation. The number of records for T1, T2 and T3 was 1,072,854, 228,992 and 265,891, respectively. The genotype file consisted of 154,318 birds genotyped for 54,713 SNP markers, and the pedigree consisted of 1,252,619 birds. The SNP data underwent quality control process as described in Lourenco et al. (2015a).

Statistical models

The statistical model for broiler chicken traits was as in Lourenco et al. (2015a), for pig traits was as in Steyn et al. (2020) and for beef traits was as in Garcia et al. (2020). The (co)variance components used in all analyses were provided by Angus Genetics Inc., PIC, and Cobb-Vantress. Both BLUP and ssGBLUP were run in a multiple-trait animal model framework. The pedigree relationship matrix (**A**) was used in BLUP and the realized relationship matrix (**H**) was used in ssGBLUP. The structure of $\mathbf{H^{-1}}$ is explained in Misztal et al. (2009) and Aguilar et al. (2010).

Genomic analysis and software

Because of the large number of genotyped animals, the algorithm for proven and young (APY) was used to create the inverse of **G** ($\mathbf{G}_{\mathrm{APY}}^{-1}$) as proposed by Misztal et al. (2014a) and Fragomeni et al. (2015). In APY, the matrix of genomic relationships among genotyped animals is partitioned based on core and noncore animals. The number of core individuals was selected based on the number of eigenvalues explaining 98% of the variance of **G** (Pocrnic et al., 2016) using PREGSF90 (Misztal et al., 2014b). The number of core individuals for broiler chickens, pigs, and beef cattle was estimated as 5030, 11,094, and 13,000, respectively.

Solutions for BLUP and ssGBLUP were obtained by using the preconditioned conjugate gradient algorithm with iteration on data as implemented in the BLUP90IOD2 (Tsuruta et al., 2001). The convergence criterion was set to 10^{-12} for all evaluations.

Criteria to investigate the starting point of Genomic preselection

Genetic trends: The point of divergence in genetic trends obtained by ssGBLUP and BLUP were used as a way to identify the onset of genomic selection. To explain how the difference between

predictions from ssGBLUP and BLUP can indicate the start of genomic selection, consider the decomposition of the (genomic) estimated breeding values ((G)EBV) of individual *i* as in Aguilar et al. (2010), VanRaden and Wright (2013), and Lourenco et al. (2015a):

$$EBV = w_1^c PA^c + w_2^c YD^c + w_3^c PC^c \quad (1)$$

and

$$GEBV = w_1^g PA^g + w_2^g YD^g + w_3^g PC^g + w_4 GI \quad (2)$$

Then, the difference between GEBV and EBV is:

$$GEBV - EBV = \left(w_1^g PA^g + w_2^g YD^g + w_3^g PC^g + w_4 GI\right) - \left(w_1^c PA^c + w_2^c YD^c + w_3^c PC^c\right) =$$

$$\left(w_1^g PA^g - w_1^c PA^c\right) + \left(w_2^g YD^g - w_2^c YD^c\right) + \left(w_3^g PC^g - w_3^c PC^c\right) + w_4 GI \quad (3)$$

where PA is the parent average, YD is yield deviation (phenotypes adjusted for the fixed effects), PC is the progeny contribution, and GI is the genomic information which is equal to GP-PP, in which GP is the genomic prediction derived using **G** and PP is the pedigree prediction derived using $\mathbf{A}_{22}$; the superscripts c and g denote components related to conventional BLUP and ssGBLUP, respectively, and w_1 *to* w_4 are weights that sum to 1.

When inbreeding is ignored in **A** and both parents are known, then, $w_1 = 2/den$, $w_2 = (n_{rec}/\alpha)/den$, $w_3 = 0.5n_{prog}/den$, and $w_4 = (g^{ii} - a_{22}^{ii})/den$, in which α is the variance ratio (residual variance over additive genetic variance), n_{prog} is the progeny size, n_{rec} is the number of records, $g^{ii}(a^{ii})$ is the diagonal element of $\mathbf{G}^{-1}(\mathbf{A}_{22}^{-1})$ for animal *i*, *den* is the sum of the numerators of w_1 to w_4.

The components of (G)EBV equations for individual *i* are as following:

$$PA_i = \left((G)EBV_{s(i)} + (G)EBV_{d(i)}\right)/2;$$

$$GI_i = \left(-\sum_{j,j\neq i}(g^{ij}/g^{ii} - a^{ij}/a^{ii})\, GEBV_j\right);$$

$$YD_i = (y_i - \sum_j x_{ij}\, \widehat{\boldsymbol{b}});$$

$$PC_i = \sum_k (2(G)EBV_k - (G)EBV_m)\, /n_{prog};$$

Where $(G)EBV_{s(i)}$ and $(G)EBV_{d(i)}$ are (genomic) breeding values of sire and dam of individual *i*, y_i is the *i*th record of animal *i*, $\widehat{\boldsymbol{b}}$ is the solutions for the level of fixed effects related to record *i*, x_{ij} is element of a design matrix relating $\widehat{\boldsymbol{b}}$ to y_i, and *k* refers to progeny and *m* indicates mate of animal *i*.

The components GP and PP are ignored under BLUP, which results in biased EBV if animals are selected based on genomic information. The bias arises not only from the lack of GP and PP, but from a combination of elements including the fact that PA, PC, and YD are not adjusted based on genomic information. For instance, if parents are non-genotyped, the difference between the predictions from BLUP and ssGBLUP originates from the contributions due to PC and GI of genotyped animals. For young animals without own and progeny records, the difference between EBV and GEBV comes from GI and PA enhanced by genomic information of parents, the latter to a smaller extent. However, as own and progeny records are added to the data, the amount of weight given especially to PC increases, and the weight of GI decreases.

When EBV or GEBV are used for selection of parents, GEBVs have higher accuracy ($r^g_{a,\hat{a}}$). This will generate a difference in amount of genetic gain (ΔG) in the next generation. Therefore, it can be shown as $\Delta G^g = ir^g_{a,\hat{a}}\sigma_a$ and $\Delta G^c = ir^c_{a,\hat{a}}\sigma_a$, and finally $\Delta G^g \geq \Delta G^c$, in which *i* is the selection intensity and σ_a is the additive genetic standard deviation. Hence, under genomic selection,

GEBVs are higher than EBV because greater accuracy of GEBV allows the selection of superior animals based on GP. Subsequently, a divergence in (G)EBV trends indicates the beginning of the genomic selection.

To obtain the genetic trend under traditional BLUP and ssGBLUP, the (G)EBVs were averaged by birth year for genotyped bulls in the beef cattle population and all genotyped individuals in the pig and chicken populations. Only animals with phenotypes were used for deriving the genetic trends. Genetic trends were obtained using a simple linear regression of (G)EBV for each trait on year of birth. For both BLUP and ssGBLUP, the genetic base was set to where more than one thousand genotyped individuals were available per year/generation. This corresponded to breeding cycle 1 in broiler chickens, and birth year 2012 in pigs and, 2007 in beef cattle. The mean GEBV from ssGBLUP was set to the same base as EBV from BLUP.

Realized Mendelian Sampling (RMS): The RMS for the genotyped individual *i* was estimated as:

$$RMS_i = (G)EBV_i - PA_i \quad (4)$$

Under some idealized evolutionary process (e.g., random mating, absence of selection, and large population size) all components are expected to be zero for the same generation.

$$E[PA] = E[YD] = E[PC] = E[GP] = E[PP] = 0$$

and consequently E[RMS]=0. When all or a random subset of young animals are used as parents of the next generation, the average RMS is close to 0. However, in the population under selection the equalities may not hold; therefore, $E(RMS) \neq 0$.

For simplicity, assume that parents and earlier generations are not genotyped. Let index *s* denotes ungenotyped animals selected for genotyping based on phenotype or BLUP (the first stage of selection), then E[YD]=δ, where $\delta = i_s r_{a,\hat{a}_s} \sigma_a$, in which i_s is the selection intensity at the first stage of selection, $r_{a,\hat{a}_s}$ is the accuracy of evaluation based on phenotype or BLUP and σ_a is the additive genetic standard deviation. Assuming young animals with neither progeny nor genotype:

$$E[(G)EBV_s] = E[w_1 PA_s + w_2 YD_s] = w_1 PA_s + w_2 \delta; \text{ with } E(RMS) = w_2 \delta \quad (5)$$

Therefore, if animals are preselected based on phenotype or BLUP, RMS from either BLUP or ssGBLUP is nonzero. Its value depends not only on the selection differential but also on the coefficient w_2, which is a function of variance ratio and the number of records.

Now assume that in the second stage of selection, the animals preselected based on phenotypes or BLUP are genotyped and reevaluated (index *sg*). On average, an animal with superior phenotype may also have a superior genomic prediction, E[GP]=τ, where $\tau = i_{sg}\sqrt{r^2_{a,\hat{a}_{sg}} - r^2_{a,\hat{a}_s}}\ \sigma_a$, with i_{sg} selection intensity in the second stage of selection and $r^2_{a,\hat{a}_{sg}}$ is the reliability of selection based on the genomic reevaluation. Then,

$$E[GEBV_{sg}] = E[w_1 PA + w_2 YD + w_4 GI] = w_1 PA + w_2 \delta + w_4 \tau, \quad E[RMS] = w_2 \delta + w_4 \tau \quad (6)$$

With many genotyped animals, the coefficient w_4 can be close to 1, with accuracy of $GEBV_{sg}$ greater than the one of EBV_s. Accordingly, RMS will be greater under genomic selection. The selective genotyping based on superior phenotypes (YD) can be replaced by superior progeny difference (PC) indicating that both have a similar effect on EBV, GEBV, and RMS.

The above derivations suggest that the RMS is close to zero when all animals are genotyped or when genotyping is at random. With selective genotyping, RMS is nonzero and is greater with

ssGBLUP than with BLUP. Because selective genotyping is the practice in livestock populations, the divergence in RMS trends obtained based on EBV and GEBV can also indicate the start point of the genomic selection. The same animals which were used for obtaining the genetic trends, were engaged in attaining the RMS trends.

Results

1) Pig production traits

Figure 1 shows the genetic trends for ADG and BF in genotyped pigs. The annual changes in average breeding values in genetic standard deviation units from 2012 to 2019 for ADG and BF were 0.27 and 0.04 for ssGBLUP and 0.18 and 0.02 for BLUP, respectively. The trends from ssGBLUP and BLUP diverged after 2013. In the last year of data (2019), the differences between average breeding values from ssGBLUP and BLUP were 0.67 SD for ADG and 0.17 SD for BF.

The genetic trend for ADG increased over time with a slightly increase in BF observed in recent years. The change in the genetic trend for BF was possibly due to the correlated response with body weight traits, as well as changes in breeding practices and in the selection objective in recent years.

The RMS (Figure 2) for ADG increased from around 0.04 in 2012, reached a peak of 0.10 in 2016, then declined. Relatively large RMS suggests preselection on a correlated trait -before genotyping. Smaller RMS for BF could be due to a correlated response to ADG.

2) Beef production traits

The genetic trends achieved by BLUP and ssGBLUP for BTW, WW, and PWG in genotyped Angus bulls are shown in Figure 3. The annual changes in (G)EBV for genotyped animals, in genetic standard deviation units, from 2006 to 2018 for BTW, WW, and PWG were -0.01, 0.11, and 0.08 for ssGBLUP and -0.01, 0.09, and 0.09 for BLUP, respectively. In the last year of data (2018), the differences between average breeding values from ssGBLUP and BLUP were 0.01, 0.23, and 0.06 SD for the three traits, respectively.

For BTW, the difference between the genetic trends for ssGBLUP and BLUP was negligible, but for WW and PWG genetic trends diverged considerably from 2016 afterward. For WW and PWG, the annual genetic gain after 2016 from ssGBLUP was 0.06 and 0.02 SD greater than BLUP, respectively. As it can be seen in Figure 3, there is a genetic improvement for all traits. However, genetic trend of BTW is downward relative to WW and PWG. Low BTW is desirable to avoid calving problems. On the other hand, BTW is positively correlated with WW and PWG; therefore, a stronger pressure is needed to keep BTW low while increasing WW and PWG. Based on the divergence, genomic selection is less important for BTW because this trait has already been recorded at the time of genotyping. Therefore, selection for BTW is based on parent average, phenotype deviation, and genomic prediction. Differently, there was a clear impact of genomic selection for WW from 2009−with an accelerated trend in 2017, and the genomic selection on PWG is slightly visible from 2017.

The RMS (Figure 4) looks very different for the 3 traits. For BTW, the trend is small and negative, at around -0.02, with small changes at the end. It suggests that the heaviest calves were not genotyped; calves are selected for lower BTW to reduce calving difficulty. For WW, RMS is large and increasing over time from 0.12 to 0.29. Such a trend suggests that the primary genotyping is after weaning and based on WW. For PWG, RMS is smaller although rising to 0.17. As the differences between EBV and GEBV were small for PWG, the values of RMS for PWG could be just a correlated response to WW as the genetic correlation between WW and PWG is high.

3) Broiler chicken traits

Trends were favorable for all traits with faster improvement in recent years. Figure 5 shows the difference between genetic trends obtained using ssGBLUP and BLUP in genetic standard deviation units for T1, T2, and T3 in genotyped birds. Divergence for the genetic trends by

ssGBLUP and BLUP occurred in breeding cycle 6 for T2 and T3. For T1, some divergence was visible from breeding cycle 2 to 16 in favor of BLUP and then from breeding cycle 20 afterwards in favor of ssGBLUP, although the divergence was reduced later. It seems that for T1 slight divergence in favor of BLUP up to breeding cycle 19 was spurious, and this divergence could represent low genomic merit of animals selected for genotyping.

The RMS trends (Figure 6) show relatively large values for T1 (up to 0.14) and small values for the other traits (0.04 or less). Animals were selected for T1 by BLUP, then superior animals were genotyped. Therefore, RMS for T1 is high. Small RMS for the other two traits measured later suggests only a correlated response from T1 because all animals measured for these traits were already genotyped.

Discussion

History of adoption of genomic selection

In this study, we used data provided by PIC, American Angus Association and Cobb-Vantress. Although each of them took different approaches when implementing genomic selection and genotyping became available, all changed to ssGBLUP after some time which corresponds to breeding cycle 6 in broiler chickens, year 2014 in pigs and year 2013 in beef cattle.

PIC started using ssGBLUP for genomic evaluations in this population in late 2013, so the first results of genomic selection were visible in 2014. Before that, selection was based on BLUP (William Herring, PIC, Hendersonville, TN, personal communication).

Angus Genetics Inc. incorporated genomic information on 15 markers in 2009 using a correlated trait approach (Kachman, 2008). The panel was updated to 384 markers in 2010 and moved to the 50k SNP chip after that. Finally, ssGBLUP was implemented for Angus cattle evaluations in 2017 (Kelli Retallick, Angus Genetics Inc., St. Joseph, MO, personal communication).

Genetic trends

We assessed the genetic trends of several traits in broiler chickens, pigs, and beef cattle to investigate the effectiveness of genomic selection. Assuming those differences in genetic basis between BLUP and ssGBLUP are correctly accounted for by the method described in Vitezica et al. (2011), the effectiveness of genomic selection can be evaluated indirectly by measuring the differences between genetic trends from BLUP and ssGBLUP. If the genetic trend by ssGBLUP is accelerating in a favorable direction and the genetic trend by BLUP is decelerating, genomic selection is likely practiced for the particular trait. If the genetic trends by both methods converge to the same point, the selection based on genotypes is not stronger than the selection based on parent average and phenotypes. The genetic trends can also be influenced by the genetic correlations among traits, especially with sequential selection, where a trend for an earlier measured trait influence a trait measured later. Based on the divergence point of genetic trends from BLUP and ssGBLUP in our study, the starting point of genomic selection in Angus cattle is 2013, in pigs is 2014, and in broiler chickens is breeding cycle 6. These starting points agree with the history of implementation of genomic selection in those populations.

If the genetic evaluations are based on ssGBLUP or GBLUP (**H** or **G** matrix), the estimates of genetic trends using BLUP (**A** matrix) are biased provided that a large portion of selected candidates are genotyped. As the correlation between the elements of **G** and $\mathbf{A}_{22}$ increases, the genetic trends by two methods will converge. However, some factors such as preselection of

320 selection candidates (Jibrila et al., 2020), incomplete pedigree information, and also the existence
321 of young animals without own and progeny records but with genotypic information (Shabalina et
322 al., 2017) makes this difference larger.

323 The main purpose in investigating genetic trends is to verify whether selection is effective and
324 whether there is an agreement with phenotypic trends. A disagreement suggests changes in the
325 environment, ineffective selection, or biased genetic trends. When there is a disagreement between
326 BLUP and phenotypic trends, but an agreement between the latter and ssGBLUP trends, there is
327 strong evidence for biased BLUP trends. Masuda et al. (2018) showed genetic trends for milk yield
328 traits based on BLUP were biased downwards for US Holstein bulls and cows. Especially for bulls,
329 the bias in EBV was because of failure in accounting for genomic preselection and underestimated
330 PC because daughters were also genotyped, and therefore, preselected before having their
331 phenotypes recorded. In the same study, the authors showed a good agreement between phenotypic
332 and ssGBLUP, meaning the latter can account for preselection and is not biased under genomic
333 selection.

334 Therefore, when the BLUP trends become biased, it means selection based on genomic
335 information became effective and BLUP EBV−or any measure derived from it, as deregressed
336 proofs−should not be used anymore. It should be noted that not only genomic preselection can
337 cause bias in BLUP evaluations, but also selection on correlated traits (Sorensen and Kennedy,
338 1984), poorly-defined unknown-parent groups (Misztal et al., 2013), preferential treatments of
339 selection candidates (Dehnavi et al., 2018) and non-random mating (Tsuruta et al., 2020) can
340 generate bias in BLUP.

341 **Realized Mendelian sampling**

342 The value and trends for RMS illustrate selective genotyping, where the decision to genotype is
343 based on phenotypes or BLUP evaluations. RMS was large for T1 in broiler chickens, for WW in
344 Angus, and for ADG in pigs where genotyping followed phenotyping. That RMS trend indicates
345 that an increasing number of piglets are being genotyped, reducing selective genotyping. As
346 genotyping becomes less expensive while the cost of phenotyping keeps constant, genotyping
347 more young animals becomes economically justified. For broiler chickens, RMS for later traits as
348 T2 and T3 was close to zero, indicating no new preselected genotyping based on these traits.

349 Although we investigated RMS and genetic trends to identify the starting point of genomic
350 selection, those two approaches are closely related. As genomic selection works by selecting
351 animals with superior Mendelian sampling, there is a sharp increase in breeding values estimated
352 under genomic methods. This increase in breeding values is evident for selected animals and also
353 their progeny (Tyrisevä et al., 2018a), where animals with large number of genotyped progenies
354 are more likely to have greater Mendelian sampling (Masuda et al., 2018). Consequently, because
355 of larger Mendelian sampling, there is an impact in genetic trends when animals are selected based
356 on genomic information, especially if the selection happens before phenotypes are recorded.

357

358 ## Conclusions

359 To detect the effective starting point of genomic selection, two possible ways included
360 divergence point of genetic trends and RMS trends obtained by ssGBLUP and BLUP using official
361 datasets from pigs, beef cattle, and broiler chickens were used. The effective starting point of
362 genomic selection in Angus cattle, pigs, and broiler chickens was determined as year 2013, 2014,
363 and breeding cycle 6, respectively. The difference between genetic and RMS trends from

ssGBLUP and BLUP is more obvious in a population under more intense selection, as in pigs and broilers compared to beef cattle. In general, the effective starting point of genomic selection can be detected by the divergence between genetic and RMS trends from BLUP and ssGBLUP, although RMS trends are present for traits recorded before genotyping and later used for genotyping decisions. The results and procedures presented here can help to evaluate the efficiency of the implementation of genomic selection in breeding programs.

Acknowledgements

This study was partially supported by Cobb-Vantress (Siloam Springs, AR), Pig Improvement Company (PIC; Hendersonville, TN), Angus Genetics Inc. (St. Joseph, MO), and by Agriculture and Food Research Initiative Competitive Grant no. 2020-67015-31030 from the US Department of Agriculture's National Institute of Food and Agriculture. We thank Vivan Breen, Rachel Hawken, Ching-Yi Chen, William Herring, Kelli Retallick, and Steve Miller for providing data access. The authors thank Andres Legarra for helpful comments.

Conflict of interest statement

The authors declare no real or perceived conflicts of interest.

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

385 evaluation of Holstein final score. J. Dairy Sci. 93(2):743-752. doi:
386 https://doi.org/10.3168/jds.2009-2730
387 Dehnavi, E., S. A. Mahyari, F. S. Schenkel, and M. Sargolzaei. 2018. The effect of using cow
388 genomic information on accuracy and bias of genomic breeding values in a simulated
389 Holstein dairy cattle population. J. Dairy Sci. 101(6):5166-5176. doi:
390 https://doi.org/10.3168/jds.2017-12999
391 Fragomeni, B. O., D. A. L. Lourenco, S. Tsuruta, Y. Masuda, I. Aguilar, A. Legarra, T. J. Lawlor,
392 and I. Misztal. 2015. Hot topic: Use of genomic recursions in single-step genomic best
393 linear unbiased predictor (BLUP) with a large number of genotypes. J. Dairy Sci.
394 98(6):4090-4094. doi: https://doi.org/10.3168/jds.2014-9125
395 Garcia, A. L. S., Y. Masuda, S. Tsuruta, S. Miller, I. Misztal, and D. Lourenco. 2020. Indirect
396 predictions with a large number of genotyped animals using the algorithm for proven and
397 young. J. Anim. Sci. 98(6):skaa154. doi: 10.1093/jas/skaa154
398 Hidalgo, J., S. Tsuruta, D. Lourenco, Y. Masuda, Y. Huang, K. A. Gray, and I. Misztal. 2020.
399 Changes in genetic parameters for fitness and growth traits in pigs under genomic selection.
400 J. Anim. Sci. 98(2)doi: 10.1093/jas/skaa032
401 Jibrila, I., J. ten Napel, J. Vandenplas, R. F. Veerkamp, and M. P. L. Calus. 2020. Investigating
402 the impact of preselection on subsequent single-step genomic BLUP evaluation of
403 preselected animals. Genet. Sel. Evol. 52(1):42. doi: 10.1186/s12711-020-00562-6
404 Koivula, M., I. Stranden, G. P. Aamand, and E. A. Mantysaari. 2018. Reducing bias in the dairy
405 cattle single-step genomic evaluation by ignoring bulls without progeny. J. Anim. Breed.
406 Genet. 135(2):107-115. doi: 10.1111/jbg.12318
407 Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and
408 genomic information. J. Dairy Sci. 92(9):4656-4663. doi: https://doi.org/10.3168/jds.2009-
409 2061
410 Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. Single Step, a general approach
411 for genomic selection. Livest. Sci. 166:54-65. doi:
412 https://doi.org/10.1016/j.livsci.2014.04.029
413 Lourenco, D., A. Legarra, S. Tsuruta, Y. Masuda, I. Aguilar, and I. Misztal. 2020. Single-Step
414 Genomic Evaluations from Theory to Practice: Using SNP Chips and Sequence Data in
415 BLUPF90. Genes 11(7):790. doi: 10.3390/genes11070790
416 Lourenco, D. A. L., B. O. Fragomeni, S. Tsuruta, I. Aguilar, B. Zumbach, R. J. Hawken, A.
417 Legarra, and I. Misztal. 2015a. Accuracy of estimated breeding values with genomic
418 information on males, females, or both: an example on broiler chicken. Genet. Sel. Evol.
419 47(1):56. doi: 10.1186/s12711-015-0137-1
420 Lourenco, D. A. L., S. Tsuruta, B. O. Fragomeni, Y. Masuda, I. Aguilar, A. Legarra, J. K. Bertrand,
421 T. S. Amen, L. Wang, D. W. Moser, and I. Misztal. 2015b. Genetic evaluation using single-
422 step genomic best linear unbiased predictor in American Angus1. J. Anim. Sci. 93(6):2653-
423 2662. doi: 10.2527/jas.2014-8836
424 Masuda, Y., P. M. VanRaden, I. Misztal, and T. J. Lawlor. 2018. Differing genetic trend estimates
425 from traditional and genomic evaluations of genotyped animals as evidence of preselection
426 bias in US Holsteins. J. Dairy Sci. 101(6):5194-5206. doi: 10.3168/jds.2017-13310
427 Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation
428 including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92(9):4648-
429 4655. doi: 10.3168/jds.2009-2064

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

Misztal, I., D. Lourenco, and A. Legarra. 2020. Current status of genomic evaluation. J. Anim. Sci. 98(4)doi: 10.1093/jas/skaa101

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. G. Vitezica. 2014b. Manual for BLUPF90 family of programs

Misztal, I., Z. G. Vitezica, A. Legarra, I. Aguilar, and A. A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. J. Anim. Breed. Genet. 130(4):252-258. doi: 10.1111/jbg.12025

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(6):342-355. doi: 10.1111/j.1439-0388.2007.00700.x

Party, C., and V. Ducrocq. 2009. Bias due to genomic selection. In: Proceedings of the Interbull technical workshop, Uppsala, Sweden

Patry, C., and V. Ducrocq. 2011a. Accounting for genomic pre-selection in national BLUP evaluations in dairy cattle. Genet. Sel. Evol. 43(1):30. doi: 10.1186/1297-9686-43-30

Patry, C., and V. Ducrocq. 2011b. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. J. Dairy Sci. 94(2):1011-1020. doi: 10.3168/jds.2010-3804

Pocrnic, I., D. A. L. Lourenco, Y. Masuda, A. Legarra, and I. Misztal. 2016. The Dimensionality of Genomic Information and Its Effect on Genomic Prediction. Genetics 203(1):573. doi: 10.1534/genetics.116.187013

Shabalina, T., E. C. G. Pimentel, C. Edel, L. Plieschke, R. Emmerling, and K. U. Götz. 2017. Short communication: The role of genotypes from animals without phenotypes in single-step genomic evaluations. J. Dairy Sci. 100(10):8277-8281. doi: 10.3168/jds.2017-12734

Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of Genetic Variances from Unselected and Selected Populations. J. Anim. Sci. 59(5):1213-1223. doi: 10.2527/jas1984.5951213x

Steyn, Y., D. A. Lourenco, C.-Y. Chen, B. D. Valente, J. Holl, W. O. Herring, and I. Misztal. 2020. Optimal definition of contemporary groups for crossbred pigs in a joint purebred and crossbred genetic evaluation. J. Anim. Sci. doi: 10.1093/jas/skaa396

Tsuruta, S., T. J. Lawlor, D. A. L. Lourenco, and I. Misztal. 2020. Bias in genomic predictions by mating practices for linear type traits in a large-scale genomic evaluation. J. Dairy Sci. doi: 10.3168/jds.2020-18668

Tsuruta, S., I. Misztal, and I. Strandén. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications1. J. Anim. Sci. 79(5):1166-1172. doi: 10.2527/2001.7951166x

Tyrisevä, A. M., W. F. Fikse, E. A. Mäntysaari, J. Jakobsen, G. P. Aamand, J. Dürr, and M. H. Lidauer. 2018a. Validation of consistency of Mendelian sampling variance. J. Dairy Sci. 101(3):2187-2198. doi: 10.3168/jds.2017-13255

Tyrisevä, A. M., E. A. Mäntysaari, J. Jakobsen, G. P. Aamand, J. Dürr, W. F. Fikse, and M. H. Lidauer. 2018b. Detection of evaluation bias caused by genomic preselection. J. Dairy Sci. 101(4):3155-3163. doi: 10.3168/jds.2017-13527

VanRaden, P. M. 2020. Symposium review: How to implement genomic selection. J. Dairy Sci. 103(6):5291-5301. doi: 10.3168/jds.2019-17684

VanRaden, P. M., and J. R. Wright. 2013. Measuring Genomic Pre-Selection in Theory and in Practice. Proceedings of the Interbull meeting 47

475 Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for
476 populations under selection. Genet. Res. 93(5):357-366. doi:
477 10.1017/S001667231100022X

478

479

Table 1. Descriptive statistics of pig data

Trait	no. Records	Mean	SD	no. Genotypes	no. Animals in Pedigree
ADG	934,148	696.86	97.45	116,943	1,310,240
BF	856,546	9.39	2.78	116,943	1,310,240

ADG: Average Daily Gain; BF: Backfat; SD: Standard Deviation

480

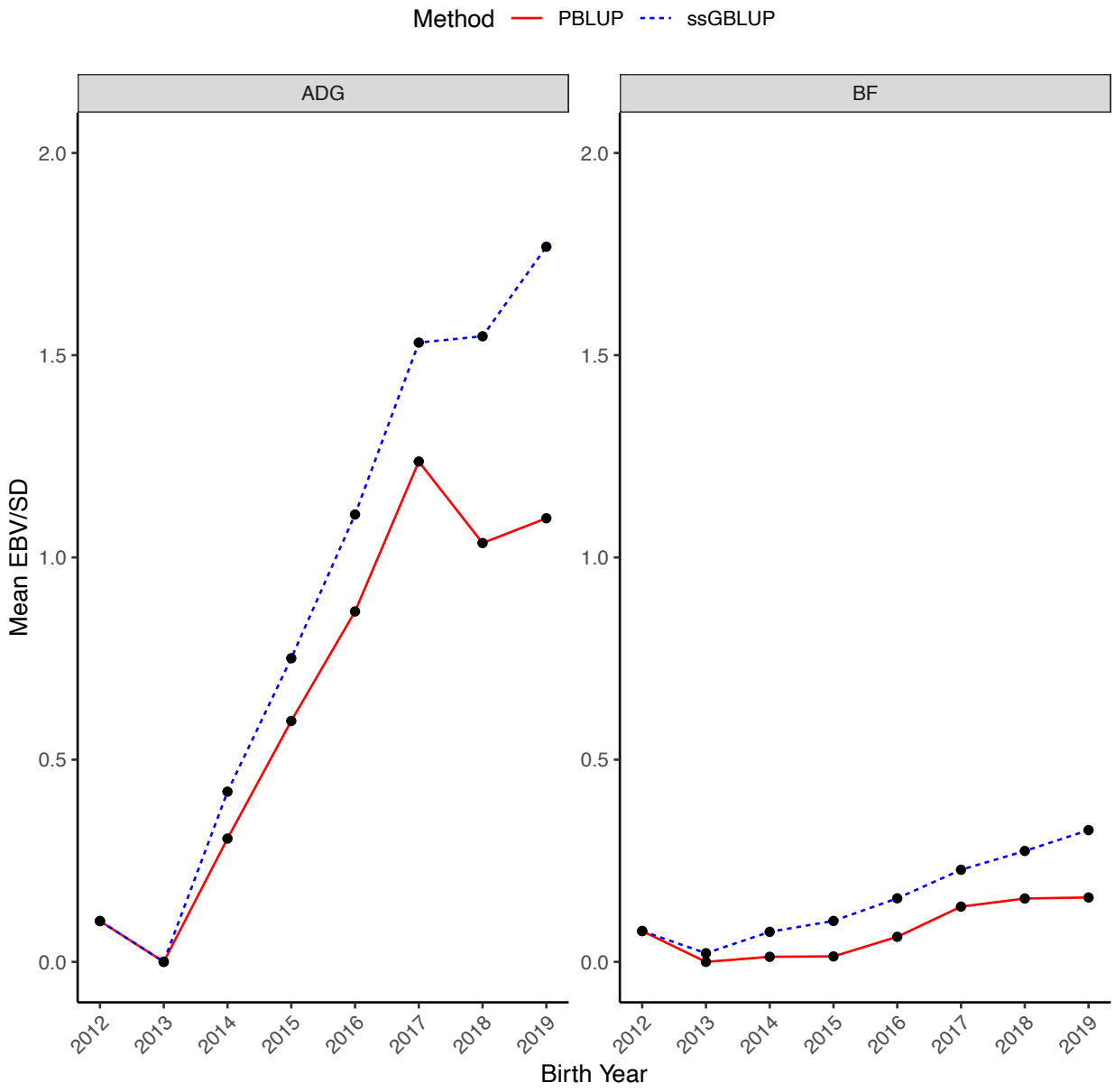
Table 2. Descriptive statistics of Angus data

Trait	no. Records	Mean	SD	no. Genotypes	no. Animals in Pedigree
BTW (lb)	9,003,125	80.57	9.87	842,199	11,573,108
WW (lb)	9,506,570	593.72	99.52	842,199	11,573,108
PWG (lb)	4,671,702	362.50	147.93	842,199	11,573,108

BTW: Birth Weight; WW: Weaning Weight; PWG: Post Weaning Gain
SD: Standard Deviation

481

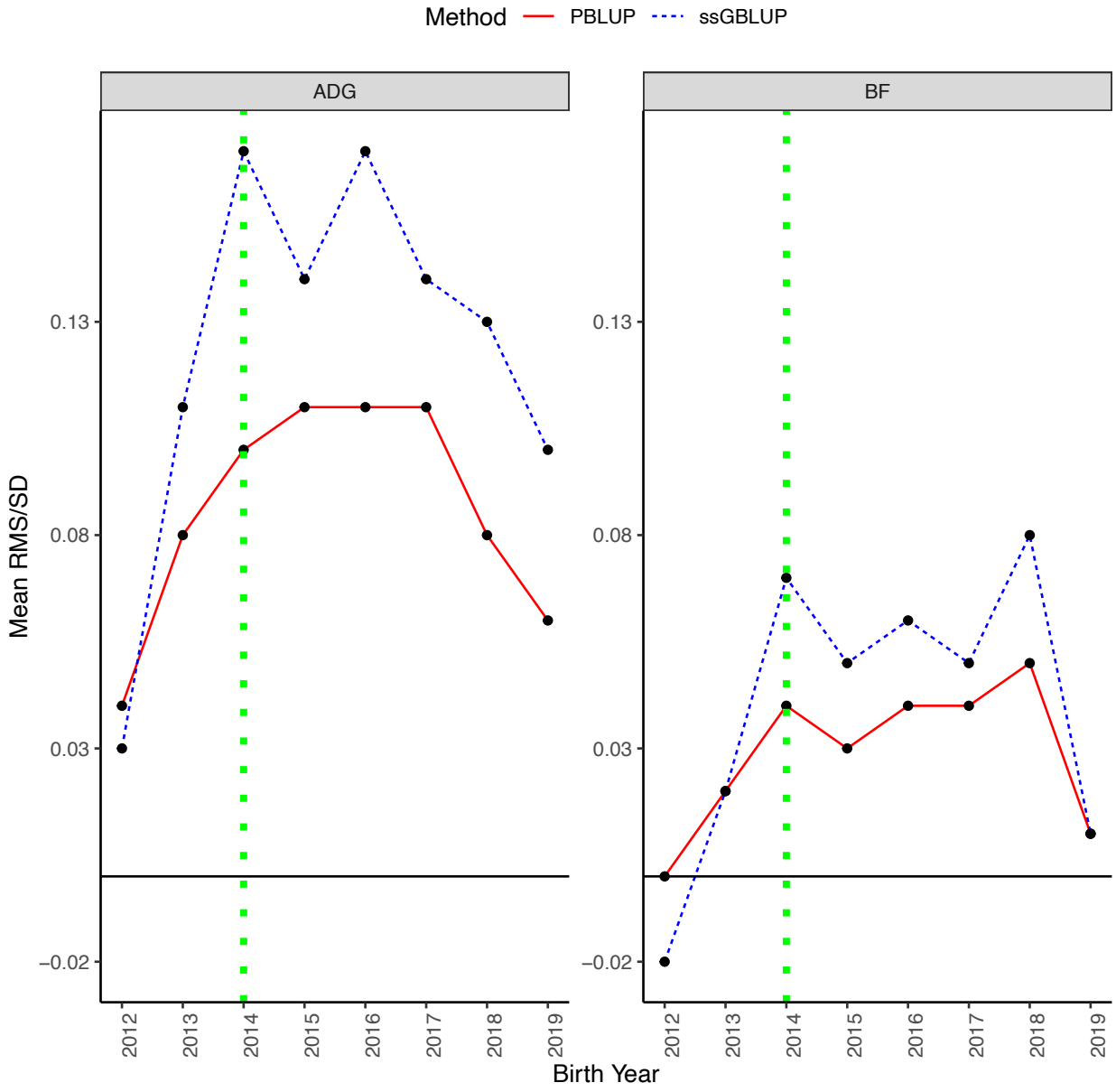
482 **Figure 1**. Genetic trends obtained using single-step GBLUP (ssGBLUP) and pedigree BLUP
483 (PBLUP) for average daily gain (ADG) and backfat (BF) in the genotyped pigs by year of birth.
484 Genetic trends are presented in additive genetic standard deviation scale and the genetic base is
485 adjusted to 2012.



486

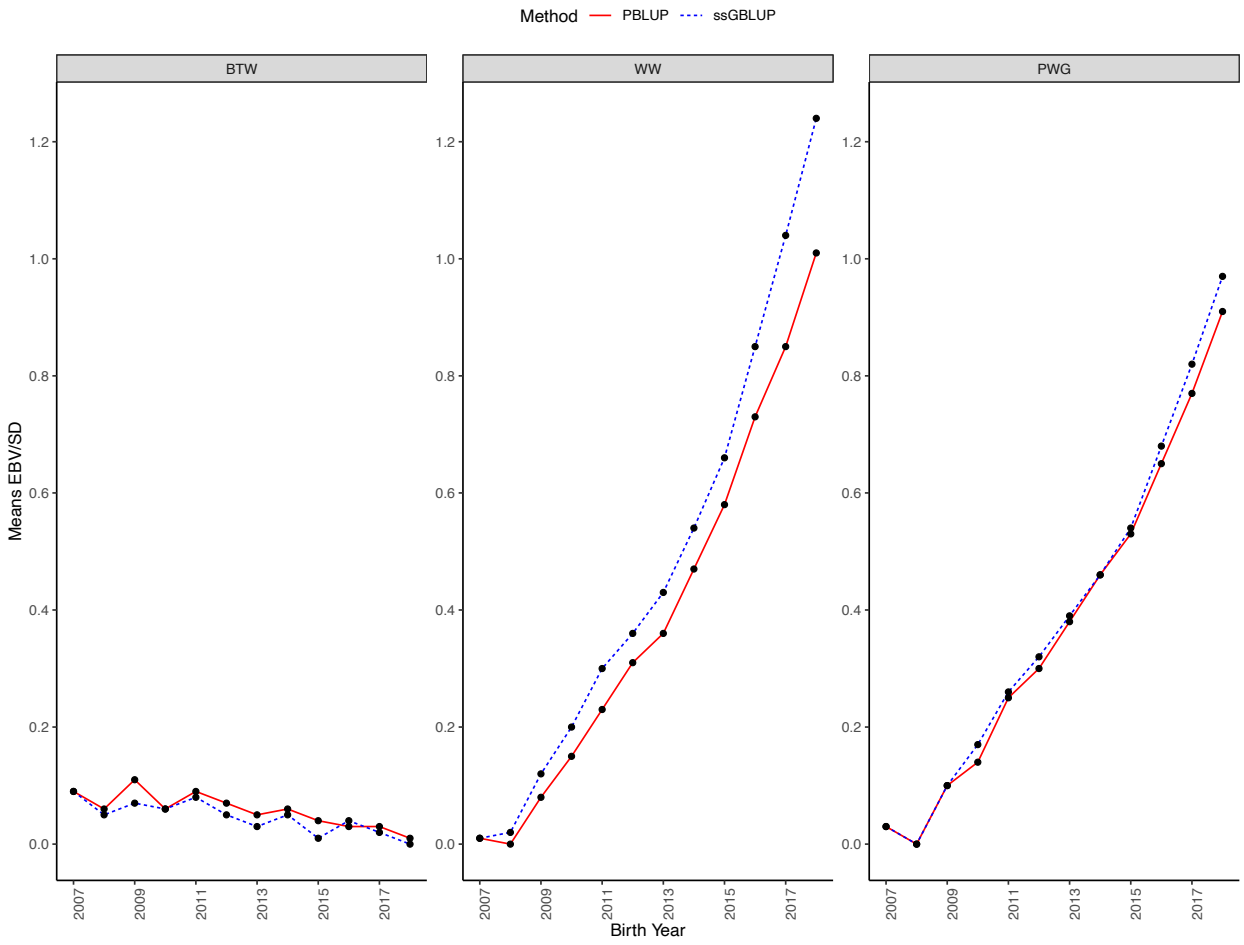
487

488 **Figure 2.** Realized Mendelian sampling (RMS) trends estimated by single-step GBLUP
489 (ssGBLUP) and pedigree BLUP (PBLUP) for average daily gain (ADG) and backfat (BF) in the
490 genotyped pigs. Mendelian sampling trends are presented in additive genetic standard deviation
491 scale. Solid black line represents the zero-base line and dotted green vertical line shows the start
492 date of genomic selection.



493
494

495 **Figure 3**. Genetic trends obtained using single-step GBLUP (ssGBLUP) and pedigree BLUP
496 (PBLUP) for birth weight (BTW), weaning weight (WW), and post weaning gain (PWG) in the
497 genotyped Angus bulls by year of birth. Genetic trends are presented in additive genetic standard
498 deviation scale and the genetic base is adjusted to 2007.



499

Figure 4. Realized Mendelian sampling (RMS) trends estimated by single-step GBLUP (ssGBLUP) and pedigree BLUP (PBLUP) for birth weight (BTW), weaning weight (WW), and post weaning gain (PWG) in the genotyped Angus bulls. Mendelian sampling trends are presented in additive genetic standard deviation scale. Solid black line represents the zero-base line and dotted green vertical line shows the start date of genomic selection.

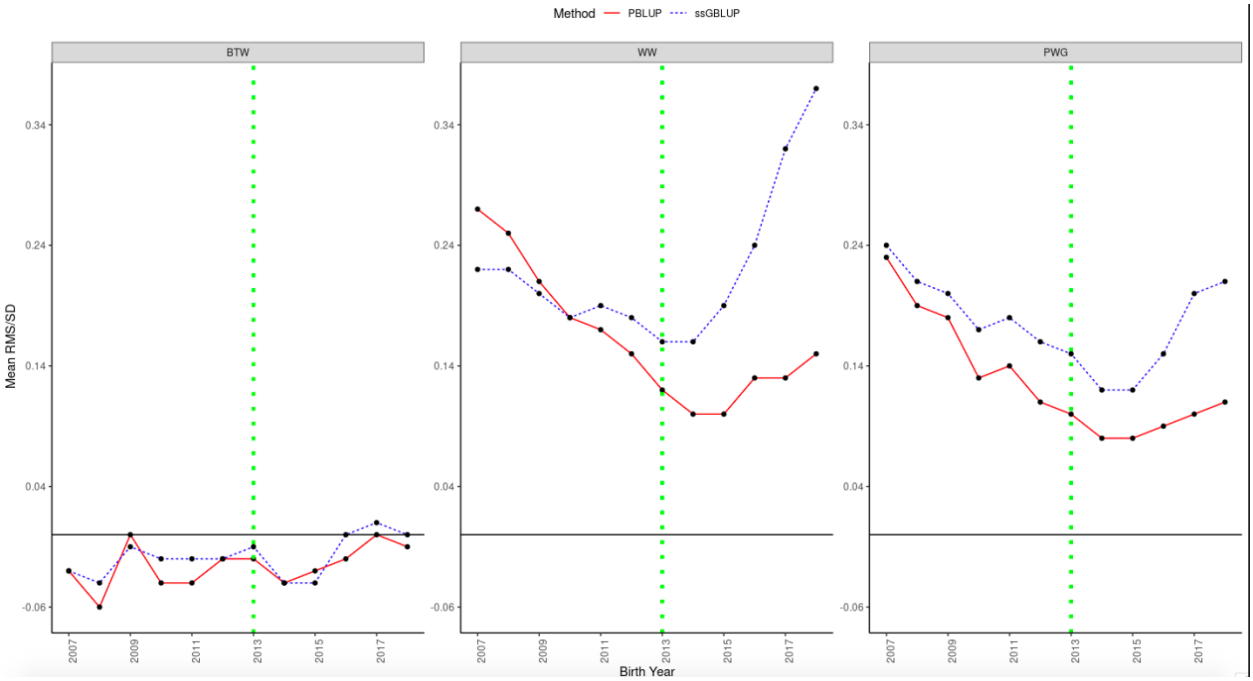
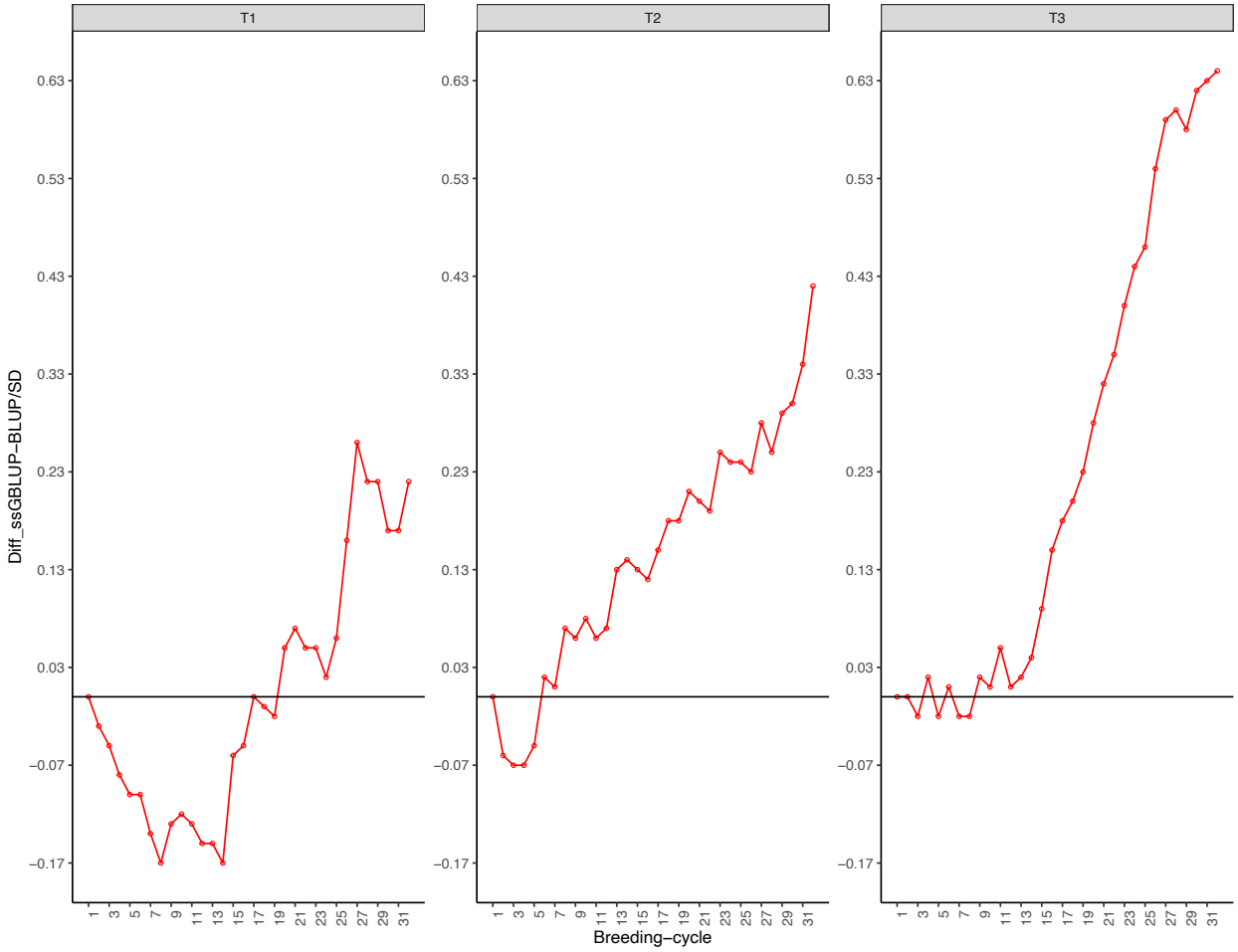
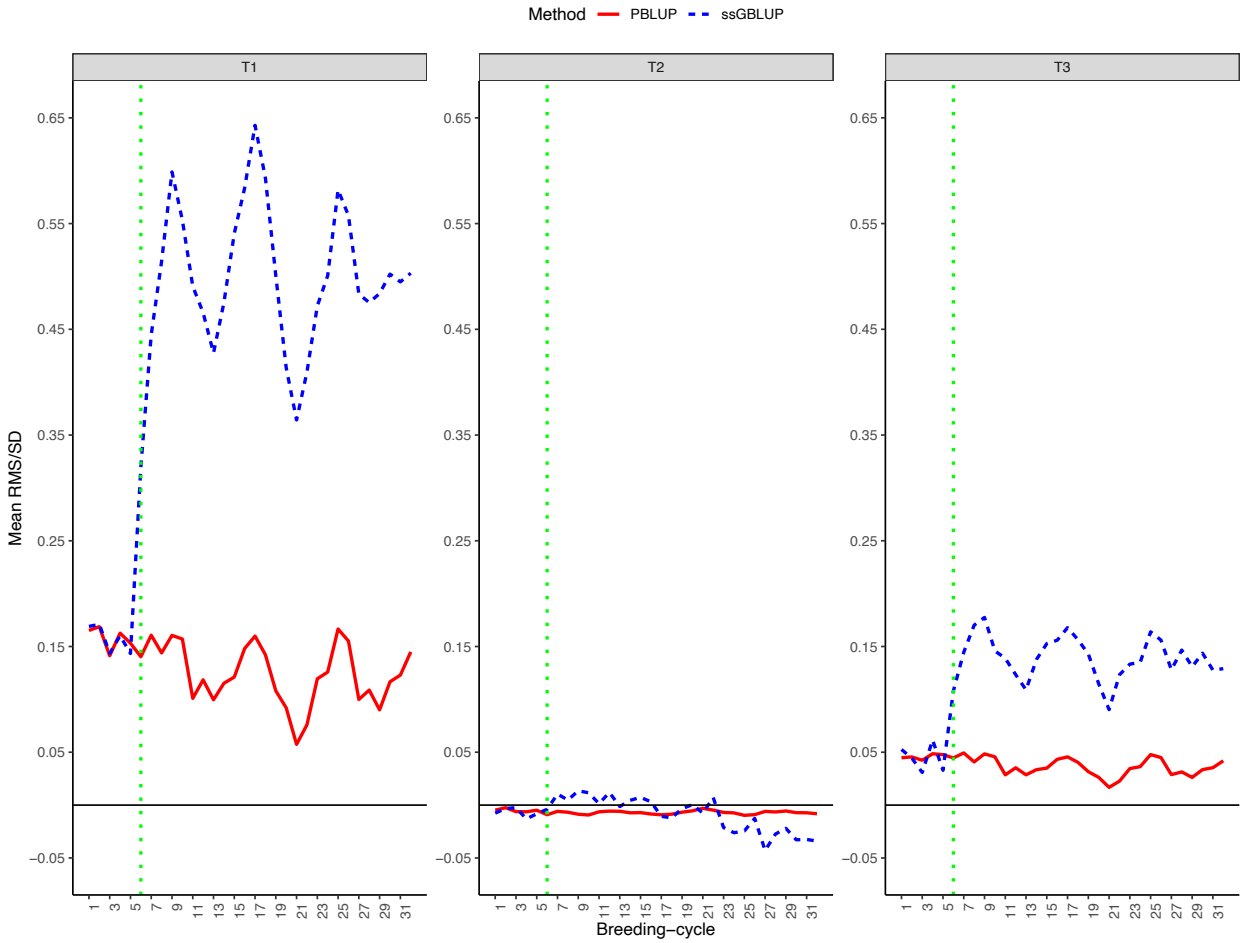


Figure 5. The difference between genetic trends obtained using single-step GBLUP (ssGBLUP) and pedigree BLUP (PBLUP) in genetic standard deviation units for three production traits referred as T1, T2, and T3 in a purebred broiler chicken line across 32 breeding cycles.



511 **Figure 6.** Realized Mendelian sampling (RMS) trends estimated by single-step GBLUP
512 (ssGBLUP) and pedigree BLUP (PBLUP) for three production traits referred as T1, T2, and T3
513 in a purebred broiler chicken line across 32 breeding cycles. Mendelian sampling trends are
514 presented in additive genetic standard deviation scale. Solid black line represents the zero-base
515 line and dotted green vertical line shows the start date of genomic selection.



516