Proc. Assoc. Advmt. Anim. Breed. Genet. 23:195-205

DEVELOPING GENOMIC STRATEGIES FOR THE LIVESTOCK INDUSTRIES: ALL IMPLEMENTATIONS ARE CHALLENGING

D.A.L. Lourenco, S. Tsuruta, Y. Masuda and I. Misztal

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

SUMMARY

Genomic selection (GS) using dense SNP panels was first implemented in 2009 for dairy cattle. Since then, GS has been extended to other livestock. However, different problems and challenges are always encountered during the implementation of GS in each population. In this paper, we show the issues in the implementation of GS and how they have been successfully solved in beef and dairy cattle, pigs, chicken, and fish. We also discuss changes in current methods and development of new algorithms to deal with large genomic data. Overall, complications for GS include, but are not limited to, selective genotyping, computing limitations, convergence problems especially for complex models, compatibility between pedigree and genomic information, among others.

BACKGROUND

Soller and Beckmann (1983) hypothesized, in the early 1980's, that DNA markers could be useful in constructing more precise genetic relationships, detecting causative variants, and determining parentage. After the first draft of the human genomic project was published in 2001 (The International SNP Map Working Group 2001), single nucleotide polymorphism (SNP) became the most important source of genome sequence variation, and therefore, the most important DNA marker. Concurrently, Meuwissen *et al*. (2001) anticipated that genomic information could help animal breeders to generate more accurate breeding values if a dense SNP assay that covered the entire genome could be constructed. It took almost 8 years for the first dense SNP assay to become available, and this was for dairy cattle (Matukumalli *et al*. 2009).

In January of 2009, researchers from AGIL-USDA released the first official genomic evaluation for Holstein and Jersey in the USA. This implementation brought a lot of excitement, especially because the top bull in the evaluation had no daughters with milking records, meaning his genetic merit was computed based on pedigree and genomic information. The superiority of this very bull was later confirmed when his progeny records became available. This endorsed the hypothesis that Meuwissen *et al*. (2001) had tested based on simulated data: the genetic merit of young animals can be computed with high accuracy if they are genotyped and SNP effects are available from a reference population.

With the release of genomic predictions based on dense SNP assays for Holsteins in the USA, the race for the implementation of genomic selection (GS) in livestock became official. Essentially, two main methods for genomic evaluation were developed: multi-step and single-step. The multi-step method was the first to be implemented (VanRaden 2008). The main advantage of the multi-step approach is that the traditional BLUP evaluation is kept unchanged and GS can be carried out by using additional analyses; however, only genotyped animals have genomic EBV (GEBV). As a result, several adjustments were proposed, especially in dairy cattle, to make EBV for non-genotyped animals comparable to GEBV under multi-step evaluations (Wiggans *et al*. 2011; Wiggans *et al*. 2012).

Intending to solve incompatibility problems and to reduce the burden in obtaining genomic predictions when only a fraction of animals is genotyped, Misztal *et al*. (2009) and Legarra *et al*. (2009) proposed a method that combines phenotypes, pedigree, and genotypes into a single evaluation. This method is called single-step genomic BLUP (ssGBLUP) and replaces the pedigree relationship matrix in the traditional BLUP by a realized relationship matrix (**H**), which combines pedigree and genomic relationships. Another class of single-step was also proposed by Fernando *et al*. (2014), which is based on a marker effect model and is called single-step Bayesian regression (ssBR). Under the same assumptions (e.g., all SNP have non-zero effect and constant variance), ssGBLUP and ssBR are equivalent models (Gao *et al*. 2018).

Over the past 4 or 5 years, ssGBLUP has become the preferred method for genomic evaluation in several species, namely beef cattle (Lourenco *et al*. 2015a), dairy cattle (Vukasinovic *et al*. 2017), pigs (Forni *et al*. 2011; Lourenco *et al*. 2016), broilers (Chen *et al.* 2011; Lourenco *et al*. 2015b), layers (Yan *et al*. 2018), dairy sheep and goat (Rupp *et al*. 2016), Australian sheep (Brown *et al*. 2018), and fish (Garcia *et al.* 2018). Possibly, in the near future, the great majority of genomic evaluations will all be based on single-step methods.

Although the idea and theory behind ssGBLUP are easily understandable, and the method seems to be simple because it just requires the change in the relationship matrix, its implementation for official genomic evaluations is quite challenging and demands several data-dependent adjustments. It is worthwhile to remember that even a small change to the genetic evaluation system can create issues. For example, a simple change in variance components can cause convergence problems and changes in scaling. Usually, the issues and challenges encountered during the change from traditional or multi-step evaluations to single-step are not disclosed. However, showing issues and strategies to solve them can help troubleshooting future implementations. In this paper, we show the problems in the implementation of GS and how they have been successfully solved in beef and dairy cattle, pigs, chicken, and fish.

GENOMIC STRATEGIES

Beef cattle. In 2009, Angus Genetics Inc. started to run multi-step genomic evaluations for American Angus Association (AAA) using a correlated approach described by Kachman (2008). In this approach, the trait phenotype and the direct genomic values (DGV) calculated based on SNP effects are used as phenotypic information in a 2-trait model, where heritability for DGV is assumed to be 0.99. Restricted maximum likelihood (REML) estimates are then obtained. Genetic correlations between each trait and DGV reflect accuracy of DGV, and solutions for the first trait are genomic-enhanced EBV. The drawback of this method is that it doubles the number of traits in the model. Additionally, genetic correlations between each trait and DGV can be overestimated, indicating the genomic information is explaining more of the genetic variance than expected, which can inflate predictions. Figure 1 shows genetic trends for marbling using traditional BLUP and the multi-step correlated approach. Inflated genetic trends for multistep predictions, together with big fluctuations for predictions every time SNP effects were recalculated (i.e., during calibration) and re-ranking of high accuracy bulls in subsequent evaluations urged Angus Genetics Inc. to find another method for their genomic evaluation.

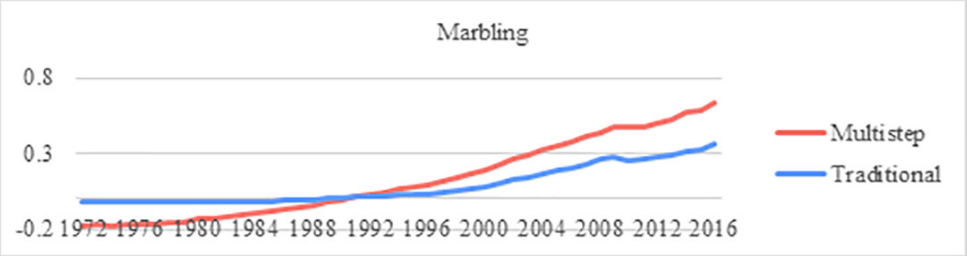

Figure 1. Genetic trends for marbling using traditional BLUP and the multi-step correlated approach in American Angus

Scaling factors and inbreeding. In 2014, we started testing ssGBLUP for growth and calving ease (CE) models in the American Angus population. Several datasets were used, but the first one was for growth traits and calving ease and comprised 8 million animals in the pedigree, along with 6 million birth (BW) and weaning weight (WW) records, 3.4 million records for post-weaning gain (PWG), 1.3M CE records, and genotypes for 52k animals. The first issue observed was the inflation of GEBV in a validation test. When adjusted phenotypes for animals considered young in 2013, but with phenotypes in 2014, were regressed on GEBV, regression coefficients were lower than 1. To solve this problem scaling parameters can be used in $\mathbf{H}^{-1}$ (Aguilar *et al*. 2010; Christensen and Lund 2010; Tsuruta *et al*. 2011):

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

where $\mathbf{A}^{-1}$ is the inverse of the pedigree relationship matrix, $\mathbf{G}^{-1}$ is the inverse of the genomic relationship matrix, and $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree relationship matrix among genotyped animals; $\boldsymbol{\tau}$ and $\boldsymbol{\omega}$ were used to rescale the amount of information in $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$, respectively. Primarily, ω controls inflation due to incompleteness of pedigree while τ controls additive genetic variance. Based on validation, we found the best combination of $\boldsymbol{\tau}$ and $\boldsymbol{\omega}$ for this data was 1.0 and 0.7. However, scaling parameters are completely *ad-hoc* and should be avoided generally. In the original implementation of ssGBLUP (Aguilar *et al*. 2010) inbreeding for $\mathbf{G}^{-1}$ and $\mathbf{A}_{22}^{-1}$ was considered, but not for $\mathbf{A}^{-1}$. After inbreeding was included in the computation of $\mathbf{A}^{-1}$, ***ω lower than 1*** was not needed anymore for this beef cattle data. Figure 2 shows coefficients for the regression of adjusted phenotypes on GEBV for BW, WW, PWG when inbreeding for $\mathbf{A}^{-1}$ is considered and $\boldsymbol{\omega}$ is lower than 1. It is clear that GEBV are deflated if $\boldsymbol{\omega}$ is lower than 1, proving that inbreeding in $\mathbf{A}^{-1}$ is enough to avoid inflation in this AAA dataset. Therefore, current official AAA evaluations use $\tau = \omega = 1$.

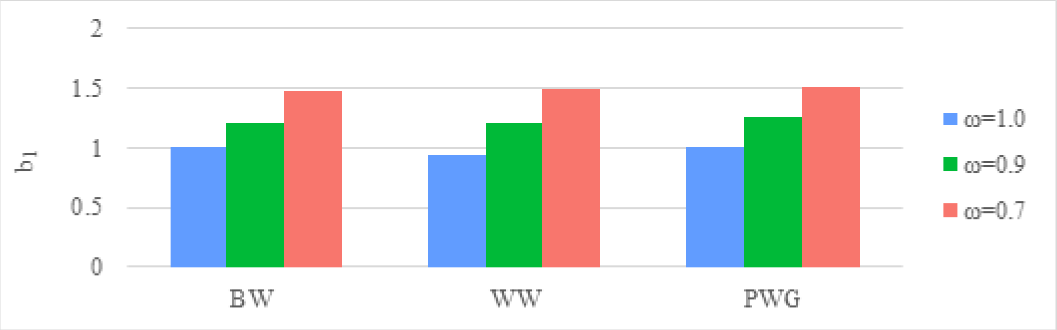

Figure 2. Regression coefficients (b_1) for birth weight (BW), weaning weight (WW) and post-weaning gain (PWG) with varying $\boldsymbol{\omega}$

Selective genotyping. Regarding the advantages of using genomic information, the average gain in predictive ability for growth traits, when moving from traditional BLUP to ssGBLUP, was 24%. In contrast, the gain in prediction accuracy for CE was only 8%, going from 0.12 to 0.13. This small increase in predictive ability is possibly because animals with difficult calving are unlikely to be retained for breeding and therefore would not be genotyped on a regular basis. In fact, only 0.35% of the animals with difficult calving were genotyped. Therefore, selective genotyping can compromise the gains that can be obtained with genomics and can also introduce some pre-selection bias into the evaluations. Similar problems occur for traits related to survival or other specific phenotypes (e.g., animals with undesirable phenotypes or dead). This means some traits might need a controlled reference set of animals, rather than relying on Industry genotyping strategies which are influenced by the perceived value of animals.

External information. For the growth model, another issue was related to the inclusion of external information in ssGBLUP. For traditional evaluations, the external EBV from Red Angus is used as prior information in the right hand side of the mixed model equations (MME), and the reliability is added to the pedigree relationships among external animals in the left hand side of the MME (Legarra *et al*. 2007). We changed the computing algorithm to support both genomic and external information, and the implementation of a genomic multi-breed model increased the computing time only by 2.5 hours, compared to a genomic single-breed model.

Many more genotyped animals. On a weekly basis, more than 2k genotyped animals are added to the AAA database. From July of 2014 to March of 2019, the number of genotyped animals increased from 82k to 627k. The most computationally expensive operation in ssGBLUP is the inversion of **G** and $\mathbf{A}_{22}$. This operation has an approximately cubic cost with the number of genotyped animals. With efficient computing algorithms, matrix inversions are feasible for up to 150,000 genotyped animals.

To overcome the limitation set by the number of genotyped animals in ssGBLUP, Misztal *et al.* (2014) proposed the algorithm for proven and young (APY) animals to construct $\mathbf{G}^{-1}$ without having to explicitly invert **G**. The logic behind the construction of $\mathbf{G}^{-1}_{\text{APY}}$ is that the genotyped animals are split into core (*c*) and noncore (*n*), and breeding values for noncore animals ($\mathbf{u}_n$) are functions of breeding values of core animals ($\mathbf{u}_c$):

$$\mathbf{u}_n = \mathbf{P}_{nc}\mathbf{u}_c + \mathbf{\Psi}_n$$

where $\mathbf{P}_{nc}$ is a matrix that relates breeding values for noncore to core animals, and $\mathbf{\Psi}_n$ is a diagonal matrix with estimation errors. The $\mathbf{G}^{-1}_{\text{APY}}$ can be constructed as:

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

with $\text{m}_{nn_{ii}} = \text{g}_{\text{ii}} - \text{g}_{\text{ic}}\text{G}^{-1}_{\text{cc}}\text{g}_{\text{ci}}$. The APY algorithm creates a generalized sparse inverse of **G** at approximately a linear cost in computing and storage. However, if $\mathbf{G}^{-1}_{\text{APY}}$ is efficiently computed but A^{-1}_{22} is not, ssGBLUP still cannot be used for over 150,000 genotyped animals. To avoid explicit inversion of A_{22}, Masuda *et al.* (2017) proposed to compute an efficient inverse indirectly as a product of sparse matrices:

$$\mathbf{A}^{-1}_{22} = \mathbf{A}^{22} - \mathbf{A}^{21}(\mathbf{A}^{11})^{-1}\mathbf{A}^{12}$$

where A^{11}, A^{21}, and A^{22} are portions of $\text{A}^{-1}$for non-genotyped, between genotyped and non-genotyped, and for genotyped animals, respectively.

Without APY, ssGBLUP would not be feasible for the AAA evaluations. However, identifying core animals was not an easy task at the beginning of the implementation of ssGBLUP for AAA. Choosing core animals randomly or based on EBV accuracy resulted in correlations >0.99 between GEBV from regular and APY ssGBLUP, providing the core group had a minimum of 10k animals. Less optimal core definitions caused convergence issues. We ultimately chose to select core animals based on EBV accuracy for the official evaluations because those animals would have more progeny recorded. No differences were found in convergence and computing time for the growth model However, for the carcass model, which combines 9 different traits, some of them are sparsely recorded (e.g., fat and ribeye area), using a random core instead almost halved the number of rounds to convergence. Since 2017, Angus Genetics Inc. has been using APY ssGBLUP for weekly evaluations of around 18 traits, using a single set of core animals assigned based on EBV accuracy.

Computing SNP effects in ssGBLUP. Besides the official national Angus evaluation, genomic predictions or direct genomic values (DGV) based on SNP effects are provided for non-registered animals, usually females, for herd management. Although ssGBLUP provides GEBV as final output, estimates of SNP effects (**a**) can be obtained by back-solving GEBV (Wang *et al*. 2012):

$$\hat{\mathbf{a}} = \mathrm{k}\mathbf{Z}'\mathbf{G}^{-1}\hat{\mathbf{u}}$$

where: k is the ratio of SNP to additive genetic variance, **Z** is a centred matrix of SNP effects, and **u** is a vector of GEBV. As the calculation of SNP effects is done by standalone software from the BLUPF90 family (postGSf90), there is a need to save $\mathbf{G}_{\mathrm{APY}}^{-1}$ to disk. However, even half stored requirements were large. To overcome this problem, we investigated the use of the subset of $\mathbf{G}_{\mathrm{APY}}^{-1}$ only for core animals ($\mathbf{G}_{\mathrm{CC}}^{-1}$). Correlations between GEBV and DGV obtained with $\mathbf{G}_{\mathrm{APY}}^{-1}$ or $\mathbf{G}_{\mathrm{CC}}^{-1}$ were greater than 0.98 using the 2014 dataset. However, as the number of genotyped animals increased, we observed a decrease in correlation when $\mathbf{G}_{\mathrm{CC}}^{-1}$ was used. Based on that, we changed the algorithm in postGSf90 to work with blocks of $\mathbf{G}_{\mathrm{APY}}^{-1}$ instead of having to allocate the full matrix in memory. The new software requires less memory and is extremely fast.

Accuracy as a measure of GEBV risk. One of the benefits of using genomic information is to increase breeding value accuracy. Accuracies are calculated based on prediction error variance (PEV) and can be obtained from the inverse of the LHS of MME. If the number of animals in the pedigree is large, the inverse is not computationally feasible and an approximation has to be used. Approximating accuracy of GEBV requires the calculation of the combined contributions due to phenotypes, pedigree, and genomic information. An algorithm to approximate genomic contributions was developed based on diagonals of **G** and the average traditional accuracy for genotyped animals. Compared to the approximation based on pedigree and phenotypes only, the increase in computing time was irrelevant. Another advantage of this algorithm is that the diagonal of **G** or $\mathbf{G}_{\mathrm{APY}}$ can be easily saved and requires a small disk space. Correlations between accuracy from the new algorithm and true accuracy from PEV were higher than 0.85 for growth traits, using a sample dataset.

Dairy cattle. Single-step GBLUP is currently used by Zoetis for genomic evaluations of wellness traits in dairy cattle. Those traits have a binary response and each one is currently analysed separately using a univariate threshold model. Heritabilities are low, ranging from 0.06 to 0.08, and trait incidences vary from 2% to 25% (Vukasinovic *et al*. 2017). As for beef cattle, changes had to be made to accommodate an increasing number of genotyped animals. Although all models were single-trait, the time to convergence with APY core animals selected based on their accuracy of EBVs, was between 24 and 50 hours, which is un-acceptable. When the core animals were randomly selected, the computing time was between 4 and 10 hours. As was previously observed for carcase traits in beef cattle, the choice of core animals becomes an issue when few genotyped animals have phenotypes. Probably, $\mathbf{G}_{\mathrm{APY}}^{-1}$ with random core is better conditioned than with high EBV accuracy core.

Although official genomic evaluations in the US are still done with multi-step methods, several tests have been done by our group to investigate the feasibility of ssGBLUP for dairy cattle using data provided by the US Holstein Association and the Council on Dairy Cattle Breeding (CDCB). A common problem is the inflation of GEBV. Although using inbreeding for the calculation of $\mathbf{A}^{-1}$ eliminated inflation in the beef cattle evaluation, the same was not true for dairy evaluations. This is because the missingness of pedigree is greater in dairy cattle data. In the initial tests before implementing inbreeding, convergence could only be reached with $\boldsymbol{\omega}$<1.

In BLUP-based methods, missing parents can be modelled by unknown parent groups (UPG). Such groups are also known as phantom parents or genetic groups, and are used to represent the average level of breeding value in a group where parents were missing. In ssGBLUP, when UPG are applied only to **A**, convergence may fail or the convergence rate can be slow. Alternatively, UPG can be assigned to both **A** and $\mathbf{A}_{22}$. For US Holstein with 18 type traits, using 10M animals in the pedigree and 570k genotyped, we observed that adding UPG for $\mathbf{A}_{22}$ helped to reduce inflation. However, the least GEBV inflation for young genotyped bulls was observed when inbreeding for UPG was also considered and the genetic variance was halved (Figure 3). The problem of reducing additive genetic variance is the shrinkage of GEBV for all animals, not only for young genotyped bulls. According to VanRaden *et al*. (2014), a reduction in genetic variance for yield traits reduced inflation caused by the inclusion of female genotypes.

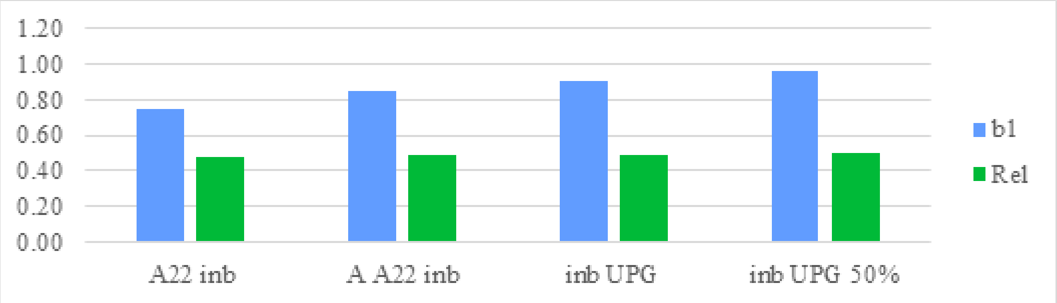

Figure 3. Average regression coefficient (b1) and reliability (Rel) for 18 type traits in US Holsteins A22 inb = inbreeding in A_{22}^{-1} + UPG; A A22 inb = inbreeding in A^{-1} and A_{22}^{-1} + UPG; inb UPG = inbreeding for UPG; Inb UPG 50% = inb UPG with 50% reduction of additive genetic variance

The US dairy industry has collected almost 3M Holstein genotypes by April 2019 (https://queries.uscdcb.com/Genotype/counts.html). Only 11% of those are for males and over 75% of the females will never have phenotypic records. Initial ssGBLUP tests using 2.3M genotyped animals, 13.5M animals in the pedigree, and 11M records on 18 type traits took 3 days to converge and required over 300 GB of memory. This was using APY with 15k randomly chosen core animals. Currently, the ssGBLUP software used to solve large systems of equations is undergoing changes for the implementation of a message-passing interface (MPI), which uses multi-processor architecture, allowing a higher level of parallelization. After these changes are completed, the convergence for the 18 type trait model is expected to be reached within 36 hours using about 30 GB of memory.

Pigs. The biggest challenge for genomic evaluations in pigs is to have accurate predictions for multi-breed or crossbred populations. In within breed ssGBLUP, **G** is constructed based on the average allele frequency. However, different breeds may have different allele frequencies, and construction of **G** must be modified. Using 2 breeds and their F1, we observed negative genomic relationships between breeds, which is an indicator of distinct allele frequencies. Breed-specific allele frequencies were subsequently used to centre and scale **G** in simulated and real pig datasets (Lourenco *et al*. 2016). Although the average relationship between the 2 breeds was zero when using breed-specific allele frequencies, accuracy of GEBV was similar to the default scenario that used across-breed allele frequencies to construct **G**. If there is a dominant breed, meaning one breed has many more genotyped animals than the other, the largest breed will likely have more accurate predictions. To avoid this issue, **G** can be constructed assuming SNP are not shared among breeds, which would create a block-diagonal **G**; however, this is less straightforward when genotypes for crossbreds are included in the evaluation (Steyn *et al*. 2019).

When APY ssGBLUP is used in multi-breed or crossbred evaluations, the choice of core animals becomes even more complex. This is because the appropriate number of core animals depends on the theory of limited dimensionality of genomic information and chromosome segments, which relies on effective population size. The suitable number of core animals in multi-breed populations can be accessed by the number of eigenvalues explaining 98% of the variance in **G**, considering all breeds together. If breeds are completely independent, the expectation for the number of eigenvalues across 2 breeds is the sum of eigenvalues within each breed (Figure 4). We observed that comparing the number of eigenvalues across and within breeds can indicate the ability to perform across-breed predictions because chromosome segments are shared (Pocrnic *et al*. 2019).

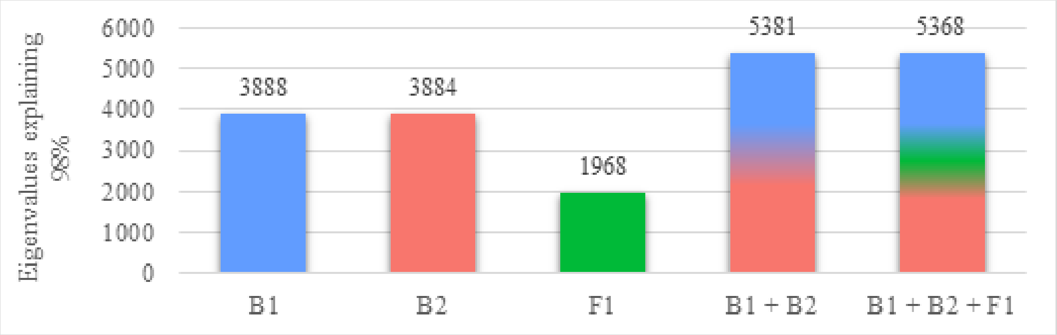

Figure 4. Number of eigenvalues explaining 98% of the variance of G across and within breeds (B1 and B2) and crossbred (F1)

Large pig breeding companies usually buy small farms/companies and combine the populations into a single evaluation. Assigning UPG for each population can help to account for the difference in base population. However, UPG are usually considered as fixed effects and a reasonable number of observations is needed for their accurate estimation. Datasets coming from small farms may have insufficient amount of information linked to UPG, leading to estimation errors and inflation of GEBV. In such a case, we observed that using random instead of fixed UPG solved estimation problems related to poor UPG connections (Pocrnic *et al*. 2018).

Poultry. Inflation of GEBV is not so evident in chicken datasets because old generations are removed and genotyped animals have complete pedigree. In fact, only 2 to 4 years of data are retained for genetic and genomic evaluation in chickens. However, in the first tests of ssGBLUP in chicken data (from Cobb-Vantress), back in 2013, we observed inflated genetic trends for GEBV compared to EBV, especially for young animals. The sources of this inflation were identified to result from the inclusion of unmapped SNP (i.e., mapped to chromosome 0) in the evaluation, the presence of imputation errors, and incorrectly labelled samples. If the imputation uses a family-based method but the pedigree has errors, the imputation can be compromised, resulting in low correlation between **G** and $\mathbf{A}_{22}$. This outcome illustrates more generally that quality control of SNP, samples, and genomic relationships is an important step before genomic evaluation, as small errors in the SNP data can be propagated, generating biased estimates.

Another issue observed in chicken data was the lower predictive ability (i.e., correlation between adjusted phenotypes and EBV or GEBV) for females compared to males for a growth trait, even though females had almost twice the number of genotypes. For an efficiency trait with the same amount of information and similar heritability, predictive ability was comparable between males and females. Separate genetic trends for males and females showed stronger selection for females than males in the growth trait but a similar trend in the efficiency trait (Figure 5). This shows that predictive ability takes the selection intensity into account and different predictive ability is expected if males and females have distinct selection differentials.

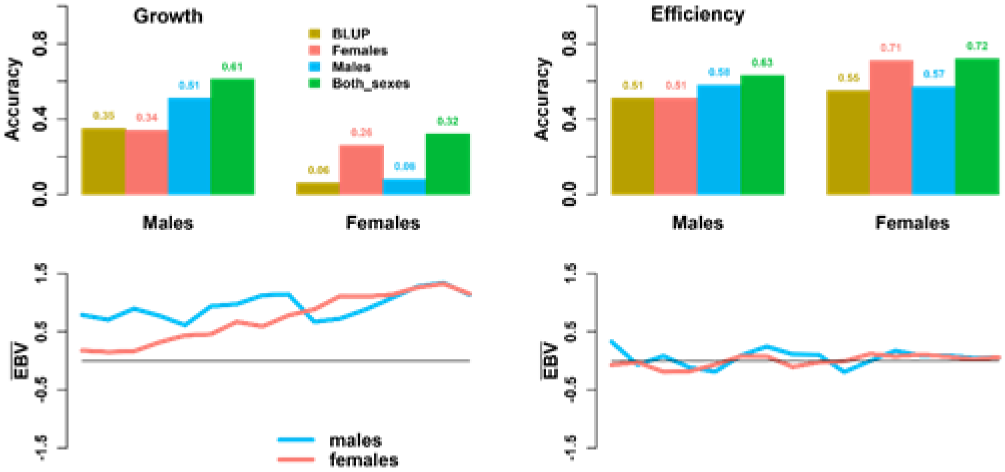

Figure 5. Accuracy and genetic trends for growth and efficiency in chickens

Fish. In fish production, several families may be raised in common ponds, making the identification of individuals difficult. Even though low density microsatellite panels are used for parentage determination with relatively high accuracy (Waldbieser and Bosworth 2012), pedigree errors still exist. As fish families are large, a single mistake in parentage assignment can be multiplied to thousands of individuals. In the implementation of genomic selection for catfish in the USA (ARS-USDA Warmwater Aquaculture Research Unit), we were able to identify mis-assigned parentage based on SNP markers. After pedigree corrections, heritability for carcass weight was adjusted from 0.27 to 0.21. Correcting variance components avoided the overestimation of genetic gains.

Another issue present in fish populations is the choice of individuals to be genotyped, given that genotyping is still expensive and full or half-sib families are large. We decided to genotype 40 fish per family, in a total of 75 families. As carcass weight is one of the most important traits in fish, genotyped individuals were also slaughtered to evaluate whether half or full-sib phenotypes would be enough to produce a high predictive ability for selection candidates. Using own genomic information combined with phenotypes and genotypes on siblings provided a 22% increase in predictive ability for carcass weight, compared to traditional BLUP.

Assessing predictive ability for disease traits either in fish or other species is quite challenging because phenotypes have a binary nature and breeding values have a normal distribution. Therefore, correlations between adjusted phenotypes and GEBV are usually very small or negative. Additionally, the regression coefficients are much lower than 1, which may not support the use of genomic selection. For binary and categorical traits, other validation methods may be more appropriate than predictive ability. For the initial tests on the feasibility of genomic selection for columnaris disease resistance in rainbow trout in the USA (ARS-USDA Cool and Cold Water Aquaculture), we adopted the LR validation (Legarra and Reverter 2018). This method is based on comparisons between complete and partial predictions. The relative increase in accuracy of GEBV compared to EBV was 40%, which encouraged the adoption of genomic selection to predict disease resistance in rainbow trout (Silva *et al*. 2019).

Common problems in large-scale genomic evaluations. Although different species usually require different strategies, common issues emerge when using large datasets. For ssGBLUP evaluations using a large number of genotyped animals, APY is one of the options. Another option is the ssGTBLUP (Mantysaari *et al*. 2017) that uses Woodbury formulas and requires only the inverse of a matrix with the size of the number of SNPs. An equivalent model that can also be used for large data is ssBR or hybrid model (Fernando *et al*. 2014), where SNP effects are estimated regardless of the number of genotyped animals.

When APY is used, even though the correlation between GEBV from regular ssGBLUP and APY ssGBLUP is greater than 0.99 when the appropriate number of core animals is used, re-ranking is still observed when different core groups are used. We investigated in beef and dairy cattle, and pig datasets different definitions of core and random core groups to identify which animals have the biggest changes in GEBV and how those changes can be minimized. In all datasets, larger changes in GEBV by using different core groups were observed for animals with lower accuracy. The observed changes relative to standard deviations of GEBV were, on average, 5%, but ranged from 0 to 100%. Increasing the number of core animals beyond the optimal value helped to asymptotically reduce changes in GEBV. Although core-dependent changes in GEBV exist, they are small and can be reduced with larger core groups.

Accounting for selected sequence variants in GBLUP-based methods. As sequence data is slowly becoming available for livestock, there is a question whether GBLUP-based methods can account for selected sequence variants and what is the possible gain in accuracy. Although the default assumption of GBLUP methods is that all SNP explain the same proportion of variance, it is possible to weight SNP differently. Recently, we observed that the increase in accuracy by SNP weighting is smaller in large populations, compared to small populations. This is because large genotyped populations allow more accurate estimation of chromosome segment effects; therefore, there is no advantage in selecting SNP and tagging segments with larger value (Lourenco *et al*. 2017).

Using a US Holstein dataset, Fragomeni *et al*. (2019) tested the performance of GBLUP and ssGBLUP when using nearly 54,000 SNP and when adding 17,000 significant variants discovered from a GWAS using sequence data involving 33 traits (VanRaden *et al*. 2017). Although VanRaden *et al*. (2017) reported an increase in reliability of GEBV of 4.3 points for stature by using non-linearA weights (i.e., a fast version of BayesA) in a multistep scenario, no gain was observed by Fragomeni *et al*. (2019) using either quadratic or non-linearA weight in GBLUP with heterogeneous residual variance and ssGBLUP. This is possibly because the amount of data used in ssGBLUP overwhelms any a priori assumption made about SNP effects, making this method less sensitive to SNP weighting in the presence of large data. Another hypothesis to explain the steady reliability is that not all causative variants were present among the 17,000 significant SNP. In a simulation study done by Fragomeni *et al*. (2017), including all simulated causative variants with respective true weights among 60k SNP, increased accuracy of ssGBLUP GEBV from 0.49 to 0.94. Although causative variants can be included in ssGBLUP assuming different weights for SNP, maximizing the accuracy of GEBV would require the true identification of all causative variants and their substitution effect.

CONCLUSIONS

Although the implementation of genomic selection seems to be straightforward, given genotypes are added to phenotypes and pedigree that are already in the evaluation system, several issues and challenges were raised during the initial application of this methodology to breeding programs of several species. Fortunately, solutions to most of the problems have come in a fast pace, enabling the widespread use of this methodology. Overall, the sources of problems include missingness of pedigree, selective genotyping, increasing number of genotyped animals, incompatibility between pedigree and genomic information, and difficulty in assessing predictive ability of genomic models for specific traits. It is expected that more issues will rise, and most of them may be related to the amount, type, and way the genomic data are being generated.

ACKNOWLEDGEMENTS

This paper is based on numerous studies done by several people from our animal breeding and genetics group at the University of Georgia, including current graduate students (Andre Garcia and Yvette Steyn), former postdocs (Ivan Pocrnic, Breno Fragomeni, and Rafael Silva), and the extended research group (Ignacio Aguilar and Andres Legarra). We thank the continued support from American Angus Association, US Holstein Association, Zoetis, PIC, Smithfield Premium Genetics, Cobb-Vantress, and USDA.

REFERENCES

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S. and Lawlor T.J. (2010) *J. Dairy Sci*. **93**: 743.

Brown D.J., Swan A.A., Li L., Gurman P.M., McMillan A.J., van der Werf J.H.J., Chandler H.R., Tier B. and Banks R.G. (2018) In Proc 11th WCGALP Auckland, New Zealand.

Chen C.Y., Misztal I., Aguilar I., Tsuruta S., Meuwissen T.H.E., Aggrey S.E., Wing T. and Muir W.M. (2011) *J. Anim. Sci*. **89**: 23.

Christensen O.F. and Lund M.S. (2010) *Genet. Sel. Evol*. **42**: 2.

Fernando R.L., Dekkers J.C.M., and Garrick D. J. (2014) *Genet. Sel. Evol*. **46**: 50.

Fragomeni B.O., Lourenco D.A.L., Masuda Y. and Misztal I. (2017) *Genet. Sel. Evol*. **49**: 59.

Fragomeni B.O., Lourenco D.A.L., Legarra A., VanRaden P.M., and Misztal I. *J. Dairy Sci*. (*Under review*).

Forni S., Aguilar I. and Misztal I. (2011) *Genet. Sel. Evol*. **43**: 1.

Gao H., Koivula M., Jensen J., Stranden I., Madsen P., Pitkanen T., Aamand G.P. and Mantysaari E.A. (2018) *J. Dairy Sci*. **101**: 10082.

Garcia A.L.S., Bosworth B., Waldbieser G., Misztal I., Tsuruta S. and Lourenco D.A.L. (2018) *Genet. Sel. Evol*. **50**: 66.

Kachman S.D., Spangler M.L., Bennett G.L., Hanford K.J., Kuehn L.A., Snelling W.M., Thallman R.M., Saatchi M., Garrick D.J., Schnabel R.D., Taylor J.F. and Pollack J. (2013) *Genet. Sel. Evol*. **45**: 30.

Legarra A., Bertrand J.K., Strabel T., Sapp R.L., Sanchez J.P. and Misztal I. (2007) *J. Anim. Breed. Genet*. **124**: 286.

Legarra A., Aguilar I. and Misztal I. (2009) *J. Dairy Sci*. **92**: 4656.

Legarra A. and Reverter A. (2018) *Genet. Sel. Evol*. **50**: 53.

Lourenco D.A.L., Tsuruta S., Fragomeni B.O., Masuda Y., Aguilar I., Legarra A., Bertrand J.K., Amen T., Wang L., Moser D.W. and Misztal .I (2015a) *J. Anim. Sci*. **93**: 2653.

Lourenco D.A.L., Fragomeni B.O., Tsuruta S., Aguilar I., Zumbach B., Hawken R.J., Legarra A. and Misztal I. (2015b) *Genet. Sel. Evol*. **47**: 56.

Lourenco D.A.L., Tsuruta S., Fragomeni B.O., Chen C.Y., Herring W.O. and Misztal I. (2016) *J. Anim. Sci*. **94**: 909.

Lourenco D.A.L., Fragomeni B.O., Bradford H.L., Menezes I.R., Ferraz J.B.S., Aguilar I., Tsuruta S. and Misztal I. (2017) *J. Anim. Breed. Genet*. **134**: 463.

Mantysaari, E. A., Evans R. D. and Stranden I. (2017) *J. Anim. Sci*. **95**: 4728.

Masuda Y., Misztal I., Legarra A., Tsuruta S., Lourenco D.A.L., Fragomeni B.O. and Aguilar I. (2017) *J. Anim. Sci*. **95**: 49.

Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F., Heaton M.P., O'Connell J., Moore S.S., Smith T.P L., Sonstegard T.S. and VanTassell C.P. (2009) *PLoS One* **4**: e5350.

Meuwissen T.H.E., Hayes B.J. and Goddard M.E. (2001) *Genetics* **157**:1819.

Misztal I., Legarra A. and Aguilar I. (2009) *J. Dairy Sci*. **92**: 4648.

Misztal I., Legarra A. and Aguilar I. (2014) *J. Dairy Sci*. **97**: 3943.

Pocrnic I., Lourenco D.A.L., Bradford H.L., Chen C.Y. and Misztal I. (2017) *J. Anim. Sci*. **95**: 3391.

Pocrnic I., Lourenco D.A.L., Chen C.Y., Herring W.O. and Misztal I. (2019) *J. Anim. Sci*. **97**: 1513.

Rupp R., Mucha S., Larroque H., McEwan J. and Conington J. (2016) *Anim. Frontiers* **6**: 39.

Saatchi M., McClure M.C., Mckay S.D., Rolf M.M., Kim J., Decker J.E., Taxis T.M., Chapple R.H., Ramey H.R., Northcutt S.L., Bauck S., Woodward B., Dekkers J.C.M., Fernando R.L., Schnabel R.D., Garrick D.J. and Taylor J.F. (2011) *Gen. Sel. Evol*. **43**: 40.

Silva R.M.O., Evenhuis J.P., Vallejo R.L., Gao G., Martin K.E., Leeds T.D., Palti Y. and Lourenco D.A.L. (2019) *Genet. Sel. Evol*. (*Under review*).

Soller M. and Beckman J.S. (1983) *Theor. Appl. Genet*. **67**: 25.

Steyn Y., Lourenco D.A.L. and Misztal I. (2019) *J. Anim. Sci*. (*Under review*).

The International SNP Map Working Group (2001) *Nature* **409**: 928.

Tsuruta S., Misztal I., Aguilar I. and Lawlor T.J. (2011) *J Dairy Sci*. **94**: 4198.

VanRaden P.M. (2008) *J. Dairy Sci*. **91**: 4414.

VanRaden P., Tooker M.E., Wright J.R., Sun C. and Hutchison J.L. (2014) *J. Dairy Sci*. **97**: 7952.

VanRaden P.M., Tooker M.E., O'Connell J.R., Cole J.B. and Bickhart D.M. (2017) *Gen. Sel. Evol*. **49**: 32.

Vukasinovic N., Bacciu N., Przybyla C.A., Boddhireddy P. and DeNise S.K. (2017) *J. Dairy Sci*. **100**: 428.

Waldbieser G.C. and Bosworth B.G. (2012) *Animal* **44**: 476.

Wang H., Misztal I., Aguilar I., Legarra A. and Muir W.M. (2012) *Genetics Research* **94**: 73.

Wiggans G.R,, Cooper T.A., VanRaden P.M. and Cole J.B. (2011) *J. Dairy Sci*. **94**: 6188.

Wiggans G.R., VanRaden P.M. and Cooper T.A. (2012) *J. Dairy Sci*. **95**: 3444.

Yan Y., Wu G., Liu A., Sun C., Han W., Li G. and Yang N. (2018) *Poult. Sci*. **97**: 397.