

Genetic evaluation using single-step genomic BLUP in American Angus

Daniela Lourenco¹, Ignacy Misztal¹, Keith Bertrand¹, Tonya Amen², and Dan Moser²

¹Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602

²Angus Genetics Inc., St. Joseph, MO 64506

Introduction

Genomic selection in beef cattle has currently been performed with multistep methods, which uses deregressed EBV to estimate SNP effects and then direct genomic value (DGV) for selection candidates based on their genotypes (Meuwissen et al., 2001; Garrick et al., 2009). The main advantage of this approach is that the traditional BLUP evaluation is kept unchanged and genomic selection can be carried out by a separate entity owning genotypes but not phenotypes. Also new animals are easily evaluated if DGV is computed as a sum of marker effects, but not if selection indexes including DGV and parent average (PA) are used.

When both phenotypes and genotypes are available jointly, single-step genomic BLUP (ssGBLUP) (Aguilar et al., 2010) is a simple alternative. This method does not rely on deregressed proofs, properly weighs information from genotyped sires and cows, thus avoiding double-counting of contributions due to relationships and records, and accounts for pre-selection bias of genomically selected parents without phenotypes (Legarra et al., 2014). In ssGBLUP it is also possible to quickly evaluate young genotyped animals without running a complete evaluation that requires several hours to converge. Quick predictions can be calculated indirectly, where genomic predictions for young animals are obtained from SNP effects. It was shown by Wang et al. (2012) that SNP effects can be derived from GEBV solutions from the main ssGBLUP evaluation.

In its current implementation, ssGBLUP uses direct inversion of genomic matrices (Aguilar et al., 2011), which has a cubic cost and a limit of 150,000 animals (Aguilar et al., 2013). Several methods were proposed to overcome that limit (Legarra and Ducrocq, 2012; Fernando et al., 2014; Liu et al., 2014), but none was successful. Recently Misztal et al. (2014) presented a method which uses an approximate inversion of genomic relationships based on recursions on a fraction of the total population; which can be suitable and inexpensive.

The goal of this paper is to discuss the feasibility of ssGBLUP for genomic evaluation in Angus cattle with reference populations of different composition. Additional goals were to evaluate the ability to predictive GEBV with genomic recursions and with indirect prediction for young animals.

Single-step genomic BLUP (ssGBLUP)

The difference between BLUP and ssGBLUP can be observed in the mixed model equations, where the inverse of pedigree-based relationship matrix (\mathbf{A}) is replaced by the inverse of the realized relationship matrix (\mathbf{H}). The \mathbf{H} matrix blends \mathbf{A} with the genomic relationship matrix constructed based on SNP genotypes (\mathbf{G}) and a pedigree-based relationship matrix for genotyped animals only (\mathbf{A}_{22}):

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

Contrary to genomic BLUP (GBLUP), ssGBLUP accounts not only for genotyped animals, but also for ungenotyped ones. In \mathbf{H} matrix, genomic relationships can influence nongenotyped animals if they are related to genotyped animals in \mathbf{A} . Even if two animals are unrelated in \mathbf{A} , they will be related in \mathbf{H} if they have genotyped descendants that are related in \mathbf{G} . This is possible because \mathbf{G} is usually identical by state, which means the genomic relationships are based on allele sharing independently of the origin.

Composition of genomic EBV (GEBV) from ssGBLUP

From traditional BLUP evaluations, EBV for an animal i can be expressed as (VanRaden and Wiggans, 1991):

$$u_i = w_1 PA + w_2 YD + w_3 PC$$

where PA is Parent Average, YD is Yield Deviation (phenotypes adjusted for the model effects' solutions other than additive genetic and error), PC is Progeny Contribution. When both parents are known, phenotype is available, and each progeny has a known mate, the weights w_1 to w_3 sum to 1. The decomposition of EBV can be derived from analyzing a row of mixed model equations for a given animal. In particular, YD is due to own phenotypic information, PA is the average of parental EBV for animal i , and PC is the sum of the differences between any progeny EBV of animal i minus one half the EBV of its dam (or the mate of animal i).

On the other hand, EBV for an animal i when genomic information is available (GEBV) is (VanRaden and Wright, 2013):

$$u_i = w_1 PA + w_2 YD + w_3 PC + w_4 GI$$

where GI contains information from animal's genotype and all weights sum to 1. According to VanRaden and Wright (VanRaden and Wright, 2013), the weight for GI is:

$$w_4 = \frac{g^{ii} - a_{22}^{ii}}{den}, \text{ with } den = 2 + n_r/\alpha + n_p/2 + g^{ii} - a_{22}^{ii}$$

Aguilar et al. (Aguilar et al., 2010) showed that in ssGBLUP, GI corresponds to two components:

$$GI = w_{4_1} DGV - w_{4_2} PP$$

with values for DGV and PP equal to:

$$DGV_i = \frac{-\sum_{j,j \neq i} g^{ij} u^j}{g^{ii}}; PP_i = \frac{-\sum_{j,j \neq i} a_{22}^{ij} u^j}{a_{22}^{ii}}$$

where DGV is the portion of prediction due to the genomic information, which comes from \mathbf{G} , and PP is pedigree prediction that comes from \mathbf{A}_{22} . The weights w_1, w_2, w_3, w_{4_1} , and w_{4_2} sum to 1.

In general, PP accounts for part of PA explained by DGV; when $\mathbf{A}=\mathbf{A}_{22}$, PA and PP cancel out and DGV explains a larger fraction of GEBV; when a genotyped animal is unrelated to the genotyped population, $PP=0$ and DGV explains a smaller fraction of GEBV; when two parents are genotyped, PP will include a large fraction of PA. Subsequently, accuracy of DGV will vary on an animal basis depending on how many ancestors of that animal are genotyped, as found by Mulder et al. (Mulder et al., 2012). When a genotyped animal has many progeny, $w_3 \approx 1$ and its GEBV is mainly driven by PC; however, genotyping those animals is useful as they are usually part of the reference population. When an animal is not genotyped, $w_4=0$ and predictions can be improved due to improved PA and PC if its relatives are genotyped. When an animal is not genotyped and has no phenotypes and no progeny, the genetic prediction is driven by PA and in most cases improvements in predictions are small (Legarra et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010).

Dataset

Datasets from American Angus Association (AAA) were available that included growth traits and calving ease (CE). Growth traits included birth weight (BW), weaning weight (WW), and post-weaning gain (PWG). Table 1 shows general statistics for all traits. The animals were genotyped for 54,609 SNP from the BovineSNP50k v2 BeadChip (Illumina Inc., San Diego, CA). Currently, no genotyping strategy is applied by AAA; therefore, the members can choose which animals are being genotyped.

Table1. Heritability (h^2) and general statistics for growth traits and CE

Trait ¹	h^2	Number of records	Average (kg)	SD (kg)	Number of genotyped animals with records
BW	0.41	6,189,661	36.47	4.45	50,784
WW	0.20	6,890,625	263.13	44.63	51,830
PWG	0.20	3,387,252	162.25	67.00	36,196
CE	0.12	1,310,684	-	-	10,558
easy	-	1,215,571	-	-	10,228
difficult	-	95,113	-	-	330

¹ BW = birth weight; WW = weaning weight; PWG = post-weaning gain; CE = calving ease.

ssGBLUP with different reference populations

The current trend in livestock genomics is to genotype young animals; however, more important animals in the reference population may give more information to the evaluations. For growth traits (CE), the first reference population tested was composed of 1,628 (1,541) top bulls with EBV accuracy (based on prediction error variance - PEV) for $BW \geq 0.85$; which we will refer hereinafter as “ref_bulls”. In this case, the **G** matrix was composed of animals in the reference population and also animals in the validation population; the last had 18,721 animals for growth traits and 13,166 for CE. The second reference population was composed of the top bulls and also top cows that had an EBV accuracy for $BW \geq 0.85$; which we will refer as ref_2k. The number of top cows was small and only 268 were added for the growth trait analysis and 323 for CE. The third reference population was composed of top bulls, top cows, and all other genotyped animals born from 1977 to 2012 (we will refer as ref_33k). This group had a total of 33,162 animals for growth and 27,380 for CE, with an average EBV accuracy for BW of 0.77 (\pm 0.05). For the latter analysis, the **G** matrix was composed of the maximum number of 51,883 genotyped animals for growth analysis and 40,546 for analysis of CE.

Method for Validation

The ability to predict future phenotypes was the validation method chosen. This method is based on Legarra et al. (2008), and predictive ability for traditional and genomic evaluations for animals born in 2013 was calculated as the correlation between (G)EBV and phenotypes corrected for fixed effects ($y-Xb$):

$$r = \text{cor}[(G)EBV, y-Xb]$$

The predictive ability or predictivity is used as an approach to compare the methods applied in this paper. For all analyses, the validation groups were kept the same to make comparisons easier. Validations involved 18,721 animals for growth traits and 16,133 animals for CE.

Results

Predictive ability on young animal when using several reference populations is shown in Figure 1. Using only top bulls as a reference population (ref_bulls) increased predictivity relative to BLUP by 0.05 for BW, 0.01 for WW, 0.04 for PWG, and 0.01 for CE. Addition of top cows to the reference population (ref_2k) did not increase the predictivity for any trait. This could be due to the small number of animals added and also because daughters of those cows already contributed through the inclusion of bulls. Addition of around 31,000 animals to the reference population provided an additional increase in predictivity of 0.05 for BW, of 0.03 for WW and of 0.02 for PWG. However, no additional increase was observed for CE by adding extra 27,000 genotyped animals, of which about 7,000 had phenotypes for that trait.

The addition of 31,000 animals with few or no progeny led to the same increase of predictivity as using only the top bulls for BW, led to an increase of 3 times for WW and an increase of 0.5 times for PWG. Among the 31,000 extra animals, almost all had phenotypes for BW and WW, but only 24,000 had phenotypes for PWG. Evidently, the composition of reference

population is also a factor that influences predictivity of GEBV besides the reference population size. Thus, genotyping strategy should take into account genotyping more important and maybe older animals with more information (higher EBV accuracy) along with genotyping large amounts of young animals.

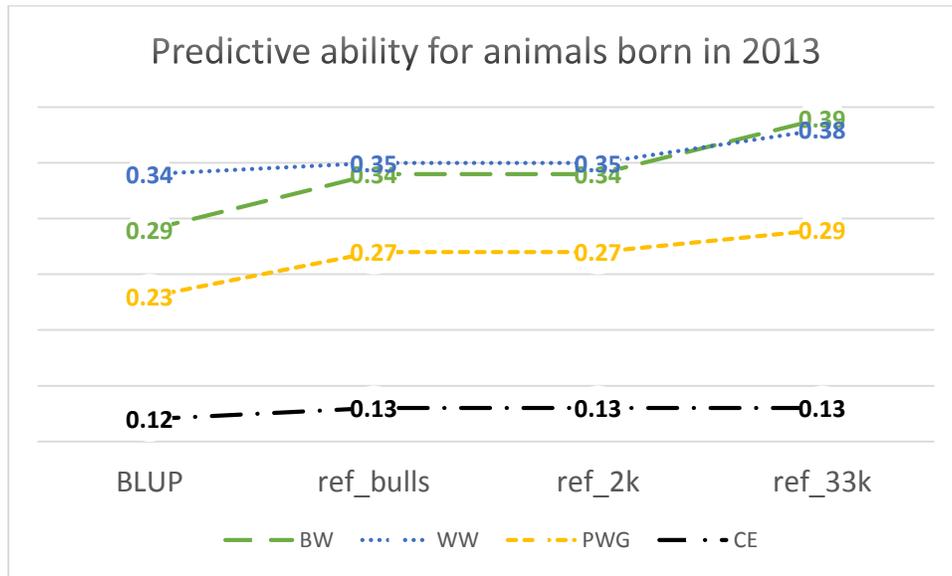


Figure 1. Predictive ability of future phenotypes for young genotyped animals born in 2013. *BW* = birth weight; *WW* = weaning weight; *PWG* = post-weaning gain; *CE* = calving ease. Predictive ability was calculated as correlation between corrected phenotypes and genomic EBV; *ref_bulls* is a reference populations that contains top bulls, *ref_2k* contains top bulls and top cows, and *ref_33k* contains all genotyped animals born up to 2012.

ssGBLUP with indirect predictions for young animals

With the increasing number of genotyped heifers and steers in beef cattle populations, the genomic methods should be able to provide predictions for young animals without phenotypes in a quick run, externally to the official evaluations. This concept is introduced here as indirect ssGBLUP, and basically mimics the mixed model equations. It would be advantageous from different perspectives: to evaluate young animals mainly for traits that are measured later in life, after the selection decisions are made; and to reduce computing costs

because the dimension of \mathbf{G} would not increase in the same proportion as the number of genotyped animals.

For ssGBLUP with indirect predictions, SNP effects can be calculated using the current run of ssGBLUP with all but young animals, and genomic predictions for young animals are obtained by multiplying the SNP content by SNP effect to obtain DGV; a more complete GEBV can also be available through a selection index that combines DGV and PA. In order to explain how it works, consider the equation for the GEBV of a single individual in ssGBLUP as a combination of equations in Aguilar et al. (2010) and VanRaden and Wright (2013):

$$\text{GEBV} = w_1\text{PA} + w_2\text{YD} + w_3\text{PC} + w_{4_1}\text{DGV} - w_{4_2}\text{PP}$$

The flow for indirect predictions in ssGBLUP is:

- 1) Run ssGBLUP with a reference population to obtain GEBV. In this step, 3 reference populations were tested:
 - a) ref_2k: reference population with top bulls and top cows (n=1,896);
 - b) ref_8k: reference population with all parents that were genotyped (n=8,285), this includes ref_2k;
 - c) ref_33k: reference population with all genotyped animals born up to 2012 (n=33,162), this includes ref_8k;
- 2) Split GEBV into all the components shown before, where DGV for an animal i in the reference population is calculated as below (Aguilar et al. (2010): $\text{DGV}_i = \frac{\sum_{j \neq i} g^{ij} \text{GEBV}^j}{g^{ii}}$ with all elements previously defined.
- 3) Calculate SNP effects using DGV from the reference population: $\hat{\mathbf{u}} = \mathbf{DZ}'\mathbf{G}^{-1}(\mathbf{DGV})$ where $\hat{\mathbf{u}}$ is a vector of estimated SNP effects, \mathbf{D} is a diagonal matrix of weights (standardized variances) for SNP (identity matrix in this case), and \mathbf{Z} is a matrix of centered genotypes for each

animal (VanRaden, 2008). A similar approach that uses GEBV instead of DGV to calculate SNP effects was proposed by Wang et al. (2012). However, for numerical purposes this involves approximations as \mathbf{G} matrix is formed as $\mathbf{G}=0.95\mathbf{ZDZ}'+0.05\mathbf{A}_{22}$ (Aguilar et al., 2010). This is done as a default approach to avoid singularity problems and may result in negligible error as shown later.

4) Calculate DGV for young genotyped animals (\mathbf{DGV}_y): $\mathbf{DGV}_y=\mathbf{Z}_y\hat{\mathbf{u}}$

where \mathbf{DGV}_y and \mathbf{Z}_y are direct genomic values and a matrix of centered genotypes for young animals not included in ssGBLUP evaluation, respectively.

5) Combine \mathbf{DGV}_y with PA for young genotyped animals: $\text{GEBV}_y \approx w_1\text{PA} + w_4\mathbf{DGV}_y$

where GEBV_y is GEBV obtained via indirect predictions for young animals, w_1 and w_4 are weights identical for all animals and calculated based on selection index.

Results

Predictive ability for indirect prediction via conversion of DGV into SNP effects is shown in Figure 2. When the reference population included top bulls and top cows (ref_2k), the predictivity of indirect \mathbf{DGV}_y was lower than predictivity for traditional EBV for the three traits (0.23 vs. 0.29 for BW; 0.28 vs. 0.34 for WW; 0.19 vs. 0.23 for PWG). Predictivity for GEBV_y calculated as an index of indirect \mathbf{DGV}_y with PA was higher than those for EBV for the three traits (0.31 vs. 0.29 for BW; 0.36 vs. 0.34 for WW; 0.24 vs. 0.23 for PWG), however, this predictivity was lower than the ones from full ssGBLUP (except for WW). With larger reference population (ref_8k), all indirect \mathbf{DGV}_y were similar or more accurate than EBV, and the index had similar predictivity as the full ssGBLUP. With the largest reference population (ref_33k), all indirect \mathbf{DGV}_y were almost as accurate as GEBV from full ssGBLUP, with the index marginally improving predictivity for WW. This marginal improvement for WW may be caused by the use

of less than optimal genetic parameters, e.g., zero covariance between direct and maternal effects (to reduce computing costs). The DGV_y obtained with ref_33k reference population were more accurate than GEBV from full ssGBLUP obtained with ref_8k reference population.

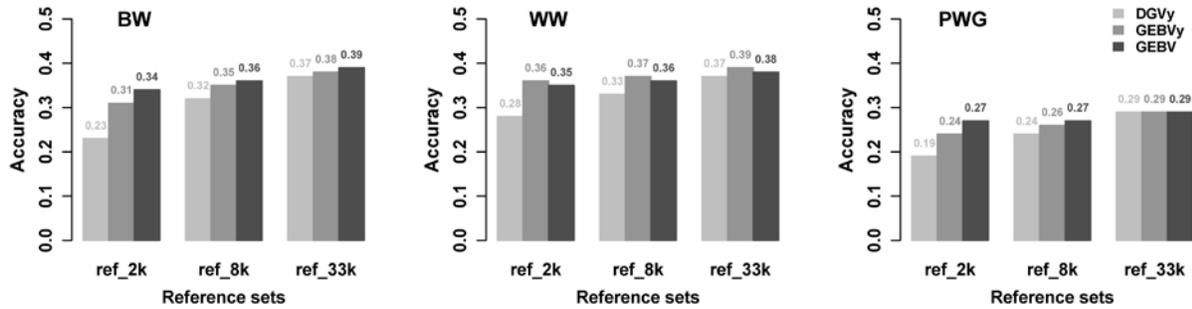


Figure 2. Predictive ability of indirect predictions on 18,721 young genotyped animals when using reference populations ref_2k, ref_8k, and ref_33k animals to run single-step genomic BLUP (ssGBLUP) and derivate SNP effects. *ref_2k* is a reference populations that contains top bulls and top cows, *ref_8k* contains all parents that were genotyped, and *ref_33k* contains all genotyped animals born up to 2012. DGV_y is direct genomic value; $GEBV_y$ is the indirect genomic EBV obtained by an index combining parent average and DGV_y ; $GEBV$ is genomic predictions obtained directly from ssGBLUP when genotypes on reference and validation animals were considered together in evaluations.

For young animals, indirect predictions via SNP effects from ssGBLUP seems a viable alternative as it can be done separately from the full evaluation. As SNP effects are calculated based on trait GEBV or DGV, indirect predictions are easily obtained for multi-trait models, as done in this study; multi-breed and crossbred evaluations are possible when the G matrix is able to account for information on all breeds. However, if young animals and particularly full-sibs are intensively selected, selection on the Mendelian sampling will not be accounted for, leading to pre-selection bias (Patry and Ducrocq, 2011). Analyses by ssGBLUP with all genotypes subject to selection are expected to account for pre-selection (VanRaden and Wright, 2013), because selection is accounted for when all information used for selection is included in the model (Henderson, 1975).

ssGBLUP with \mathbf{G} inverted by a recursive algorithm

When the number of genotyped animals is large and there is a desire for using all of them in ssGBLUP evaluations to get direct predictions for all, including young animals, an algorithm that splits genotypes into proven and young animals and uses recursion to approximate the inverse of the \mathbf{G} matrix was proposed by Misztal et al. (2014). This algorithm is known as APY, and \mathbf{G}^{-1} containing all genotyped animals can be expressed as:

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathbf{G}_{pp}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{pp}^{-1} \mathbf{G}_{py} \\ \mathbf{I} \end{bmatrix} \mathbf{M}_g^{-1} \begin{bmatrix} -\mathbf{G}_{yp} \mathbf{G}_{pp}^{-1} & \mathbf{I} \end{bmatrix}$$

where the subscript pp stands for proven animals and py for the covariance between proven and young animals; each element of \mathbf{M}_g is obtained (for the i^{th} young animal) as $m_{g,i} = g_{ii} -$

$\mathbf{G}_{ip} \mathbf{G}_{pp}^{-1} \mathbf{G}_{pi}$ and is called genomic Mendelian sampling. In APY, the only direct inversion needed is for part of \mathbf{G} that contains relationships among proven animals (\mathbf{G}_{pp}), whereas all other coefficients are obtained through recursions.

For this analysis, four definitions of proven animals were tested that included the 3 definitions used for indirect predictions (ref_2k, ref_8k, and ref_33k), plus one more definition where 3,872 genotyped parents of genotyped animals were considered as proven (ref_4k). This last group was added to test if proven animals would have strong links with the young genotyped population.

The greatest advantages of this algorithm are the reduction of computing cost, which is still cubic for proven animals, but can be linear for young animals; and the possibility of using large amounts of genotyped animals in ssGBLUP evaluations. The secondary advantage is numerical stability as the regular \mathbf{G} matrix is singular when the number of animals is greater than the number of SNP markers and cannot be inverted without blending with \mathbf{A}_{22} .

Results

Predictive ability of GEBV when the inverse of \mathbf{G} is computed with APY is shown in Figure 3. When the recursions were conditioned on ref_2k, ref_4k, ref_8k, and ref_33k, the procedure accounted for 67%, 88%, 97%, and 100% of predictivity gains of ssGBLUP over BLUP, respectively. Therefore, in ssGBLUP, using genomic recursion to invert \mathbf{G} while conditioning on enough number of animals, in this case about 8,000, has the same prediction power as \mathbf{G} using direct inversion. The amount of memory necessary for APY \mathbf{G}^{-1} using ref_2k, ref_4k, ref_8k, and ref_33k was approximately 0.8, 1.6, 3.2, and 13.7 Gbytes, respectively, whereas the amount of memory for the regular \mathbf{G}^{-1} is 21.6 Gbytes. Therefore, using APY \mathbf{G}^{-1} makes computations less costly and faster.

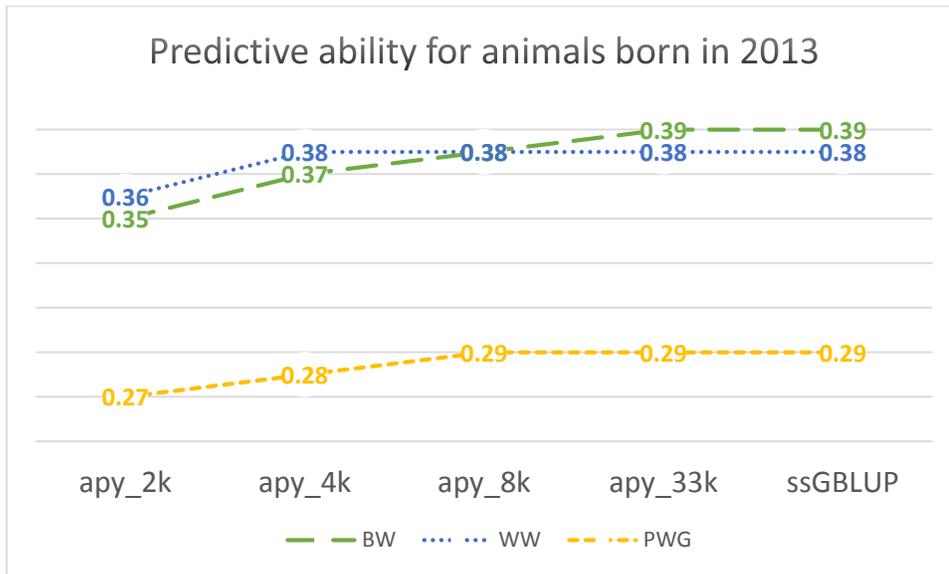


Figure 3. Predictive ability of GEBV for 18,721 young genotyped animals when using APY (algorithm for proven and young animals) to invert \mathbf{G} matrix (genomic-based relationship matrix) with different definitions of proven animals: ref_2k, ref_4k, ref_8k, and ref_33k. *ref_2k* is a reference populations that contains top bulls and top cows, *ref_4k* contains genotyped parents of genotyped animals, *ref_8k* contains all parents that were genotyped, and *ref_33k* contains all genotyped animals born up to 2012.

The main advantages of APY are low computing costs and numerical stability. With conditioning on 8,000 animals, for example, the only inverse required is for a block of \mathbf{G} for 8,000 animals, and additional genotypes require only linear storage and computations. Subsequently, computations with a large number of genotyped animals may be feasible with similar predictivity as in the regular inversion. APY would be the algorithm of choice for regular evaluations with very large number of genotyped animals.

Conclusions

Genomic evaluation in beef cattle using single-step genomic BLUP is feasible for either linear or categorical traits. Gains in predictive ability over BLUP are dependent on the composition of the reference population, and are greater for growth traits and small for CE. With a sufficient number of animals in the reference population, indirect prediction for young animals via SNP effects provides similar predictivity to full single-step genomic BLUP, allowing for quick genomic predictions without running a complete evaluation. Use of the algorithm for proven and young animals in single-step genomic BLUP allows for incorporation of large number of genotyped animals at low cost without compromising the predictive ability.

Literature cited

- 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.
- Aguilar, I., I. Misztal, A. Legarra, and S. Tsuruta. 2011. Efficient computation of the genomic relationship matrix and other matrices used in single-step evaluation. *J. Anim. Breed. Genet.* 128: 422–428.
- Aguilar, I., A. Legarra, S. Tsuruta, and I. Misztal. 2013. Genetic evaluation using unsymmetric single step genomic methodology with large number of genotypes *Interbull Bull.* 47: 222–225.
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42: 2.

- Fernando, R. L., J. C. M. Dekkers, and D. Garrick. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet. Sel. Evol.* 46: 50.
- Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41: 55.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31: 423–447.
- Legarra, A., C. R. Granie, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. *Genetics* 180: 611–618.
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92: 4656–4663.
- Legarra, A., and V. Ducrocq. 2012. Computational strategies for national integration of phenotypes, genomic, and pedigree data in single-step best linear unbiased prediction. *J. Dairy Sci.* 95: 4629–4645.
- Legarra, A., O. F. Chistensen, I. Aguilar, and I. Misztal. 2014. Single step, a general approach for genomic selection. *Livest. Prod. Sci.* 166: 54–65.
- Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. A single-step genomic model with direct estimation of marker effect. *J. Dairy Sci.* 97: 5833–5850.
- 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.
- 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.
- Mulder, H. A., M. P. L. Calus, T. Druet, and C. Schrooten. 2012. Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. *J. Dairy Sci.* 95: 876–889.
- Patry, C., and V. Ducrocq. 2011. Evidence of biases in genetic evaluations due to genomic preselection in dairy cattle. *J. Dairy Sci.* 94: 1011–1020.
- VanRaden, P. M., and G. R. Wiggans. 1991. Deviation, calculation, and use of national animal model information. *J. Dairy Sci.* 74: 2737–2746.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91: 4414–4423.
- VanRaden, P. M., and J. R. Wright. 2013. Measuring genomic pre-selection in theory and in practice. *Interbull Bull.* 47: 147–150.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res.* 94: 73–83.