Core-dependent changes in genomic predictions using the algorithm for proven and young in single-step genomic best linear unbiased prediction

Journal of Animal Science, 2020, Vol. 98, No. 12, 1–8

doi:10.1093/jas/skaa374
Advance Access publication November 19, 2020
Received: 4 August 2020 and **Accepted:** 11 November 2020
Animal Genetics and Genomics

ANIMAL GENETICS AND GENOMICS

Core-dependent changes in genomic predictions using the Algorithm for Proven and Young in single-step genomic best linear unbiased prediction

Ignacy Misztal,[†,1] Shogo Tsuruta,[†] Ivan Pocrnic,[‡] and Daniela Lourenco[†]

[†]Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602
[‡]The Roslin Institute, The University of Edinburgh, Edinburgh, UK

[1]Corresponding author: ignacy@uga.edu

ORCiD numbers: 0000-0002-0382-1897 (I. Misztal); 0000-0002-6897-6363 (S. Tsuruta); 0000-0001-5246-7428 (I. Pocrnic).

Abstract

Single-step genomic best linear unbiased prediction with the Algorithm for Proven and Young (**APY**) is a popular method for large-scale genomic evaluations. With the APY algorithm, animals are designated as core or noncore, and the computing resources to create the inverse of the genomic relationship matrix (**GRM**) are reduced by inverting only a portion of that matrix for core animals. However, using different core sets of the same size causes fluctuations in genomic estimated breeding values (**GEBV**s) up to one additive standard deviation without affecting prediction accuracy. About 2% of the variation in the GRM is noise. In the recursion formula for APY, the error term modeling the noise is different for every set of core animals, creating changes in breeding values. While average changes are small, and correlations between breeding values estimated with different core animals are close to 1.0, based on the normal distribution theory, outliers can be several times bigger than the average. Tests included commercial datasets from beef and dairy cattle and from pigs. Beyond a certain number of core animals, the prediction accuracy did not improve, but fluctuations decreased with more animals. Fluctuations were much smaller than the possible changes based on prediction error variance. GEBVs change over time even for animals with no new data as genomic relationships ties all the genotyped animals, causing reranking of top animals. In contrast, changes in nongenomic models without new data are small. Also, GEBV can change due to details in the model, such as redefinition of contemporary groups or unknown parent groups. In particular, increasing the fraction of blending of the GRM with a pedigree relationship matrix from 5% to 20% caused changes in GEBV up to 0.45 SD, with a correlation of GEBV > 0.99. Fluctuations in genomic predictions are part of genomic evaluation models and are also present without the APY algorithm when genomic evaluations are computed with updated data. The best approach to reduce the impact of fluctuations in genomic evaluations is to make selection decisions not on individual animals with limited individual accuracy but on groups of animals with high average accuracy.

Key words: APY algorithm, genomic selection, single-step GBLUP, stability of genomic predictions

Abbreviations

APY	Algorithm for Proven and Young
BLUP	best linear unbiased prediction
EBV	estimated breeding value
GBLUP	genomic BLUP
GEBV	genomic EBV
GRM	genomic relationship matrix
PEV	prediction error variance
SDa	Additive genetic standard deviation
SEP	standard error of prediction
SNP	single-nucleotide polymorphism
ssGBLUP	single-step GBLUP
UPG	unknown parent group

Introduction

Genomic evaluations by genomic best linear unbiased prediction (**GBLUP**) and single-step GBLUP (**ssGBLUP**) require the inverse of the genomic relationship matrix (**GRM**). If the number of genotyped individuals is large, usually the inverse is difficult to obtain for two reasons. First, the matrix is not positive definite, which is usually addressed by blending with the pedigree relationship matrix (VanRaden, 2008); second, the cost is cubic for computing and quadratic for storage when dense matrix techniques are applied. Consequently, computations with greater than 100,000 genotyped individuals are expensive.

One way to reduce the cost of computing the inverse of GRM is by exploiting the limited dimensionality of GRM. Assume that the genome is divided into N independent chromosome segments (Stam, 1980) and that breeding values of any N animals contain all the information about the values of the segments. Then, a generalized inverse of the GRM can be obtained by recursion on N animals, mimicking the algorithm to invert the pedigree relationship matrix (**A**) (Henderson, 1976; Quaas, 1988). In the Algorithm for Proven and Young (**APY**) (Misztal, 2016), the recursion is on N animals called core and involves the remaining animals called noncore. In simulation tests, the prediction accuracy using APY approached the maximum when N was set to $4N_eL$, the formula described by Stam (1980), where N_e is effective population size and L is genome length in Morgans. The number of segments was approximately equal to the number of the largest eigenvalues of GRM explaining 98% of its variation (a number hereinafter referred to as eigen98), suggesting that the remaining 2% is noise (Pocrnic et al., 2016a). Studies with field datasets agreed with simulation results, and the number of independent chromosome segments ranged from about 4,000 for pigs and chickens to about 16,000 for Holsteins (Pocrnic et al., 2016b). In a multibreed pig population, the addition of crossbred animals did not increase the number of independent chromosome segments (Pocrnic et al., 2019). The choice of core animals was shown to have a minimal impact on prediction accuracy, with a random choice preferred for computational reasons (Bradford et al., 2017).

The APY algorithm has been used to construct the inverse of GRM successfully for up to 2.3 million genotyped individuals (Masuda et al., 2019). Additionally, it is currently being used in commercial genomic evaluations to reduce computing time for beef (Lourenco et al., 2015b) and dairy (Gonzalez-Peña et al., 2019) cattle, pigs (Pocrnic et al., 2019), and broiler chickens (R. Hawken, Cobb-Vantress, Siloam Springs, AR, personal communication) and was also found useful for crossbred populations (Mäntysaari et al., 2017). However, multiple reports from commercial users of APY have indicated that the use of different sets of animals in the recursions results in isolated changes of genomic estimated breeding value (**GEBV**) of nearly one additive genetic standard deviation (**SD_a**), even though the mean change is <30% of one SD_a. Changes in GEBV with different core animals were also noticed by Stranden et al. (2017) and Mäntysaari et al. (2017).

The stability of GEBV when more data are included is a desirable feature of commercial genetic evaluations. An additional requirement is the least possible change in GEBV over time without additional information. However, GEBV fluctuations over time are artifacts of limited accuracy or reliability of individual GEBV. For example, if the reliability of GEBV is 90%, the prediction error variance (**PEV**) is 10% of the additive genetic variance. Assuming a normal distribution, the mean difference from true breeding value is about 0.3 SD_a (i.e., $\sqrt{0.1}$ SD_a), whereas the probability of observing changes that are ±1 SD_a and ±1.5 SD_a is one in 1,000 animals and one in 1,000,000, respectively.

The objectives of this study were to 1) develop a theory to explain GEBV changes when recursion in APY is done on different sets of animals; 2) determine the magnitude of changes in commercial populations of pigs and beef and dairy cattle; 3) determine the magnitude of changes in relation to the accuracy of GEBV; 4) propose steps to reduce the changes, 5) find out whether larger changes of GEBV over time are properties of genomic predictions; and, if so, 6) suggest ways to cope with increased changes.

Materials and Methods

Theoretical aspects of APY

The APY is based on recursion of breeding values $\mathbf{u}$ of noncore (n) on core (c) animals (Misztal, 2016):

$$\mathbf{u}_n = \mathbf{P}_{nc}\mathbf{u}_c + \boldsymbol{\varepsilon},$$

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

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

Following, the GRM ($\mathbf{G}$) is:

$$\mathrm{var}(\mathbf{u}) = \mathbf{G}\sigma_a^2 = \begin{bmatrix}\mathbf{I} & 0 \\ \mathbf{P} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathrm{var}(\mathbf{u}_c) \\ \mathrm{var}(\boldsymbol{\varepsilon})\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{P} \\ 0 & \mathbf{I}\end{bmatrix},$$

where σ_a^2 is additive variance, and the inverse of GRM is:

$$\mathbf{G}^{-1} = \begin{bmatrix}\mathbf{I} & -\mathbf{P}' \\ 0 & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathrm{var}(\mathbf{u}_c)^{-1} & 0 \\ 0 & \mathrm{var}(\boldsymbol{\varepsilon})^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{I} & 0 \\ -\mathbf{P} & \mathbf{I}\end{bmatrix}/\sigma_a^2$$

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

$$\mathbf{G}^{-1} = \begin{bmatrix}\mathbf{G}_{cc}^{-1} & 0 \\ 0 & 0\end{bmatrix} + \begin{bmatrix}-\mathbf{G}_{cc}^{-1}\mathbf{G}_{cn} \\ \mathbf{I}\end{bmatrix}\mathbf{M}^{-1}\left[-\mathbf{G}_{nc}\mathbf{G}_{cc}^{-1}\mathbf{I}\right].$$

In general, the matrix $\mathbf{M}$ is not diagonal. Therefore, ignoring its off-diagonal elements may result in an approximation and, consequently, in a reduction in the accuracy of (ss)GBLUP. However, when the number of core animals is equal or greater than eigen98, that accuracy no longer increases, indicating that the off-diagonal elements of $\mathbf{M}$ no longer carry useful information and can be ignored (Bradford et al., 2017).

For noncore animal i:

$$\mathbf{u}_i = \mathbf{P}_{ic}\mathbf{u}_c + \boldsymbol{\varepsilon}_i, \quad (1)$$

And the variances are $\text{var}(u_i) \approx \sigma_a^2$, $\text{var}(\mathbf{P}_{ic}\mathbf{u}_c) \approx \eta\sigma_a^2$, and $\text{var}(\varepsilon_i) \approx (1-\eta)\sigma_a^2$, where η is the fraction of variance considered in GRM, and the approximation is due to ignored inbreeding and assuming the same η for all animals. In a study by Pocrnic et al. (2016a), the optimal η for maximizing prediction accuracy was 0.98.

For an animal with reliability r^2, $var(\widehat{u_i}) = r^2\text{var}(u_i)$. Then, $var(\widehat{u_i}) = r^2\text{var}(\mathbf{P}_{ic}\mathbf{u}_c) + r^2\text{var}(\varepsilon_i)$ and $\text{var}(\widehat{\varepsilon_i}) \approx r^2(1-\eta)\sigma_a^2$. When the recursion formula is used for young animals (i.e., without phenotypes and offspring), the variance of the difference between GEBV for the same animal obtained with two different core sets is at most $2r^2(1-\eta)\sigma_a^2$ when the errors are uncorrelated and are smaller otherwise; the errors are likely less correlated when the two core sets have minimum overlapping, which is likely the case for a very large number of genotyped animals. Therefore, the changes are mainly attributable to the amount of ε in the recursion to compute $\mathbf{u}$ for noncore animals.

The above formula can be used to predict absolute mean change for all noncore animals or for specific outliers when the number of core animals increases and consequently $\text{var}(\mathbf{P}_{ic}\mathbf{u}_c)$ increases. The mean of outliers for the absolute changes can be calculated based on the normal distribution, assuming a random variable x from a standardized normal distribution with mean 0 and variance 1. If the mean for |x| is 0.8, then the mean for the most extreme outliers of 1 in a 100 is 2.9, for 1 in 10,000 is 4.1, and for 1 in 1 million is 5.1. Based on that, theoretical trends for mean absolute change for all animals and for outliers can be drawn together with prediction accuracies (Figure 1). Prediction accuracy increases to a point ($\eta = 0.98$) and then slightly decreases. At that point, if prediction accuracy is assumed to be 0.6, the mean change is 0.12 SD_a, but it increases to about 0.6 SD_a for the top 0.01% outliers. Increasing the number of core animals to that corresponding to the number of eigenvalues that explains 99.5% of variance ($\eta = 0.995$) reduces mean change by half.

Data and models

Three commercial datasets were used to investigate changes in GEBV from ssGBLUP when using APY with different core sets: a dairy cattle dataset from Holstein Association USA and the Council on Dairy Cattle Breeding, a beef dataset from the American Angus Association, and a pig dataset from GENUS-PIC. The dairy data set used a single-trait (udder depth) version of the 18 type trait, repeatability model, and validation similar to Tsuruta et al. (2019). The beef dataset used a single-trait (postweaning gain) version of a three-trait model shown in (Lourenco et al., 2015b), with similar validation. The pig data set used the single-trait (trait 1) version of a model described in Pocrnic et al. (2019), using the same validation method. The three data sets are described in Table 1.

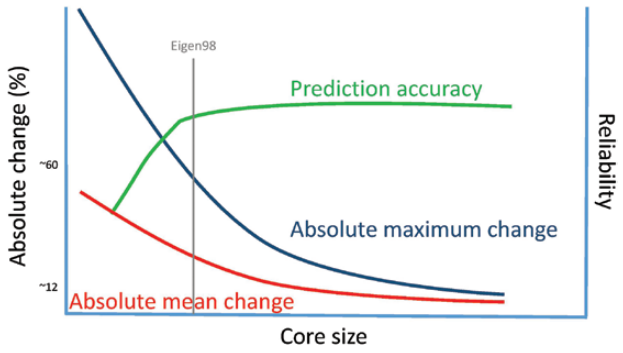

Figure 1. Theoretical mean change in GEBVs and accuracy for two evaluations with a different core set. Mean change was for all animals and maximum change included the 0.01% of animals with the largest change. Eigen98 indicates the number of core animals equivalent to the number of eigenvalues that explain 98% of the variance of the GRM. The approximate values of 12% and 60% were derived assuming the reliability of 0.6.

Analyses

For all the datasets, GEBVs were computed using ssGBLUP with APY and the BLUP90IOD program (Misztal et al., 2014) using default settings, which included the construction of **G** based on VanRaden (2008), blending of **G** with 5% of the pedigree relationship matrix for genotyped animals ($\mathbf{A}_{22}$) to avoid singularity problems, and rescaling of **G** to match $\mathbf{A}_{22}$ by means of diagonals and off-diagonals.

According to Pocrnic et al. (2016b), the number of core animals needed in APY to maximize prediction accuracy of GEBV should be equivalent to the number of eigenvalues that explain 98% of the variance of GRM. For the dairy, beef, and pig single-breed datasets, that number was 15, 13, and 6.9 k, respectively. However, to investigate the effect of core size on GEBV change, several core sizes were used, and each core size was randomly sampled twice ($core_1$ to $core_2$) to compute two sets of GEBV ($GEBV_1$ and $GEBV_2$). For animal i, the absolute change (Δ) was computed as:

$$\Delta_i = |GEBV_{1_i} - GEBV_{2_i}|.$$

Mean absolute change (or difference) between $GEBV_1$ and $GEBV_2$ and maximum absolute change were obtained for all datasets. In addition to investigating mean and maximum changes between $GEBV_1$ and $GEBV_2$ for each core size, mean and maximum changes in $\text{var}(\varepsilon) = \mathbf{M} = \text{diag}(g_{ii} - g_{i,c}\mathbf{G}_{cc}^{-1}g_{c,i})$ were also evaluated for the dairy cattle data.

In the beef data, Δ_i were compared with GEBV accuracy and possible changes based on a 95% confidence interval. Accuracy of GEBV was computed using accf90GS (Tsuruta et al., 2016) for a core size of 20 k, which has just above the number of eigenvalues that correspond to 98% of the variation in GRM. Accuracy of GEBV (*acc*) was then backsolved to the standard error of prediction (*SEP*):

$$SEP = \sqrt{(1 - acc^2)\,\sigma_a^2}.$$

The possible change based on a 95% confidence interval was computed as 1.96 times *SEP* and reported as absolute values.

To find out whether larger changes in predictions over time are intrinsic to genomic evaluations, we mimicked the real scenario where official evaluations are run twice a year (i.e., evaluations A and B). The only difference between evaluations A and B was the number of phenotypes for the beef cattle population in the evaluation system, that is, 124,794 records were added. Solutions from evaluations A and B were obtained using both ssGBLUP and BLUP. Changes in solutions were compared for all genotyped animals.

For the pig dataset, mean and maximum absolute changes were also investigated when different blending was used. By default, GRM in ssGBLUP is blended with 5% of the portion of the **A** for genotyped animals ($\mathbf{A}_{22}$) to avoid singularity problems. To show that changes in GEBV can happen when small modifications are made to GRM, regular ssGBLUP without APY

was used with GRM blended with 1%, 5%, 10%, and 20% of $\mathbf{A}_{22}$ and 1 and 10% of **I**.

Results and Discussion

Trends for mean and maximum absolute changes between $GEBV_1$ and $GEBV_2$ as a function of the number of core animals for Holsteins, Angus, and pigs are shown in Figure 2a–c. In general, the predictive ability or reliability at the number of core animals equal to eigen98 (denoted as a gray, vertical line) was close to the peak, as expected from the theory (Pocrnic et al., 2016a). At that number, the average change for animals being noncore in the random sets was about 7% of SD_a for Holsteins and about 10% of SD_a for Angus and pigs. The maximum changes were 45% of SD_a for Holsteins and about 60% of SD_a for Angus and pigs. These numbers were close to values presented in the theoretical aspects of APY, with somewhat smaller changes for Holsteins, possibly due to the presence of animals with higher average reliability. In fact, the observed ratio of the maximum change to the average change of around 6 is very close to the theoretical ratio of the mean of the 1 in 1 million cases with the mean change.

Mean changes decreased when increasing the number of core animals. Therefore, one way to reduce the changes in APY when the core animals change is to increase the core size beyond the number of eigenvalues that explain 98% of the variance in GRM. However, using more core animals requires increased computing resources without increased prediction accuracy or reliability. For example, in the Angus evaluation, the genomic setup that included computation of $\mathbf{G}^{-1}_{APY}$ took 8.5 h

Table 1. Number of animals in the pedigree, phenotypes available in the complete and partial datasets, total number of genotyped animals and animals in the validation population, SNP count, and genetic parameters for the three datasets used in this study

Data	Pedigree	Phenotypes complete	Phenotypes partial	Genotyped animals	Validation animals	Number of SNPs	h^2	σ^2_a
Holstein	9,730,943	10,067,745	9,235,355	569,404	1,711	58,990	0.33	21.2
Angus	10,661,517	4,386,184	4,298,754	509,072	33,318	39,774	0.24	729
Pigs	2,633,266	771,828	756,973	43,195	2,757	43,195	0.27	1,565

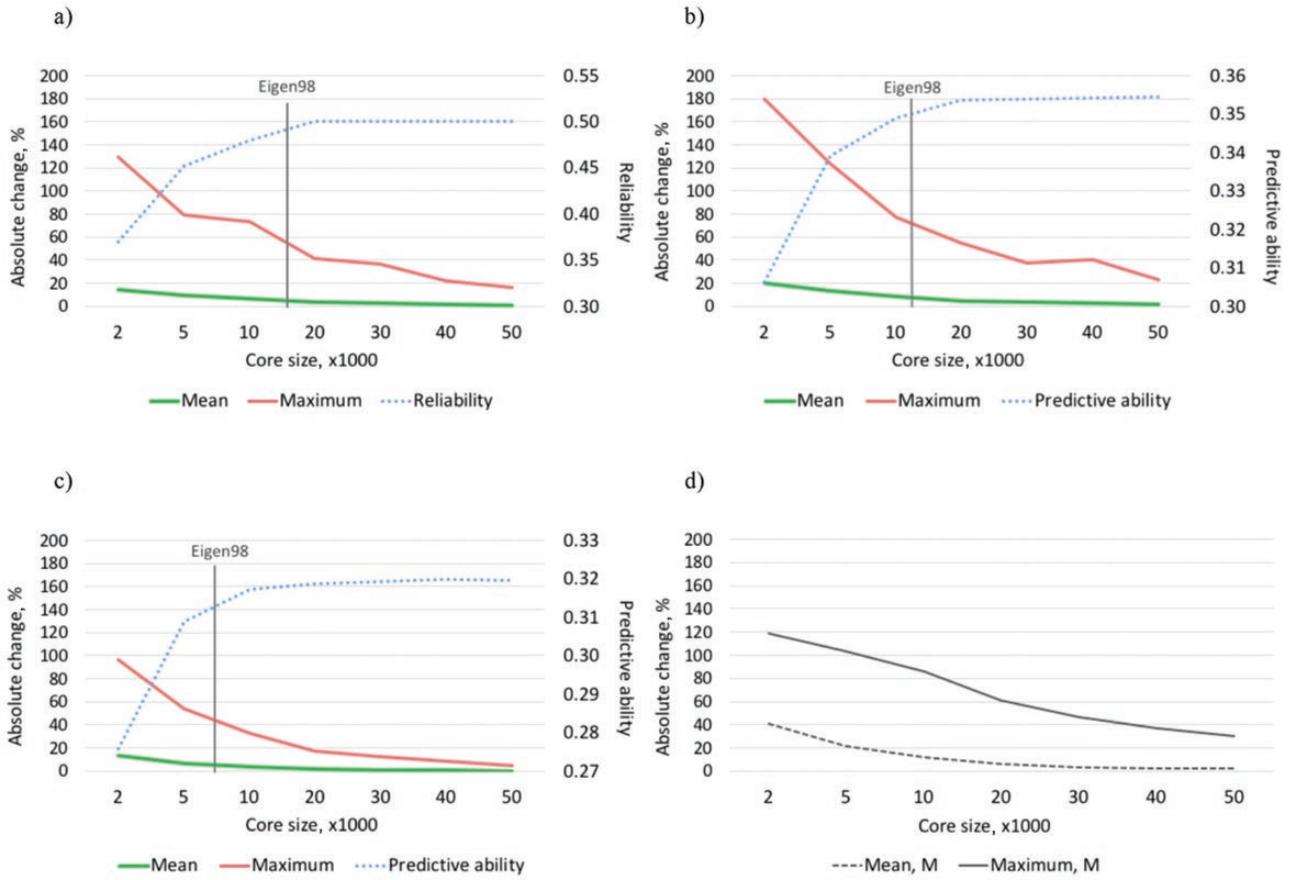

Figure 2. Changes in GEBV and reliability or predictivity for (a) Holstein udder depth, (b) Angus postweaning gain, and (c) pig trait 1 using different core groups. Mean and maximum absolute changes were expressed as a percentage of the additive genetic standard deviation of (a) 4.6, (b) 27.0, and (c) 39.5. Eigen98 indicates the number of core animals equivalent to the number of eigenvalues that explain 98% of the variance of the GRM; (d) absolute changes in the variance of estimation error for Holstein udder depth with increasing core animals. Error was estimated for the move from one core to another, and absolute change was expressed as a percentage of the additive genetic standard deviation of 4.60. **M**, relative variance of estimation error.

when using a 20 k core and 21.5 h with a 50 k core; however, the increase in number of iterations to reach convergence was not extreme (i.e., from 745 to 855).

For the animals in the noncore group, recursions to compute $\mathbf{G}^{-1}_{APY}$ have the following form:

$$\mathbf{G}^{nn} = \mathbf{M}^{-1}_{nn},$$

where $\mathbf{G}^{nn}$ is the portion of $\mathbf{G}^{-1}_{APY}$ that refers to noncore animals and $\mathbf{M}$ depends on the relationships among core animals and also between core and noncore. As more core animals are used, $g_{i,c}\mathbf{G}^{-1}_{cc}g_{c,i}$ increases, consequently reducing error and its variance because $1 - \eta$ is reduced. Using Holstein udder depth as an example, trends for absolute changes in $\mathbf{M}$ when moving from $core_1$ to $core_2$ with increasing numbers of core animals are shown in Figure 2d. Mean and maximum changes in $\mathbf{M}$ decreased when the number of core animals increased. Correlations between changes in $\mathbf{M}$ and changes in GEBV for noncore animals were greater than 0.98, which indicated a strong relationship between estimation error and GEBV changes, in agreement with the theory.

To investigate which animals had the largest GEBV changes when the core group was changed, the absolute values for observed numerical changes in Angus GEBV for postweaning gain were compared with GEBV accuracy based on PEV for a core size of 20 k (Figure 3). All the changes were considerably smaller than the possible changes based on a confidence interval of 95% (i.e., 1.96 times *SEP*). The largest differences between $GEBV_1$ and $GEBV_2$ were observed for animals with accuracy lower than 0.7; on the other hand, animals with higher accuracy had noticeably smaller changes. This agrees with the definition of accuracy, which measures the possible change in breeding value when more data are available. Larger changes between $GEBV_1$ and $GEBV_2$ for animals with lower accuracy imply that information is redistributed in $\mathbf{G}^{-1}_{APY}$ when the core group changes. This redistribution is clear in the recursion formula to compute $\mathbf{G}^{-1}_{APY}$ because all coefficients are computed based on relationships for core animals and between core and noncore animals. Therefore, changing core animals modifies relationships in the GRM inverse when fewer core animals are used. After increasing the core size to 50 k, the relationships stabilize, and further increasing core size no longer reduces the changes because they are already minimal.

Changes over time under genomic and nongenomic evaluation

Figure 4 shows the distribution of changes in breeding values for Angus postweaning gain in two BLUP and ssGBLUP subsequent evaluations. Those evaluations mimicked the real scenario where phenotypes were added only to animals that went through the postweaning gain test. The changes were observed for all genotyped animals. In BLUP evaluations, around 58% of the animals had small changes that varied from 0% to 5% of one SD_a. This was true for all genotyped animals, independently of their phenotyping status. For genomic evaluations, 31% of the animals had GEBV that changed from 0% to 5% of one SD_a, whereas 58% changed from 5% to 20% of one SD_a, indicating that changes in GEBV are larger than in EBV when new data are added. In BLUP, there is little or no change in EBV for one animal when no information is added for itself or its close relatives, but there are larger changes when the information is shareable through the pedigree. This is because the reliability of animals in BLUP is lower. In ssGBLUP, there are always changes in GEBV because the GRM allows the added phenotypes to influence all genotyped animals.

With BLUP, changes in genetic evaluations over time are mainly for younger animals with new data, whereas changes for older animals are usually small because the impact of new animals on old animals through pedigree relationships is small. Exceptions would be the result of redefinition of fixed effects that are applicable to older animals, such as contemporary groups, age adjustments, or unknown parent groups (**UPGs**). For example, large EBV fluctuations in Holsteins were traced to the redefinition of contemporary groups every round of evaluation (P. VanRaden, USDA-AGIL, Beltsville, MD, personal communication). Small changes for older animals, regardless of individual accuracy, create the impression that BLUP evaluations are stable.

In genomic evaluations, the addition of new genotyped animals affects evaluations of older genotyped animals because

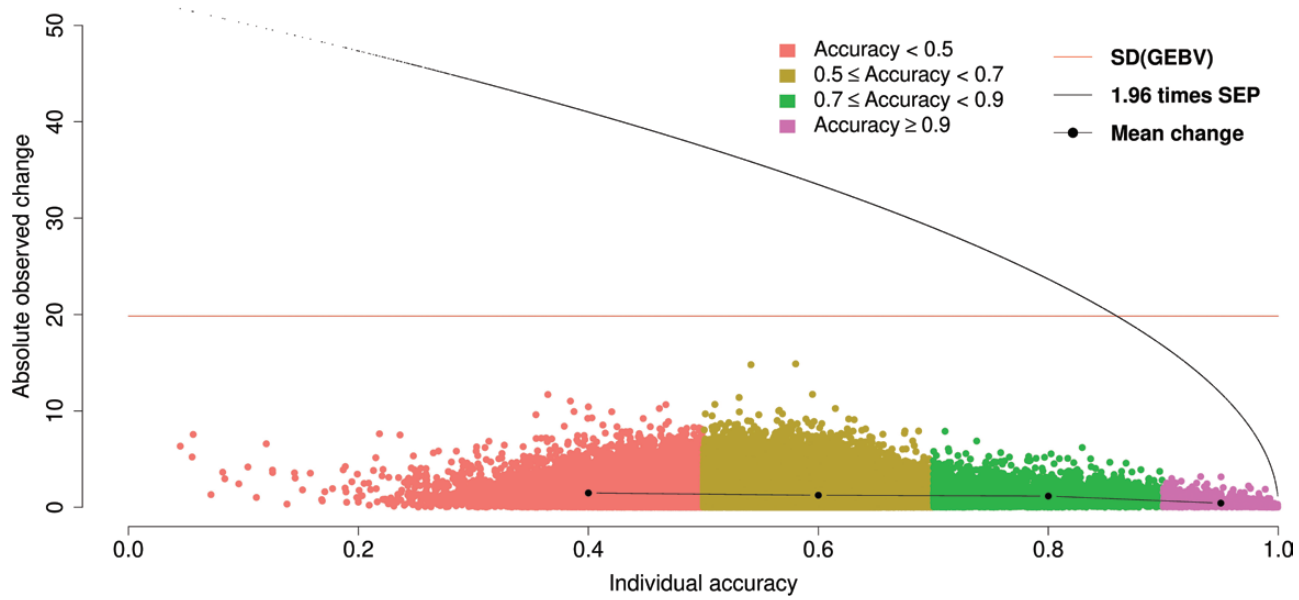

Figure 3. Absolute changes in GEBV for Angus postweaning gain by individual accuracy. Evaluations were based on two different core groups, and accuracy was based on PEV.

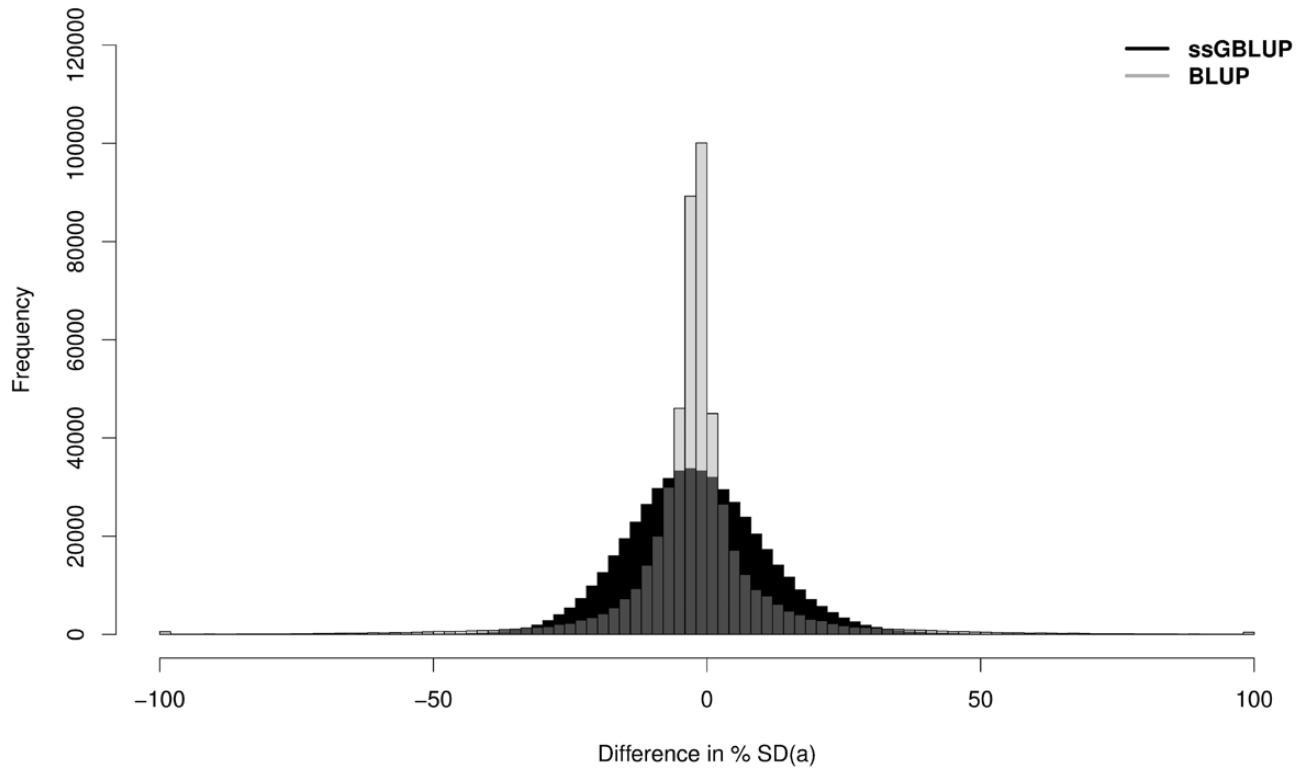

Figure 4. Changes in breeding values for postweaning gain in two subsequent BLUP and ssGBLUP evaluations, where extra 124,794 phenotypes were added from one evaluation to the other. Individual absolute changes were expressed as a percentage of the additive genetic standard deviation of 27.01. Outliers were binned to −100 or 100.

genomic relationships create stronger ties than pedigree relationships. The correlation of GEBV for young bulls between consecutive (4 mo) evaluations is 0.99 (P. VanRaden, USDA-AGIL, Beltsville, MD, personal communication). While the average change in met merit (NM$) for young bulls was about 10% of one SD_a, the maximum change was close to 1.0 SD_a (T. Lawlor, US Holstein Association, Brattleboro, VT, personal communication). Large changes for individual bulls initially ranked as top and priced accordingly create a loss of faith in the genomic evaluations although the changes are in line with individual reliabilities (J. Mabry, Iowa State University, Ames, IA, personal communication).

Changes in GEBV can also happen with small changes in GRM even if APY is not used. For instance, GRM is usually blended to avoid singularity, although the amount of blending is mostly arbitrary, with the fraction of blending varying from 1% to 20%. Figure 5 shows the absolute change in GEBV when using regular $\mathbf{G}^{-1}$ and assuming different blending proportions compared with using the default blending of 5% of $\mathbf{A}_{22}$. Although mean absolute changes were small and varied from 2.8% to 7.5% of one SD_a, the maximum absolute changes reached 44.7% and 87.5% of one SD_a when 20% of $\mathbf{A}_{22}$ and 10% of $\mathbf{I}$ were used instead of 5% of $\mathbf{A}_{22}$, respectively. Despite large changes, correlations between GEBV from different blendings were >0.99. Large individual changes of GEBV can occur with small modifications in genomic relationships. Any modification in the model or variance components also would result in GEBV changes.

Another source of changes of (G)EBV in either BLUP or ssGBLUP are definitions of UPGs if those are based on the year of birth and when a substantial fraction of animals has incomplete pedigree. Over time, new groups are added for new animals, and older ones may be eliminated if the old data are truncated. Subsequently, solutions of UPGs will fluctuate directly influencing (G)EBV.

Methods not using APY formula

Fluctuations due to the choice of core animals do not exist in methods that do not use the APY algorithm. However, GEBV could be different depending on details or approximations in the model. For example, in ssGTBLUP (Mäntysaari et al., 2017), the inverse of GRM is obtained using the Woodbury formula, which requires only the inverse of a matrix with the size of the number of single-nucleotide polymorphisms (**SNP**s). The formula requires full-rank GRM, and such GRM is obtained by blending with an identity matrix or the pedigree relationship matrix among genotyped animals. In practice, the amount of blending does not affect prediction accuracy although blending at 5% reduced the accuracy of GEBV compared with blending at 1% in a simulation study with quantitative trait nucleotide in the data (Fragomeni et al., 2017). As shown before, GEBV with different amount of blending would be slightly different but no changes due to blending will result over time if the amount of blending is unchanged. The ssGBLUP with APY as implemented in the BLUPF90 software uses blending for stability; however, with careful selection of core animals, convergence has been achieved without blending and with marginally increased prediction accuracy (results not provided).

The costs with APY depend on the dimensionality of the genomic information related to the effective population size and less on the number of SNP. The cost of ssGTBLUP depends on the number of SNP. Subsequently, ssGTBLUP is a good choice in situations when the dimensionality of GRM is close to the number of SNP, like for Irish cattle multibreed populations (Mäntysaari et al.,

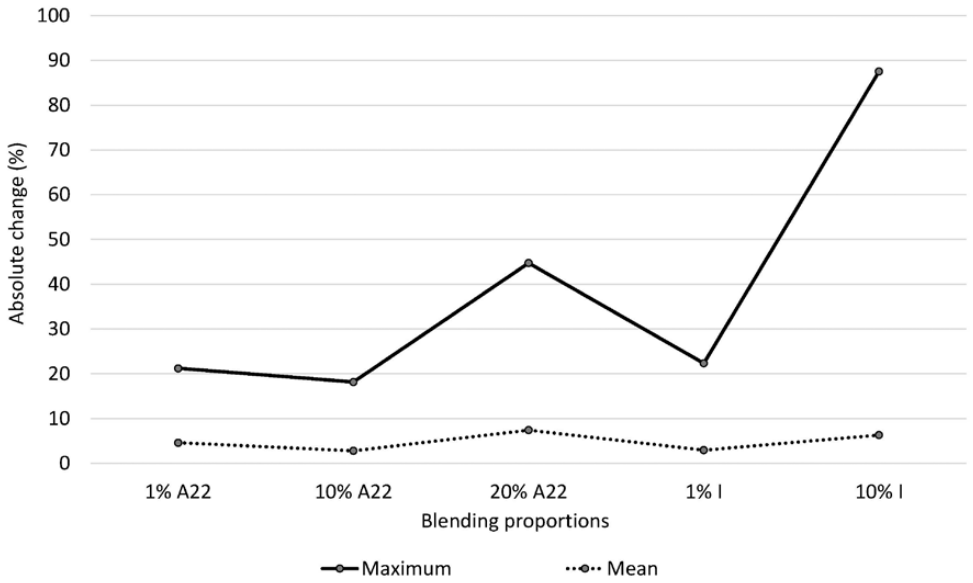

Figure 5. Absolute changes in GEBV for different blending proportions. Genomic evaluations were calculated using the regular inverse of the GRM, and different blending proportions of the portion of the pedigree relationship matrix for genotyped animals and the identity matrix were compared with using 5% of the portion of the pedigree relationship matrix for genotyped animals. Mean and maximum absolute changes were expressed as a percentage of the additive genetic standard deviation of 39.56. $\mathbf{A}_{22}$, portion of the pedigree relationship matrix for genotyped animals; **I**, identity matrix.

2017), although a single relationship matrix may be insufficient for accurate modeling of a large number of breeds (Steyn et al., 2019). The ssGBLUP using APY would be more efficient when the dimensionality is smaller than the number of SNP, or the number of SNPs is very large. Another method that can potentially be used with large datasets is single-step Bayesian regression (Fernando et al., 2014), where the same number of SNP effects is estimated regardless of the number of genotyped animals.

How to minimize the change in GEBV when APY is used

When the evaluation is performed frequently, for example, weakly, with little additional data but different core sets every time, most of the changes are due to changing the core set. Several options are available to try to minimize GEBV changes or their effect when the APY algorithm is used. The first option is to ignore changes, as they do not affect prediction accuracy and subsequently genetic gain and trends because GEBVs obtained with different cores sizes are correlated at >0.99 when each core size is at least eigen98. Such an option is obvious for species where individual nucleus animals are not marketed (chicken and mostly pigs). A second option is to use the same core animals for an extended period of time (e.g., 1 yr); using core animals from previous generations in the simulation study (Bradford et al., 2017) or the core set 2 yr old in Angus (results not shown) did not reduce prediction accuracy. Finally, with large data when the APY is useful, young animals may be removed from the main ssGBLUP evaluation and their GEBV calculated as indirect predictions based on backsolved SNP effects (Lourenco et al., 2015), also for an extended period of time. In general, GEBVs include both parent average and genomic predictions (VanRaden and Wiggans, 1991; Lourenco et al., 2015a), and the use of indirect predictions that exclude parent average can lead to lower prediction accuracy. With large data, the fraction of parent average in GEBV is small, and indirect predictions based on SNP effects are accurate. In the Angus data, correlations of indirect predictions with regular GEBV calculated up to 1 yr later were >0.99 (Hidalgo et al., 2020).

Overall, fluctuations in GEBV due to either additional genomic data or computing details are not necessarily detrimental but instead illustrate limited reliability and discretion in extensively using an animal as a breeder until its reliability increases. In fact, low fluctuations in EBV for an animal with low reliability may falsely create an impression that the EBVs are reliable while they are not. When GEBV reaches high reliability (i.e., when the animal has many progenies and phenotypic records), the portion of GEBV that is contributed by genomic information is negligible. Therefore, the expected changes in GEBV when $\mathbf{G}^{-1}_{APY}$ is modified are close to zero.

Even without fluctuations induced by APY, EBV and GEBV changes are inevitable with model adjustments and additional data. An educational effort is needed to explain changes in GEBV for genomically evaluated animals because of the industry trend to put a high economic value on animals tentatively evaluated as best (Lourenco et al., 2015b). One way, the dairy industry deals with fluctuations in GEBV is by marketing a group instead of individual sires (Gottardo et al., 2019). If each sire has a repeatability of 70%, a mean of 30 sires, assuming independent errors, would have a repeatability of 99%. Using groups of top-ranked sires instead of individual sires has a positive impact on diversity.

Conclusions

Fluctuations in GEBV from ssGBLUP with different core animals are due to the error term that models noise in GRM. On average, the GEBV changes are small relative to the possible changes based on SEP, although outliers are often observed. The fluctuations can be minimized, without affecting prediction accuracy, by keeping the same set of core animals over extended periods of time or by using indirect predictions based on periodically calculated SNP effects for young animals. Regardless of methodology, genomic predictions appear less

stable than those by BLUP because the genomic relationships tie all the genotyped animals. Outliers can have changes in GEBV that are several times larger than the average changes, and the reasons for extreme values include adding of new data and changing contemporary groups, among others. Problems may arise when outliers are ranked as top animals and priced ignoring the accuracy of GEBV and subsequently later changes. The best way to deal with the changes/fluctuations is to base breeding programs on groups of animals instead of on top-ranked individual animals solely.

Acknowledgments

This study was supported by grants from Holstein Association USA, American Angus Association, Pig Improvement Company, and USDA National Institute of Food and Agriculture (Agriculture and Food Research Initiative competitive grant 2020-67015-31030). We thank Steve Miller and Dan Moser from Angus Genetics Inc., Ching-Yi Chen and William Herring from Pig Improvement Company, Tom Lawlor from Holstein Association USA Inc., Paul VanRaden from AGIL-USDA, ARS-USDA, and CDCB for providing data. Editing by Suzanne Hubbard is gratefully acknowledged.

Conflict of interest statement

The authors declare no real or perceived conflicts of interest.

Literature Cited

Bradford, H. L., I. Pocrnic, B. O. Fragomeni, D. A. L. Lourenco, and I. Misztal. 2017. Selection of core animals in the Algorithm for Proven and Young using a simulation model. *J. Anim. Breed. Genet* **134**:545–552. doi:10.1111/jbg.12276

Fernando, R. L., J. C. Dekkers, and D. J. 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. doi:10.1186/1297-9686-46-50

Fragomeni, B. O., D. A. L. Lourenco, Y. Masuda, and I. Misztal. 2017. Incorporation of causative quantitative trait nucleotides in single-step GBLUP. *Genet. Sel. Evol.* **49**:1. doi:10.1186/s12711-017-0335-0

Gonzalez-Peña, D., N. Vukasinovic, J. J. Brooker, C. A. Przybyla, and S. K. DeNise. 2019. Genomic evaluation for calf wellness traits in Holstein cattle. *J. Dairy Sci.* **102**:2319–2329. doi:10.3168/jds.2018-15540

Gottardo, P., G. Gorjanc, M. Battagin, R. C. Gaynor, J. Jenko, R. Ros-Freixedes, C. Bruce A. Whitelaw, A. J. Mileham, W. O. Herring, and J. M. Hickey. 2019. A strategy to exploit surrogate sire technology in livestock breeding programs. *G3 (Bethesda)*. **9**:203–215. doi:10.1534/g3.118.200890

Henderson, C. R. 1976. A simple method for computing the inverse of a relationship matrix used in prediction of breeding values. *Biometrics* **32**:69–83. doi:10.2307/2529339

Hidalgo, J., D. Lourenco, S. Tsuruta, S. Miller, A. Garcia, Y. Masuda, and I. Misztal. 2020. Changes in genomic predictions when new data is included. *J. Anim. Sci.* **98**(Suppl. 4):7–8. doi:10.1093/jas/skaa278.014

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

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

Mäntysaari, E. A., R. D. Evans, and I. Strandén. 2017. Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals. *J. Anim. Sci.* **95**:4728–4737. doi:10.2527/jas2017.1912

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

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

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. *Manual for BLUPF90 family of programs*. Available from http://nce.ads.uga.edu/ wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Pocrnic, I., D. A. L. Lourenco, C. Y. Chen, W. O. Herring, and I. Misztal. 2019. Crossbred evaluations using single-step genomic BLUP and algorithm for proven and young with different sources of data. *J. Anim. Sci.* **97**:1513–1522. doi:10.1093/jas/skz042

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

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

Quaas, R. L. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* **71**:1338–1345.

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

Steyn, Y., D. A. L. Lourenco, and I. Misztal. 2019. Genomic predictions in purebreds with a multibreed genomic relationship matrix. *J. Anim. Sci.* **97**:4418–4427. doi:10.1093/jas/skz296

Stranden, I., K. Matilainen, G. P. Aamand, and E. A. Mantysaari. 2017. Solving efficiently large single-step genomic best linear unbiased prediction models. *J. Anim. Breed. Genet.* **134**:164–274. doi:10.1111/jbg.12257

Tsuruta, S., D. A. L. Lourenco, Y. Masuda, I. Misztal, and T. J. Lawlor. 2019. Controlling bias in genomic breeding values for young genotyped bulls. *J. Dairy Sci.* **102**:9956–9970. doi:10.3168/jds.2019-16789

Tsuruta, S., D. Lourenco, Y. Masuda, D. W. Moser, and I. Misztal. 2016. Practical approximation of accuracy in genomic breeding values for a large number of genotyped animals. *J. Anim. Sci.* **94**(Suppl. 5):162. doi:10.2527/jam2016-0337

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

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