Efficient single-step genomic evaluation for a multibreed beef cattle population having many genotyped animals[1]

E. A. Mäntysaari,*[2] R. D. Evans,† and I. Strandén*

*Natural Resources Institute Finland (Luke), Green Technology, FI-31600 Jokioinen, Finland;
and †Irish Cattle Breeding Federation, Highfield House, Newcestown Road, Bandon, Cork, Ireland

ABSTRACT: An equivalent computational approach called ssGTBLUP was formulated for the original single-step GBLUP (ssGBLUP). In ssGTBLUP, the genomic relationship matrix has the form $\mathbf{G} = \mathbf{ZZ}' + \mathbf{C}$, where the (centered and scaled) **Z** marker matrix has size n x m (numbers of genotypes and markers), and the **C** matrix can be easily inverted. The inverse can be written as $\mathbf{G}^{-1} = \mathbf{C}^{-1} - \mathbf{T}'\mathbf{T}$ where **T** is an m by n matrix. When the preconditioned conjugate gradient (PCG) method is used to solve the mixed model equations, a matrix vector product $\mathbf{G}^{-1}\mathbf{d}$ needs to be computed. In ssGBLUP, this requires n^2 multiplications, but in ssGTBLUP, the product $\mathbf{T}'\mathbf{Td}$ has $2nm$ multiplications and $\mathbf{C}^{-1}\mathbf{d}$ has cn multiplications with the constant c independent of n or m. In an approximate approach called ssGTBLUP(p), the eigendecomposition of $\mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z}$ is used to reduce the number of rows in the **T** matrix. Here, p is the percentage of total variance explained by the accepted eigenvalues. The objective of this study was to compare the performance of ssGBLUP, ssGTBLUP, ssGTBLUP(p), and the APY (algorithm for proven and young) method. In APY, the core had 50,000 (APY50K), 30,000 (APY30K), or 10,000 (APY10K) animals. The approaches were tested on the Irish beef carcass conformation genetic evaluation which has a heterogeneous multibreed population. The pedigree had 13.3 million animals. There were m = 54,620 markers available from n = 163,277 genotyped animals. For genotyped animals, the correlations of breeding values between ssGBLUP and ssGTBLUP(p) for the 11 traits in the model ranged from 0.999–1.000 for p = 99, 0.998–1.000 for p = 98, and 0.992–0.998 for p = 95 but were 0.994–1.000 for APY50K, 0.969–0.997 for APY30K, and 0.899–0.967 for APY10K. Computing times per iteration were 4.43, 3.30, 2.69, 2.29, 1.55, 1.76, 1.27, and 0.55 min for ssGBLUP, ssGTBLUP, ssGTBLUP(99), ssGTBLUP(98), ssGTBLUP(95), APY50K, APY30K, and APY10K, respectively. The ssGTBLUP(p) approach allowed a well-defined approximation to ssGBLUP and fast computations

Key words: beef cattle, breeding value, genomic evaluation

J. Anim. Sci. 2017.95:4728–4737
doi:10.2527/jas2017.1912

INTRODUCTION

Single-step genomic BLUP (**ssGBLUP**) is the preferred method for genomic evaluations (Aguilar et al., 2010; Christensen and Lund, 2010) because it includes genomic and pedigree information simultaneously and, thus, tries to avoid bias in estimation of breeding values. A central challenge in solving single-step mixed model equations (**MME**) is the inversion of a genomic relationship matrix **G**. First, the computational cost of matrix inversion increases cubically with the number of genotyped animals. Second, the number of floating point operations due to the $\mathbf{G}^{-1}$ matrix in the iterative solving of the MME increases quadratically with the number of genotyped animals. Typically, solving time of the MME becomes dominated by the $\mathbf{G}^{-1}$ matrix when the number of genotyped animals surpasses 30,000.

To ease these computational challenges Legarra and Ducrocq (2012) presented several equivalent single-step MME, some of which avoid inversion of **G**, or even making it. Some alternative formulations

[1]The authors acknowledge the Irish Cattle Breeding Federation (ICBF) for providing the data.
[2]Corresponding author: esa.mantysaari@luke.fi
Received July 12, 2017.
Accepted September 10, 2017.

account for the genomic information through marker effects (e.g., Liu et al., 2014; Fernando et al., 2014, 2016; Taskinen et al., 2017). Misztal et al. (2014) and Fragomeni et al. (2015) suggested using an approximate sparse inverse of the matrix called the algorithm for proven and young (**APY**) which allow computations for large genotyped populations (Masuda et al., 2016; Strandén et al., 2017).

In iterative solving by the preconditioned conjugate gradient (**PCG**) method the original single-step MME has typically required the least number of iterations (Legarra and Ducrocq, 2012; Strandén and Mäntysaari, 2014). We propose an exact approach named ssGTBLUP that has the same convergence properties in PCG as the original MME, but is computationally less demanding. Solving time can be further reduced by an approximation using eigendecomposition with rank reduction. The approaches are tested on a multibreed beef cattle evaluation and compared with the original ssGBLUP with or without APY.

MATERIALS AND METHODS

MME for ssGBLUP

Consider a univariate ssGBLUP model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wa} + \mathbf{e},$$

where incidence matrix **X** relates fixed effects **b** and incidence matrix **W** relates breeding values **a** to appropriate observations in vector **y**, and **e** is random residual vector. Assume that $\text{Var}(\mathbf{e}) = \mathbf{R}\sigma^2_e$, where **R** is a positive definite matrix, such as the identity matrix **I,** or a diagonal matrix with weights (VanRaden, 2008). In ssGBLUP, covariance structure for the breeding values is $\text{Var}(\mathbf{a}) = \mathbf{H}\sigma^2_a$, where σ^2_a is the genetic variance and **H** has both pedigree (**A**) and genomic (**G**) relationship matrix information (Aguilar et al., 2010; Christensen and Lund, 2010).

Animals can be assigned to 2 groups: group 1 has non-genotyped animals; group 2 has genotyped animals. Then, the relationship matrix and its inverse can be expressed by submatrices using these 2 groups:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \text{ and } \mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix}.$$

Mixed model equations for the ssGBLUP are

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{W} \\ \mathbf{W}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \lambda\mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}y \\ \mathbf{W}'\mathbf{R}^{-1}y \end{bmatrix}, \quad [1]$$

where $\lambda = \frac{\sigma^2_e}{\sigma^2_a}$ and $\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - (\mathbf{A}_{22})^{-1} \end{bmatrix}$.

Introducing ssGTBLUP

Solving MME like [1] by PCG iteration requires a coefficient matrix of the MME times a vector product. When an iteration on data approach is used, this product can be calculated easily for all the other parts (e.g., Strandén and Lidauer, 1999; Strandén et al., 2017), except $\mathbf{G}^{-1}$ which requires making and inverting the **G** matrix.

Consider a genomic relationship matrix of form

$$\mathbf{G}_0 = \mathbf{ZZ}',$$

where $\mathbf{Z} = (\mathbf{M} - \mathbf{P})(c\mathbf{B})^{0.5}$ is an $n \times m$ matrix of centered and scaled marker genotypes, **M** is an $n \times m$ matrix of marker genotypes, $\mathbf{P} = 2\mathbf{1}_n \mathbf{p}'$ is an $n \times m$ matrix for centering, **p** is an $m \times 1$ vector of allele frequencies, c is a scaling constant, and **B** is an $m \times m$ diagonal scaling matrix. Each genotype value in the **M** matrix is a count of the first allele. Thus, a homozygous genotype for the first allele has value 0, heterozygote has 1, and homozygous for the second allele has 2. Base population allele frequencies are typically used in the **p** vector (VanRaden, 2008). In VanRaden (2008) method 1, the scaling matrix $\mathbf{B} = \mathbf{I}$ and the scaling constant $c = \frac{1}{k}$, where $k = 2\sum_{i=1}^{m} p_i(1 - p_i)$.

Matrix $\mathbf{G}_0$ is singular when the allele frequencies are averages from the genotyped animals in matrix **M**, i.e., $\mathbf{p} = \frac{1}{2n}\mathbf{M}'\mathbf{1}_n$ (e.g., Strandén and Christensen, 2011), or always when $n > m$. To guarantee nonsingularity of the **G** matrix, a regularization matrix **C** can be added:

$$\mathbf{G}_C = \mathbf{G}_0 + \mathbf{C}.$$

For example, $\mathbf{C} = \varepsilon\mathbf{I}$, where ε is a small number. Or, $\mathbf{C} = w\mathbf{A}_{22}$, where w is the proportion of polygenic variance not accounted by the markers, and the proportion of variance accounted by the markers (1-w) can be included in the scaling constant c in $\mathbf{G}_0$.

Inverse of the $\mathbf{G}_C$ matrix can be calculated using the matrix inversion lemma (Householder, 1964):

$$\mathbf{G}_C^{-1} = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{Z}\left(\mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z} + \mathbf{I}\right)^{-1}\mathbf{Z}'\mathbf{C}^{-1}. \quad [2]$$

This can be written

$$\mathbf{G}_C^{-1} = \mathbf{C}^{-1} - \mathbf{T}_C'\mathbf{T}_C, \quad [3]$$

where $\mathbf{T}_C = \mathbf{L}_C^{-1}\mathbf{Z}'\mathbf{C}^{-1}$, and the lower triangle matrix $\mathbf{L}_C$ is the Cholesky decomposition of $\mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z} + \mathbf{I}$, i.e., $\mathbf{L}_C\mathbf{L}_C' = \mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z} + \mathbf{I}$.

Consider 2 special cases of the **C** matrix. When $\mathbf{C} = \varepsilon\mathbf{I}$, we have

$$\mathbf{G}_{\varepsilon}^{-1} = \frac{1}{\varepsilon}\mathbf{I} - \frac{1}{\varepsilon^2}\mathbf{Z}\left(\frac{1}{\varepsilon}\mathbf{Z}'\mathbf{Z} + \mathbf{I}\right)^{-1}\mathbf{Z}',$$

which can be written

$$\mathbf{G}_{\varepsilon}^{-1} = \frac{1}{\varepsilon}\mathbf{I} - \mathbf{T}_{\varepsilon}'\mathbf{T}_{\varepsilon},$$

where $\mathbf{T}_{\varepsilon} = \frac{1}{\varepsilon}\mathbf{L}_{\varepsilon}^{-1}\mathbf{Z}'$, and the lower triangle matrix $\mathbf{L}_{\varepsilon}$ is the Cholesky decomposition of $\frac{1}{\varepsilon}\mathbf{Z}'\mathbf{Z} + \mathbf{I}$, i.e., $\mathbf{L}_{\varepsilon}\mathbf{L}_{\varepsilon}' = \frac{1}{\varepsilon}\mathbf{Z}'\mathbf{Z} + \mathbf{I}$. Similarly, when $\mathbf{C} = w\mathbf{A}_{22}$ we have

$$\mathbf{G}_{w}^{-1} = \frac{1}{w}\mathbf{A}_{22}^{-1} - \frac{1}{w^2}\mathbf{A}_{22}^{-1}\mathbf{Z}\left(\frac{1}{w}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z} + \mathbf{I}\right)^{-1}\mathbf{Z}'\mathbf{A}_{22}^{-1},$$

which can be written

$$\mathbf{G}_{w}^{-1} = \frac{1}{w}\mathbf{A}_{22}^{-1} - \mathbf{T}_{w}'\mathbf{T}_{w},$$

where $\mathbf{T}_{w} = \frac{1}{w}\mathbf{L}_{w}^{-1}\mathbf{Z}'\mathbf{A}_{22}^{-1}$ and the lower triangle matrix $\mathbf{L}_{\mathrm{W}}$ is the Cholesky decomposition of $\frac{1}{w}\mathbf{Z}'\mathbf{A}_{22}^{-1}\mathbf{Z} + \mathbf{I}$.

We call ssGBLUP using formulation [3] in the computations single-step GBLUP with a **T** factoring (**ssGTBLUP)**. When the PCG method is used to solve ssGTBLUP having many genotyped animals, computationally the most demanding step is the product $\mathbf{T}_{\mathrm{C}}'\mathbf{T}_{\mathrm{C}}\mathbf{d}_2$ where the $\mathbf{T}_{\mathrm{C}}$ is an *m* by *n* matrix. Computing this product in 2 steps from right to left requires 2*mn* multiplications. When the number of genotyped animals *n* is more than twice the number of markers *m*, the multiplication $\mathbf{T}_{\mathrm{C}}'\mathbf{T}_{\mathrm{C}}\mathbf{d}_2$ has less multiplications than $\mathbf{G}^{-1}\mathbf{d}_2$ and, thus, may be faster.

In practice, calculation of the $\mathbf{T}'\mathbf{T}\mathbf{d}_2$ product can be performed like in iteration on data, where the **T** matrix is read from disk and processed 1 line at a time (Strandén and Lidauer, 1999). Let $\mathbf{t}_i$ be row *i* of **T** matrix. The product can be written

$$\mathbf{T}'\mathbf{T}\mathbf{d}_2 = \begin{bmatrix}\mathbf{t}_1' & \mathbf{t}_2' & \cdots & \mathbf{t}_m'\end{bmatrix}\begin{bmatrix}\mathbf{t}_1\\ \mathbf{t}_2\\ \cdots\\ \mathbf{t}_m\end{bmatrix}\mathbf{d}_2 = \sum_{i=1}^{m}\mathbf{t}_i'\left(\mathbf{t}_i\mathbf{d}_2\right),$$

where every row $\mathbf{t}_i$ has *n* elements. Thus, for each line 2 steps are performed: 1) product $\mathbf{t}_i\mathbf{d}_2$ which gives a scalar constant, and 2) the constant times a vector product is accumulated to the result.

Approximation through Eigendecomposition

Low rank approximation of the inverse $\mathbf{G}_{\mathrm{C}}$ of matrix in ssGTBLUP can be derived using the eigen-decomposition. Equation [2] can be written

$$\mathbf{G}_{\mathrm{C}}^{-1} = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{Z}\left(\mathbf{V}\mathbf{D}\mathbf{V}' + \mathbf{I}\right)^{-1}\mathbf{Z}'\mathbf{C}^{-1}, \qquad [4]$$

where $\mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z} = \mathbf{V}\mathbf{D}\mathbf{V}'$ is the eigendecomposition with the diagonal matrix **D** having the eigenvalues in decreasing order of size, and columns of the **V** matrix having corresponding orthogonal eigenvectors of $\mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z}$. Note that both **D** and **V** have size *m*. Equation [4] can be further rewritten

$$\mathbf{G}_{\mathrm{C}}^{-1} = \mathbf{C}^{-1} - \mathbf{T}_{\mathrm{E}}'\mathbf{T}_{\mathrm{E}}, \qquad [5]$$

where $\mathbf{T}_{\mathrm{E}} = \left(\mathbf{D} + \mathbf{I}\right)^{-\frac{1}{2}}\mathbf{V}'\mathbf{Z}'\mathbf{C}^{-1}$.

Mathematically, Eq. [3] and [5] give the same result. However, in [5] a rank reduction can be made to decrease the number of computations. Let $\mathbf{D}_r$ include the *r* highest eigenvalues in **D**, and $\mathbf{V}_r$ corresponding eigenvectors or columns in the **V** matrix. Thus, $\mathbf{D}_r$ is an *r* by *r* matrix, and $\mathbf{V}_r$ is an *m* by *r* matrix. Then, the approximate $\mathbf{T}_{\mathrm{E}}$ matrix can be calculated as $\mathbf{T}_r = \left(\mathbf{D}_r + \mathbf{I}_r\right)^{-\frac{1}{2}}\mathbf{V}_r'\mathbf{Z}'\mathbf{C}^{-1}$, which is an *r* by *n* matrix. Thus, the number of multiplications due to $\mathbf{T}_r'\mathbf{T}_r\mathbf{d}_2$ is 2*rn* instead of 2*mn*.

When $\mathbf{C} = \varepsilon\mathbf{I}$, we have $\mathbf{T}_{\mathrm{E}} = \frac{1}{\varepsilon}\left(\mathbf{D} + \mathbf{I}\right)^{-\frac{1}{2}}\mathbf{V}'\mathbf{Z}'$, where $\frac{1}{\varepsilon}\mathbf{Z}'\mathbf{Z} = \mathbf{V}\mathbf{D}\mathbf{V}'$ with **D** and **V** from the eigende-composition. Now, the rank *r* approximation of the $\mathbf{T}_{\mathrm{E}}$ matrix is

$$\mathbf{T}_r = \frac{1}{\varepsilon}\left(\mathbf{D}_r + \mathbf{I}\right)^{-\frac{1}{2}}\mathbf{V}_r'\mathbf{Z}'.$$

The rank *r* approximation can be similarly calculated for $\mathbf{C} = w\mathbf{A}_{22}$ as well.

Size of rank *r* can be based on the proportion of total variance explained by the accepted eigenvalues. Let d_i be eigenvalue *i* in the diagonal of **D** where the eigenvalues have been ordered in decreasing order, i.e., $\mathrm{d}_i \geq \mathrm{d}_{i+1}$. Then, the proportion explained by *r* eigenvalues is $\frac{\sum_{i=1}^{r} d_i}{tr(\mathbf{D})}$, where $tr(\mathbf{D})$ is the trace or **D** or sum of diagonal elements in **D**.

Data

Genomic evaluation approaches were tested using data from the routine beef carcass conformation genetic evaluation performed by the Irish Cattle Breeding Federation (ICBF). The ssGBLUP was based on the same multitrait animal model and variance components as the current official routine breeding value evaluation described in more detail in Evans et al. (2014) and McHugh et al. (2011). However, our model had no genetic groups and the single-step evaluations included genomic information. The model had 9 correlated traits: WQ = farmer-assessed weanling quality score (scale of 1 to 5), CP = mart calf price (0 to 42 d) from calves sold in single lots through

auction houses, WP = weanling price (150–300 d) from weanlings sold in single lots through auction houses, PWP = post-weanling price (300–600 d) from weanlings sold in single lots through auction houses, MS = muscle composite score for animals linear scored by a trained technician, CC = carcass conformation score from abattoirs on a scale of 1 to 15, CCC = cull cow carcass conformation score from abattoirs on a scale of 1 to 15, MSF = progeny yield deviation for muscle score for foreign sires, CCF = progeny yield deviation for carcass conformation score for foreign sires. The data had 8.33 million animals with records (Table 1). Additional information pertaining to the traits in the model can be found in Pabiou et al. (2012) for WQ and MS; in McHugh et al. (2010) for CP, WP, and PWP; and in Pabiou et al. (2011, 2012) for CC.

Pedigree included 13.35 million animals of which 163,277 were genotyped. Genomic data consisted of genotypes of animals from 41 breeds. Breed proportion over 96% for a breed was observed for 52,566 animals, and breed proportions of 50%–96% for 80,011 animals. The most common pure breeds, having more than 1,000 animals with breed proportion over 96%, were Charolais (n = 16,382), Limousine (n = 16,512), Angus (n = 8,322), Hereford (n = 4,573), and Simmental (n = 3,228). The animals had been genotyped using Illumina Bovine SNP50 Bead Chip (Illumina, San Diego, CA). All the data were provided by ICBF (Cork, Ireland). After the quality edits and imputation of missing markers there were 54,620 markers from 29 bovine autosomes available for the analysis.

Table 1. Trait, number of observations (N), average (Mean), standard deviation (Std), and heritability (h^2) for the analyzed data

Trait[1]	N	Mean	Std	h^2
WQ	1,420,555	3.6	0.8	0.28
CP	117,290	153.3	77.3	0.51
WP	886,755	715.6	180.7	0.44
PWP	840,785	865.0	233.1	0.32
MS	225,997	42.4	7.5	0.41
CC	5,683,908	7.2	2.4	0.35
CCC	1,330,415	3.9	2.2	0.24
MSF	29,078	-1.2	8.5	0.07
CCF	20,428	-0.2	11.7	0.06

[1]WQ = farmer-assessed weanling quality score (scale of 1 to 5), CP = mart calf price (0 to 42 d) from calves sold in single lots through auction houses, WP = weanling price (150–300 d) from weanlings sold in single lots through auction houses, PWP = post-weanling price (300–600 d) from weanlings sold in single lots through auction houses, MS = muscle composite score for animals linear scored by a trained technician, CC = carcass conformation score from abattoirs on a scale of 1 to 15, CCC = cull cow carcass conformation score from abattoirs on a scale of 1 to 15, MSF = progeny yield deviation for muscle score for foreign sires, CCF = progeny yield deviation for carcass conformation score for foreign sires

Solving Single-Step Evaluations

Single-step genomic evaluations were performed by 4 different approaches. In the first approach, the $\mathbf{G}_\varepsilon$ matrix was formed and inverted using LAPACK (Anderson et al., 1999) subroutines (dpotri after dpotrf) available in the MKL library (Intel, 2014). In the second approach, the $\mathbf{G}_\varepsilon$ matrix was not formed fully, but instead its inverse was approximated using $\mathbf{G}^{-1}_{\text{APY}}$ (Fragomeni et al., 2015). In the third approach, the $\mathbf{T}_\varepsilon$ matrix was calculated for the ssGTBLUP approach, and neither $\mathbf{G}_\varepsilon$ nor its inverse was explicitly formed. In the fourth approach, eigendecomposition was used to reduce the number of multiplications during the PCG. The eigendecomposition was computed using a LAPACK subroutine (dsyevr) available in the MKL library. The routine includes a highly optimized algorithm to compute eigenvalues and eigenvectors of a real symmetric matrix. We considered 3 reduced cases where the percentage of total variance explained by the accepted eigenvalues was 99%, 98%, or 95%. In addition to the single-step evaluations, the MME were solved without any genomic information to obtain ordinary animal model based breeding values (**AMBLUP**).

In every approach tested, the genomic relationship matrix was defined as $\mathbf{G}_\varepsilon = \mathbf{ZZ}' + \varepsilon\mathbf{I}$, where $\mathbf{Z}$ was VanRaden (2008) method 1 (centered and scaled) marker genotyped matrix and ε was 10^{-3}. The centering in the $\mathbf{Z}$ matrix used base population allele frequencies which were estimated using the method described in McPeek et al. (2004) and Strandén et al. (2017).

All models were solved using MiX99 software (Strandén and Lidauer, 1999) which uses PCG iteration in solving the MME. The main computational cost in the PCG method is a matrix times a vector product where within each iteration round a so-called direction vector is multiplied by the coefficient matrix. In solving the MME of ssGBLUP this multiplication includes computing $\mathbf{x} = \left(\mathbf{G}_\varepsilon^{-1} - \left(\mathbf{A}_{22}\right)^{-1}\right)\mathbf{d} = \mathbf{Cd}$. The product $\left(\mathbf{A}_{22}\right)^{-1}\mathbf{d}$ was computed by the formula $\left(\mathbf{A}^{22} - \mathbf{A}^{21}\left(\mathbf{A}^{11}\right)^{-1}\mathbf{A}^{12}\right)\mathbf{d}$, where all terms of $\mathbf{A}^{-1}$ can be easily formed using pedigree information. However, there are alternatives (Strandén et al., 2017) for computing $\left(\mathbf{A}^{11}\right)^{-1}\mathbf{v}$, where $\mathbf{v} = \mathbf{A}^{12}\mathbf{d}$. We relied on Cholesky decomposition as implemented in the CHOLMOD library (Davis and Hager, 2009; Chen et al., 2008). The details of implementation are explained in Strandén et al. (2017).

Critical for the computing load in the APY algorithm is the number of genotyped core animals. Moreover, the choice of representative core animals is critical for the correlation of GEBV by the original ssGBLUP and the APY approach to be high (Misztal et al., 2014). In a complex multibreed scenario, such as with the Irish beef cattle evaluations, both of these decisions can be difficult.

Table 2. Computing times and peak memory needed by the solver for the animal model BLUP (AM) and the original single-step genomic model (ssGBLUP), and APY based approaches with 50,000 (50K), 30,000 (30K), and 10,000 (10K) genotyped core animals[1]

	P_w (h)	P (h)	Matrix (GB)	I (h/1000 iterations)	N (*n*)	S (h)
AM	0.9	0.9	–	3	1,309	4
ssGBLUP	14.7	93.2	50	74	1,312	97
ssGTBLUP, D	4.0	21.9	34	55	1,316	72
ssGTBLUP, M	4.0	21.9	–	17	1,300	22
ssGTBLUP(99)	7.2	60.1	23	45	1,325	59
ssGTBLUP(98)	7.2	60.1	20	38	1,343	51
ssGTBLUP(95)	7.2	60.1	15	26	1,388	36
APY50K[2]	4.1	25.2	26	29	1,775	52
APY30K[2]	2.2	11.2	17	21	1,618	34
APY10K[2]	0.9	2.6	6	9	1,425	13

[1]Single-step genomic model using the T matrix approach (ssGTBLUP) accessed the T matrix from file (D) or from memory (M) every iteration round. Eigendecomposition with rank reduction in ssGTBLUP had 99%, 98%, and 95% proportion of variance explained, and read T matrix from the disk. Wall clock time in hours for the preprocessing (P_w), CPU times in hours for the preprocessing (P), and CPU time per 1,000 iterations in hours (I), size of external matrix read by the solver in gigabytes (Matrix), number of iterations (N), and total solver CPU time (S). In preprocessing, 10 processors were used when making the required matrices for ssGBLUP, ssGTBLUP, and APY calculations

[2]Average of 3 random core replicates.

We tested 3 different core sizes: 10,000 (10K), 30,000 (30K), and 50,000 (50K). Core animals were selected randomly and 3 replicates were made. This approach worked well for a Holstein dairy cattle population (Strandén et al., 2017). In all APY calculations, the term $\left(\mathbf{A}_{22}\right)^{-1}\mathbf{d}$ was calculated implicitly using the CHOLMOD library as already described.

RESULTS AND DISCUSSION

The original ssGBLUP had the longest preprocessing time of 14.7 h, and all the other approaches used less than half of that time (Table 2). Making the necessary computations using the eigendecompostion approach took 7.2 h. Preprocessing time for the ssGTBLUP was only 4.0 h. Note that the matrix inversion needed in the ssGBLUP and ssGTBLUP approaches is only one step that affects the total preprocessing time. Although the eigendecomposition is more time consuming than inverting a matrix of same size, here eigendecomposition was needed for a much smaller matrix of size 54,620, i.e., number of markers, whereas the inversion in the original ssGBLUP was for the **G** matrix of size 163,277, i.e., number of genotyped animals. The APY approaches require even less preprocessing time than the ssGTBLUP because APY works with small submatrices (Masuda et al., 2016; Strandén et al., 2017).

The preprocessing programs used parallel computing through the MKL library. Thus, the computations to make the **G** matrix, invert the **G** matrix, make the **T** matrix, and calculate the eigendecomposition used parallel computing. We restricted the number of available processors to 10 in all runs. There were clear advantages of using parallel computing (Table 2). The wall clock time shows amount of elapsed time while the CPU time has total time needed by all the processors. Consequently, the CPU times could be divided by 10 to give average CPU time per processor.

The size of the external file having either the **G** inverse or the **T** matrix approximates to the number of computations needed in the PCG iteration (Table 2). For the original ssGBLUP, this file consists of a dense lower triangle matrix having $n(n+1)/2$ elements, but n^2 multiplications are made. For the ssGTBLUP, the file is a rectangular *n* by *m* matrix but $2nm$ computations are made. According to the file sizes, the original ssGBLUP is expected to take the most time and APY10K the least. For the rank-reduced ssGTBLUP, the number of rows in the **T** matrix was reduced from 54,620 to 36,705 in ssGTBLUP(99), to 32,142 for ssGTBLUP(98), and to 24,602 for ssGTBLUP(95). In addition to being an indicator for the number of computations, the external file needs to be read in every iteration. Operating system may, however, buffer it to the RAM memory. In our case, the **T** matrix was relatively small, and it was possible to read the file only once and keep it in the memory. With the **T** matrix in memory, we were able to do the matrix times vector multiplications using a highly optimized BLAS (Basic Linear Algebra Subprograms; Dongarra et al., 1990) subroutine in the MKL library which reduced the computing time further. The CPU time for each iteration round was only 0.99 min per iteration (ssGTBLUP M in Table 2). Note that the rank-reduced ssGTBLUP runs were timed with iterating with disk data file and could be also run with the rank-reduced **T** matrix in memory. An added bonus on the use of the memory-stored **T** matrix is an easy implementation of parallel computing through the parallel MKL library as

was done in the preprocessing step. However, for easier comparison between all approaches, we did not use parallel computing in the solver.

Computing times per 1,000 iterations were lower in ssGTBLUP than in the original ssGBLUP (Table 2). However, the ratio of 1.35 in computing times (ssGBLUP divided by ssGTBLUP) is not equal to the ratio in the number of computations due to the external file which is $n^2/(2nm) = 1.49$. There are many reasons for this, the most important being natural variation in the computing time when usage of the computer varies due to other users. However, in general, the computing times follow roughly the expected ratio for the number of computations. Rank reduction by the eigendecomposition was able to decrease computing times, as was the APY approach. The decrease in the computing time depended on the level of approximation.

Based on the results, ssGTBLUP had similar convergence properties as the original ssGBLUP (Table 2). All APY based approaches required slightly more iterations than the ssGTBLUP approach or the original ssGBLUP which has been reported also in other studies (Masuda et al., 2016; Strandén et al., 2017). In particular, the number of iterations increased in APY when more animals were used in the core, i.e., the closer the APY approach was to the original ssGBLUP. A slight increase in the number of iterations was observed for the ssGTBLUP approach when the rank was reduced using the eigendecomposition. Thus, the 2 approaches behaved differently when the level of approximation is changed. In general, the differences in the numbers of iterations between the approaches were small and may be affected by the used convergence criterion. Nevertheless, the total solving time by ssGTBLUP(98) and APY50K, and by ssGTBLUP(95) and APY30K were relatively closer than their per iteration computing times.

Essentially, the solutions from the original ssGBLUP and the ssGTBLUP approach were the same. Correlations of GEBV for the original ssGBLUP and ssGTBLUP were 1.000 for all the traits. When the eigendecomposition with rank reduction was used, the correlations for the genotyped animal GEBVs had a range of 0.999–1.000 for ssGTBLUP(99), 0.998–1.000 for ssGTBLUP(98), and 0.992–0.998 for ssGTBLUP(95), i.e., lower correlations with fewer included eigenvalues. Similarly, correlations of GEBV solutions by APY with the original ssGBLUP depended on the core size. A core size of 50K animals gave highly correlated solutions with the original ssGBLUP, having a range of 0.999–1.000 across traits. Correlations decreased when the size of the core decreased, having a range of 0.952–0.996 for APY30K and 0.899–0.964 for APY10K. According to these results, APY10K seemed to deviate too much from the original ssGBLUP, and even APY30K results could be borderline acceptable but APY50K seemed to be very similar to the original ssGBLUP.

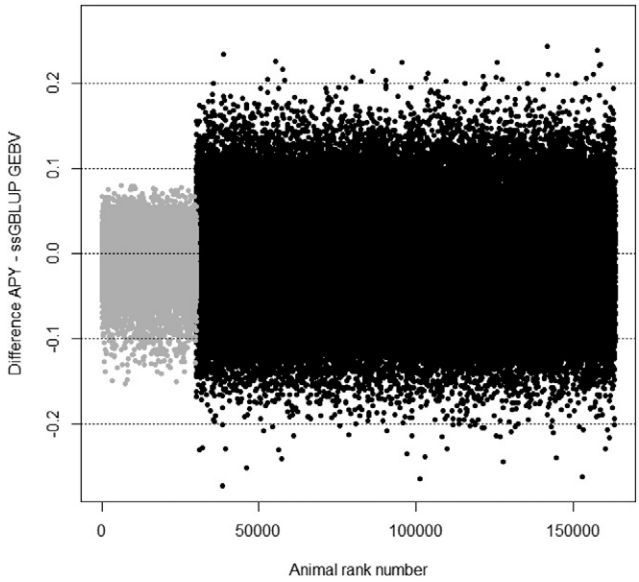

Figure 2. Difference in GEBV of cull cow carcass conformation score (CCC) for genotyped animals between the original ssGBLUP and APY randomly chosen 30,000 animals in the core. X-axis has animal rank by the core selection number. Gray dots represent the core animals, and black dots the non-core animals.

Even though the correlations of the GEBV solutions between APY30K and the original ssGBLUP were high, there were large differences with some GEBV solutions, e.g., for the cull cow carcass conformation score (Fig. 1). The standard deviation of GEBV of the genotyped animals was 0.58 with using the original ssGBLUP and 0.57 with ssGBLUP-APY30K. The number of GEBV deviating more than 0.2 standard deviations was 4,734, corresponding to 2.9% of the genotyped animals, deviating more than 0.15 was 14,717 or 9.0%, and more than 0.1 was 37,517 or 23.0%. These numbers are relatively high in comparison to those in a Holstein population where even in the case of animals deviating the least amount (i.e., > 0.10 SD) these were less than 3% with a core of 10,000 animals (Strandén et al., 2017). The correlation between APY10K and the original ssGBLUP in that study was 0.999, when here the highest correlation for APY30K was 0.996.

The data analyzed here provide a big challenge for implementation of the APY algorithm. The recommended number of animals in the core has been associated with the effective size of the population. This can be further addressed by examining the non-zero eigenvalues in the **G**-matrix. In the analyses of Pocrnic et al. (2016a), the number of largest eigenvalues that explained 98% of the sum of all eigenvalues, was 14,026 and 11,500 for Holstein and Jersey, respectively. From our data, we estimated that

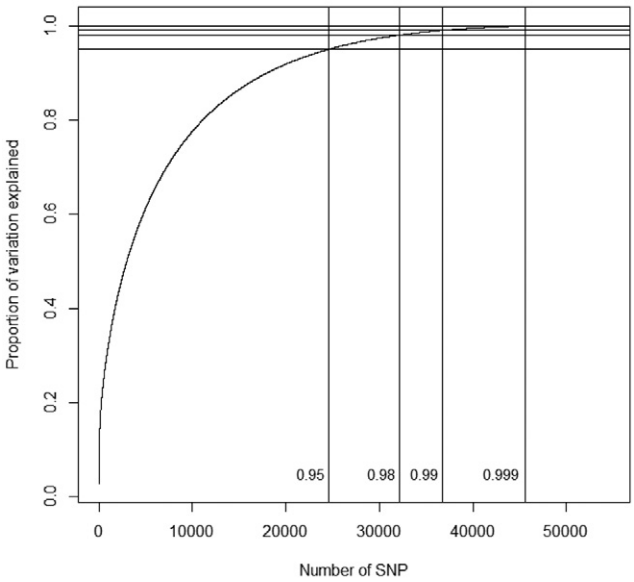

Figure 2. Cumulative sum of eigenvalues proportional to total variation for the genomic relationship matrix of 163,277 animals in Irish Beef evaluations 2016. The lines in the X-axis mark the number of eigenvalues needed to describe 0.95, 0.98, 0.99, and 0.999 of the total sum of eigenvalues in the Y-axis.

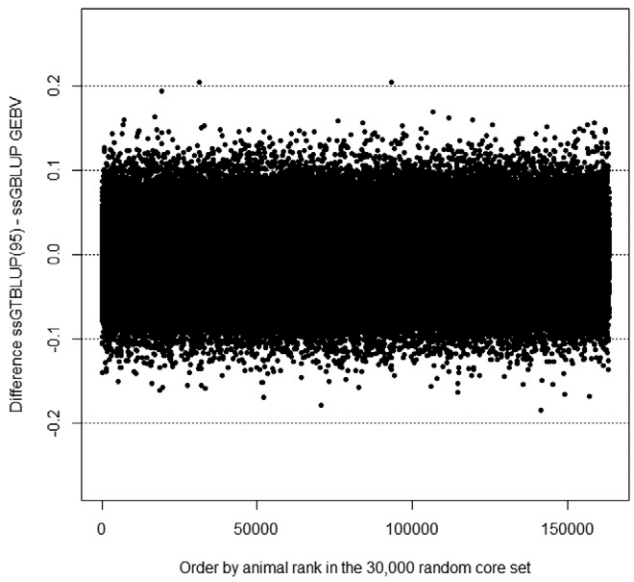

Figure 3. Difference in GEBV of cull cow carcass conformation score (CCC) for genotyped animals between the original ssGBLUP and ssGTBLUP with the 95% variance explained. X-axis has animal rank by number of progeny.

98% of the variation was covered by 32,142 of the largest eigenvalues (Fig. 2). Pocrnic et al. (2016b) summarized that the dimensionality of the genomic information (number of significant eigenvalues in the **G** matrix) is limited either by number of independent segregating genome segments (M_e), number of SNP in the genotyping chip (m), or the number of genotyped individuals (n). Considering that the genotyped animals represented genetic material from 41 breeds and that the large differences in GEBV were seen using APY10K compared to the original ssGBLUP, it seems that the number of segregating genome segments was too large to be described by an arbitrarily chosen set of 10,000 core animals. Even APY30K showed uneven deviations from the original ssGBLUP that depended on the chosen set of core animals.

Figures 1 and 3 illustrate differences to the original ssGBLUP solutions for the genotyped animals for APY30K and ssGTBLUP(95), respectively. The correlation of GEBV for CCC between APY30K and the original ssGBLUP was 0.996, while the equivalent comparison for ssGTBLUP(95) was 0.998, i.e., about the same. The correlations, however, do not reveal that the GEBV by these 2 approaches behaved differently in comparison to predictions by the original ssGBLUP. Standard deviation of GEBV in the original ssGBLUP was 0.58 for this trait. It seems that APY30K had more negative deviations than ssGTBLUP(95). There were 1,817 GEBV deviating more than 0.12 units above, i.e., more than 0.2 standard deviation units of GEBV higher values, by APY30K, but 301 such for ssGTBLUP(95). However, the numbers for deviating 0.12 units below were 2,917 and 281 for APY30K and ssGTBLUP(95), respectively. Thus, the total number of GEBV deviating more than 0.12 units was 4,734 for APY30K and 582 for ssGTBLUP(95). So, ssGTBLUP(95) seemed to deviate from ssGBLUP quite evenly between a positive or negative direction, but APY30K seemed to give GEBV values deviating more toward the negative than the positive direction. For ssGTBLUP(99) and ssGTBLUP(98), there were no GEBV deviating more than 0.12 units from the original ssGBLUP solutions.

An option for APY is to increase the number of core animals. However, the rank of the genotype matrix **Z** is never more than the number of markers. The rank of **Z** matrix for the core animals is always less than the number of core animals and the number of markers. Thus, when the number of core animals is beyond the number of markers, the genotype matrix has redundant information. It may be possible to select a suitable core for a multibreed population that is close to the number of markers. However, the larger the core size the more iterations are needed to attain convergence (Table 2). This is due to numerical values in the $\mathbf{G}^{-1}_{APY}$ matrix becoming larger. In the $\mathbf{G}^{-1}_{APY}$ matrix, the submatrix of the noncore animals is diagonal, and the diagonal elements approach infinity the more animals are in the core (Strandén et al., 2017). For our data, the average diagonal of the $\mathbf{G}^{-1}_{APY}$ matrix was 4 for APY10K, 31 for APY30K, and 430 for APY50K. Because the core size needs to be large in a heterogeneous population, selection of the APY core animals may require some more thought in a multibreed population than in a single-

breed population. However, the realized correlations suggest that the use of random core animals works reasonably well when the core size is large.

Selection of the core animals is an important step in APY. The random selection of the core animals has been shown to work well in single-breed populations (e.g., Masuda et al., 2016; Bradford et al., 2017; Strandén et al., 2017). We decided to follow this approach for the multibreed population because it selects core animals from the breeds proportionally to their amounts in the population. However, it may be that this led to insufficient presentation of the smaller breeds in the core group. Thus, the APY core selection approaches may need to be further investigated for multibreed populations. In contrast, the eigendecomposition approach presented for the ssGTBLUP gives a well-defined and automatic approach to reduce the number of computations that can be expected to work for any population structure.

The $\mathbf{T}$ matrix has size $m \times n$. The $\mathbf{G}^{-1}$ matrix has size $n \times n$. In iterative solving of the MME by PCG, product $\mathbf{G}^{-1}\mathbf{d}$ has to be computed. The number of multiplications needed to perform this operation using $\mathbf{G}^{-1}$ matrix is n^2. Product $\mathbf{T'Td}$ needs $2nm$ multiplications. Thus, the product $\mathbf{G}_{\varepsilon}^{-1}\mathbf{d}$ needs $2nm + n$ multiplications. The product $\mathbf{G}_{w}^{-1}\mathbf{d}$ involves additional work due to $\frac{1}{w}\mathbf{A}_{22}^{-1}\mathbf{d}$, which can be performed efficiently without making or inverting $\mathbf{A}_{22}$ using sparse matrix operations (Strandén and Mäntysaari, 2015; Masuda et al., 2016; Strandén et al., 2017) and the work can be combined in the other product of $\mathbf{A}_{22}^{-1}\mathbf{d}$ needed in the original ssGBLUP without further increasing computational work. When the number of genotyped animals n increases, it can be expected that the number of multiplications in the product $\mathbf{T'Td}$ increases linearly, but in the product $\mathbf{G}^{-1}\mathbf{d}$ the increase is quadratic. For $\mathbf{G}_{\varepsilon}^{-1}\mathbf{d}$, it can be predicted that this product is computationally less demanding in the ssGTBLUP than in the original ssGBLUP when the number of genotyped animals exceeds 100,000 when the number of markers m equals 50,000.

The product $\mathbf{G}_{\text{APY}}^{-1}\mathbf{d}$ requires $n_c(2n\text{-}n_c\text{-}1)+n$ multiplications, where n_c is the number of core animals. This is always less than needed in the product $\mathbf{G}^{-1}\mathbf{d}$ for the original ssGBLUP (n^2) because n_c is less or equal to n. For the ssGTBLUP, the number of computations to perform $\mathbf{G}_{\varepsilon}^{-1}\mathbf{d}$ is ($2nm+n$), which is more than is done in $\mathbf{G}_{\text{APY}}^{-1}\mathbf{d}$ when the number of the core animals (n_c) is less than number of markers (m). However, the relative advantage of APY diminishes as the number of genotyped animals (n) increases. With the most challenging multibreed single-step evaluations, the n_c might have to be higher than m. It can be shown that the number of multiplications in $\mathbf{G}_{\text{APY}}^{-1}\mathbf{d}$ is more than that in $\mathbf{G}_{\varepsilon}^{-1}\mathbf{d}$ when the number of core animals in APY is more than $\frac{1}{2}\left(2n-1-\sqrt{(2n-1)^2-8nm}\right)$ and n is at least $2m$. This lower bound to n_c decreases when n increases and m is unchanged. With 100,000 animals, 50,000 markers, and 50,000 core animals, the original ssGBLUP and ssGTBLUP have about the same number of multiplications, but APY50K has about 25% fewer. With 500,000 genotyped animals, the ssGTBLUP has ~20% and APY50K has ~19% of the calculations of the original ssGBLUP. The difference in the calculations is about the same as the difference in non-zeroes in the external structure matrix stored in the ssGTBLUP and APY50K approaches. When the number of core animals equals the number of markers, the difference is the size of the number of elements in the upper triangle of matrix for the core animals, i.e., $n_c(n_c-1)/2$, because the $(\mathbf{G^{-1}})_{\text{APY}}$ matrix is a squared symmetric matrix but the T matrix is rectangular.

We analyzed a multibreed population using ssGTBLUP which was shown to work faster than the original ssGBLUP when the number of genotyped animals is more than twice the number of markers. Undoubtedly, this result can be replicated for a single-breed population as well. However, in a single-breed population the APY approach is expected to perform well because the core size can be limited typically to numbers much less than the number of markers (e.g., Masuda et al., 2016, Strandén et al., 2017). A possibility is to use the eigendecomposition approach in ssGTBLUP to reduce the rank of the $\mathbf{T}$ matrix. It is likely that the rank-reduced $\mathbf{T}$ and APY will give similar results as in here for the multibreed case. Moreover, to achieve a similar agreement with the GEBV from original ssGBLUP, the rank of the $\mathbf{T}$ matrix is expected to be less than the number of core animals in APY. Therefore, in practice, the approaches would require about the same solving times. Construction of $\mathbf{T}_r$ requires eigendecomposition and is thus computationally more demanding than the construction of $\mathbf{G}_{\text{APY}}^{-1}$ with a core size much less than the number of markers. This needs to be considered in the total computing time in ssGTBLUP(98), but the difference is dispensed when more traits are to be evaluated with the same $\mathbf{G}$-matrix.

We performed eigendecomposition on a matrix of form $\mathbf{Z'C}^{-1}\mathbf{Z}$ in $\mathbf{G}_{\text{C}}^{-1} = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{Z}\left(\mathbf{Z'C}^{-1}\mathbf{Z}+\mathbf{I}\right)^{-1}\mathbf{Z'C}^{-1}$, where $\mathbf{C} = \varepsilon\mathbf{I}$. Alternatively, the decomposition could have been done to $\left(\mathbf{Z'C}^{-1}\mathbf{Z}+\mathbf{I}\right)$. For this case, we can note the following two consequences due to the different approaches. First, the two approaches have different eigenvalues. Our approach gives lower eigenvalues. However, the eigenvalues of our approach can be used to calculate eigenvalues of the other approach by a simple formula: $d_i^* = 1 + d_i$, where d_i is the eigenvalue

in $\mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z}$. Second, our approach leads to the $\mathbf{T}$ matrix of $\mathbf{T}_r = \frac{1}{\varepsilon}\left(\mathbf{D}_r + \mathbf{I}\right)^{-\frac{1}{2}}\mathbf{V}_r'\mathbf{Z}'$, but the other approach leads to $\mathbf{T}_r = \frac{1}{\varepsilon}\left(\mathbf{D}_r^*\right)^{-\frac{1}{2}}\mathbf{V}_r'\mathbf{Z}'$, where both approaches give the same eigenvector matrix $\mathbf{V}_r$. Thus, practical differences between these approaches are minimal.

Conclusions

We derived an equivalent computational approach for the original ssGBLUP called ssGTBLUP which needs fewer computations than ssGBLUP in iterative solving by the PCG method when the number of genotyped animals (n) is more than twice the number of used markers (m). The most challenging task is a computation of product $\mathbf{T}'\mathbf{T}\mathbf{d}$ in each iteration of ssGTBLUP instead of $\mathbf{G}^{-1}\mathbf{d}$ in ssGBLUP where $\mathbf{G}$ has size n x n and $\mathbf{T}$ has size m x n. An approximate approach of ssGTBLUP uses the most important eigenvalues and eigenvectors in the eigendecomposition of the $\mathbf{T}$ matrix. The approximation allows a decrease in the number of rows in the $\mathbf{T}$ matrix. The ssGTBLUP gave the same solutions as the original ssGBLUP but used less computing time in analysis of our data set. The approximation approach performed logically when the degree of approximation was changed. In the challenging multibreed population, solutions by the APY algorithm suggested asymmetric deviation from the original ssGBLUP but the eigendecomposition based ssGTBLUP performed symmetrically. The ssGTBLUP with or without eigendecomposition approach seems to offer a computationally competitive approach for solving genomic breeding values with the single-step method when the number of genotyped animals is very large even when a genetically heterogeneous multibreed population is analyzed.

LITERATURE CITED

Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenny, and D. Sorensen. 1999. LAPACK users' guide, 3rd ed. SIAM, Philadelphia, PA. http://www.netlib.org/lapack/lug/. (Accessed 11 July 2017.) doi:10.1137/1.9780898719604

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. doi:10.3168/jds.2009-2730

Bradford, H. L., I. Pocrnic, B. O. Fragomeni, D. A. L. Lourenco, and I. Misztal. 2017. Optimum selection of core animals in the efficient inversion of the genomic relationship matrix. J. Anim. Sci. 95(Suppl. 4):90–91. doi:10.2527/asasann.2017.183

Chen, Y., T. A. Davis, W. W. Hager, and S. Rajamanickam. 2008. Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate. ACM Trans. Math. Software 35(3):22. doi:10.1145/1391989.1391995

Christensen, O., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. doi:10.1186/1297-9686-42-2

Davis, T. A., and W. W. Hager. 2009. Dynamic supernodes in sparse Cholesky update/downdate and triangular solves. ACM Trans. Math. Software 35(4):27. doi:10.1145/1462173.1462176

Dongarra, J. J., J. Du Croz, I. S. Duff, and S. Hammarling. 1990. A set of level-3 basic linear algebra subprograms. ACM Trans. Math. Software 16(1):1–17. doi:10.1145/77626.79170

Evans, R. D., J. F. Kearney, J. McCarthy, A. Cromie, and T. Pabiou. 2014. Beef performance evaluations in a multi-layered and mainly crossbred population. Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada. p. 732.

Fernando, R. L., J. C. M. 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

Fernando, R. L., H. Cheng, B. L. Golden, and D. J. Garrick. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. Genet. Sel. Evol. 48:96. doi:10.1186/s12711-016-0273-2

Fragomeni, B. O., D. A. L. Lourenco, S. Tsuruta, Y. Masuda, I. Aguilar, A. Legarra, T. J. Lawlor, and I. Misztal. 2015. Hot topic: Use of genomic recursions in single-step genomic best linear unbiased predictor (BLUP) with a large number of genotypes. J. Dairy Sci. 98:4090–4094. doi:10.3168/jds.2014-9125

Householder, A. S. 1964. The theory of matrices in numerical analysis. Dover, New York.

Intel. 2014. Intel Math Kernel Library reference manual. https://software.intel.com/en-us/mkl-reference-manual-for-fortran. (Accessed 11 July 2017.)

Legarra, A., and V. Ducrocq. 2012. Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. J. Dairy Sci. 95:4629–4645. doi:10.3168/jds.2011-4982

Liu, Z., M. E. Goddard, F. Reinhardt, and R. Reents. 2014. A single-step genomic model with direct estimation of marker effects. J. Dairy Sci. 97:5833–5850. doi:10.3168/jds.2014-7924

Masuda, Y., I. Misztal, S. Tsuruta, A. Legarra, I. Aguilar, D. A. L. Lourenco, B. O. Fragomeni, and T. J. Lawlor. 2016. Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. J. Dairy Sci. 99:1968–1974. doi:10.3168/jds.2015-10540

McPeek, M. S., X. Wu, and C. Ober. 2004. Best linear unbiased allele-frequency estimation in complex pedigrees. Biometrics 60:359–367. doi:10.1111/j.0006-341X.2004.00180.x

McHugh, N., A. G. Fahey, R. D. Evans, and D. P. Berry. 2010. Factors associated with selling price of dairy calves at livestock marts. Adv. Anim. Biosci. 1:154. doi:10.1017/S2040470010002979

McHugh, N., R. D. Evans, P. R. Amer, A. G. Fahey, and D. P. Berry. 2011. Genetic parameters for cattle price and body weight from routinely collected data at livestock auctions and commercial farms. J. Anim. Sci. 89:29–39. doi:10.2527/jas.2010-3044

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. doi:10.3168/jds.2013-7752

Pabiou, T., W. F. Fikse, A. R. Cromie, M. G. Keane, A. Näsholm, and D. P. Berry. 2011. Use of digital images to predict carcass cut yields in cattle. Livest. Sci. 137:130–140. doi:10.1016/j.livsci.2010.10.012

Pabiou, T., W. F. Fikse, P. R. Amer, A. R. Cromie, A. Näsholm, and D. P. Berry. 2012. Genetic relationships between carcass cut weights predicted from video image analysis and other performance traits in cattle. Animal 6:1389–1397. doi:10.1017/S1751731112000705

Pocrnic, I., D. A. L. Lourenco, Y. Masuda, and I. Misztal. 2016a. 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

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

Strandén, I., and M. Lidauer. 1999. Solving large mixed models using preconditioned conjugate gradient iteration. J. Dairy Sci. 82:2779–2787. doi:10.3168/jds.S0022-0302(99)75535-9

Strandén, I., and O. F. Christensen. 2011. Allele coding in genomic evaluation. Genet. Sel. Evol. 43:25. doi:10.1186/1297-9686-43-25

Strandén, I., and E. A. Mäntysaari. 2014. Comparison of some equivalent equations to solve single-step GBLUP. Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada. p. 069.

Strandén, I., K. Matilainen, G. P. Aamand, and E. A. Mäntysaari. 2017. Solving efficiently large single-step genomic best linear unbiased prediction models. J. Anim. Breed. Genet. 134:264–274. doi:10.1111/jbg.12257

Taskinen, M., E. A. Mäntysaari, and I. Strandén. 2017. Single-step SNP-BLUP with on-the-fly imputed genotypes and residual polygenic effects. Genet. Sel. Evol. 49:36. doi:10.1186/s12711-017-0310-9

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