J. Dairy Sci. 104:3298–3303
https://doi.org/10.3168/jds.2020-19460

Technical note: Genetic groups in single-step single nucleotide polymorphism best linear unbiased predictor

Jeremie Vandenplas,[1]* Herwin Eding,[2] and Mario P. L. Calus[1]
[1]Animal Breeding and Genomics, Wageningen UR, PO 338, 6700 AH, Wageningen, the Netherlands
[2]CRV BV, Wassenaarweg, 20, 6843 NW, Arnhem, the Netherlands

ABSTRACT

Genetic groups, also called unknown or phantom parents groups, are often used in dairy cattle genetic evaluations to account for selection that cannot be accounted for by known genetic relationships. With the advent of genomic evaluations, the theory of genetic groups was extended to the so-called single-step genomic BLUP (ssGBLUP). In short, genetic groups can be fitted in ssGBLUP through regression effects, or by including them in the pedigree and computing the adequate combined pedigree and genomic relationship matrix. In this study, we applied the so-called Quaas and Pollak transformation to a system of equations for single-step SNP BLUP (ssSNPBLUP), such that genetic groups can thereafter be included in the pedigree. The example in this study showed that including the genetic groups in the pedigree for ssSNPBLUP allowed reduced memory burden and computational costs in comparison to genetic groups fitted as covariates.
Key words: unknown parents groups, phantom parents, single-step, genomic evaluation

Technical Note

Genetic groups, also called unknown or phantom parents groups, are allocated to animals to account for selection that cannot be accounted for by known genetic relationships (Quaas, 1988). Genetic groups can be defined based on, among others, the birth year of animals, sex, selection path, and breed (Quaas, 1988; Westell et al., 1988; Legarra et al., 2007). Developed for pedigree-based BLUP, the theory of genetic groups was extended to single-step genomic BLUP (**ssGBLUP**; Misztal et al., 2013). An alternative implementation of genetic groups is to fit the genetic group contributions as regression effects (Misztal et al., 2013). While this approach is straightforward and can be easily implemented in current software for genomic evaluations, it is expected to lead to considerable memory and computational costs for genetic evaluations using a very large pedigree (i.e., containing millions of animals) and a relatively large number of genetic groups (i.e., higher than 100), which is often the case in dairy cattle genetic evaluations (e.g., Matilainen et al., 2018). Another equivalent implementation is to include the genetic groups in the pedigree as "phantom parents" and to set up the combined pedigree and genomic relationship matrix adequately for ssGBLUP, as proposed by Misztal et al. (2013), who applied the Quaas and Pollak (1981) transformation (**QP** transformation) to the mixed model equations of ssGBLUP. In practice, this latter implementation based on the QP transformation seems to be more feasible for dairy cattle ssGBLUP (Mäntysaari et al., 2020). However, as far as we know, the inclusion of the genetic groups based on the QP transformation has not been derived for single-step SNP BLUP (**ssSNPBLUP**) yet. In contrast to ssGBLUP, the ssSNPBLUP approaches directly estimate the SNP effects. They also avoid some drawbacks of ssGBLUP, such as requiring an inverse of the genomic relationship matrix or a representation thereof (Liu et al., 2014; Fernando et al., 2016). Therefore, the aim of this study was to derive the system of equations for ssSNPBLUP proposed by Liu et al. (2014) that includes the genetic groups in the pedigree.

For the derivation, we first assumed that genetic groups were explicitly fitted as fixed effects in the model associated with the system of equations proposed by Liu et al. (2014). Ignoring all fixed effects, except the genetic groups, the resulting ssSNPBLUP system of equations following Liu et al. (2014) can be written as follows:

Received August 12, 2020.
Accepted October 2, 2020.
*Corresponding author: Jeremie.vandenplas@wur.nl

$$\begin{bmatrix} \mathbf{Q}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}\mathbf{Q} & \mathbf{Q}_n'\mathbf{W}_n'\mathbf{R}_n^{-1}\mathbf{W}_n & \mathbf{Q}_g'\mathbf{W}_g'\mathbf{R}_g^{-1}\mathbf{W}_g & \mathbf{0} \\ \mathbf{W}_n'\mathbf{R}_n^{-1}\mathbf{W}_n\mathbf{Q}_n & \mathbf{W}_n'\mathbf{R}_n^{-1}\mathbf{W}_n + \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} & \boldsymbol{\Sigma}^{13} \\ \mathbf{W}_g'\mathbf{R}_g^{-1}\mathbf{W}_g\mathbf{Q}_g & \boldsymbol{\Sigma}^{21} & \mathbf{W}_g'\mathbf{R}_g^{-1}\mathbf{W}_g + \boldsymbol{\Sigma}^{22} & \boldsymbol{\Sigma}^{23} \\ \mathbf{0} & \boldsymbol{\Sigma}^{31} & \boldsymbol{\Sigma}^{32} & \boldsymbol{\Sigma}^{33} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{t}} \\ \hat{\mathbf{u}}_n \\ \hat{\mathbf{u}}_g \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{W}_n'\mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{W}_g'\mathbf{R}_g^{-1}\mathbf{y}_g \\ \mathbf{0} \end{bmatrix}, \quad [1]$$

where the subscripts g and n refer to genotyped and nongenotyped animals, respectively, $\mathbf{y}$ is the vector of records, $\hat{\mathbf{t}}$ is the vector of estimated genetic group effects, $\hat{\mathbf{u}}_n$ and $\hat{\mathbf{u}}_g$ are the vectors of estimated additive genetic effects for nongenotyped animals and for genotyped animals, respectively, $\hat{\mathbf{g}}$ is the vector of estimated SNP effects, the matrix $\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_n \\ \mathbf{Q}_g \end{bmatrix}$ is a matrix that assigns animals to genetic groups, the matrix $\mathbf{W} = \begin{bmatrix} \mathbf{W}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_g \end{bmatrix}$ is an incidence matrix relating records to the corresponding effects, the matrix $\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{R}_n^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_g^{-1} \end{bmatrix}$ is the inverse of the residual (co) variance structure matrix, and the matrix $\boldsymbol{\Sigma}^{-1}$ is defined as

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} & \boldsymbol{\Sigma}^{13} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} & \boldsymbol{\Sigma}^{23} \\ \boldsymbol{\Sigma}^{31} & \boldsymbol{\Sigma}^{32} & \boldsymbol{\Sigma}^{33} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{nn} & \mathbf{A}^{ng} & 0 \\ \mathbf{A}^{gn} & \mathbf{A}^{gg} + \left(\frac{1}{w} - 1\right)\mathbf{A}_{gg}^{-1} & -\frac{1}{w}\mathbf{A}_{gg}^{-1}\mathbf{Z} \\ 0 & -\frac{1}{w}\mathbf{Z}'\mathbf{A}_{gg}^{-1} & \frac{1}{w}\mathbf{Z}'\mathbf{A}_{gg}^{-1}\mathbf{Z} + \frac{m}{1-w}\mathbf{I} \end{bmatrix} \sigma_u^{-2}$$

(Liu et al., 2014; equation [A17]). The matrix $\mathbf{Z}$ contains the SNP genotypes (coded as 0 for one homozygous genotype, 1 for the heterozygous genotype, or 2 for the alternate homozygous genotype) centered by their observed means. The scalar σ_u^{-2} is the inverse of the additive genetic variance, w is the proportion (strictly between 0 and 1) of the additive genetic variance due to residual polygenic effects, and $m = 2\Sigma p_j\left(1 - p_j\right)$, with p_j being the observed allele frequency of the jth SNP. The matrix $\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{nn} & \mathbf{A}_{ng} \\ \mathbf{A}_{gn} & \mathbf{A}_{gg} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{nn} & \mathbf{A}^{ng} \\ \mathbf{A}^{gn} & \mathbf{A}^{gg} \end{bmatrix}$ is the inverse of the pedigree relationship matrix among all animals, and $\mathbf{A}_{gg}^{-1} = \mathbf{A}^{gg} - \mathbf{A}^{gn}\left(\mathbf{A}^{nn}\right)^{-1}\mathbf{A}^{ng}$ is the inverse of the pedigree relationship matrix among genotyped animals. Each row of $\mathbf{Q}$ contains the genetic group composition of an animal.

Following Quaas and Pollak (1981), we first defined the matrix $\mathbf{P}$ as

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q_n} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{Q_g} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Then, pre-multiplying both sides of the system of Equation [1] by $\left(\mathbf{P}^{-1}\right)'$ and inserting $\mathbf{P}^{-1}\mathbf{P}$ between the coefficient matrix and the vector of solutions (which is known as the QP transformation) resulted, after some algebra, in

$$\begin{bmatrix} \boldsymbol{\Sigma}_{QP}^{11} & \boldsymbol{\Sigma}_{QP}^{12} & \boldsymbol{\Sigma}_{QP}^{13} & \boldsymbol{\Sigma}_{QP}^{14} \\ \boldsymbol{\Sigma}_{QP}^{21} & \mathbf{W}_n'\mathbf{R}_n^{-1}\mathbf{W}_n + \boldsymbol{\Sigma}_{QP}^{22} & \boldsymbol{\Sigma}_{QP}^{23} & \boldsymbol{\Sigma}_{QP}^{24} \\ \boldsymbol{\Sigma}_{QP}^{31} & \boldsymbol{\Sigma}_{QP}^{32} & \mathbf{W}_g'\mathbf{R}_g^{-1}\mathbf{W}_g + \boldsymbol{\Sigma}_{QP}^{33} & \boldsymbol{\Sigma}_{QP}^{34} \\ \boldsymbol{\Sigma}_{QP}^{41} & \boldsymbol{\Sigma}_{QP}^{42} & \boldsymbol{\Sigma}_{QP}^{43} & \boldsymbol{\Sigma}_{QP}^{44} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{t}} \\ \hat{\mathbf{a}}_n \\ \hat{\mathbf{a}}_g \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{W}_n'\mathbf{R}_n^{-1}\mathbf{y}_n \\ \mathbf{W}_g'\mathbf{R}_g^{-1}\mathbf{y}_g \\ \mathbf{0} \end{bmatrix}, \text{with} \quad [2]$$

$$\hat{\mathbf{a}}_n = \mathbf{Q}_n\hat{\mathbf{t}} + \hat{\mathbf{u}}_n,$$

$$\hat{\mathbf{a}}_g = \mathbf{Q}_g\hat{\mathbf{t}} + \hat{\mathbf{u}}_g,$$

$$\boldsymbol{\Sigma}_{QP}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{QP}^{11} & \boldsymbol{\Sigma}_{QP}^{12} & \boldsymbol{\Sigma}_{QP}^{13} & \boldsymbol{\Sigma}_{QP}^{14} \\ \boldsymbol{\Sigma}_{QP}^{21} & \boldsymbol{\Sigma}_{QP}^{22} & \boldsymbol{\Sigma}_{QP}^{23} & \boldsymbol{\Sigma}_{QP}^{24} \\ \boldsymbol{\Sigma}_{QP}^{31} & \boldsymbol{\Sigma}_{QP}^{32} & \boldsymbol{\Sigma}_{QP}^{33} & \boldsymbol{\Sigma}_{QP}^{34} \\ \boldsymbol{\Sigma}_{QP}^{41} & \boldsymbol{\Sigma}_{QP}^{42} & \boldsymbol{\Sigma}_{QP}^{43} & \boldsymbol{\Sigma}_{QP}^{44} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{*tt}\sigma_u^{-2} & \mathbf{A}^{*tn}\sigma_u^{-2} & \mathbf{A}^{*tg}\sigma_u^{-2} & \mathbf{0} \\ \mathbf{A}^{*nt}\sigma_u^{-2} & \boldsymbol{\Sigma}^{12} & \boldsymbol{\Sigma}^{12} & \boldsymbol{\Sigma}^{13} \\ \mathbf{A}^{*gt}\sigma_u^{-2} & \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} & \boldsymbol{\Sigma}^{23} \\ \mathbf{0} & \boldsymbol{\Sigma}^{32} & \boldsymbol{\Sigma}^{32} & \boldsymbol{\Sigma}^{33} \end{bmatrix} +$$

$$\begin{bmatrix} \left(\frac{1}{w}-1\right)\mathbf{Q}_g'\mathbf{A}_{gg}^{-1}\mathbf{Q}_g & \mathbf{0} & -\left(\frac{1}{w}-1\right)\mathbf{Q}_g'\mathbf{A}_{gg}^{-1} & \frac{1}{w}\mathbf{Q}_g'\mathbf{A}_{gg}^{-1}\mathbf{Z} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\left(\frac{1}{w}-1\right)\mathbf{A}_{gg}^{-1}\mathbf{Q}_g & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{1}{w}\mathbf{Z}'\mathbf{A}_{gg}^{-1}\mathbf{Q}_g & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}\sigma_u^{-2}, \text{ and}$$

$$\mathbf{A}^{*} = \begin{bmatrix} \mathbf{A}^{*tt} & \mathbf{A}^{*tn} & \mathbf{A}^{*tg} \\ \mathbf{A}^{*nt} & \mathbf{A}^{nn} & \mathbf{A}^{ng} \\ \mathbf{A}^{*gt} & \mathbf{A}^{gn} & \mathbf{A}^{gg} \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} & -\mathbf{Q}_n'\mathbf{A}^{nn} - \mathbf{Q}_g'\mathbf{A}^{gn} & -\mathbf{Q}_n'\mathbf{A}^{ng} - \mathbf{Q}_g'\mathbf{A}^{gg} \\ -\mathbf{A}^{nn}\mathbf{Q}_n - \mathbf{A}^{ng}\mathbf{Q}_g & \mathbf{A}^{nn} & \mathbf{A}^{ng} \\ -\mathbf{A}^{gn}\mathbf{Q}_n - \mathbf{A}^{gg}\mathbf{Q}_g & \mathbf{A}^{gn} & \mathbf{A}^{gg} \end{bmatrix},$$

created by including genetic groups using the rules of Quaas (1988).

In our implementation, before running the iterative algorithm for solving the system of Equation [2], the matrix $\mathbf{Q}_g$ and the submatrices of the inverse of a pedigree relationship matrix for the genotyped animals and their ancestors, $\mathbf{A}_{anc}^{gg}$, $\mathbf{A}_{anc}^{nn}$, and $\mathbf{A}_{anc}^{ng}$, were first set up such that $\mathbf{A}_{gg}^{-1}\mathbf{Q}_g$ could be computed as $\mathbf{A}_{gg}^{-1}\mathbf{Q}_g = \left[\mathbf{A}_{anc}^{gg} - \mathbf{A}_{anc}^{gn}\left(\mathbf{A}_{anc}^{nn}\right)^{-1}\mathbf{A}_{anc}^{ng}\right]\mathbf{Q}_g$. The matrices $\left(\frac{1}{w}-1\right)\mathbf{Q}_g'\mathbf{A}_{gg}^{-1}\mathbf{Q}_g$, $-\left(\frac{1}{w}-1\right)\mathbf{A}_{gg}^{-1}\mathbf{Q}_g$, and $\frac{1}{w}\mathbf{Z}'\mathbf{A}_{gg}^{-1}\mathbf{Q}_g$ were then computed and stored in a file for their use by the iterative solver.

It is worth noting that absorbing the equations associated with $\hat{\mathbf{g}}$ in the system of Equation [2] leads to the system of equations derived by Misztal et al. (2013) for ssGBLUP, demonstrating the equivalence between both models. Furthermore, the same approach could be applied to other ssSNPBLUP linear systems (e.g., Fernando et al., 2016).

For illustrating the performances of the systems of Equations [1] and [2], the data set and associated variance components from the 4-trait routine genetic evaluation of August 2019 for temperament and milking speed of dairy cattle for the Netherlands and the Flemish region of Belgium (CRV Animal Evaluation Unit, 2020a,b) were used. Performance in both countries were considered as different traits. The data file included 4,058,154 records with a single record per animal. The pedigree included 6,344,482 animals. A total of 441 genetic groups were defined based on selection path, breed, country of origin, and year of birth (CRV Animal Evaluation Unit, 2020c). On average, each genetic group replaced unknown parents of 1,516 animals, varying from 373 to 9,362 animals. The number of animals genotyped for 37,995 segregating SNPs was 123,644. The 4-trait mixed model included random effects (additive genetic and residual), fixed covariables (heterosis and recombination), and fixed cross-classified effects [herd × year × season at classification, age (in month) at classification, lactation stage at classification, milk yield, and month of calving; CRV Animal Evaluation Unit, 2020a,b]. An additional fixed covariate that estimates the mean breeding value in genotyped animals was fitted (Hsu et al., 2017). This additional effect aimed to correct for the fact that the genotypes were centered using the observed allele frequencies instead of using the base population allele frequencies (Hsu et al., 2017). The genetic groups were fitted in ssSNPBLUP with covariates (for the system of Equation [1]) or with the QP transformation (for the system of Equation [2]). Additionally, to avoid problems with estimating their effects, genetic groups were considered as random effects in both cases by adding 1 times the additive genetic covariance matrix to their block-diagonal elements (Misztal et al., 2013; Schaeffer, 2018). Practically, the matrix $\mathbf{Q}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}\mathbf{Q}$ was replaced by $\mathbf{Q}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}\mathbf{Q} + \mathbf{I}\sigma_u^{-2}$ in the system of Equation [1], and the matrix $\left(\frac{1}{w}-1\right)\mathbf{Q}_g'\mathbf{A}_{gg}^{-1}\mathbf{Q}_g\sigma_u^{-2}$ was replaced by $\left[\left(\frac{1}{w}-1\right)\mathbf{Q}_g'\mathbf{A}_{gg}^{-1}\mathbf{Q}_g + \mathbf{I}\right]\sigma_u^{-2}$ in the system of Equation [2]. The proportion of additive genetic variance due to residual polygenic effects, w, was assumed to be equal to 0.10. To illustrate the additional memory and computational costs of fitting genetic groups, ssSNPBLUP without genetic groups was also performed.

The different ssSNPBLUP systems were solved using a Fortran 2003 program detailed in Vandenplas et al. (2020). The matrix-free version of the software was used, and all the data needed to perform the multiplication of the coefficient matrix by a vector was stored in-memory. A 2-level preconditioned conjugate gradient (**PCG**) method was used as iterative solver (Vandenplas et al., 2019). The first-level preconditioner $\mathbf{M}$ included only the diagonal elements of the coefficient matrix for the fixed effects, and a block-diagonal

structure for the random effects with blocks corresponding to equations for different traits within a level. The second-level preconditioner was a diagonal matrix with all diagonal elements equal to 1, except for the diagonal elements that were equal to 100, and for the diagonal elements corresponding to the genetic group equations fitted as covariates that were equal to 10 (Vandenplas et al., 2019). The value of 10 for the genetic group equations was determined following the ratio between the largest eigenvalues of the preconditioned coefficient matrices of ssSNPBLUP with and without genetic groups fitted as covariates (Vandenplas et al., 2019). The 2-level PCG method iterated until the relative difference between 2 consecutive solutions was $\leq 10^{-5}$ (Lidauer et al., 1999). The smallest and largest Ritz values, that were approximations of the smallest and largest eigenvalues that influenced the convergence, were also computed using the Lanczos method based on information obtained from the PCG method (Kaasschieter, 1988). The effective spectral condition numbers of the different preconditioned coefficient matrices were computed as the ratio of the largest to the smallest eigenvalue that influence the convergence (Nabben and Vuik, 2006). All computations were performed on a computer with 376 GB and an Intel Xeon Gold 6130 (2.10 GHz) processor with 32 cores (Intel, Santa Clara, CA). The number of OpenMP (www.openmp.org) threads used for all computations was equal to 5. All reported times are indicative because they may have been influenced by other jobs running simultaneously on the computer.

All the characteristics and results for the ssSNPBLUP systems without and with genetic groups are in Table 1 and Figure 1. The ssSNPBLUP system without genetic groups included 26,861,588 equations, while the 2 ssSNPBLUP systems with genetic groups included 26,863,352 equations. The Pearson correlations between the estimates for all fixed effects, additive direct effects, and SNP effects of the 2 ssSNPBLUP systems with genetic groups were higher than 0.99, as expected, because the same model underlies the 2 ssSNPBLUP systems with genetic groups. For the ssSNPBLUP systems with genetic groups, the matrices $\mathbf{Q}$ (of size 6,344,482 by 441), or $\left(\frac{1}{w}-1\right)\mathbf{Q}_g'\mathbf{A}_{gg}^{-1}\mathbf{Q}_g$ (of size 441 by 441), $-\left(\frac{1}{w}-1\right)\mathbf{Q}_g'\mathbf{A}_{gg}^{-1}$ (of size 441 by 123,644), and $\frac{1}{w}\mathbf{Q}_g'\mathbf{A}_{gg}^{-1}\mathbf{Z}$ (of size 441 by 37,995), were stored in-memory using double precision reals. The differences in size among these matrices explained the main increase of random access memory (RAM) needed by ssSNPBLUP with genetic groups as covariates (25,211 MB; Table 1) in comparison to ssSNPBLUP with QP transformation (4,726 MB; Table 1). These differences could also largely explain the differences in wall clock times needed for the whole process of the 2 ssSNPBLUP with genetic groups: about 9,660 s for ssSNPBLUP with genetic groups as covariates, and about 4,600 s for ssSNPBLUP with the QP transformation. These differences are expected to be even larger for evaluations with larger pedigree.

The 3 ssSNPBLUP systems converged in about the same number of iterations (i.e., around 1,200–1,300 iterations; Table 1; Figure 1). This agreed with the extremal eigenvalues and the effective spectral condition numbers that were similar for the 2 systems (Table 1). It is worth noting that similar effective spectral condition numbers for the 3 ssSNPBLUP were possible after applying the second-level preconditioner also to the equations corresponding to the genetic group covariates. Without applying the second-level preconditioner to the genetic group covariates, around 150 additional iterations were needed to reach convergence.

In this study, we derived a system of equations for ssSNPBLUP that included genetic groups in the pedigree and that required less memory and computational costs

Table 1. Characteristics of systems for single-step SNP BLUP without genetic groups, with genetic groups fitted as covariates and fitted using the Quaas and Pollak (1981) transformation (QP transformation)

Characteristic	No genetic group	Covariate	QP transformation
Number of equations	26,861,588	26,863,352	26,863,352
Number of iterations	1,221	1,352	1,292
Smallest eig	1.488×10^{-5}	1.297×10^{-5}	1.357×10^{-5}
Largest eig	5.948	7.421	5.948
κ[1]	3.997×10^{5}	5.722×10^{5}	4.382×10^{5}
Software peak memory[2] (MB)	4,130	25,211	4,727
Wall clock time[3] (s)	4,017	9,657	4,602

[1]κ = effective spectral condition number of the preconditioned coefficient matrix defined as the ratio of the largest to the smallest eigenvalue.

[2]The software peak memory is defined as the peak resident set size ("high water mark," VmHWM) obtained from the Linux /proc virtual file system.

[3]Wall clock time of the whole process.

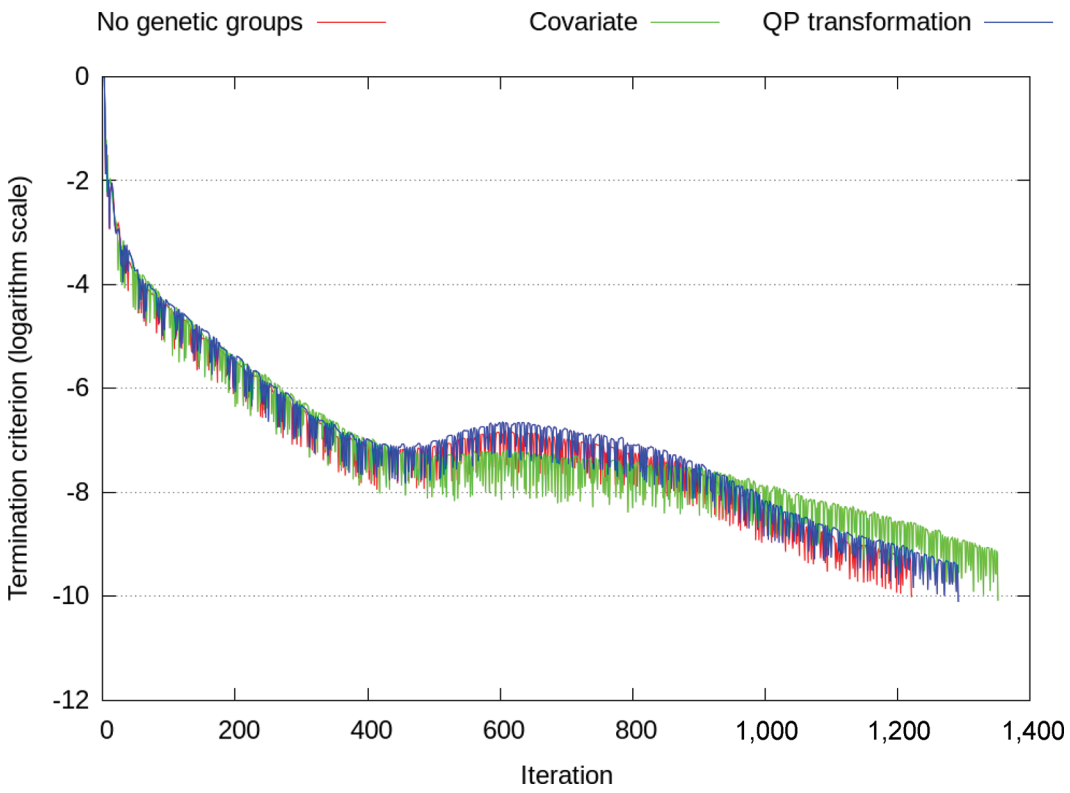

Figure 1. Termination criterion for single-step SNP BLUP without genetic groups, with genetic groups fitted as covariates, and fitted using the Quaas and Pollack (1981) transformation (QP transformation).

than an equivalent system with genetic groups fitted as covariates. Additional research should be performed to evaluate if fitting genetic groups in single-step evaluations (ssSNPBLUP or ssGBLUP) is beneficial (e.g., in terms of accuracy and bias) for such dairy cattle data sets. As already mentioned, the system of Equation [2] is equivalent to the ssGBLUP system proposed by Misztal et al. (2013), and may therefore encounter the same issues as its equivalent ssGBLUP (e.g., Bradford et al., 2019; Tsuruta et al., 2019). The concept of metafounders could be also easily implemented in ssSNPBLUP to replace the genetic groups, similar to studies with ssGBLUP (e.g., Macedo et al., 2020).

ACKNOWLEDGMENTS

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri & Food Project 16022; The Hague, the Netherlands) and the Breed4Food partners Cobb Europe (Colchester, Essex, United Kingdom), CRV (Arnhem, the Netherlands), Hendrix Genetics (Boxmeer, the Netherlands), and Topigs Norsvin (Helvoirt, the Netherlands). The use of the high-performance cluster was made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR, Wageningen, the Netherlands). The authors have not stated any conflicts of interest.

REFERENCES

Bradford, H. L., Y. Masuda, P. M. VanRaden, A. Legarra, and I. Misztal. 2019. Modeling missing pedigree in single-step genomic BLUP. J. Dairy Sci. 102:2336–2346. https://doi.org/10.3168/jds.2018-15434.

CRV Animal Evaluation Unit. 2020a. Statistical indicators, E16: Breeding value-Temperament during milking. Accessed Aug. 8, 2020. https://cooperatiecrv-be6.kxcdn.com/wp-content/uploads/2020/04/E_16-Gedrag-April-2020-Engels.pdf

CRV Animal Evaluation Unit. 2020b. Statistical indicators, E-15: Breeding value milking speed. Accessed Aug. 8, 2020. https://cooperatiecrv-be6.kxcdn.com/wp-content/uploads/2020/04/E_15-Melksnelheid-April-2020-Engels.pdf

CRV Animal Evaluation Unit. 2020c. Statistical indicators, E-7: Breeding value estimation of milk production traits with test-day model. Accessed Aug. 8, 2020. https://cooperatiecrv-be6.kxcdn.com/wp-content/uploads/2020/04/E_07-Melkproductie-April-2020-Engels.pdf

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. https://doi.org/10.1186/s12711-016-0273-2.

Hsu, W.-L., D. J. Garrick, and R. L. Fernando. 2017. The Accuracy and Bias of Single-Step Genomic Prediction for Populations Under Selection. G3: Genes, Genomes. G3 (Bethesda) 7:2685–2694.

Kaasschieter, E. F. 1988. A practical termination criterion for the conjugate gradient method. BIT 28:308–322. https://doi.org/10.1007/BF01934094.

Legarra, A., J. K. Bertrand, T. Strabel, R. L. Sapp, J. P. Sanchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. J. Anim. Breed. Genet. 124:286–295. https://doi.org/10.1111/j.1439-0388.2007.00671.x.

Lidauer, M., I. Stranden, E. A. Mantysaari, J. Poso, and A. Kettunen. 1999. Solving large test-day models by iteration on data and preconditioned conjugate gradient. J. Dairy Sci. 82:2788–2796. https://doi.org/10.3168/jds.S0022-0302(99)75536-0.

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

Macedo, F. L., O. F. Christensen, J.-M. Astruc, I. Aguilar, Y. Masuda, and A. Legarra. 2020. Bias and accuracy of dairy sheep evaluations using BLUP and ssGBLUP with metafounders and unknown parent groups. Genet. Sel. Evol. 52:47. https://doi.org/10.1186/s12711-020-00567-1.

Mäntysaari, E. A., M. Koivula, and I. Strandén. 2020. Symposium review: Single-step genomic evaluations in dairy cattle. J. Dairy Sci. 103:5314–5326. https://doi.org/10.3168/jds.2019-17754.

Matilainen, K., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2018. Single step genomic evaluation for female fertility in Nordic Red dairy cattle. J. Anim. Breed. Genet. 135:337–348. https://doi.org/10.1111/jbg.12353.

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

Nabben, R., and C. Vuik. 2006. A comparison of deflation and the balancing preconditioner. SIAM J. Sci. Comput. 27:1742–1759. https://doi.org/10.1137/040608246.

Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:91–98. https://doi.org/10.1016/S0022-0302(88)79986-5.

Quaas, R. L., and E. J. Pollak. 1981. Modified equations for sire models with groups. J. Dairy Sci. 64:1868–1872. https://doi.org/10.3168/jds.S0022-0302(81)82778-6.

Schaeffer, L. R. 2018. Necessary changes to improve animal models. J. Anim. Breed. Genet. 135:124–131. https://doi.org/10.1111/jbg.12321.

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. https://doi.org/10.3168/jds.2019-16789.

Vandenplas, J., M. P. L. Calus, H. Eding, and C. Vuik. 2019. A second-level diagonal preconditioner for single-step SNPBLUP. Genet. Sel. Evol. 51:30. https://doi.org/10.1186/s12711-019-0472-8.

Vandenplas, J., H. Eding, M. Bosmans, and M. P. L. Calus. 2020. Computational strategies for the preconditioned conjugate gradient method applied to ssSNPBLUP, with an application to a multivariate maternal model. Genet. Sel. Evol. 52:24. https://doi.org/10.1186/s12711-020-00543-9.

Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. J. Dairy Sci. 71:1310–1318. https://doi.org/10.3168/jds.S0022-0302(88)79688-5.

ORCIDS

Jeremie Vandenplas https://orcid.org/0000-0002-2554-072X
Herwin Eding https://orcid.org/0000-0002-6763-5110
Mario P. L. Calus https://orcid.org/0000-0002-3213-704X