Chapter 13

Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values

Samuel A. Clark and Julius van der Werf

Abstract

Genomic best linear unbiased prediction (gBLUP) is a method that utilizes genomic relationships to estimate the genetic merit of an individual. For this purpose, a genomic relationship matrix is used, estimated from DNA marker information. The matrix defines the covariance between individuals based on observed similarity at the genomic level, rather than on expected similarity based on pedigree, so that more accurate predictions of merit can be made. gBLUP has been used for the prediction of merit in livestock breeding, may also have some applications to the prediction of disease risk, and is also useful in the estimation of variance components and genomic heritabilities.

Key words Genomic, Prediction, Genomic selection, BLUP, Genetics, Genome-wide association studies, Methods

1 Introduction

1.1 DNA Markers for the Prediction of Merit

The availability of dense DNA marker information has enabled the large-scale genotyping of individuals for prediction of an individual's genetic merit. The most common markers used for the prediction of disease risk and genetic merit are called single nucleotide polymorphisms (SNPs) and are abundant on the genome. These genetic markers have been used for various purposes in human, livestock, and plant genetics. Some uses include: the detection of areas of the genome that have a significant effect on quantitative trait variation (quantitative trait loci—QTL), the prediction of an individual's risk to disease infection, and the estimation of heritability and genetic variance components [1]. In animal and plant industries, genetic markers have also been used to determine the genetic value of individuals so that they can be selected for breeding purposes.

Numerous statistical methods have been useful in helping make these predictions more accurate. A common, simple approach is single marker linear regression, which is used to identify significant

Cedric Gondro et al. (eds.), *Genome-Wide Association Studies and Genomic Prediction*, Methods in Molecular Biology, vol. 1019, DOI 10.1007/978-1-62703-447-0_13,

regions of the genome and has been extensively used for finding QTL (commonly referred to as association studies). However, there is an important statistical problem that the number of SNP effects is usually much larger than the number of observed phenotypes. One solution is to model SNP effects as random effects and make prior assumptions about the variance explained due to their effects. Some nonlinear methods such as Bayes A, Bayes B [2, 3], and Bayes C [4] give more emphasis to some genomic regions by allowing the variance to differ between SNP loci, whereas the genomic best linear unbiased prediction (gBLUP) method assigns the same variance to all loci and essentially treats them all as equally important.

1.2 Results Using gBLUP in Genomic Research

gBLUP has been examined in many research articles and has been shown to obtain as accurate or more accurate breeding values in livestock breeding programs than pedigree-based BLUP. VanRaden et al. [5] reported increases in breeding value accuracies of 20–50 %, similarly Harris et al. [6] used gBLUP to generate genomic predictions on 4,500 dairy cattle and found that reliabilities were 16–33 % higher than the breeding values based on parent average information for milk production traits. Moser et al. [7] also showed that there was very little difference between using gBLUP and the nonlinear models (i.e., Bayes A, B). However, Habier et al. [8] showed that these predictions quickly erode when the relationship between individuals with phenotypic information and those being predicted reduces. Clark et al. [9] showed that predictions using the Bayes B method would be more accurate if significant QTL exist, but accuracies become more equal to gBLUP when there are many QTL each with a small effect. All of these methods fit all SNP effects simultaneously in a prediction model, and in the context of a genome wide association study, Yang et al. [1] showed that in estimating the variance components and heritability of human height that common SNPs explain a larger proportion of genetic variance than the sum of significant SNPs obtained from single marker regressions.

2 Methods

2.1 Incorporating Marker Information into Best Linear Unbiased Prediction (BLUP)

The use of best linear unbiased prediction (BLUP) has enabled large amounts of genetic gain to be achieved in many livestock breeding programs. The traditional BLUP methodology relies on pedigree information to define the covariance between known relatives. This covariance can also be defined by using large amounts of DNA marker information, most commonly a large number of SNP markers. This matrix is termed the genomic relationship matrix (GRM). We will discuss two methods to incorporate genomic information into BLUP: ridge regression BLUP

(RR-BLUP) and genomic BLUP (gBLUP), and we will show how they are equivalent.

2.1.1 Ridge Regression BLUP

This method was examined by Meuwissen et al. [2] and Habier et al. [8] which assume the model

$$y = 1_n\mu + \sum_i \boldsymbol{W} q_i + e$$

where μ is the mean, $\boldsymbol{W}$ is a matrix that contains genotypes coded as 0, 1, or 2, and q_i is the effect of each SNP. The elements in $\boldsymbol{W}$ in each column j have an amount $2p_j$ (where p_j is the minor allele frequency of marker j) subtracted from the genotype code to achieve that the sum of coefficients in each column is zero. Here, the SNP effects are treated as random and summed over all segments. The genetic variance explained by the SNP effects is given by $\boldsymbol{WW}\prime\sigma_q^2$ and the residual variance is $I\sigma_e^2$, and the variance–covariance matrix among observations is therefore $\boldsymbol{WW}\prime\sigma_q^2 + I\sigma_e^2$. The variance for each SNP can be assumed equal. This method has also been termed RR (ridge regression or random regression) BLUP [8] or SNP BLUP. Alternatively, this variance has to be estimated for each SNP, in which a prior distribution of the SNP effects has to be assumed. Bayesian methods have been proposed to achieve this (see, e.g., [2–4]).

2.1.2 Genomic Best Linear Unbiased Prediction

The second method used to combine genomic information into BLUP is using a GRM as a substitute for the numerator relationship matrix and is called gBLUP. The gBLUP method was introduced by VanRaden [10] and Habier et al. [8]. In practice, the model used to implement gBLUP is:

$$y = \boldsymbol{X}b + \boldsymbol{Z}g + e$$

where y is a vector of phenotypes, $\mathbf{X}$ is a design matrix relating the fixed effects to each animal, b is a vector of fixed effects, $\mathbf{Z}$ is a design matrix allocating records to genetic values, g is a vector of additive genetic effects for an individual, and e is a vector of random normal deviates with variance σ_e^2. Furthermore $\text{var}(g) = \mathbf{G}\sigma_g^2$ where $\mathbf{G}$ is the genomic relationship matrix, and σ_g^2 is the genetic variance for this model. Note that the vector g contains animals with phenotypic data but can be extended to animals with no phenotypes. The first group is then referred to as the training or reference population, whereas the latter is the test population or a set of individuals to be predicted.

gBLUP has three important features that make it more desirable to use than RR-BLUP: (1) the dimensions of the genetic effects in the mixed model equations is reduced from $m \times m$ (where m is the number of markers) in RR-BLUP to $n \times n$ (where n is the number of individuals in the population) in

gBLUP, which is computationally more efficient; (2) the accuracy of an individual's genomic estimated breeding value (GEBV) can be calculated in the same way as in pedigree-based BLUP; and (3) gBLUP information can be incorporated with pedigree information in a single step method [11].

2.1.3 Equivalence Between gBLUP and RR-BLUP

Habier et al. [8] showed that gBLUP and RR-BLUP are actually equivalent models. The model for gBLUP (ignoring fixed effects) is given by:

$$y = 1_n\mu + \boldsymbol{Z}g + e$$

where y is a vector of phenotypes, 1_n is a vector of ones, μ is the mean, $\mathbf{Z}$ is a design matrix allocating records to genetic values, g is a vector of additive genetic effects for an individual, and e is a vector of random normal deviates σ_e^2. The variance of y in this model is given by $(\boldsymbol{ZGZ}\prime\sigma_g^2 + \boldsymbol{I}\sigma_e^2)$ where σ_e^2 is the residual variance.

Similarly the model for RR-BLUP is given by:

$$y = 1_n\mu + \sum_i \boldsymbol{W}q_i + e$$

where μ is the mean, $\boldsymbol{W}$ is an incidence matrix linking observations to SNP genotypes, q_i is the effect of each SNP which is treated as random, and the variance of y is $\text{var}(y)\,\boldsymbol{WW}'\sigma_q^2 + \boldsymbol{I}\sigma_e^2$ (assuming equal variance for each SNP), therefore the variance of y in gBLUP and RR-BLUP is the same (*see* **Note 1**).

2.2 Building the Genomic Relationship Matrix

Combining the information from genetic markers into a relationship matrix was first suggested by Nejati Javaremi et al. [12]. Similarly, Villaneuva et al. [13] examined the use of a GRM as a method of genomic evaluation and suggested that when genetic variation is explained by many QTL of small effect, BLUP using a GRM can be used to produce higher accuracy estimates than pedigree-based BLUP by representing additive relationships between individuals based on information using shared DNA markers. Relationship estimates in the GRM can deviate from the expected relationship given in the numerator relationship matrix **A**. For example, variation in the relationship between two full siblings may range from 0.4 to 0.6 instead of the expectation of 0.5 given in **A** [15, 16]. This exploitation of the variation in relationships is what makes the GRM a useful tool in genomic evaluations. Estimates of the GRM can be formed using different methods and various ways to make the GRM have been proposed [1, 10, 14]. Some of these will be presented in this section.

2.2.1 VanRaden [10]

The method presented by VanRaden [10] essentially develops the matrix $\boldsymbol{W}$ as presented in Subheading 2.1.1. He defined an incidence matrix **M**, coded as −1, 0, 1 that specifies which alleles each

individual has inherited. The minor allele frequency (the SNP allele with the lowest frequency) at locus i is p_i, and the matrix **P** contains the allele frequencies expressed as a difference from 0.5 and multiplied by 2, such that column i of **P** is $2(p_i - 0.5)$. Subtraction of **P** from **M** gives exactly $\boldsymbol{W}$ (termed *matrix* $\boldsymbol{Z}$ in [10]). The minor allele frequency correction forces the sum of coefficients across animals to be zero for each marker. I also give more weighting to rare alleles than to common alleles when calculating genomic relationships. The GRM is calculated as $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{W}'/[2\sum p_i(1-p_i)]$. The division by $2\sum p_i(1-p_i)$ places $\boldsymbol{G}$ on the same scale to the numerator relationship matrix ($\boldsymbol{A}$) which is used widely in livestock breeding; however, this is only if the allele frequencies used to scale $\boldsymbol{G}$ are referring to the same base population as used in **A** (*see* **Note 2**).

2.2.2 Yang et al. [1]

Other genomic matrices have been proposed, such as the one used by Yang et al. [1]. Here, they combined the information on all N SNPs (i) (coded 0, 1, 2) to calculate the relationship between individuals j and k into the GRM ($\boldsymbol{G}_{ijk}$) using a weighting scheme based on allele frequencies similar to VanRaden [8]. Weighting the off-diagonal and diagonal elements differently, when $j \neq k$ then:

$$\boldsymbol{G}_{jk} = \frac{1}{N}\sum_i \boldsymbol{G}_{ijk} = \frac{1}{N}\sum_i \frac{(w_{ij} - 2p_i)(w_{ik} - 2p_i)}{2p_i(1-p_i)}$$

When j is equal to k (i.e., the relationship of an individual to itself), then:

$$\boldsymbol{G}_{jk} = \frac{1}{N}\sum_i \boldsymbol{G}_{ijk} = 1 + \frac{1}{N}\sum_i \frac{w_{ij}^2 - (1-2p_i)w_{ij} + 2p_i^2}{2p_i(1-p_i)}$$

where w_{ij} is the element of $\boldsymbol{W}$ pertaining to marker i and individual j. These estimates of relationship are all relative to a base population in which the average relationship between individuals is zero (all individuals are completely unrelated). Yang et al. [1] used the individuals in the sample as the base so that the average relationship between all pairs of individuals is 0 and the average relationship of an individual with him- or herself is 1. N is the number of markers.

2.2.3 Goddard et al. [14]

The matrix $\boldsymbol{G}_m$ can be constructed as $\boldsymbol{G}_m = \boldsymbol{W}\boldsymbol{W}'/M$ where each element of matrix **W** is formed as in Yang et al. ([1], see above) and $\boldsymbol{M} = \sum 2p_j(1-p_j)$. The matrix **A** is the matrix with expected numerator relationships as derived from the pedigree information. Then, $\hat{\boldsymbol{G}}$ can be calculated as

$$\hat{\boldsymbol{G}} = [\boldsymbol{A} + b(\boldsymbol{G}_m - \boldsymbol{A})]$$

where σ_a^2 is the additive genetic variance and σ_g^2 is the variance of each of the marker effects, $b = \sigma_g^2/\sigma_a^2$. The regression of $\hat{\boldsymbol{G}}$ back toward $\boldsymbol{A}$ is said to now remove some of the error associated with

estimating genomic relationships from a finite number of markers, therefore acknowledging that $\hat{\boldsymbol{G}}$ is an estimate of the true genomic relationship $\boldsymbol{G}$.

2.3 How gBLUP Works

In the model used for RR-BLUP, the variance of known phenotypes can be written as $\mathbf{WW}' + \lambda \boldsymbol{I}$ where $\mathbf{W}$ links individual phenotypes to the marker effects, which is a matrix of animal genotypes coded 0, 1, or 2 (for the number of copies of a specific allele the animal has), and $\lambda = \sigma_e^2/\sigma_g^2$. The multiplication of $\mathbf{WW}'$ gives the correlation between the genomes of two individuals and the elements of the corresponding matrix have the same expected values as the numerator relationship matrix ($\boldsymbol{A}$) in the traditional BLUP equations. This is important because it further solidifies that even if there is no linkage disequilibrium (LD) then genomic estimates of merit will have a nonzero value because of the relationships between animals in $\mathbf{WW}'$. However, this also may mean that if LD is low then predictions of merit may quickly erode as the relationship between animals reduces [3, 8, 17, 18].

Using the GRM to compute genomic breeding values has simplified how genomic predictions are estimated and can be easily completed in software such as ASReml [19] (*see* **Note 3**) and R. The GRM combines data on all n animals with phenotypes and genotypes and links it to animals that have genotypes collected but no phenotypic information. ASReml [19] allows for an animal model to be fitted where the inverse of the G matrix is used to fit the covariance structure among the animal effects (*see* **Notes 4** and **5**). The mixed model looks like:

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{X} & 0 \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{G}^{11} & \boldsymbol{G}^{12} \\ 0 & \boldsymbol{G}^{21} & \boldsymbol{G}^{22} \end{bmatrix} \begin{bmatrix} b \\ g1 \\ g2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'y \\ \boldsymbol{Z}'y \\ 0 \end{bmatrix}$$

The positions of the GRM $\boldsymbol{G}^{11}$ is the subset of individuals that have phenotypic and genotypic information recorded on them, positions $\boldsymbol{G}^{12}$ and $\boldsymbol{G}^{21}$ pertain to the relationships between the animals with phenotypic data and those without, and $\boldsymbol{G}^{22}$ represents the animals without phenotypic measurements. The breeding values of animals without phenotypes are therefore estimated as:

$$\hat{g}_2 = -\left(\boldsymbol{G}^{22}\right)^{-1} \boldsymbol{G}^{21} \hat{g}_1$$

This is the genomic regression of breeding values of animals without data on the breeding value of animals with data or gBLUP.

2.4 A Simple Example Using Genomic BLUP

Let us assume we have five animals, four have phenotypes and we wish to use genotypic information to predict the breeding value of the fifth animal. Let us also assume that animal 1 is the parent of 2, 3, and 5, with there being no information about the ancestry of animal 4. If we assume that fixed effects are known and that each value of y is a deviation from the mean using BLUP

based on pedigree, we obtain estimates of each animal as $\hat{u} = (\boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{A}^{-1})^{-1}\boldsymbol{Z}'y$. To obtain estimates using genomic information $\boldsymbol{G}^{-1}$ replaces $\boldsymbol{A}^{-1}$.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0.5 & 0.5 & 0 & 0.5 \\ 0.5 & 1 & 0.25 & 0 & 0.25 \\ 0.5 & 0.25 & 1 & 0 & 0.25 \\ 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0.25 & 0.25 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{G} = \begin{bmatrix} 1.0 & 0.50 & 0.50 & 0.02 & 0.50 \\ 0.50 & 1.0 & 0.20 & 0.015 & 0.20 \\ 0.5 & 0.20 & 1.0 & 0.025 & 0.30 \\ 0.02 & 0.015 & 0.025 & 1.0 & 0.20 \\ 0.5 & 0.20 & 0.30 & 0.02 & 1.0 \end{bmatrix}$$

The $\boldsymbol{A}$ matrix is derived from the path coefficients from the pedigree. Whereas the $\boldsymbol{G}$ is an arbitrary example in which animal 3 is more similar to animal 5 than animal 2 (based on the expected degrees of relationship for half siblings). In the pedigree example, animal 4 is completely unrelated to the other animals; however, now with genomic data animal 4 shares some information with the other ones.

When assuming a heritability of 0.25, the breeding value of animal 5 would be estimated as;

$\widehat{u_5} = 0.5\widehat{u_1}$ (Note that $\widehat{u_1}$ contains also phenotypic information from animals 2 and 3).

Whereas under gBLUP, the prediction would be according to:

$$\widehat{g_5} = 0.499\widehat{g_1} - 0.026\widehat{g_2} + 0.0622\widehat{g_3} + 0.0144\widehat{g_4}$$

2.5 What Information Is Used

The regression coefficients presented above may not make sense at first because it seems illogical that the weight on the breeding value of animal 2 is negative and the weight on animal 4, which is far less related, is positive. The reason for this is that most of the information in 2 is used to predict the breeding value of 1. The importance of different sources of information can also be illustrated by the regression of the breeding values on phenotypes. These can be calculated as $\widehat{u_5} = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{V}'^{-1}y$ and for animal 5 the result shows that

$$\widehat{u_5} = 0.1136\widehat{y_1} - 0.0455\widehat{y_2} + 0.0455\widehat{y_3}$$

$$\widehat{g_5} = 0.1135\widehat{y_1} + 0.0328\widehat{y_2} + 0.0591\widehat{y_3} + 0.0519\widehat{y_4}$$

This has important consequences for interpreting the results of gBLUP. It illustrates that gBLUP and Pedigree BLUP are very similar and share similar sources of information. It also illustrates that information on unrelated individuals may now be included in an estimate of a breeding value. Given that animal 4 contributed no information in pedigree BLUP and now influences the prediction

in gBLUP. In this example, this would have a small effect, as the coefficient is small, but in a larger reference population there may be thousands of records contributing a small amount of information, altogether contributing to a large increase in accuracy. Another important implication of using genomic relationships in BLUP is that now known siblings can contribute different amounts of information to breeding value estimates, as there is now some ability to differentiate between these animals and access some of the within family variation, due to Mendelian sampling.

Example 1 VanRaden 2008 implementation of the GRM and gBLUP

```
#Making the genomic relationship matrix

nmarkers=1000
data = matrix(scan("genotypes.txt"),ncol=nmarkers,byrow=-
   TRUE);
sumpq=0
freq=dim(data)[1]
P=freq
lamda=ncol(data)
for(i in 1:ncol(data)){(freq[i]<-((mean(data[,i])/2)))
(P[i]=(2*(freq[i]-0.5)))
(sumpq=sumpq+(freq[i]*(1-freq[i])))}

Z<-data
for(i in 1:nrow(data)){
   for(j in 1:ncol(data)){(Z[i,j]<-((data[i,j]-1)-(P[j])))}
}
Zt=t(Z)
ZtZ=Z%*%Zt
G=ZtZ/(2*sumpq)
G

#gBLUP

for(i in 1:nrow(G)){
   (G[i,i]<-((G[i,i]+0.01)))}
y=matrix(scan("phen.txt"),ncol=1,byrow=TRUE);
I=matrix(1,100,1)
   EBV=(solve(1+(lamda*solve(G))))%*%y
```

3 Notes

1. The gBLUP method that uses a GRM is equivalent to the RR-BLUP method only when markers in $\boldsymbol{WW}'$ in the RR model are scaled the same as those used to calculate $\boldsymbol{G}$, i.e., $\boldsymbol{XX}' = \boldsymbol{X}_{ij}/2\sum p_i(1-p_i)$. Furthermore gBLUP and RR - BLUP are only equivalent when $\boldsymbol{G}\sigma_g^2 = \boldsymbol{WW}'\sigma_q^2$

2. When building the GRM, it may be important to examine the allele frequencies used to scale the matrix. In the original gBLUP method of VanRaden [10], it was proposed that frequencies should be of the base population and therefore needed to be estimated. However, recent work by Forni et al. [20] suggests that similar results can be obtained using the allele frequencies of the current population. The definition of the allele frequencies affects when the base animals are observed and is mainly important when gBLUP is used in the single step method of Misztal et al. [11]. The single step method combines pedigree and genomic information so that this information can be used to predict a single breeding value. Combining this information requires the base populations of genomic and pedigree relationship populations to be the same.
3. The ASReml software is easily used to implement gBLUP and can be downloaded from: http://www.vsni.co.uk/downloads/asreml.
4. To undertake gBLUP using ASReml, the user must provide a predefined GRM. This can be inverted and loaded into ASReml as a .giv file.
5. Often large data sets in ASReml can require more memory to be allocated to the program, this can be achieved by entering –s8 into the command line (see ASReml users guide for more details).

References

1. Yang J, Benyamin B, McEvoy BP et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–571
2. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829
3. Habier D, Tetens J, Seefried FR et al (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol 42:5
4. Habier D, Fernando RL, Kizilkaya K et al (2010) Extension of the Bayesian alphabet for genomic selection. In: Proceedings of the ninth congress on genetics applied to livestock production, Leipzig, 1–6 Aug 2010
5. VanRaden PM, Van Tassell CP, Wiggans GR et al (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92:16–24
6. Harris BL, Johnson DL, Spelman RJ (2008) Genomic selection in New Zealand and the implications for national genetic evaluation. In: Sattler JD (ed) Proceedings of the 36th ICAR session, Niagara Falls, New York, pp 325–330
7. Moser G, Tier B, Crump RE et al (2009) A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet Sel Evol 41:56
8. Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397
9. Clark S, Hickey JM, van der Werf JHJ (2011) Different models of genetic variation and their effect on genomic evaluation. Genet Sel Evol 43:18
10. VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

11. Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92:4648–4655
12. Nejati-Javaremi A, Smith C, Gibson JP (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. J Anim Sci 75:1738–1745
13. Villanueva B, Pong-Wong R, Fernandez J et al (2005) Benefits from marker-assisted selection under an additive polygenic genetic model. J Anim Sci 83:1747–1752
14. Goddard ME, Hayes BJ, Meuwissen TH (2011) Using the genomic relationship matrix to predict the accuracy of genomic selection. J Anim Breed Genet. doi:10.1111/j.1439-0388.2011.00964.x
15. Visscher PM, Medland SE, Ferreira MA et al (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2:e41
16. Hill WG, Weir BS (2011)Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genet Res doi:10.1017/S0016672310000480. http://dx.doi.org/
17. Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res 91:47–60
18. Clark SA, Hickey JM, Daetwyler H et al (2012) The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet Sel Evol 44:4
19. Gilmour AR, Gogel BJ, Cullis BR et al (2009) ASReml user guide release 30. VSN International, Hemel Hempstead
20. Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet Sel Evol 43:1