



## A predictive assessment of genetic correlations between traits in chickens using markers

Mehdi Momen, Ahmad Ayatollahi Mehrgardi, Ayoub Sheikhy, Ali Esmailizadeh, Masood Asadi Fozi, Andreas Kranis, Bruno D. Valente, Guilherme J. M. Rosa, Daniel Gianola

### ► To cite this version:

Mehdi Momen, Ahmad Ayatollahi Mehrgardi, Ayoub Sheikhy, Ali Esmailizadeh, Masood Asadi Fozi, et al.. A predictive assessment of genetic correlations between traits in chickens using markers. *Genetics Selection Evolution*, 2017, 49 (1), pp.16. 10.1186/s12711-017-0290-9 . hal-01479145

**HAL Id: hal-01479145**

**<https://hal.science/hal-01479145>**

Submitted on 28 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Open Access



# A predictive assessment of genetic correlations between traits in chickens using markers

Mehdi Momen<sup>1</sup>, Ahmad Ayatollahi Mehrgardi<sup>1\*</sup>, Ayoub Sheikhy<sup>2</sup>, Ali Esmailizadeh<sup>1,3</sup>, Masood Asadi Fozi<sup>1</sup>, Andreas Kranis<sup>4</sup>, Bruno D. Valente<sup>5</sup>, Guilherme J. M. Rosa<sup>5,6</sup> and Daniel Gianola<sup>5,6,7</sup>

## Abstract

**Background:** Genomic selection has been successfully implemented in plant and animal breeding programs to shorten generation intervals and accelerate genetic progress per unit of time. In practice, genomic selection can be used to improve several correlated traits simultaneously via multiple-trait prediction, which exploits correlations between traits. However, few studies have explored multiple-trait genomic selection. Our aim was to infer genetic correlations between three traits measured in broiler chickens by exploring kinship matrices based on a linear combination of measures of pedigree and marker-based relatedness. A predictive assessment was used to gauge genetic correlations.

**Methods:** A multivariate genomic best linear unbiased prediction model was designed to combine information from pedigree and genome-wide markers in order to assess genetic correlations between three complex traits in chickens, i.e. body weight at 35 days of age (BW), ultrasound area of breast meat (BM) and hen-house egg production (HHP). A dataset with 1351 birds that were genotyped with the 600 K Affymetrix platform was used. A kinship kernel (**K**) was constructed as  $\mathbf{K} = \lambda \mathbf{G} + (1 - \lambda) \mathbf{A}$ , where **A** is the numerator relationship matrix, measuring pedigree-based relatedness, and **G** is a genomic relationship matrix. The weight ( $\lambda$ ) assigned to each source of information varied over the grid  $\lambda = (0, 0.2, 0.4, 0.6, 0.8, 1)$ . Maximum likelihood estimates of heritability and genetic correlations were obtained at each  $\lambda$ , and the "optimum"  $\lambda$  was determined using cross-validation.

**Results:** Estimates of genetic correlations were affected by the weight placed on the source of information used to build **K**. For example, the genetic correlation between BW–HHP and BM–HHP changed markedly when  $\lambda$  varied from 0 (only **A** used for measuring relatedness) to 1 (only genomic information used). As  $\lambda$  increased, predictive correlations (correlation between observed phenotypes and predicted breeding values) increased and mean-squared predictive error decreased. However, the improvement in predictive ability was not monotonic, with an optimum found at some  $0 < \lambda < 1$ , i.e., when both sources of information were used together.

**Conclusions:** Our findings indicate that multiple-trait prediction may benefit from combining pedigree and marker information. Also, it appeared that expected correlated responses to selection computed from standard theory may differ from realized responses. The predictive assessment provided a metric for performance evaluation as well as a means for expressing uncertainty of outcomes of multiple-trait selection.

\*Correspondence: mehrgardi@uk.ac.ir

<sup>1</sup> Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman (SBUK), Kerman, Iran

Full list of author information is available at the end of the article

**Background**

The increasing availability of genome-wide dense molecular markers [e.g., single nucleotide polymorphisms (SNPs)] has opened new avenues for obtaining additional genetic gain in breeding of elite animals and plants by exploiting “genomic selection” methods. These techniques have become important tools in modern breeding programs [1, 2]. Many statistical methods with parametric or non-parametric formulations have been proposed to predict either genomic estimated breeding values (GEBV) of animals or yet-to-be observed phenotypes [1, 3–6].

Most prediction studies have been based on single-trait (uni-variate) statistical models. However, in practice, animals and plants often must be evaluated for several economically important traits. Multiple-trait model predictions have been typically regarded as better than uni-variate predictions [7]. For example, milk yield and composition in dairy cattle or grain yield and resistance to disease in plants are often analyzed with multiple-trait methods [8, 9]. Multi-trait models based on pedigree information represent the typical modeling strategy used to capitalize on genetic evaluation of several correlated traits before genomic selection methods became popular [10]. A multiple-trait analysis requires knowledge of phenotypic and genetic correlations among characters [7]. These correlations indicate the extent to which measurements on one trait inform about other traits [11], and predictions based on single-trait models do not exploit the extra information provided by other traits.

Multiple-trait genomic selection models (MT-GS) have been explored and tested in research only to a limited extent [12]. A genome-based multiple-trait analysis may also offer insight into mechanisms that create trait associations, such as pleiotropy and linkage disequilibrium (LD) between quantitative trait loci (QTL) and markers [13]. One hypothesis is that correlation parameters that are inferred using whole-genome dense molecular markers may give a novel picture of the genetic correlation between traits [14, 15]. However, the sources of genetic and genomic correlations may be distinct [13, 16]. Genomic correlations depend in part on linkage disequilibrium (LD) relationships between markers and QTL, which are unknown, while genetic correlations are in part a function of LD between QTL. Multivariate genome-based models may produce “missing”: situation in which the genetic correlation is undetected by the markers, “excessive”: LD between markers increase the magnitude of the pleiotropy effects of the QTL, or even “spurious”: there is no pleiotropy but LD between markers and/or pairs of QTL may produce pseudo pleiotropy (abbreviated as MES) genetic correlations and, as a consequence, distort expectations about outcomes of multiple-trait selection.

The objective of this study was to infer genetic and genomic correlations between three traits measured in broilers by exploring linear combinations of pedigree-based (genealogical) and marker-based relationship matrices. As advocated by [17], a predictive approach was used to gauge parameter estimates and to provide an empirical test of the extent of genetic association between traits.

**Methods**

**Data**

The data consisted of records on 1351 birds from a commercial broiler chicken line that had undergone several generations of selection using the traditional multiple-trait genetic evaluations at the Aviagen Ltd Company (Aviagen Ltd, Newbridge, UK). The traits considered were body weight at 35 days of age (BW), ultrasound area of breast meat (BM), and hen-house production (HHP, total number of eggs laid between weeks 28 and 54). Some features of the dataset and pedigree information are in Table 1. All birds had phenotype records and a known sire and dam, and there were 326 and 274 paternal half-sib and full-sib groups in the sample, respectively. This dataset has also been used in other studies by Abdollahi-Arpanahi et al. [18] and Morota et al. [19].

**Phenotype correction**

Prior to implementing the genome-enabled trivariate prediction model, we pre-corrected phenotypes to eliminate all known nuisance non-genetic sources of variation. This correction was based on uni-variate mixed effects models; BW and BM were corrected for a combined effect of sex, hatch week, contemporary group of parents, and pen in the growing farm. HHP was corrected for random hatch effects, with a general mean as the sole fixed effect. Figure 1 shows a scatter plot of pre-corrected phenotypes for these traits. A positive association between BW and BM is suggested, whereas the scatter plots for the pairs BM-HHP and BW-HHP do not indicate concomitant variation.

**Table 1 Pedigree information and features of the chicken data used**

Total birds in the pedigree	1675
Number of sires	326
Number of dams	592
Number of full-sib groups	274
Number of progeny with records and known sire and dam	1351
Number of inbreds (pedigree-based inbreeding >0)	159
Inbreeding coefficient range all birds in the pedigree (%)	0.4 to 10.9

### Genotyping

The 1351 birds were genotyped using an 600 K Affymetrix SNP chip. SNPs with a minor allelic frequency (MAF) lower than 1% and a call frequency lower than 0.95 were filtered out. Missing genotypes were imputed locus by locus using the Beagle software version 3.3.2 [20]. After quality control, 354,364 SNPs remained for statistical analyses.

### Whole-genome prediction models

Tri-variate linear models were used for estimating (co) variance components and for predicting genomic breeding values. Such models were an extension of a typical single-trait model with random pedigree or genome-based effects, which can be represented as:

$$\mathbf{y}_t = \mathbf{1}\mu_t + \mathbf{Z}\mathbf{g}_t + \boldsymbol{\epsilon}_t; \quad t = 1, 2, 3, \quad (1)$$

where,  $\mathbf{y}_t$  is a vector of  $m \times 1$  pre-corrected phenotypes for trait  $t$  ( $m = 1351$ );  $\mu_t$  is a general constant and  $\mathbf{1}$  is a vector of ones;  $\mathbf{Z}$  is an incidence matrix (an identity matrix in all cases) that allocates records to breeding values;  $\mathbf{g}_t$  is a vector of additive genetic effects or of direct genomic breeding values, and  $\boldsymbol{\epsilon}_t$  is a vector of residuals for trait  $t$ . It was assumed that  $\mathbf{g}_t \sim (0, \mathbf{K}\sigma_{\mathbf{g}_t}^2)$  where  $\sigma_{\mathbf{g}_t}^2$  is the additive genetic or genomic variance of trait  $t$ , and  $\mathbf{K}$  ( $m \times m$ ) reflects a covariance structure that results from the combined use of pedigree and marker information, as described later. Random residuals were assumed to follow a normal distribution  $\boldsymbol{\epsilon}_t \sim N(0, \mathbf{I}_t\sigma_{\epsilon_t}^2)$ , where  $\mathbf{I}_t$  is an  $m \times m$  identity matrix and  $\sigma_{\epsilon_t}^2$  is the residual variance for trait  $t$ ; this term represents variation of pre-corrected phenotypes that is not explained by additive genomic effects. The vectors  $\mathbf{g}_t$  and  $\boldsymbol{\epsilon}_t$  were assumed to be independent. The multi-variate model was:

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{pmatrix} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}_3 \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \end{bmatrix}, \quad (2)$$

where,  $\mathbf{y}_t$ ,  $\mu_t$ ,  $\mathbf{Z}_t$ ,  $\mathbf{g}_t$  and  $\boldsymbol{\epsilon}_t$  are as before. The vector of multi-trait additive genetic or genomic breeding values was distributed as

$\begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \end{bmatrix} \sim N(0, \mathbf{K} \otimes \mathbf{Q})$ , where  $\mathbf{K}$

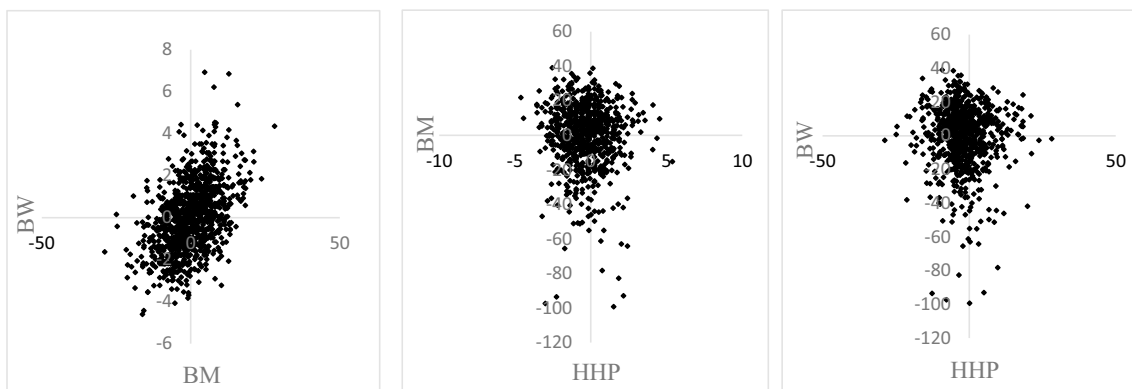
is a kinship or kernel matrix (described later) and  $\mathbf{Q}$  is the  $(3 \times 3)$  matrix of pedigree- or marker-based covariances among traits. The multivariate residual distribution was assumed to be

$\begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \end{bmatrix} \sim N(0, \mathbf{I} \otimes \mathbf{R})$ , where  $\mathbf{R}$  is

the  $(3 \times 3)$  residual covariance matrix among traits. The Kronecker product ( $\otimes$ ) notation applies to the residual covariance since all traits were measured on all birds.

### Pedigree-based and whole-genome relationship matrices

In a genomic best linear unbiased prediction model (GBLUP), a genomic relationship matrix ( $\mathbf{G}$ ) computed from marker data replaces the pedigree-based matrix ( $\mathbf{A}$ ) of standard BLUP applications. The genomic relationship matrix intends to measure the realized fraction of alleles shared, rather than the expected fraction, as is the case for  $\mathbf{A}$  [21, 22]. Genomic relationship matrices can be calculated in different ways (e.g., [23]) but here we used two known alternatives, as described next. First, VanRaden [22] proposed the  $m \times m$  matrix  $\mathbf{G}_V = \frac{\mathbf{W}\mathbf{W}'}{2\sum p_i q_i}$ , which renders  $\mathbf{G}$  analogous to the numerator relationship matrix  $\mathbf{A}$  due to the denominator,  $2\sum p_i q_i$ . Here,  $\mathbf{W}$  is a  $m \times p$  centered matrix of SNP genotype codes 0, 1 and 2 ( $p = 354,364$ ) and  $p_i$  is the MAF at locus  $i$ . Second, Forni



**Fig. 1** Scatter plots of phenotypes pre-corrected for non-genetic sources of variation for body weight (BW), breast muscle area (BM) and hen-house production (HHP)

et al. [21] suggested a modification of the denominator,  $\mathbf{G}_F = \frac{\mathbf{W}\mathbf{W}'}{\{\text{trace}[\mathbf{W}\mathbf{W}']\}/m}$ , which attempts to attain compatibility of the genomic relationship matrix with  $\mathbf{A}$  when either the average level inbreeding is low or when the number of generations back to the base population is small.

An alternative to using any given  $\mathbf{G}$  is to combine genomic and pedigree information into a single kinship “kernel” matrix (in the sense of [24]). A “kernel” matrix that exploits genealogy information together with marker-based information could potentially capture parts of the genetic covariance among traits that is not accounted for by either  $\mathbf{A}$  or  $\mathbf{G}$  alone. We followed multiple-kernel ideas [25] and used the kernel  $\mathbf{K} = \lambda\mathbf{G} + (1 - \lambda)\mathbf{A}$ , where  $\lambda$  is a parameter (weight) bounded between 0 and 1, and  $\mathbf{G} = \mathbf{G}_V$  or  $\mathbf{G}_F$ . For example, if  $\lambda = 0$ , pedigree information “dominates” in the model, which retrieves a traditional pedigree-based BLUP. Our expectation was that a specific combination of  $\mathbf{A}$  and  $\mathbf{G}$  matrices would provide the “best” estimates of parameters, as gauged by prediction of outcomes, as opposed to using either  $\mathbf{A}$  or  $\mathbf{G}$  alone or both, with (co)variance components estimated in training samples. To assess the best value of  $\lambda$ , we applied the grid  $\lambda = (0, 0.2, 0.4, 0.6, 0.8, 1)$  and evaluated the ensuing predictive abilities over such a grid.

When using marker- and pedigree-based relationship matrices together, scaling of genomic relationship matrices is needed for interpretation of parameters in the context of theory, e.g., in relation to a base population [26]. Estimates of parameters may be distorted if a genomic relationship matrix is not on the same scale as the pedigree-based relationship matrix. A reasonable rescaling may be achieved by using genomic relationship matrices with elements that range between 0 and 2, which are the minimum and maximum values of  $\mathbf{A}$ , respectively. To render  $\mathbf{G}$  on the same scale as  $\mathbf{A}$ , we used the *map min-max-function* that is widely used in machine learning, e.g. [27], as follows:

$$Gs_{ij} = \frac{(Gs_{\max} - Gs_{\min}) \times (G_{ij} - G_{\min})}{G_{\max} - G_{\min}} + Gs_{\min}. \quad (3)$$

Here,  $Gs_{ij}$  is a scaled element of the  $\mathbf{G}_V$  or  $\mathbf{G}_F$  matrix and  $G_{ij}$  is typical element of  $\mathbf{G}_V$  or  $\mathbf{G}_F$ ;  $Gs_{\max} = 2$  and  $Gs_{\min} = 0$  are the minimum and maximum values elements that the scaled matrix is allowed to take, respectively, and  $G_{\min}$  and  $G_{\max}$  are the maximum and minimum entries of the  $\mathbf{G}_V$  or  $\mathbf{G}_F$  matrix, respectively. While  $\mathbf{G}_V$  and  $\mathbf{G}_F$  may contain negative off-diagonals, this is not the case for the scaled matrices used here.

### Model fitting and validation

Variance and covariance components were estimated with multiple-trait restricted maximum likelihood

(REML) via an average information algorithm (AI) implemented in the WOMBAT program [28]. The software provides point estimates of (co)variance components and their asymptotic standard errors. Matrix  $\mathbf{K} = \lambda\mathbf{G} + (1 - \lambda)\mathbf{A}$  was used as kinship matrix, where  $\mathbf{G}$  was either the unscaled or scaled versions  $\mathbf{G}_V$  or  $\mathbf{G}_F$ .

We used a cross-validation scheme with 20 randomly constructed training and testing sets to assess predictive ability over the grid of  $\lambda$  values. We randomly partitioned the whole data into training (60%) and testing (40%) sets in each of the 20 repetitions. After a model was fitted to the training set data, we compared its predictions against realized values in the test set. Predictive ability was measured by mean squared error (MSE) and by the correlation between predicted and observed phenotypes in the testing set.

### Realized versus expected genetic regressions between traits

We also evaluated predictive relationships between pairs of traits, i.e., BW-BM, BW-HHP and BM-HHP, according to the cross-validation scheme described earlier. Over the predefined grid  $\lambda = (0, 0.2, 0.4, 0.6, 0.8, 1)$ , we computed least-squares estimates of the regression of the phenotype for trait  $x$  on DGV for trait  $y$ , and vice versa, for each pair of traits for each of the 20 validation sets. These realized regressions were compared to expected genetic regressions deduced from REML (co)variance component estimates as:

$$b_{(x,y)}(\lambda) = r_{G(x,y)}(\lambda) \times \sqrt{\sigma_{G(x)}^2(\lambda) / \sqrt{\sigma_{G(y)}^2(\lambda)}},$$

where,  $r_{G(x,y)}(\lambda)$  is the estimated genetic correlation between traits  $x$  and  $y$ , and  $\sigma_{G(x)}^2(\lambda)$  and  $\sigma_{G(y)}^2(\lambda)$  are the genetic variances estimated by REML over the predefined grid of  $\lambda$ .

## Results

### Heritability

Table 2 shows the heritability estimates obtained for each  $\lambda$  value, both for unscaled and scaled genomic relationship matrices. A low to moderate heritability was found for BW, BM and HHP. When using pedigree-based information only, heritability estimates (standard errors in parenthesis) were 0.187 (0.049), 0.244 (0.052) and 0.315 (0.074), respectively (Table 2). These estimated heritabilities changed to 0.165 (0.039), 0.255 (0.042) and 0.196 (0.052) with an unscaled  $\mathbf{G}_F$ , and to 0.156 (0.06), 0.243 (0.04) and 0.174 (0.04) with an unscaled  $\mathbf{G}_V$ . Scaling the genomic relationship matrices increased heritability estimates relative to those obtained from unscaled matrices. Estimated heritabilities in the present study were lower than in [29, 30] using the same population from which



**Table 2** Estimates of heritability for body weight (BW), ultrasound area of breast meat (BM) and hen-house egg production (HHP) obtained by placing varying weights ( $\lambda$ ) on the pedigree-based relationship matrix ( $\mathbf{A}$ ) and on Forni's ( $\mathbf{G}_F$ ) or VanRaden's ( $\mathbf{G}_V$ ) relationship matrix

Regularization parameter ( $\lambda$ )	$\mathbf{G}_F$						$\mathbf{G}_V$					
	Unscaled			Scaled			Unscaled			Scaled		
	$h^2_{BW}$	$h^2_{BM}$	$h^2_{HHP}$	$h^2_{BW}$	$h^2_{BM}$	$h^2_{HHP}$	$h^2_{BW}$	$h^2_{BM}$	$h^2_{HHP}$	$h^2_{BW}$	$h^2_{BM}$	$h^2_{HHP}$
<b>A</b> ( $\lambda = 0$ )	0.187	0.244	0.315	0.187	0.244	0.315	0.187	0.244	0.315	0.187	0.244	0.315
$\lambda = 0.20$	0.226	0.291	0.309	0.234	0.297	0.340	0.223	0.291	0.299	0.230	0.295	0.311
$\lambda = 0.40$	0.232	0.303	0.285	0.278	0.348	0.360	0.227	0.299	0.270	0.247	0.318	0.295
$\lambda = 0.60$	0.219	0.297	0.258	0.315	0.395	0.374	0.213	0.290	0.239	0.245	0.323	0.272
$\lambda = 0.80$	0.197	0.281	0.230	0.335	0.431	0.377	0.189	0.272	0.209	0.227	0.316	0.245
<b>G</b> ( $\lambda = 1$ )	0.165	0.255	0.196	0.313	0.442	0.361	0.156	0.243	0.174	0.193	0.293	0.214

The largest estimates are italics

our dataset was drawn but with a larger sample size from four generations of three commercial lines, at varying intensities of selection in the Aviagen UK breeding program. For example, in [29, 30] estimates for BW ranged from 0.326 (0.011) to 0.399 (0.015), whereas in our study they ranged from 0.156 (0.06) to 0.187 (0.049). Our result is based on a small subset of birds taken from the overall population; therefore it is expected that estimated heritabilities  $h^2$  would differ from those obtained using all available data, which would account for past selection.

Here, we used scaled kinship matrices to obtain “genetic parameters” which do not necessarily correspond to only those from standard pedigree-based additive genetic relationships or realized genomic pairwise similarities. Following VanRaden [22], if the expectation of  $\mathbf{G}$  is  $\mathbf{A}$ , then;  $E(\mathbf{K}) = E(\lambda\mathbf{G} + (1 - \lambda)\mathbf{A}) = \mathbf{A}$ . However, if one uses a scaled  $\mathbf{G}_V$ , it follows from the scaling formula that  $E(G_{s_{ij}}) = 2E(G_{ij} \times G_{min}/(G_{max} - G_{min}))$ . The latter expectation cannot be written in a closed form, because this requires knowledge of the distributions of  $G_{min}$  and  $G_{max}$ .

Our multiple-trait GBLUP analysis indicated that the highest heritability estimates were not obtained at the extremes (0 or 1) of the  $\lambda$  grid. For example, the highest genomic heritability for BW was obtained at  $\lambda = 0.4(0.8)$ , for unscaled (scaled)  $\mathbf{G}_F$ , and at  $\lambda = 0.4$  for the two versions of  $\mathbf{G}_V$ . Scaling  $\mathbf{G}_F$  and  $\mathbf{G}_V$  always increased heritability estimates. For BM, higher heritabilities were obtained when scaling was applied. More specifically, the highest estimates were obtained at  $\lambda = 1$  ( $\mathbf{G}_F$ ) and  $\lambda = 0.6$  ( $\mathbf{G}_V$ ). With unscaled matrices, the highest heritabilities were obtained at  $\lambda = 0.4$  and  $\lambda = 0.2$ . For HHP, the highest heritabilities were found at the extreme values of  $\lambda$ :  $\lambda = 0$  for unscaled  $\mathbf{G}_F$  and  $\mathbf{G}_V$  (scaled or unscaled) and  $\lambda = 1$  for scaled  $\mathbf{G}_F$ . Adding genomic information had little impact on heritability estimates of HHP, except with scaled  $\mathbf{G}_F$ .

Our findings illustrate a fairly obvious point made by Legarra et al. [26]: genomic heritability and its estimates are not invariant with respect to how  $\mathbf{G}$  is constructed. Hence, inferences and comparisons between results from different studies must be done with care. In short, our results with a multiple-trait model indicated that a pedigree-marker based kernel ( $\mathbf{K}$ ) had an impact on heritability estimates and that scaling of the genomic relationship matrix led to higher “heritability” estimates, especially for  $\mathbf{G}_F$ .

### Genetic correlations

Estimates of correlations are in Table 3 and Figs. 2 and 3. Estimates of residual and phenotypic correlations were less sensitive to  $\lambda$  than genetic correlations, so our discussion concentrates on the latter. All parameters were estimated for each  $\lambda$  and for each of the two genomic relationship matrices. When using a pedigree-marker based kinship matrix ( $\mathbf{K}$ ), estimates of genetic correlations for BW-HHP and BM-HHP changed gradually when  $\lambda$  increased from 0 to 1. Results are shown graphically in Figs. 2 and 3 for  $\mathbf{G}_F$  and  $\mathbf{G}_V$ , respectively. Changes were more pronounced for the genetic correlation between BW and HHP, which decreased in absolute value from  $-0.192$  ( $\lambda = 0$ ) to  $-0.02$  ( $\mathbf{G}_F$ , unscaled),  $-0.019$  ( $\mathbf{G}_F$ , scaled), and  $0.033$  ( $\mathbf{G}_V$  scaled or unscaled) with  $\lambda = 1$ . Estimates of the genetic correlation between BM and HHP were always negative and tended to decrease in absolute value as  $\lambda$  increased. They decreased from about  $-0.206$  when only pedigree-based information ( $\lambda = 0$ ) was used to  $-0.154$  when only genomic information ( $\lambda = 1$ ) was used to construct  $\mathbf{K}$  from scaled or unscaled versions of  $\mathbf{G}_F$ . BW and BM presented large positive genetic correlation estimates that ranged from 0.484 with the pedigree-based model to 0.497 (0.525) when only  $\mathbf{G}_F$  ( $\mathbf{G}_V$ ) was used. It was insensitive to scaling

**Table 3** Phenotypic ( $r_p$ ) and environmental ( $r_e$ ) correlations between body weight (BW), ultrasound area of breast meat (BM) and hen-house egg production (HHP) from a tri-variate analysis with varying weights ( $\lambda$ ) on the pedigree-based relationship matrix (**A**) and on Forni's (**G<sub>F</sub>**) or VanRaden's (**G<sub>V</sub>**) relationship matrix

Regularization parameter ( $\lambda$ )	$r_{e(BW,BM)}$	$r_{e(BW,HHP)}$	$r_{e(BM,HHP)}$	$r_{p(BW,BM)}$	$r_{p(BW,HHP)}$	$r_{p(BM,HHP)}$
<b>G<sub>F</sub></b>						
<i>Unscaled</i>						
<b>A</b> ( $\lambda = 0$ )	0.480 (0.034)	−0.026 (0.058)	−0.010 (0.063)	0.480 (0.023)	−0.066 (0.036)	−0.065 (0.039)
$\lambda = 0.20$	0.481 (0.036)	−0.023 (0.059)	−0.003 (0.064)	0.479 (0.023)	−0.067 (0.037)	−0.065 (0.039)
$\lambda = 0.40$	0.482 (0.035)	−0.034 (0.057)	−0.011 (0.061)	0.479 (0.023)	−0.065 (0.037)	−0.063 (0.039)
$\lambda = 0.60$	0.482 (0.033)	−0.047 (0.053)	−0.020 (0.057)	0.479 (0.023)	−0.062 (0.036)	−0.061 (0.039)
$\lambda = 0.80$	0.481 (0.031)	−0.058 (0.049)	−0.028 (0.053)	0.479 (0.023)	−0.059 (0.036)	−0.059 (0.038)
<b>G</b> ( $\lambda = 1$ )	0.479 (0.029)	−0.068 (0.045)	−0.035 (0.049)	0.479 (0.023)	−0.059 (0.036)	−0.062 (0.038)
<i>Scaled</i>						
$\lambda = 0.20$	0.484 (0.035)	−0.022 (0.059)	−0.006 (0.064)	0.479 (0.024)	−0.074 (0.039)	−0.073 (0.041)
$\lambda = 0.40$	0.484 (0.036)	−0.024 (0.059)	−0.006 (0.063)	0.478 (0.027)	−0.079 (0.042)	−0.080 (0.045)
$\lambda = 0.60$	0.485 (0.035)	−0.034 (0.057)	−0.012 (0.061)	0.477 (0.031)	−0.080 (0.047)	−0.084 (0.049)
$\lambda = 0.80$	0.484 (0.033)	−0.050 (0.052)	−0.062 (0.078)	0.477 (0.035)	−0.072 (0.052)	−0.084 (0.055)
<b>G</b> ( $\lambda = 1$ )	0.479 (0.029)	−0.068 (0.045)	−0.035 (0.049)	0.481 (0.038)	−0.052 (0.056)	−0.082 (0.059)
<b>G<sub>V</sub></b>						
<i>Unscaled</i>						
$\lambda = 0.20$	0.480 (0.036)	−0.033 (0.058)	−0.004 (0.063)	0.480 (0.023)	−0.066 (0.037)	−0.066 (0.039)
$\lambda = 0.40$	0.480 (0.035)	−0.046 (0.055)	−0.011 (0.059)	0.480 (0.023)	−0.064 (0.036)	−0.064 (0.039)
$\lambda = 0.60$	0.479 (0.033)	−0.060 (0.051)	−0.018 (0.055)	0.480 (0.023)	−0.061 (0.036)	−0.063 (0.039)
$\lambda = 0.80$	0.476 (0.030)	−0.070 (0.047)	−0.024 (0.051)	0.481 (0.023)	−0.059 (0.036)	−0.062 (0.038)
<b>G</b> ( $\lambda = 1$ )	0.474 (0.028)	−0.079 (0.044)	−0.030 (0.047)	0.481 (0.023)	−0.060 (0.036)	−0.065 (0.038)
<i>Scaled</i>						
$\lambda = 0.20$	0.484 (0.036)	−0.030 (0.059)	−0.006 (0.063)	0.479 (0.024)	−0.070 (0.037)	−0.071 (0.040)
$\lambda = 0.40$	0.483 (0.035)	−0.043 (0.056)	−0.010 (0.060)	0.479 (0.024)	−0.069 (0.038)	−0.072 (0.041)
$\lambda = 0.60$	0.482 (0.033)	−0.057 (0.053)	−0.018 (0.056)	0.479 (0.025)	−0.066 (0.039)	−0.072 (0.041)
$\lambda = 0.80$	0.478 (0.031)	−0.070 (0.048)	−0.025 (0.051)	0.481 (0.025)	−0.061 (0.039)	−0.072 (0.042)
<b>G</b> ( $\lambda = 1$ )	0.474 (0.028)	−0.078 (0.044)	−0.030 (0.047)	0.483 (0.025)	−0.056 (0.039)	−0.073 (0.042)

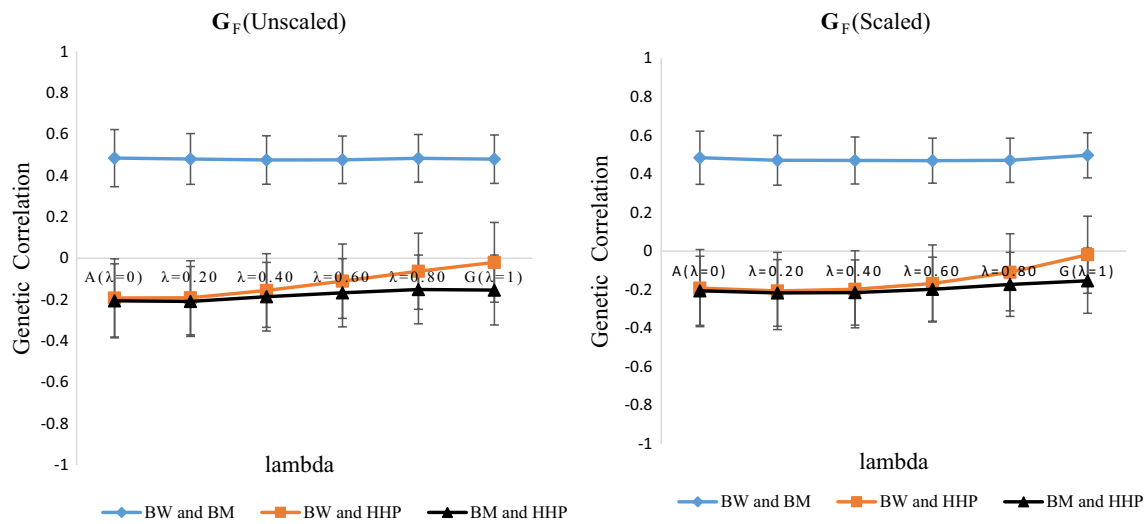
**A**: numerator relationship matrix, **G<sub>F</sub>**: Forni's relationship matrix, **G<sub>V</sub>**: VanRaden's relationship matrix

of the **G** matrix. Standard errors of estimates for BM-BW (results not shown) tended to decrease when  $\lambda$  increased. There were no clear tendencies for the standard errors of estimates of the genetic correlation of BW with HHP and BM with HHP. In short, classical genetic correlations (based on **A**) and genomic correlations (based on **G**) were distinct, depending on the pairs of traits considered. However, varying  $\lambda$  from 0 to 1 produced very minor changes in estimates of the genetic correlation between BM and HHP, but large changes in estimates of the genetic correlation between BW and HHP. Estimates of the genetic correlations between BW and BM were insensitive to  $\lambda$ .

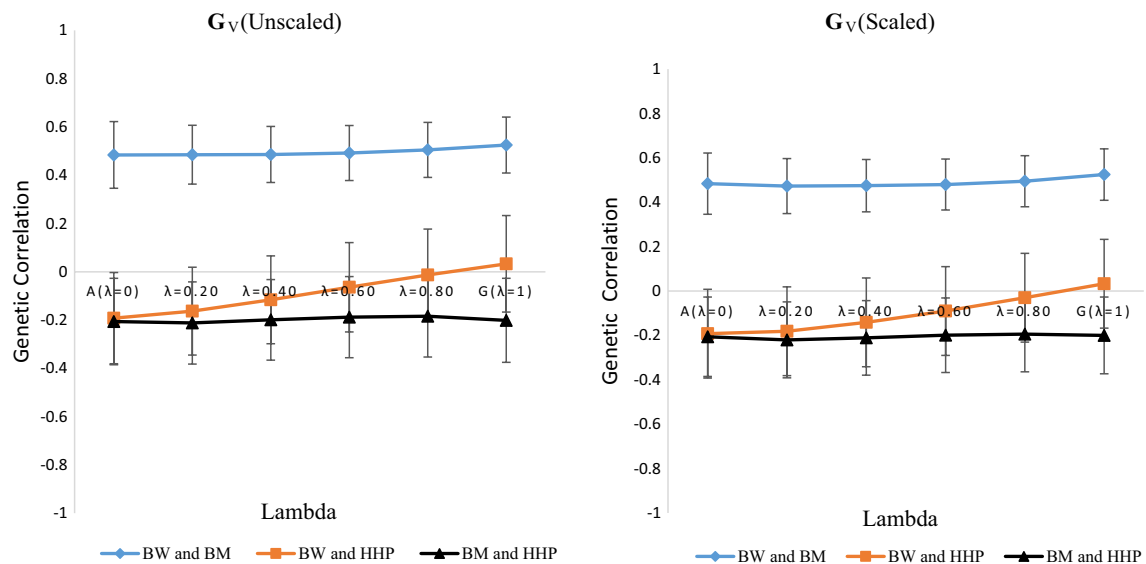
The differences that were observed in estimates of genetic correlations depended on the type of information used. From theory, standard pedigree-based linear models capture expected genetic covariation, whereas marker-based models capture genetic covariation that is

marked by SNPs. Our results are important from the perspective of multiple-trait genomic analysis because they indicate that estimates of genetic correlations between some traits may depend on the type of information used. This was clearly the case for the genetic correlation between BW and HHP.

Multiple-trait pedigree or marker-based prediction was designed to exploit genetic correlations between target characters and indicator traits [16], especially when a lowly heritable target trait is genetically correlated with an indicator that has a higher heritability. Our results indicate that estimates of genomic correlation between characters may reaffirm or disagree with expectations that are developed from a pedigree-based analysis. For example, on the one hand, the genetic and genomic correlations between BW and BM were insensitive to  $\lambda$  values, i.e., estimates of the genomic correlation and of the genetic correlation derived from the infinitesimal model were the



**Fig. 2** Average and standard errors estimates of genetic and genomic correlations across 20 replicates between body weight (BW), breast meat (BM) and hen-house production (HHP) as a function of the weight placed on Forni's genomic relationship matrix  $G_F(\lambda)$



**Fig. 3** Average and standard errors estimates of genetic and genomic correlations across 20 replicates between body weight (BW), breast meat (BM) and hen-house production (HHP) as a function of the weight placed on VanRaden's genomic relationship matrix  $G_V(\lambda)$

same. On the other hand, when considering BW and HHP, the estimate of the pedigree-based genetic correlation was equal to 0.2, whereas the estimate of the genomic correlation was close to 0. This illustrates a situation where part of the covariance between a pair of traits was not detected by SNPs ("missing correlation"). Sources of genetic correlation may be lost in a multiple-trait marker-based analysis. In the case of BM and HHP, the classical genetic correlation was estimated at  $-0.20$  and the genomic

correlation at  $-0.15$ . The pedigree-based analysis suggested a stronger genetic correlation.

Care should be exercised when interpreting and using genetic parameters that are assessed via molecular markers, as predictions for complex traits based on pedigree data may differ significantly from those based on SNP data. For this reason, we explored whether the two sources of information could be combined in some "optimal" manner.



### Predictive ability

The question of how to arrive at a “best” estimate of a genetic correlation (i.e., for which the greatest advantage of predicting ability is obtained) was examined and, to accomplish this objective, we used the predictive approach advocated by Lo et al. [17]. Figure 4 shows boxplots with the distributions of predictive correlations and mean squared errors for the cross-validation with 20 random repetitions. Some of the plots (e.g., BW) show a mild advantage of using a linear combination of **G** and **A** as kinship kernel. For BW, the largest correlation and lowest MSE were obtained with unscaled **G<sub>F</sub>** and **G<sub>V</sub>**. In terms of the predictive correlation for BW, the largest values were obtained with scaled **G<sub>V</sub>** and unscaled **G<sub>F</sub>**, both at  $\lambda = 0.8$ . For BM, the largest predictive correlation was achieved with unscaled **G<sub>V</sub>** and scaled **G<sub>F</sub>**, at  $\lambda = 0.4$  and  $0.8$ , respectively. For HHP, both scaled **G<sub>V</sub>** and **G<sub>F</sub>** resulted in better performance, and the largest predictive correlations were obtained at  $\lambda = 0.2$ .

The lowest MSE for BW was achieved for unscaled **G<sub>F</sub>** and scaled **G<sub>V</sub>** at  $\lambda = 0.6$ . For BM, the lowest MSE was obtained with  $\lambda$  close to 1 using scaled **G<sub>V</sub>** and **G<sub>F</sub>**. In addition, the scaled **G<sub>V</sub>** and **G<sub>F</sub>** produced the lowest MSE for HHP, with a slight superiority for values of  $\lambda$  close to 1. In terms of MSE, except for BW with **G<sub>F</sub>**, scaling of genomic relationship matrices yielded better results. Our findings are in agreement with Rodríguez-Ramilo et al. [31], who reported that when a larger weight was assigned to the numerator relationship matrix (**A**), the predictive correlation was lower than when assigning more weight to the genomic relationship matrix (**G**); a similar behavior was found for MSE. Rodríguez-Ramilo et al. [31] estimated  $\lambda$  by using Bayesian methods and reported that the posterior mean of  $\lambda$  depended on training sample size and the trait.

Our results indicate that multiple-trait genome-enabled predictions may be improved in some cases by combining **A** and **G** to quantify kinship. This result may also hold when prediction involves multiple selection lines or crossbred animals. Combining kernels can be viewed as a form of model averaging [25], with markers and pedigree playing complementary roles in prediction, e.g., markers may exploit similarity in state and LD, with **A** informing about similarity by descent.

Our results using dense SNPs (600 K Affymetrix platform) indicate that GBLUP with scaled or unscaled relationship matrices typically performed better than pedigree-based BLUP. However, in most cases, the largest correlation and lowest MSE were achieved using a linear combination of **A** and **G**.

### Regression coefficients

Figures 5, 6 and 7 show scatter plots and average (red dotted line) genetic regression coefficients of the three traits on the estimated direct genomic values (DGV) of other traits calculated from REML estimates of (co)variance components. The realized regression coefficients were computed at each  $\lambda$  value for the 20 cross-validation random samples and their medians are depicted as dark blue dotted lines on each plot. The REML regressions express the expected change in genetic value of trait  $i$  if the direct genomic value for trait  $j$  changes by one unit.

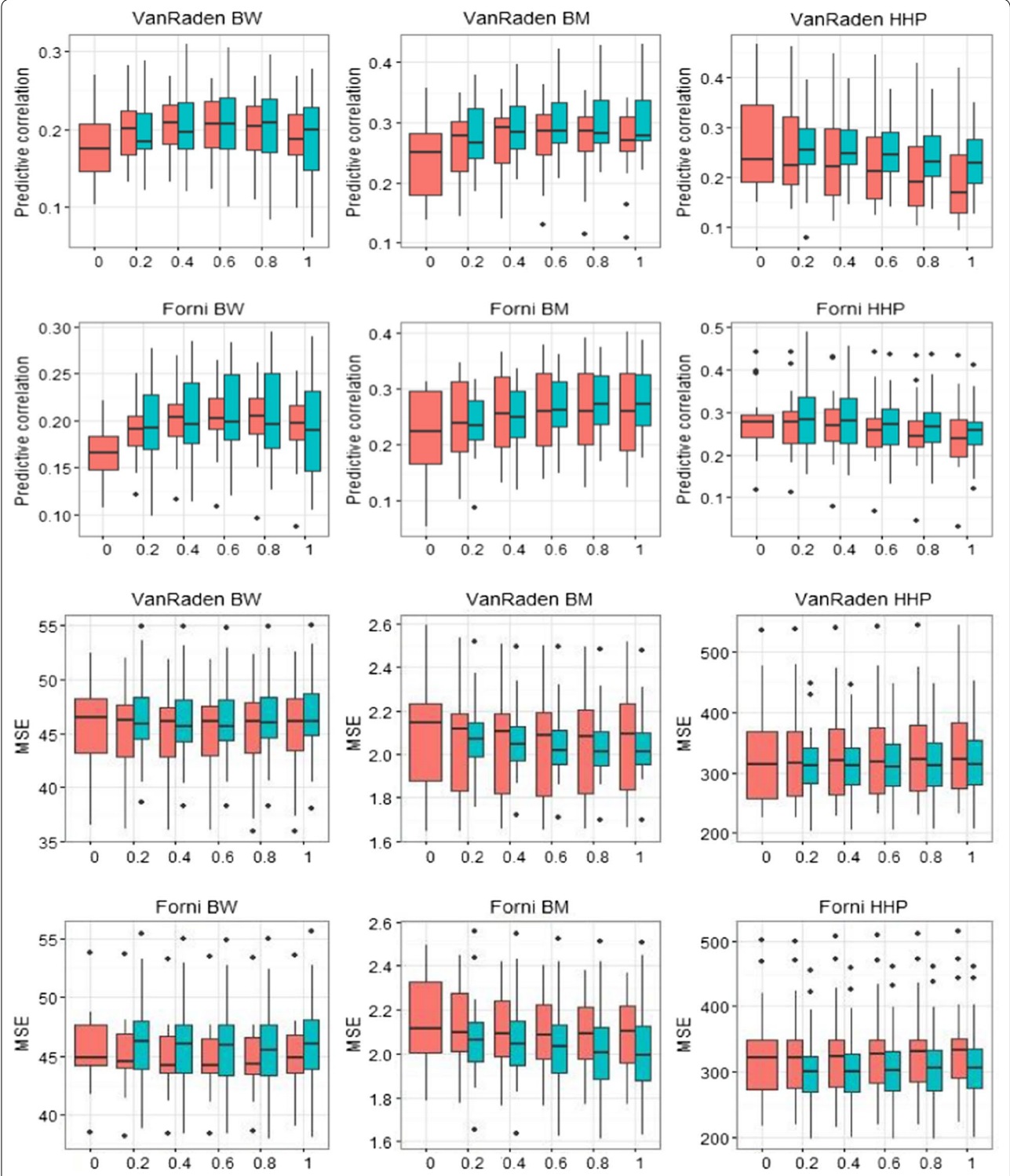
For BW and BM (Fig. 5), the expected and realized regressions were larger than 0 for all values of  $\lambda$ . In general, there was reasonable agreement between expected and realized regressions. However, for BW and HHP (Fig. 6), the expected genetic regressions were negative and moved toward 0 as  $\lambda$  increased, but the realized regressions (blue dotted lines) varied around 0 for all  $\lambda$  values. There was some apparent inconsistency between the expectations based on REML estimates and the cross-validation regression.

Figure 7 indicated a disagreement between expected genetic regressions and cross-validation regressions of BM on HHP when  $\lambda$  was close to 0. The expected regressions based on pedigree information were negative, while the cross-validation regressions were positive. The expected regressions of HHP phenotypes on DGV of BM and variances tended towards 0 as  $\lambda$  tended to 1, i.e., when more weight was placed on SNPs. The cross-validation regressions were much more affected by the value of  $\lambda$  than the expected regressions based on REML estimates.

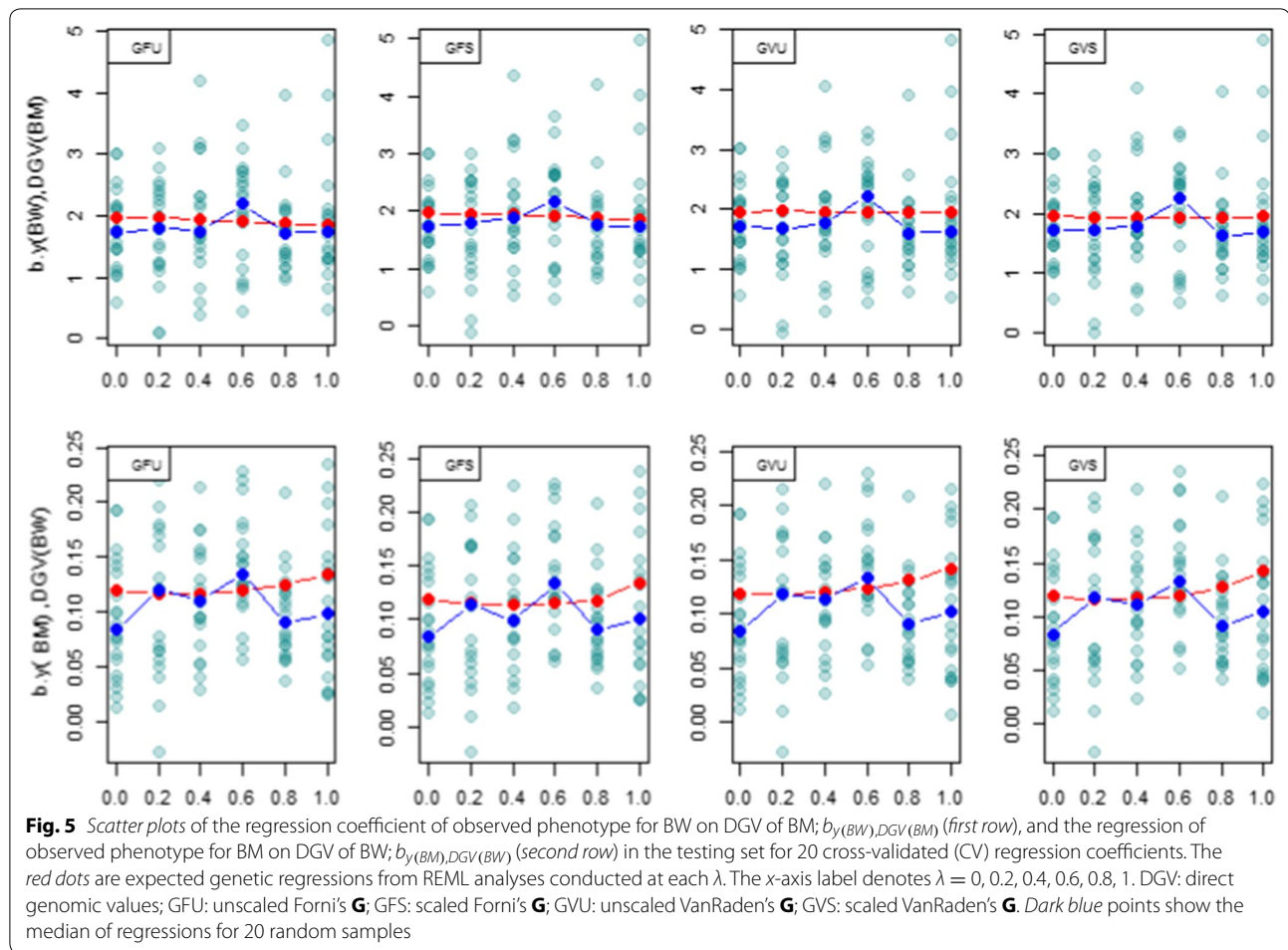
### Discussion

In genome-enabled prediction, there are different ways of incorporating molecular marker information into parametric and non-parametric models [24, 32]. Research with simulated and real data has consistently shown that single-trait GBLUP displays slightly better prediction accuracy when a trait is affected by a large number of QTL with small effects and as well as other genomic prediction methods for most traits [33, 34]. However, few studies on multiple-trait genomic prediction have been carried out with GBLUP, or have assessed estimates of genetic correlations when genomic or pedigree data were used. Similar to traditional pedigree-based genetic evaluations, the use of multiple-trait GBLUP is expected to increase the accuracy of predictions via “borrowing” of information from genetically correlated traits [35].

In order to explore a multiple-trait GBLUP model that also makes use of pedigree information, we constructed a pedigree-marker based kinship matrix (**K**) as a linear



**Fig. 4** Boxplot of predictive correlations across 20 replicates between phenotypes and predicted breeding values (upper two rows), and of mean squared errors (MSE) (bottom two rows) in testing sets. Red and light blue colors denote values for unscaled and scaled relationship matrices of Forni or VanRadon, respectively. Outliers are denoted as black dots, and the x-axis label denotes  $\lambda = 0, 0.2, 0.4, 0.6, 0.8, 1$



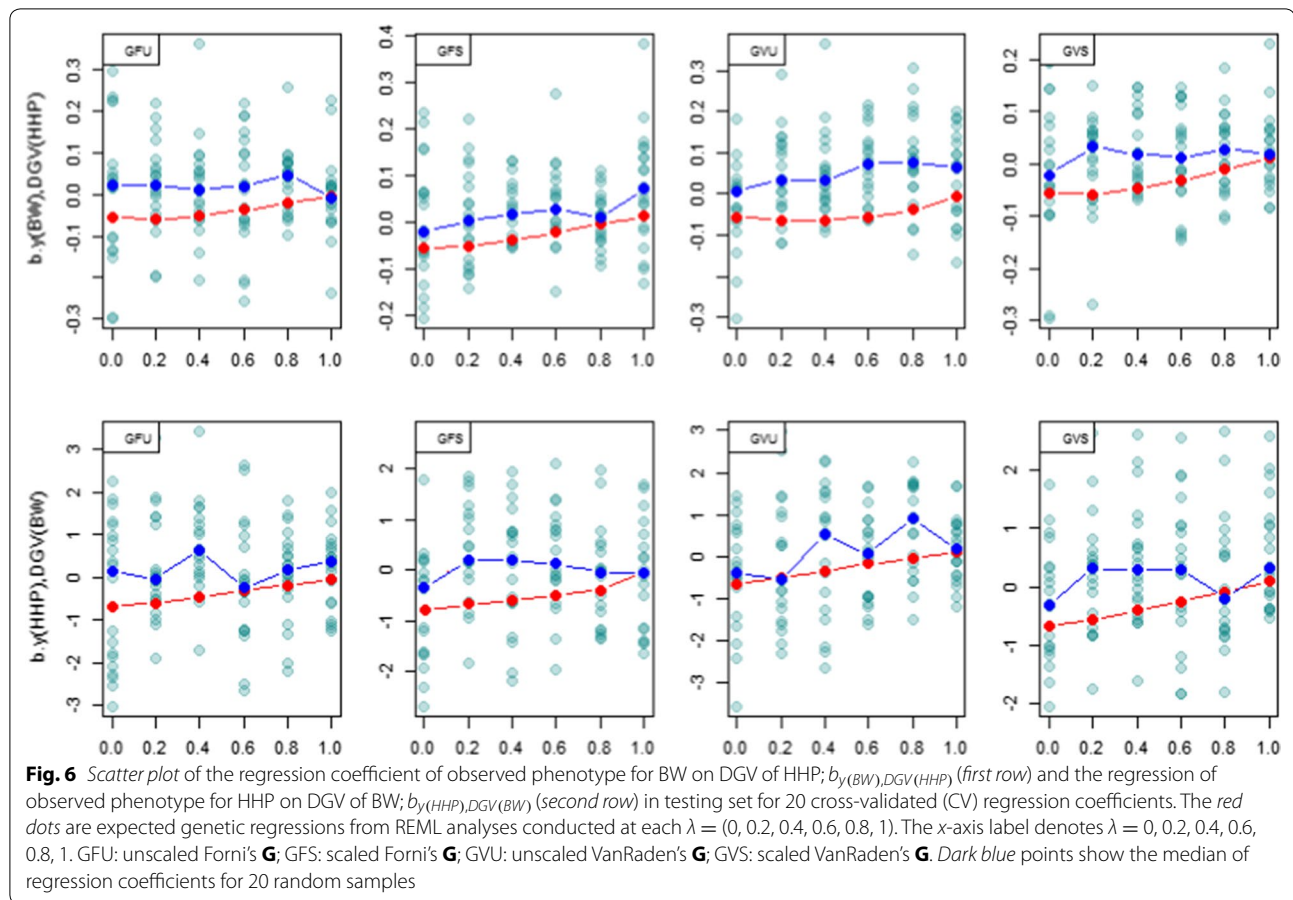
combination of pedigree and marker-based relationships between animals, defined as  $\mathbf{K} = \lambda \mathbf{G} + (1 - \lambda) \mathbf{A}$ . Predictive ability of the model and parameter estimates were obtained over a grid of values of  $\lambda$  that varied between 0 and 1, e.g.,  $\lambda = 0$  implied that all weight was assigned to pedigree, and none to SNPs.

One important factor to take into account when combining marker- and pedigree-based relationship matrices is that such matrices are on the same scale. The elements of the additive relationship matrix are the numerators of Wright's correlation coefficients that represent the relative genetic variances and covariances among individuals. Consequently, the diagonals of  $\mathbf{A}$  can be as large as 2, and relationships between two individuals can be greater than 1.

Traditionally, to quantify coefficients of relationship with respect to a base (reference) population, as discussed in [36, 37], the probability that alleles are identical by descent (IBD) was derived from pedigree information and from a base population consisting of founders. However, for relationships estimated from genetic markers

there is no obvious base population, and they estimate the proportion of the genome that is identical by state (IBS). In our data, genomic relationships measured by unscaled  $\mathbf{G}_V$  and  $\mathbf{G}_F$  can take negative values, whereas pedigree relationships are non-negative. In our data, no negative values were observed for full-sib genomic relationships but negative genomic additive relationships with small values near 0 were observed for unrelated individuals based on the pedigree (i.e., pedigree based relationship = 0). It remains to be seen whether genomic relationship measures can detect true 'negative genomic correlations' (if such correlations exist), which may be detectable using deep pedigree information and a definition of a base population. The genomic relationship matrices in our analyses were based on (IBS) information and on frequencies of alleles to build the GRM.

Our results suggest that multiple-trait genetic predictions depended on the weight assigned to genomic data. Better predictions were often obtained when pedigree and SNP information were used simultaneously. Earlier studies using simulated or real data have explored



the potential superiority of multiple-trait over single-trait genomic prediction with a focus on the relationship between traits in terms of differences in heritability, genetic correlations and number of indicator traits (e.g., [35, 38, 39]). De Los Campos et al. [40] indicated that potential problems may emerge when trying to infer genetic parameters using molecular markers that are imperfectly associated with genotypes at causal loci. Gianola et al. [13] showed that correlation parameters that are inferred from markers (i.e., genomic correlations) can give a distorted picture of the genetic correlation between traits. The sources of genetic correlation are pleiotropy (i.e., the same QTL affects more than one trait) and LD between QTL. When markers are used, marker-QTL LD and LD relationships among markers intervene in the genomic correlation.

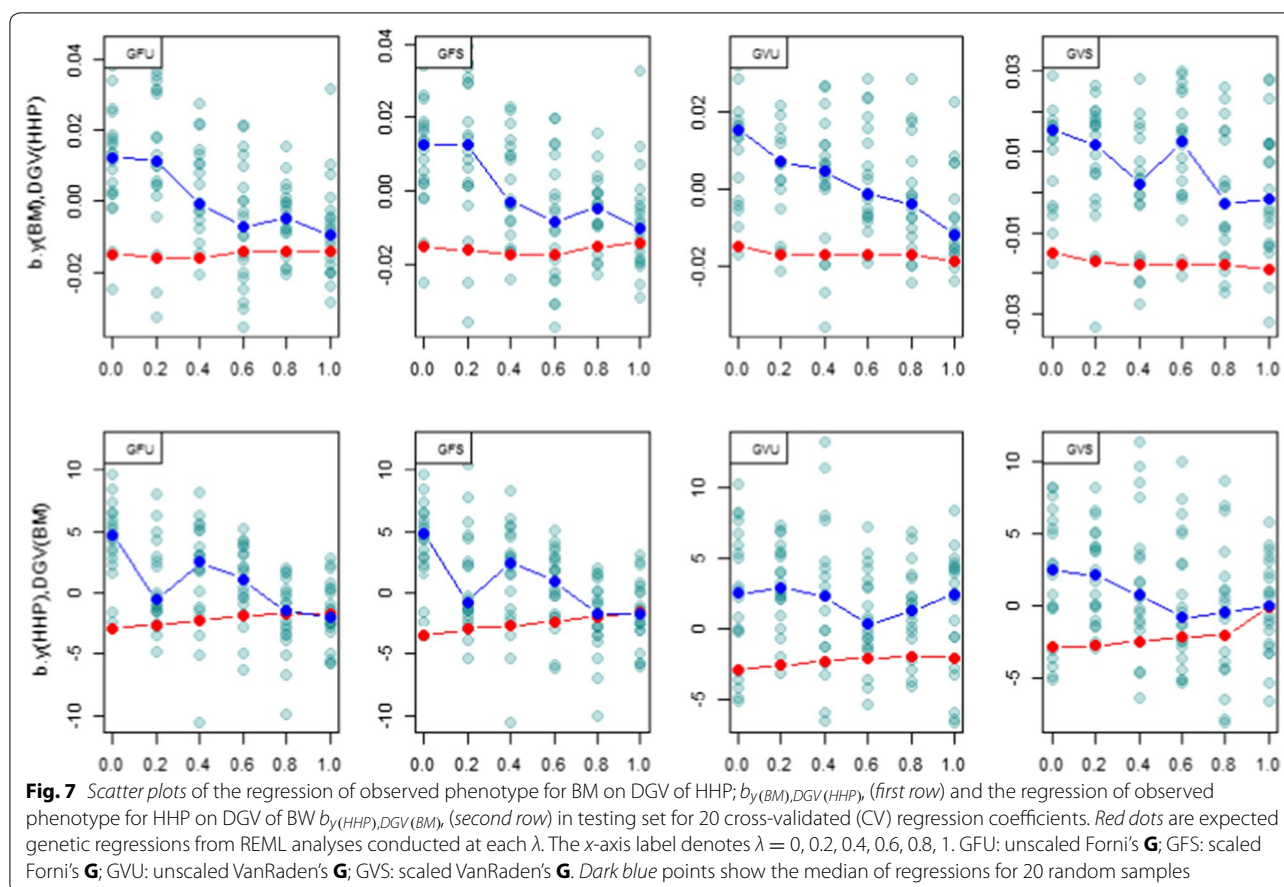
Here, we examined the impact of combining **A** and **G** on estimates of the genomic correlation between three chicken traits and evaluated outcomes using a predictive framework. Some studies [12, 41] have shown superiority of multiple-trait prediction over single-trait prediction, and combining pedigree with marker information was

found to be better than when using either **A** and **G** alone [32].

Our estimates of genetic correlations depended on the choice of  $\lambda$ . For example, on the one hand for BW and HHP, when using pedigree as the only measure of similarity ( $\lambda = 0$ ), the genetic correlation was  $-0.20$ , but it shifted to near 0 or was even positive ( $G_V$ ) when only marker information was used. On the other hand, the estimate of the genetic correlation between BW and BM was stable with respect to  $\lambda$ , while the estimate of the genetic correlation between BM and HHP was only slightly affected, i.e. changing from  $-0.21$  to  $-0.15$  for  $G_F$ , and from  $-0.22$  to  $-0.20$  for  $G_V$  for  $\lambda = 0$  and  $\lambda = 1$ , respectively. Clearly, genomic data provide a distinct measure of similarity between individuals, and this translates into differential capturing of genetic signals. For instance, most off-diagonal entries of **A** were zero but all entries of **G** were non-null.

In order to increase the accuracy of predictions by using pedigree and genomic information jointly, Legarra et al. [42] proposed a single-step procedure that enhances relationship information for non-genotyped animals,





without requiring major changes in the implementation of a standard BLUP model. In the study of Aguilar et al. [43], a three-fold increase in accuracy of GEBV was found for traits related to conception rate in Holstein dairy cows, with low heritability, when using a genomic-based relationship combined with a pedigree-based relationship in a multiple trait model [43]. Using 18 quantitative traits in Holstein dairy cattle, Tsuruta et al. [44] reported that prediction accuracies increased when a multiple-trait genomic prediction model was used compared to a single-trait model, but the increase depended on the trait being predicted. However, Bao et al. [45] did not observe clear benefits when four traits were included in a multiple-trait genomic prediction model for soybeans compared to a single-trait model. However, these studies did not examine the impact of using combined genetic and genomic relationships. This shows that the effect of genetic correlations on multi-trait genomic prediction depends on the information type being used to construct **K**, with an impact on accuracy of prediction.

In our study, predictive ability was measured using the correlation between predicted genetic values and observed phenotypic values, and mean squared error

of these predictions. As shown in Fig. 4, the optimum weight placed on genomic relationships was trait-dependent. We took the view point that a “best” estimate of the genetic correlation would correspond to the linear combination of **A** and **G** (with a specific weight on each one) that delivered the best predictive ability, which was found by searching the weight ( $\lambda$ ) placed on genomic versus pedigree relationships. This type of predictive approach has been advocated in the statistical literature [17, 46, 47].

In general, combinations of **A** and **G** kernels yielded better predictions than when only **G** was used. In a GBLUP model, the entries of **G** reflect the actual extent of IBS relationships between individuals, but without making a clear reference to a base population [48]. This implies that genomic (co)variance parameters do not necessarily have the same meaning as standard classical multiple-trait models genetic parameters, such as the infinitesimal genetic correlation. According to [49], pedigree information, co-segregation and population-wide LD are three sources of genetic information that contribute to the predictive ability of genomic selection models. Co-segregation information can be captured by

IBD or IBS relationships and, when a pedigree is not deep enough, relatedness among individuals that is inferred from markers may improve prediction. However, how does one decide if an estimate of genetic correlation derived from genomic data is better than a pedigree-based estimate?

In classical quantitative genetics, a genetic correlation between traits arises due to either genes that have an effect on both traits (pleiotropy), or due to LD between genes that affect different traits [50]. When investigating the basis of a genetic correlation, an important question is to determine the extent to which these two forces act on the genetic parameters [13, 51]. Multiple-trait QTL mapping methods may help distinguishing pleiotropy from linkage [52], but any such dissection in the absence of knowledge on QTL would be speculative.

Estimates of genetic correlations obtained from pedigree or from markers may differ either due to chance or other reasons, such as extent of LD between markers and the unknown QTL. One possible way of testing if such differences are systematic, is to examine pairs of estimates of pedigree- and marker-based correlations in re-samples from the dataset and constructing a paired comparison, by using either a parametric or a non-parametric approach. For example, the estimates of correlations could be z-transformed and a paired t test conducted.

In summary, combining pedigree- and marker-based information had an impact on predictive performance of multiple-trait models. Discerning the optimum weight placed on genomic and genealogical information is an important issue, and a grid-search scheme was used for that purpose. We found that estimates of genetic correlation obtained with **A** and **G** matrices were different, but depended on the trait. This indicates that multiple-trait marker-based prediction may be enhanced by the combined use of genealogy and marker information in the models.

## Conclusions

To our knowledge, this is the first study with animal breeding data that explores how the weight placed on pedigree and marker information affects multiple-trait predictions. We designed a tri-variate genomic prediction model that exploited pedigree and marker information simultaneously. Use of a kinship matrix that is formed as a linear combination of pedigree- and marker-based relationships may enhance genome-enabled prediction, but the optimal weight placed on the two sources of information will differ between traits. Genetic correlation estimates from pedigree-based models may differ from those obtained from marker-based models, at least in some cases. Cross-validation was useful for gauging the genetic correlation in multiple-trait models.

## Authors' contributions

MM carried out the study and wrote the first draft of the manuscript. DG and GJMR designed the experiment, supervised the study and critically contributed to the final version of manuscript. BDV and AAM participated in discussion and reviewed the manuscript. AS, AE and MAF contributed materials and revised the manuscript. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman (SBUK), Kerman, Iran. <sup>2</sup> Department of Statistical, Faculty of Mathematic and Computer Science, Shahid Bahonar University of Kerman (SBUK), Kerman, Iran. <sup>3</sup> State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. <sup>4</sup> Roslin Institute, University of Edinburgh, Midlothian, UK. <sup>5</sup> Department of Animal Sciences, University of Wisconsin, Madison, WI, USA. <sup>6</sup> Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA. <sup>7</sup> Department of Dairy Science, University of Wisconsin, Madison, WI, USA.

## Acknowledgements

The first author wishes to acknowledge Aviagen (Midlothian, United Kingdom) for providing the data, and the Ministry of Science, Research and Technology of Iran for financially supporting his visit to the University of Wisconsin-Madison. Work was partially supported by the Wisconsin Agriculture Experiment Station under hatch Grant 142-PRJ63CV to DG.

## Competing interests

The authors declare that they have no competing interests.

Received: 16 July 2016 Accepted: 16 January 2017

Published online: 01 February 2017

## References

1. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819–29.
2. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci*. 2009;92:433–43.
3. de Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 2013;193:327–45.
4. Gianola D, Fernando RL, Stella A. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*. 2006;173:1761–76.
5. Habier D, Fernando RL, Dekkers JC. The impact of genetic relationship information on genome-assisted breeding values. *Genetics*. 2007;177:2389–97.
6. González-Recio O, Rosa GJM, Gianola D. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest Sci*. 2014;166:217–31.
7. Henderson CR, Quaas RL. Multiple trait evaluation using relatives records. *J Anim Sci*. 1976;43:1188–97.
8. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics*. 2010;9:166–77.
9. Kadarmideen HN. Genomics to systems biology in animal and veterinary sciences: progress, lessons and opportunities. *Livest Sci*. 2014;166:232–48.
10. Gianola D, Rosa GJM. One hundred years of statistical developments in animal breeding. *Annu Rev Anim Biosci*. 2015;3:19–56.
11. Thompson R, Meyer K. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest Prod Sci*. 1986;15:299–313.
12. Jia Y, Jannink JL. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*. 2012;192:1513–22.
13. Gianola D, de los Campos G, Toro MA, Naya H, Schön CC, Sorensen D. Do molecular markers inform about pleiotropy? *Genetics*. 2015;201:23–9.
14. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet*. 2012;44:1066–71.
15. Maier R, Moser G, Chen GB, Ripke S, Coryell W, Potash JB, et al. Joint analysis of psychiatric disorders increases accuracy of risk prediction for



- schizophrenia, bipolar disorder, and major depressive disorder. *Am J Hum Genet.* 2015;96:283–94.
16. Calus MP, Veerkamp RF. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol.* 2011;43:26.
  17. Lo A, Chernoff H, Zheng T, Lo SH. Why significant variables aren't automatically good predictors. *Proc Natl Acad Sci USA.* 2015;112:13892–7.
  18. Abdollahi-Arpanahi R, Morota G, Valente BD, Kranis A, Rosa GJM, Gianola D. Differential contribution of genomic regions to marked genetic variation and prediction of quantitative traits in broiler chickens. *Genet Sel Evol.* 2016;48:10.
  19. Morota G, Abdollahi-Arpanahi R, Kranis A, Gianola D. Genome-enabled prediction of quantitative traits in chickens using genomic annotation. *BMC Genomics.* 2014;15:109.
  20. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* 2007;81:1084–97.
  21. Forni S, Aguilar I, Misztal I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol.* 2011;43:1.
  22. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91:4414–23.
  23. Astle W, Balding DJ. Population structure and cryptic relatedness in genetic association studies. *Stat Sci.* 2009;24(4):451–71.
  24. Morota G, Gianola D. Kernel-based whole-genome prediction of complex traits: a review. *Front Genet.* 2014;5:363.
  25. De Los Campos G, Gianola D, Rosa GJM, Weigel KA, Crossa J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet Res (Camb).* 2010;92:295–308.
  26. Legarra A, Christensen OF, Vitezica ZG, Aguilar I, Misztal I. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. *Genetics.* 2015;200:455–68.
  27. Gianola D, Okut H, Weigel KA, Rosa GJ. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genet.* 2011;12:87.
  28. Meyer K. WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J Zhejiang Univ Sci B.* 2007;8:815–21.
  29. Kapell DN, Hill WG, Neeteson AM, McAdam J, Koerhuis AN, Avendaño S. Genetic parameters of foot-pad dermatitis and body weight in purebred broiler lines in 2 contrasting environments. *Poult Sci.* 2012;91:565–74.
  30. Kapell DN, Hill WG, Neeteson AM, McAdam J, Koerhuis AN, Avendaño S. Twenty-five years of selection for improved leg health in purebred broiler lines and underlying genetic parameters. *Poult Sci.* 2012;91:3032–43.
  31. Rodríguez-Ramilo ST, García-Cortés LA, González-Recio Ó. Combining genomic and genealogical information in a reproducing kernel Hilbert spaces regression model for genome-enabled predictions in dairy cattle. *PLoS One.* 2014;9:e93424.
  32. Crossa J, de Los Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics.* 2010;186:713–24.
  33. Wientjes YC, Veerkamp RF, Calus MP. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics.* 2013;193:621–31.
  34. Wimmer V, Lehermeier C, Albrecht T, Auinger HJ, Wang Y, Schön CC. Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics.* 2013;195:573–87.
  35. Guo G, Zhao F, Wang Y, Zhang Y, Du L, Su G. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 2014;15:30.
  36. Powell JE, Visscher PM, Goddard ME. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet.* 2010;11:800–5.
  37. Forneris NS, Steibel JP, Legarra A, Vitezica ZG, Bates RO, Ernst CW, et al. A comparison of methods to estimate genomic relationships using pedigree and markers in livestock populations. *J Anim Breed Genet.* 2016;133:452–62.
  38. Hayashi T, Iwata H. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics.* 2013;14:34.
  39. Schultess AW, Wang Y, Miedaner T, Wilde P, Reif JC, Zhao Y. Multiple-trait and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theor Appl Genet.* 2015;129:273–87.
  40. de Los Campos G, Sorensen D, Gianola D. Genomic heritability: what is it? *PLoS Genet.* 2015;11:e1005048.
  41. Clark SA, Hickey JM, Daetwyler HD, van der Werf JH. 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.* 2012;44:4.
  42. Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci.* 2009;92:4656–63.
  43. Aguilar I, Misztal I, Tsuruta S, Wiggins G, Lawlor T. Multiple trait genomic evaluation of conception rate in Holsteins. *J Dairy Sci.* 2011;94:2621–4.
  44. Tsuruta S, Misztal I, Aguilar I, Lawlor T. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *J Dairy Sci.* 2011;94:4198–204.
  45. Bao Y, Kurler JE, Anderson G, Young ND. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. *Mol Breed.* 2015;35:128.
  46. Shmueli G. To explain or to predict? *Stat Sci.* 2010;25(3):289–310.
  47. Geisser S. Predictive inference: an introduction. New York: Chapman & Hall; 1993.
  48. Román-Ponce SI, Samoré AB, Dolezal MA, Bagnato A, Meuwissen TH. Estimates of missing heritability for complex traits in Brown Swiss cattle. *Genet Sel Evol.* 2014;46:36.
  49. Habier D, Fernando RL, Garrick DJ. Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics.* 2013;194:597–607.
  50. Falconer D, Mackay T. Introduction to quantitative genetics. Harlow: Longman Group Ltd.; 1995.
  51. Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet.* 2012;8:e1002637.
  52. Stich B, Piepho HP, Schulz B, Melchinger A. Multi-trait association mapping in sugar beet (*Beta vulgaris* L.). *Theor Appl Genet.* 2008;117:947–54.

Submit your next manuscript to BioMed Central  
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

