

# JOURNAL OF ANIMAL SCIENCE

*The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science*

## **New method to combine molecular and pedigree relationships**

E. Bömcke, H. Soyeurt, M. Szydlowski and N. Gengler

*J Anim Sci* 2011.89:972-978.

doi: 10.2527/jas.2010-3135 originally published online Nov 19, 2010;

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://jas.fass.org/cgi/content/full/89/4/972>



**American Society of Animal Science**

[www.asas.org](http://www.asas.org)

# New method to combine molecular and pedigree relationships<sup>1</sup>

E. Bömcke,<sup>\*†2</sup> H. Soyeurt,<sup>\*‡</sup> M. Szydlowski,<sup>\*§</sup> and N. Gengler<sup>\*‡</sup>

<sup>\*</sup>University of Liege, Gembloux Agro-Bio Tech, Animal Science Unit, 5030 Gembloux, Belgium;

<sup>†</sup>Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA), 1000 Brussels, Belgium;

<sup>‡</sup>National Fund for Scientific Research (FNRS), 1000 Brussels, Belgium;

and <sup>§</sup>Poznan University of Life Sciences, 60-637 Poznan, Poland

**ABSTRACT:** Relationship coefficients are traditionally based on pedigree data. Today, with the development of molecular techniques, they are often completely replaced by coefficients calculated from molecular data. Examples are relationships from microsatellites for biodiversity studies but also genomic relationships from SNP as currently used in genomic prediction of breeding values. There are, however, many situations in which optimal combination of both sources would be the best solutions. Obviously, this is the case for incompletely genotyped populations, but also when pedigree information is sparse. Also, markers, even dense ones, do not reflect the whole genome and therefore give only an incomplete picture of relationships. The main objective of this study was therefore to develop a method

to calculate a relationship matrix by the combination of molecular and pedigree data. It will be useful for all situations where pedigree and molecular data are available. In this study, based on simulations of pedigree and marker data, we used partial least squares regression and linear regression to combine total allelic relationship coefficients calculated for each marker with additive relationship coefficients calculated from incomplete pedigree. The results showed that the greatest advantage of this method, compared with the one that replaces a part of the pedigree-based relationship matrix by a genomic relationship matrix, is that adding the partial pedigree data allows for the correction of the molecular coefficient for the ungenotyped part of the genome.

**Key words:** molecular information, pedigree, relationship matrix

©2011 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2011. 89:972–978  
doi:10.2527/jas.2010-3135

## INTRODUCTION

Relationship coefficients correspond to genetic covariances between related individuals expressed relatively, independent from the considered traits. Many authors contributed in a significant way to the theoretical development of these coefficients that led to the existence of various definitions and denominations (Wright, 1922; Malécot, 1948). Knowledge of relationships among animals is useful for the study of wild populations and the genetic management of captive or threatened popula-

tions or both (Glaubitz et al., 2003). It is therefore probably one of the principal tools used to optimize conservation strategies (Caballero and Toro, 2002; Verrier et al., 2005).

Relationship coefficients are traditionally based on pedigree data. But any pedigree is somewhat incomplete for various reasons. The most important are, first, that cut-off dates for recording the pedigree might exist (e.g., 1950) and, second, that animals of unknown origins enter the herdbooks (e.g., animals coming from another breed or country). Often this leads to the loss of relationships among individuals even if they have common ancestors because they are considered not related and their descendants not inbred (VanRaden, 1992). With molecular data being available, pedigree-based relationship coefficients are often completely replaced by coefficients calculated from molecular data. Examples are relationships from microsatellites for biodiversity studies (Caballero and Toro, 2002; Oliehoek et al., 2006), and genomic relationships from SNP as currently used in genomic prediction of breeding values (Zhang et al., 2007; VanRaden, 2008). However, the limit is that genotyping an entire population is gener-

<sup>1</sup>The authors gratefully acknowledge the financial support of the Ministry of Agriculture of the Walloon Region of Belgium [Service Public Wallon, Direction générale opérationnelle "Agriculture, Ressources naturelles et Environnement" (DGO3), Direction du Développement et de la Vulgarisation, Namur, Belgium; Project D31-1168]. Nicolas Gengler, a research associate of the National Fund for Scientific Research (Brussels, Belgium), acknowledges their support. Additional support was provided through grants 2.4507.02F (2) and F.4552.05 of the National Fund for Scientific Research.

<sup>2</sup>Corresponding author: elisabeth.bomcke@ulg.ac.be  
Received May 5, 2010.

Accepted November 11, 2010.

ally impossible due to its high cost or for logistic restrictions (i.e., culled, slaughtered, or foreign animals).

Consequently, there are many situations in which optimal combination of pedigree and molecular data in a single relationship matrix would be the best solution. This is the case for incompletely genotyped populations, but also when pedigree information is sparse. An example of this situation is the Skyros pony, a Greek indigenous horse breed that was used hereafter as a reference population for the simulations. The objective of this study was therefore to develop a new method to estimate relationships by combining molecular data with pedigree data. The following paper will put the method in a biodiversity setting, but it could be extended to other field of genetics, with further developments, such as genomic selection.

## MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because the research was done on simulations.

### *Pedigree and Molecular Relationships*

Definitions of pedigree relationships are numerous (Wright, 1922; Malécot, 1948; Henderson, 1976; Van Vleck et al., 1987; Minvielle, 1990). Combinations of molecular relationships with pedigree relationships can only be done using compatible coefficients. The additive relationship coefficient (Henderson, 1976) between 2 animals  $x$  and  $y$ , referenced hereafter as  $a_{xy}$ , is directly related to the concept of identity by descent (**IBD**; Minvielle, 1990). Among the numerous molecular relationships coefficients, the total allelic relationship coefficient (Nejati-Javaremi et al., 1997), referenced as  $ta_{xy}$  between 2 animals  $x$  and  $y$ , seems to be the closest match. Indeed,  $a_{xy}$  corresponds to  $2 \times$  Malecot's relationship coefficient (Malécot, 1948), which is the probability for any locus that 1 allele drawn at random among the 2 carried by animal  $x$  and that 1 allele drawn identically in animal  $y$  are IBD. The  $ta_{xy}$  coefficient, on the other hand, corresponds to  $2 \times$  molecular coancestry, which is the probability that for 2 alleles taken at random, 1 from each individual are identical. This molecular coefficient has the advantage that it applies the Malecot's definition to the markers (Toro et al., 2002). However, it makes no distinction between IBD and identity-by-state (**IBS**) alleles (Fernández and Toro, 2006). The value of molecular coancestry ( $fM$ ) for every pair of individuals  $x$  and  $y$  for locus  $l$  is computed as

$$fM_{xy,l} = 1/4[S_{ac} + S_{ad} + S_{bc} + S_{bd}],$$

where the subscript  $l$  indicates the locus, the subscripts  $a$ ,  $b$ ,  $c$ , and  $d$  indicate allelic position 1 of  $l$  of individual  $x$ , allelic position 2 of  $l$  of  $x$ , allelic position 1 of  $l$  of in-

dividual  $y$ , and allelic position 2 of  $l$  of  $y$ , respectively, and  $S..$  refers to values depending on whether alleles at the allelic positions in the subscript are the same ( $S.. = 1$ ) or not ( $S.. = 0$ ). The computations of the total allelic relationship ( $ta_{xy}$ ) of the 2 alleles of an individual ( $x$ ) with the 2 alleles of the other individual ( $y$ ) are done for each locus ( $l$ ) as

$$ta_{xy,l} = 2 \times fM_{xy,l}$$

where  $fM_{xy,l}$  is the molecular coancestry between individuals  $x$  and  $y$  for locus  $l$  (Caballero and Toro, 2000; Eding and Meuwissen, 2001). Relationship matrices based on total allelic relationships are called **TA** hereafter.

### *New Method to Combine Molecular and Pedigree Relationships*

Marker and pedigree information are already used simultaneously and independently in small populations to minimize inbreeding and genetic drift (Toro et al., 1999; Wang, 2001). In QTL analysis, marker information is combined with pedigree information to modify the genealogical coancestries that allow obtaining coancestries conditional to marker information (Fernando and Grossman, 1989). However, the method is computationally very expensive. Recently, in field of genomic selection, some authors (e.g., VanRaden, 2008) combined genomic SNP-based relationship coefficients with pedigree-based coefficients using arbitrarily chosen weighting ( $w$ ). VanRaden (2008) presented a regression-based approach, where the inverted formula was used to regress genomic coefficients toward pedigree-based coefficients. The rationale behind this was that pedigree-based relationships would represent the correct expected value. However, genomic coefficients also do not distinguish IBS and IBD well, and the same is true for total allelic relationships used in this study. Therefore, by regressing molecular relationships on pedigree relationships, the starting point for considering IBD from IBS is based on the value of the constant term. VanRaden (2008) used some heuristics to obtain these regression coefficients. An alternative regression method, partial least square regression (**PLSR**), which corresponds better to these objectives, will be used in this study. The PLSR method can be described as a method in which both the independent and dependent variables are projected toward a common space to explain the maximum variance of the dependent variable (SAS Inst. Inc., Cary, NC). A second modification of the regression approach of VanRaden (2008) is the creation of a reference population to establish genealogical relationships, reflecting expected relationships in a complete pedigree and used to derive combining equations.

In a preliminary step, reference relationships are computed with a tabular method (Emik and Terrill,

1949) for animals having the greatest completeness of pedigree. In this study, these relationships were calculated on simulated pedigree, considered to be complete (100 yr of simulation, 15 known generations for living animals). These coefficients correspond to the theoretical expected values of relationships; the exact values are expected to be distributed around these values.

In a first step, prediction equations for reference relationships were obtained by PLSR with single locus **TA** matrix coefficients as independent variables. The resulting equation combining the number of genotyped loci ( $nl$ ) can be summarized as

$$a_{xy,CPED} = b_0 + \sum_{l=1}^{nl} (b_l ta_{xy,l}) + e_{xy}, \quad [1]$$

where  $a_{xy,CPED}$  is the additive relationship coefficient between individuals  $x$  and  $y$  calculated on the complete pedigree (CPED) of the reference population,  $ta_{xy,l}$  is the total allelic relationship coefficient between individuals  $x$  and  $y$  for locus  $l$ ,  $b_0$  and  $b_l$  are respectively the intercept and the regression coefficients for the  $l$ th locus, and  $e_{xy}$  is the corresponding residual. Use of this PLSR procedure in a multiple-marker situation has the great advantage of giving a different weight to each marker. It takes into account that each marker has a different degree of informativeness. As explained before, this equation can be considered as a generalization of one of the methods described by VanRaden (2008), which is based on a single regression. Furthermore, in this method, the nature of the single locus relationship coefficients can be different from a total allelic relationship as long as they are compatible.

As the markers describe only the regions of the genome, which are in linkage disequilibrium with the marker, the residuals were considered to capture the part of the genome that is not explained by the markers. Therefore, in addition to VanRaden (2008), we explained, in a second step, these residuals further by using least squares to regress additive relationship coefficients obtained from the incomplete pedigree, created by random deletion of known parents, on adjusted relationships  $a_{xy,CPED}^*$  using Eq. [2].

$$a_{xy,CPED}^* = a_{xy,CPED} - \sum_{l=1}^{nl} (b_l ta_{xy,l}) = b_0^* + b_P a_{xy,IP} + e_{xy}^*, \quad [2]$$

where  $a_{xy,IP}$  is the additive relationship coefficient between individuals  $x$  and  $y$  calculated on the incomplete pedigree (**IP**), and  $b_0^*$  and  $b_P$  are the additional linear regression coefficients.

Based on the combination of Eq. [1] and Eq. [2], the prediction of combined relationship coefficients between animals  $x$  and  $y$  was obtained, in a third step, through formula Eq. [3]:

$$\hat{a}_{xy,combined} = \hat{b}_0 + \sum_{l=1}^{nl} (\hat{b}_l ta_{xy,l}) + \hat{b}_0^* + \hat{b}_P a_{xy,IP}. \quad [3]$$

The matrix, which contains these coefficients, is called  $\mathbf{A}_{combined}$ . The hypothesis behind this is that pedigree, even if incomplete, contains information on the proportion of the genome that is shared by  $x$  and  $y$  for the ungenotyped part of the genome.

To test this last hypothesis, Eq. [3], hereafter called the full prediction equation (**FPE**), was compared with a reduced prediction equation (**RPE**; i.e., without the terms that include the pedigree ( $\hat{b}_0^* + \hat{b}_P a_{xy,IP}$ )). For this comparison, the obtained regressions coefficients and intercept calculated on simulated pedigree and genotypes are applied to other data (i.e., to other pedigree and genotypes having similar characteristics). Correlation coefficients between predicted values and expected values obtained from CPED, considered as the true theoretical value of the relationship among individuals, were computed. The residuals obtained with the 2 equations were also compared.

### Integration of Ungenotyped Animals

The procedure as described will only combine relationship coefficients of genotyped animals. However, the relationships between ungenotyped and genotyped animals as well as among related ungenotyped animals could be affected by the modification of relationships among genotyped animals. Modification of the complete **A** matrix, calculated on incomplete pedigree, is therefore required. The resulting relationship matrix will be called the modified relationship matrix and written  $\mathbf{A}_{modified}$ . Based on a suggestion by Gengler et al. (2007) that gene content could be modeled using the usual mixed model methodology, the propagation of relationships can be derived. As shown recently by several authors (e.g., Legarra et al., 2009; Christensen and Lund, 2010), by writing conditional distributions and predicting ungenotyped from genotyped animals, the following equation providing the inverted complete  $\mathbf{A}_{modified}$  matrix can be derived, animals being considered sorted (ungenotyped before genotyped):

$$\left(\mathbf{A}_{modified}\right)^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{A}_{combined}\right)^{-1} - \left(\mathbf{A}_{genotyped}\right)^{-1} \end{bmatrix}, \quad [4]$$

where the elements of  $\left(\mathbf{A}_{combined}\right)^{-1}$  are obtained through inversion of the combine matrix obtained using the strategy explained before, and  $\mathbf{A}_{genotyped}$  is the pedigree-based relationship matrix among genotyped animals. This modified matrix can then be directly used in genetic evaluation programs or be inverted back to  $\mathbf{A}_{modified}$  and used in programs for the management of small

breeds (e.g., GENCONT; Meuwissen, 2002). Equivalent equations on a noninverted scale were established by Legarra et al. (2009), and their equivalence to Eq. [4] was shown by Aguilar et al. (2010) and Christensen and Lund (2010).

To test this new method to combine relationships, the relative errors between true relationship values, calculated on the simulated CPED, and estimations obtained in  $\mathbf{A}_{IP}$  or in  $\mathbf{A}_{modified}$  were compared. Calculation of relative error was based on the Frobenius norm, as described by Misztal et al. (1995). This method has the advantage to be convenient for comparison of covariance matrices with similar diagonal elements. The biases of both estimations as well as the probability density function of the residuals were also calculated.

### *Simulations of Complete and Incomplete Pedigree and of Genotypes*

This study was realized on simulated pedigree and marker data, 5 pedigrees were simulated and considered as complete (CPED1 to CPED5). Population parameters of the Skyros pony breed (e.g., age of sire/dam at first calving, number of descendants per sire/dam, maximum death age), were used as input for the simulation programs. We selected these CPED to have around 200 living animals (and around 900 animals in total), as the Skyros pony breed, and similar number of founders genome (between 15 and 20) than estimated for this breed. The genotypes of the living animals were considered to be known. Level of inbreeding was consequently high in the 5 pedigrees, as is the case of small populations or population with bottlenecks (between 13.57 and 33.08%); this situation gives also the opportunities to test the influence of inbreeding on the effectiveness of this new method. For these pedigrees, we have an above average knowledge of the pedigree of living animals [i.e., over 15 generation-equivalents (**GEQ**) for the 5 simulations]. For each CPED, 10 IP were also obtained by putting randomly a fraction of the parents to unknown. In 5 cases (IP A1 to A5), 25% of paternal records and 25% of maternal records were randomly deleted (with 5% in common). In the 5 other cases (IP B1 to B5), 40% of paternal records and 30% of maternal records were randomly deleted (with 20% in common) to be consistent with what we observed for the Skyros pony. This allows obtaining an average pedigree knowledge for the living animals of about 3.06 and 1.84 GEQ, respectively. Based on this simulated CPED and IP records,  $a_{xy,CPED}$  and  $a_{xy,IP}$  were calculated using the tabular method (Emik and Terrill, 1949).

Genotypes were distributed across the simulation of CPED by gene dropping. For each founder, 25 markers were simulated. The allelic frequencies were considered equal to 1 divided by the number of alleles in the founder population. The number of alleles was randomly selected between 3 and 15. For each CPED, 10 repetitions of marker transmission were realized (data1 to data10). Genotypes were considered as known only for the living

animals. Based on this simulation, 25 TA matrices were calculated, one for each simulated locus. Consequently, the intercept and 25, for RPE, and 26, for FPE, regression coefficients were obtained using CPED1, data1 of CPED1 and, only for FPE, IP A1 of CPED1. The obtained coefficients were tested, first, on the other data and incomplete pedigree of CPED1. Second, those coefficients were tested on the 4 other pedigrees (2 with lower inbreeding level and 2 with greater inbreeding level) and the corresponding incomplete pedigrees and data.

## RESULTS AND DISCUSSION

### *Combination of Molecular and Pedigree Relationships*

The regression coefficients  $b_1$  to  $b_{25}$  obtained using CPED1 and data 1 of CPED1 for RPE and FPE are equal by definition and ranged, in this case, from 0.006 to 0.045. We observed, as expected, that the coefficients were following the same pattern as the informativeness of markers (results not shown), calculated by the polymorphism information content (Botstein et al., 1980). The PLSR gave the least weight to the less informative markers and the greatest weight to the more informative markers. Through their derivation, the differences between the 2 equations were that FPE had an additional regression coefficient ( $b_p$ ), obtained using IP A1 of CPED1, corresponding to the addition of a new term and that the intercept was modified by adding  $b_0^*$  to  $b_0$ . The  $b_p$  was equal to 0.378; this value was bigger than the weights obtained for each marker because the pedigree-based coefficient, even if incomplete, reflects the unexplained part of the genome. The other difference is that adding  $b_0^*$  to the intercept decreased the total value of the intercept, what is also a consequence of the inclusion of more information in the estimation.

The equations RPE and FPE were tested on other data and the other IP of CPED1, we obtained a significant increase of the correlation with FPE as well as a decrease of all parameters describing the residuals (Table 1). This was a consequence of the inclusion of more information in FPE. The high range of values of the residuals expressed that the residuals capture the difference between theoretical value of relationship, which measure the probable proportion of genes that are alike for 2 individuals due to their common ancestry, and “real” relationship, which is affected by Mendelian sampling. For example, 2 full-sibs share theoretically 50% of their genome, but in reality, they can share 0%, if by chance they get a completely different part of the genome of their parents, to 100% of their genome, if they are twins. The mean residual was slightly negative because molecular-based relationship coefficients have the tendency to overestimate the relationships. Explanation of this overestimation was that molecular-based estimators account not only for the IBD that arises during the population history (genealogical coancestry), but also

**Table 1.** Correlation between true-simulated and predicted relationships and parameters describing the residuals obtained with the reduced prediction equation (RPE) and the full prediction equation (FPE) applied on complete pedigree 1

Parameter	RPE	FPE
Correlation (SD)	0.647 (0.023)	0.814 (0.015)
Residual		
Mean (SD)	-0.040 (0.079)	-0.034 (0.063)
Range	0.816	0.577
IR <sup>1</sup>	0.090	0.085

<sup>1</sup>IR = interquartile range.

for the IBS present in the founder population (Toro et al., 2002; Fernández and Toro, 2006). However, accounting for the pedigree relationship value in the estimation decreased this overestimation of about 15%. The differences in the range and in the interquartile range of the residuals showed that the use of FPE also decreased the dispersion of these residuals. No significant differences were detected between the results obtained with IP of case A and IP of case B (results not shown). So, a pedigree of poor quality can give information about the theoretical part of the genome shared by individuals, independent of the depth of this one, at least for the genotyped individuals.

In practice, RPE could be computed from genotyped animals having complete pedigrees down to the base generation. In these cases, the method assumes that a sufficient number of animals with complete pedigree and well-known relationships were genotyped in addition to the animals of interest (e.g., animals with large number of descendants and unknown parents). For the Skyros pony, this was not the case, so it was necessary to simulate a pedigree and genotypes with similar characteristics as the one of the population under study.

We tested these 2 equations on other pedigrees with greater or lower inbreeding level (Table 2). Again, the results were better when we used FPE. When applied on pedigree with lower inbreeding level than the one used for the development of the equation, the results are equivalent or slightly better. When applied on pedigree

with a greater inbreeding level, results for all parameters decreased but remained significantly better than the one obtained with RPE. The range of values of the residuals has the tendency to stay similar between pedigrees. A possible explanation is that the inbreeding level introduced only a bias in the results. If confirmed by further experiments, this will allow for the calculation of the weight of the markers independent from the inbreeding level. The inbreeding level could then be used to calculate the intercept.

### Integration of Ungenotyped Animals

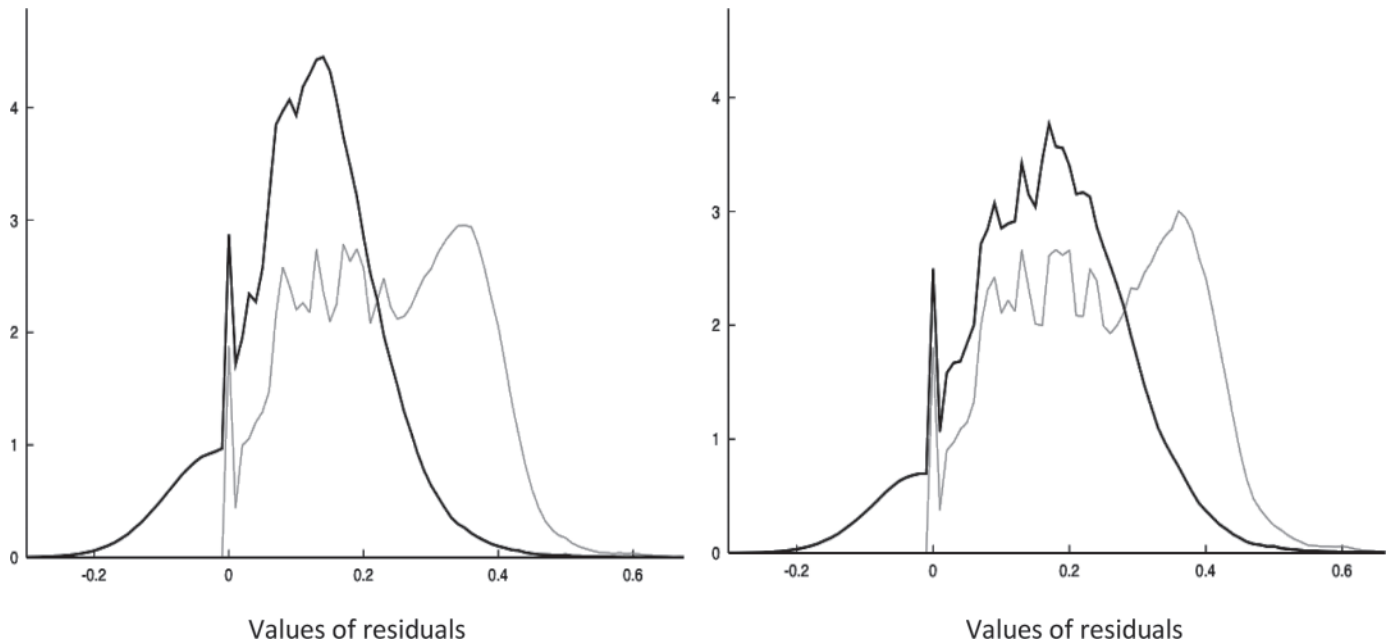
After addition of this combined information in the complete relationship matrix, the mean relative error between true relationship values, calculated on CPED, and estimations obtained in  $A_{IP}$  is equal to 0.931, whereas the mean relative error between true relationship values and estimations obtained in  $A_{modified}$  is equal to 0.607. These values quantified the dispersion of the results; there is thus a decrease of 34.83% of the dispersion of the results when we combined pedigree and genotypes in a single matrix.

The total mean bias of estimations calculated with classical tabular method on IP was equal to 0.243, whereas bias of estimations obtained with modified matrix was equal to 0.149. Again, there is a strong decrease (39.07%) between the 2 values. These observations are reflected in Figure 1. With IP, the estimated values are always inferior (i.e., the curve starts at 0) to the true values, which creates an increased dispersion of the results and an increased bias. On the other hand, the presence of residuals inferior to 0 with modified estimations expressed that relationships can by chance be not only less but also greater than the theoretical value. If the estimations were perfect, the probability density function would have been a normal curve centered on 0. This is not the case because some animals have no genotyped descendants; it is consequently impossible to transmit the molecular information to these animals. Figure 1 also shows that the results are better when the method is developed on IP with the same depth. When the method is tested with other pedigree of type

**Table 2.** Correlation between true-simulated and predicted relationships and parameters describing the residuals obtained with the reduced prediction equation (RPE) and the full prediction equation (FPE) applied on pedigrees with greater or less inbreeding than pedigree 1

Parameter	Less inbreeding		Greater inbreeding	
	RPE	FPE	RPE	FPE
Correlation (SD)	0.665 (0.016)	0.827 (0.009)	0.478 (0.032)	0.707 (0.022)
Residual				
Mean (SD)	-0.062 (0.075)	-0.057 (0.060)	0.120 (0.074)	0.125 (0.067)
Range	0.809	0.578	0.746	0.571
IR <sup>1</sup>	0.082	0.079	0.093	0.096

<sup>1</sup>IR = interquartile range.



**Figure 1.** Probability density function of the residuals obtained with new modified matrix (in black) and with classical tabular method (in gray) applied on incomplete pedigree of type A (left) and of type B (right).

A (on the left), the curve is much closer to a normal curve than when we test the method on IP of type B (on the right) and the center of the curve is also closer to 0, which expressed a smaller bias. This observation showed that, if no significant differences were detected between the results obtained with IP of case A and IP of case B when we considered only the genotyped individuals, differences appear when ungenotyped individuals are integrated. The explanation could be that the small difference that we considered as not significant is exponentially increased when the information is transmitted to the ancestors of the genotyped individuals. Because there are fewer links in the pedigree, some links are not re-created and residuals increased. If the method is, as in this case, developed on simulated pedigree, the solution is always to calculate the regressions coefficient on a simulated pedigree with worse characteristics than what we observed in reality. The need to simulate pedigree occurs when there are not enough animals with required completeness of pedigree (i.e., having complete pedigrees down to the base generation).

In both cases, dispersion and bias, there were strong differences between values obtained with incomplete pedigree of type A and of type B (Figure 1), showing that even if there are no differences in case of the genotyped animals, there is one when the genotyped matrix is integrated in the complete matrix.

### Conclusions

In conclusion, this method is a promising strategy to combine molecular information with genealogical information into a single relationship matrix, which is similar to **A**. Compared with the methods of VanRaden

(2008) and Aguilar et al. (2010), this new method used PLSR to assign a weight to each marker, which seems to be linked to their informativeness, which solves the problem of giving an arbitrary weight (e.g.,  $w$  in VanRaden, 2008) to the genomic matrix before combining it to the pedigree matrix. Even if  $w$  is related to  $b_P$ , because both are a way (one empirical, the other calculated) to account for the polygenic effect not explained by the marker, their precise theoretical relationship is yet unclear. This method thus modifies the additive relationship matrix for “genomic” information but also corrected this molecular information for the ungenotyped part of the genome.

### LITERATURE CITED

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.
- Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314–331.
- Caballero, A., and M. A. Toro. 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet. Res.* 75:331–343.
- Caballero, A., and M. A. Toro. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.* 3:289–299.
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.
- Eding, H., and T. H. E. Meuwissen. 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J. Anim. Breed. Genet.* 118:141–159.
- Emik, L. O., and C. E. Terrill. 1949. Systematic procedures for calculating inbreeding coefficients. *J. Hered.* 40:51–55.
- Fernández, J., and M. A. Toro. 2006. A new method to estimate relatedness from molecular markers. *Mol. Ecol.* 15:1657–1667.

- Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. *Genet. Sel. Evol.* 21:467–477.
- Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. *Animal* 1:21–28.
- Glaubitz, J. C., O. E. Rhodes Jr., and J. A. Dewoody. 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol. Ecol.* 12:1039–1047.
- Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix use in prediction of breeding values. *Biometrics* 32:69–83.
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92:4656–4663.
- Malécot, G. 1948. *Les mathématiques de l'hérédité*. Masson et Cie, Paris, France.
- Meuwissen, T. H. E. 2002. GENCONT: An operational tool for controlling inbreeding in selection and conservation schemes. Pages 769–770 in *Proc. 7th World Congr. Genet. Applied to Livest. Prod.*, Montpellier, France.
- Minvielle, F. 1990. *Principes d'amélioration génétique des animaux domestiques*. INRA, Paris & PUL, Québec, Canada.
- Misztal, I., K. Weigel, and T. J. Lawlor. 1995. Approximation of estimates of (co)variance components with multiple-trait restricted maximum likelihood by multiple diagonalization for more than one random effect. *J. Dairy Sci.* 78:1862–1872.
- Nejati-Javaremi, A., C. Smith, and J. P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Anim. Sci.* 75:1738–1745.
- Oliehoek, P. A., J. J. Windig, J. A. M. van Arendonk, and P. Bijma. 2006. Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics* 173:483–496.
- Toro, M., C. Barragán, C. Óvilo, J. Rodrigañez, C. Rodriguez, and L. Silió. 2002. Estimation of coancestry in Iberian pigs using molecular markers. *Conserv. Genet.* 3:309–320.
- Toro, M., L. Silió, J. Rodrigañez, C. Rodriguez, and J. Fernández. 1999. Optimal use of genetic markers in conservation programmes. *Genet. Sel. Evol.* 31:255–261.
- VanRaden, P. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J. Dairy Sci.* 75:3136–3144.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- Van Vleck, L. D., E. J. Pollak, and E. A. B. Oltenacu. 1987. *Genetics for the Animal Sciences*. Freeman, New York, NY.
- Verrier, E., X. Rognon, H. de Rochambeau, and D. Laloe. 2005. Les outils et méthodes de la génétique pour la caractérisation, le suivi et la gestion de la variabilité génétique des populations animales. *Ethnozootecnie* 76:67–82.
- Wang, J. 2001. Optimal marker-assisted selection to increase the effective size of small populations. *Genetics* 157:867–874.
- Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Naturalist* 56:330–338.
- Zhang, Z., R. J. Todhunter, E. S. Buckler, and L. D. Van Vleck. 2007. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *J. Anim. Sci.* 85:881–885.



## References

This article cites 24 articles, 5 of which you can access for free at:  
<http://jas.fass.org/cgi/content/full/89/4/972#BIBL>