Estimating Between And Within Line Variation Based On Pedigree And Genomic Relationship Matrix In Laying Hens

F. Biscarini[*], H.A. Mulder[†], H. Bovenhuis[*] and M.P.L. Calus[†]

Introduction

In genetic evaluations of livestock information on relatives is included using a relationship matrix. The classic numerator relationship matrix reports the average proportion of alleles identical by descent (IBD) shared by a pair of individuals predicted on the basis of their pedigree relationship (Henderson, 1976). However, due to Mendelian sampling the true proportion of genes shared by a pair of individuals can be different from the average value inferred from genealogy (Hill (1993); Guo (1996)). For instance, full-sibs are expected to have on average 50% of the genes in common but, depending on which alleles they inherit from their parents at the moment of meiosis, this proportion can theoretically vary from 0 to 100%. Around the mean value of 0.5, Guo (1996) estimated a standard deviation of 0.04 for additive relationships between full-sibs in a species with 30 chromosomes of 1M each. Marker information can be used to estimate the actual proportion of alleles IBD for each pair of individuals and to estimate true relationships with a certain degree of accuracy. The use of true instead of average relationships can be of importance in genetic evaluations, since it leads to higher accuracies of estimated breeding values (Villanueva et al. (2005); Hayes et al. (2009a)) and genetic parameters, thus increasing the response to selection. Genomic relationships may reveal also relatedness between different populations, like breeds or genetic lines, that appear unrelated based on the registered pedigree relations. This information can be used to estimate the genetic variance between populations, and not only within them. In this study we built the matrix of genetic relationships using a panel of 1536 SNP markers in a population of 675 laying hens from 9 different lines. We estimated the within- and between-line variation for body weight measured at 19 wk: results were compared with estimates from the classical approach based on the numerator relationship matrix derived from the pedigree.

Material and methods

Description of data. The animal population used in this study consisted of 675 laying hens from 9 genetic lines (Table 1). Lines were either of Rhode Island Red type or of White Leghorn type and were not related based on the relationships registered in the pedigree file. Hens were housed in battery cages in a single stable from wk 19 to 69 (laying period of 51 wk); cages comprised 4 hens of the same line, either full-sibs or randomly mixed. Hens were

---

[*] ABGC, Wageningen University, PO Box 338, 6700 AH Wageningen, The Netherlands
[†] ABGC, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad, The Netherlands

genotyped for a panel of 1536 SNP markers selected to cover candidate genes and QTL regions for immune and behavioural traits, identified in previous mapping studies. The SNPs were distributed over 24 of the 39 chromosome pairs of the chicken genome. Only the 1031 SNPs with a minor allele frequency > 5% were used in the analysis. Details on the SNPs used and on the editing procedure can be found in Biscarini et al. (2010). The phenotypes consisted of individual records of the BW of the hens at 19 wk of age (BW19) (Table 1). From the pedigree file 4 generations of ancestors were extracted for genetic analysis. Within each line, roosters were on average mated with 2.3 females, whereas females were mated with only 1 male.

Table 1: Number of hens per line and mean and std. dev. of body weight at 19 wk.

trait	line	B1	B2	B3	BB	W1	WA	WB	WC	WF	tot
	n	83	77	77	67	74	78	77	63	79	675
BW19	mean	1542	1472	1570	1562	1201	1267	1343	1301	1238	1390
	sd	120	99	150	188	83	134	123	155	118	192

Statistical analyses. Variance components for BW19 were estimated using either the classical numerator relationship matrix $\mathbf{A}$ (Henderson (1976)), or the genomic relationship matrix $\mathbf{G}$. The two methods are hereafter named A and G. A third model considered a G-matrix in which relationships between lines were set to 0, and was named G*. The genomic relationship matrix $\mathbf{G}$ was obtained with the following formula:

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum p_i(1-p_i)} \tag{1}$$

where $\mathbf{Z}$ is a matrix that relates SNP alleles to individuals, and p_i is the second allele frequency of SNP_i (Van Raden (2008)). Allele frequencies were calculated across lines. The classical MME were set up using the following single-trait animal model:

$$y_{ij} = \mu + line_i + a_j + e_{ij} \tag{2}$$

where y_{ij} is the BW of hen j of line i; $line_i$ is the random effect of genetic line; a_j is the random additive genetic effect of the j_{th} hen; and e_{ij} is the residual. In model A, $Var(\mathbf{a}) = \mathbf{A}\sigma^2_{a\_WITHIN}$, where $\sigma^2_{a\_WITHIN}$ is a pooled estimate of the within-line genetic variance, and $Var(\mathbf{e}) = \mathbf{I}\sigma^2_e$. In model G, Var(e) remains unchanged while $Var(\mathbf{a}) = \mathbf{G}\sigma^2_{a\_TOTAL}$, with $\mathbf{G}$ being the genomic relationship matrix derived from equation [1]: this is a combination of the within- and between-line genetic variation. The line effect was not considered in model G because the G-matrix contains relationships among lines. Since lines are unrelated in the pedigree file, the $\mathbf{A}$ matrix is in fact a block-diagonal matrix with all off-diagonal block elements equal to 0. In the $\mathbf{G}$ matrix relationships between

lines are instead inferred from the marker information. These between lines relationships in the **G** matrix where then set to 0 to create the **G*** matrix, which was used in model G*, where all elements are as specified in equation [2], but $Var(\mathbf{a}) = \mathbf{G*}\sigma^2_{a\_WITHIN}$, with **G*** being the genomic relationship matrix in which between lines relationship have been set to 0. The Asreml software package was used for the estimation of variance components and breeding values (Gilmour et al. (2002)).

Results and discussion

G matrix. In the genomic relationship matrix the phylogenetic clusters corresponding to the White Leghorns and the Rhode Island Reds were clearly distinguished. Also the different lines could be identified, especially for the white layers that have been selected for longer times and show higher homozygousity (Hillel et al. (2003); Biscarini et al. (2010)). Overall these results were comparable with the genetic distances estimated by Biscarini et al. (2010) with the method of Nei (1972).

Between and within line variance. Estimates of variance components are reported in Table 2. With models A and G* separate estimates of the pooled within-line additive genetic variance and of the between-line variance are obtained. In model G the total genetic variance, combination of the within- and between-line variances, is estimated. In this case, a rough estimate of the between-line variance can be obtained by subtracting the additive genetic variance estimated with the A or G* models from that estimated with the G model: this gives values of 6536 or 9443, approximately 30% to 45% of the between-line variance estimated with models A and G*. There is evidence of a substantial contribution of the between-line variation to the total variance: it constitutes from 20/30% (model G) to over half of the total variance (54% and 55% in models A and G* respectively). Mixed populations are classically analysed by including the fixed effect of lines or genetic groups; genomic relationships offer an alternative for modeling population stratification. Heritability estimates for BW19 differed in the three models, and are difficult to compare given the different definitions of variance components. In models A and G*, the variance of the random line effect was not considered in calculating the heritability. However, the standard error of the estimate was much lower with the G model, being approximately half of the SE of the A model estimate. This is due to the fact that estimated relationships between animals are more accurate in the genomic than in the additive relationship matrix. In the G* model, the SE of estimated heritability was still lower than in the A model, but quite higher than in the G model, probably because relationships between lines were set to 0. Although their ratio remained more or less constant, the genetic and residual variances estimated with the G model were more than 1.5 to 2 times bigger than those estimated with the A and G* models. This is most likely because the G model considers also between line variation, whereas the A and G* models with line effect look only at within line variation. The G-matrix was obtained from multiple lines of layers using the across-line allele frequencies. It could be more appropriate to use within-line allele frequencies to derive within-line genomic relationships, and across-line allele frequencies to estimate between-line genomic relationships.

Table 2: Residual, within-line and between-line variance components for BW at 19 wk estimated with the 3 models.

model	var(A)$^{\alpha}$	var(E)	var(line)	h^2	s.e. of h^2
A	9117	8972	21149	0.504	0.103
G	15653	15299		0.506	0.054
G*	6210	10998	21356	0.359	0.095

[a]Var(a) has different definitions in the three models: it is the within-line genetic variance in models A and G*, and the combination of within- and between-line variances in model G

Conclusion

Using genomic information, the relationships between lines can be inferred. This allows for the estimation of both within- and between-line variation, showing that a considerable proportion of the total genetic variance is actually due to variation between lines. Besides, the genomic relationships matrix allows for the estimation of GEBVs, and it can be used also to reconstruct phylogenetic relationships between lines.

Acknowledgements

References

Biscarini, F., Bovenhuis, H., van Arendonk, J.A.M. *et al.* (2010). *Anim. Genet.*, 41:26–38.

Guo, S.W. (1996*). Hum. Hered.*, 46:61–70.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. *et al.* (2002) ASReml User Guide

Hayes, B.J., Visscher, P.M., and Goddard, M.E. (2009) *Genet. Res.,* 91:47–60.

Henderson, C.R., and Quaas, R.L. (1976). *J. Anim. Sci.,* 43:1188–1197.

Hill, W.G. (1993). *Heredity*, 71:652–653.

Hillel, J., Groenen, M.A.M., Tixier-Boichard, M. *et al*. (2003) *Genet. Sel. Evol*. 35:533-557.

Nei, M. (1972) *Am. Nat*. 106:283-292.

Villanueva, B., Pong-Wong, R., Fernández, J. *et al*. (2005). *J. Anim. Sci*., 83:1747-1752.

Van Raden, P.M. (2008) *J. Dairy. Sci*., 91:4414–4423.