Genome Wide Marker Assisted Selection in Chicken: Making the Most of All Data, Pedigree, Phenotypic, and Genomic in a Simple One Step Procedure

C. Y. Chen[*], I. Misztal[*], I. Aguilar[*†], S. Tsuruta[*], T. H. E. Meuwissen[‡], S. E. Aggrey[§], and W. M. Muir[**]

Introduction

Genomic selection methods are centered on what assumptions are considered valid, the most critical being the assumed distribution of gene effect (Meuwissen *et al*., 2001). Several studies have found that an assumption of the infinitesimal model, equivalent to BLUP with a genomic relationship matrix (GRM), performed as well as others (Bayes A or Bayes B, Hayes *et al.*, 2009; VanRaden *et al*., 2009). In addition, a GRM approach allows inclusion of pedigree performance data that cannot be included using those other procedures. However, use of a GRM normally requires a multi-step procedure: deregressed evaluations and estimation of genomic effects followed by combining with traditional parent averages and genomic solutions (VanRaden, 2008; Hayes *et al.*, 2009; VanRaden *et al*., 2009). In chickens, phenotypes on genotyped animals have been used directly (González -Recio *et al.*, 2008, 2009), ignoring contributions from ungenotyped animals. Misztal *et al.* (2009) proposed a single-step procedure (SSP) which utilizes joint information provided by a full pedigree and genomic data by modification of the usual relationship matrix for genomic selection. Modifications were shown by Legarra *et al.* (2009) and implemented by Aguilar *et al.* (2010) in Holsteins. The objective of this study was to apply the SSP for genomic evaluation in broiler chickens to determine if accuracy of prediction could be enhanced using phenotypic data from pedigreed animals to augment information obtained on animals genotyped (FULL), as compared to a subset which included only phenotypes on those animals genotyped (SUB), i.e. the procedure used by González -Recio *et al.* (2008, 2009).

Material and methods

Data. Body weight at 6 weeks (BW, 100g), breast meat area (BM, cm^2), and leg score (LEG, 1=no and 2=yes for defect) for two pure lines of broiler chickens were provided from Cobb-Vantress, Inc. A full data set of all animals (FULL, n=183,784 and 164,246 birds for lines 1 and 2) and a subset of genotyped animals (SUB, n=3,284 and 3,098 birds for lines 1 and 2) were analyzed separately for each line. Genotypes were assayed using the poultry 60k SNP chip developed by the Chicken Genomic Selection Project. Descriptions of phenotypic records are shown in Table 1. A total of 57,636 SNP were informative. The training population consisted of records from generations 1 and 2. The validation population contained 799 genotyped animals in generation 3.

[*] Department of Animal and Dairy Science, University of Georgia, Athens 30602, USA
[†] Instituto Nacional de Investigación Agropecuaria, Las Brujas 90200, Uruguay
[‡] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, NO-1432 As , Norway
[§] Department of Poultry Science, University of Georgia, Athens 30602, USA
[**] Department of Animal Science, Purdue University, West Lafayette 47907, USA

Statistical analyses. The single-trait model used for BW, BM, and LEG was:

$$y = Xb + Zu + Wmp + e,$$

where y is the vector of observations; b is the vector of fixed effects including hatch and sex; u and mp are vectors of random additive genetic and maternal permanent environmental effects; X, Z, and W are incidence matrices; e is the vector of residuals. Maternal permanent environmental effect was not considered for BM and LEG. In a regular BLUP, the (co)variance matrix was assumed to

$$\text{var}\begin{bmatrix} u \\ mp \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_u^2 & 0 & 0 \\ 0 & I\sigma_{mp}^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix},$$

where A is the numerator relationship matrix, and σ_u^2, σ_{mp}^2, and σ_e^2 were additive, maternal permanent and residual variances, respectively. In SSP with genomic information, the A matrix was replaced by the H matrix with the following inverse (Aguilar *et al.*, 2010):

$$H^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & G^{-1} + A^{22} - A_{22}^{-1} \end{bmatrix},$$

where H is a modified relationship matrix incorporating genomic information, indices 1 and 2 correspond to ungenotyped and genotyped animals, respectively, and G is a genomic relationship matrix that created as in Aguilar *et al.* (2010). Genetic evaluations were done by modified BLUP90IOD (Tsuruta *et al.*, 2001; Misztal *et al.*, 2002; Aguilar *et al.*, 2010) using regular BLUP and SSP with FULL and SUB. Bayes A approach (Meuwissen *et al.*, 2001) was used with SUB only. Predictive ability was estimated as the correlation between predicted breeding value and the sum of true breeding value and residual, $r(\hat{u}, u + e)$. Accuracy was estimated as the correlation between predicted and true breeding values, $r(\hat{u}, u) = r(\hat{u}, u + e)/h$ where h is the square root of heritability.

Table 1: Description of phenotypic records of two lines in each data set

Trait	FULL[1]		SUB[2]	
	Line 1	Line 2	Line 1	Line 2
BW, 100g				
No. of records	183,784	164,246	3,284	3,098
Mean	24.50	23.53	25.09	23.36
SD	3.22	3.17	2.94	2.63
BM, cm^2				
No. of records	40,914	40,576	3,099	2,993
Mean	42.81	41.09	42.87	41.00
SD	5.35	5.10	5.47	5.12
LEG, score				
No. of records	183,784	164,246	3,284	3,098
Mean	1.19	1.16	1.07	1.12
SD	0.39	0.37	0.25	0.33

[1]Phenotypic records of all animals from three generations.
[2]Phenotypic records of genotyped animals from three generations.

Results and discussion

Estimates of variance components using FULL are in Table 2. Heritability for BW, BM, and LEG were 0.20, 0.30, and 0.11 for line 1 and 0.17, 0.35, and 0.09 for line 2. Significant changes in estimates using SUB (not shown) indicated strong and line-specific selection of genotyped animals with heritability of 0.25, 0.21, and 0.09 for line 1 and 0.24, 0.29, and 0.20 for line 2. The accuracies of prediction are in Table 3. In general, accuracies were lower for BM than for BW despite its higher heritability. This is likely caused by incomplete recording on BM. Accuracies were very low for LEG/SUB, especially in line 1. Proportion of LEG=2 were 19% (FULL) or 7% (SUB) in line 1, and 16% and 12% in line 2, respectively, indicating that SUB were preselected stronger in line 1 with accuracies in SSP/SUB much lower than in BLUP/FULL. Pollak *et al.* (1994) showed that preselection causes upward bias for the worst animals and downward bias for the best animals.

Table 2: Estimates of variance components using FULL for the two lines

Estimates	Line 1			Line 2		
	BW	BM	LEG	BW	BM	LEG
σ^2_u	1.03	4.04	0.02	0.85	4.34	0. 01
σ^2_{mp}	0.40	—	—	0.32	—	—
σ^2_e	3.69	9.61	0.13	3.83	7.95	0.12
h^2	0.20	0.30	0.11	0.17	0.35	0.09

Table 3: Accuracy obtained using BLUP, SSP, and two-step Bayes A

Item	No genomic Information		Genomic		
	BLUP		SSP		Bayes A
	SUB	FULL	SUB	FULL	SUB
Line 1					
BW	0.46	0.51	0.60	0.61	0.60
BM	0.30	0.34	0.34	0.40	0.36
LEG	<0	0.28	0.06	0.37	0.09
Line 2					
BW	0.39	0.24	0.50	0.44	0.47
BM	0.27	0.33	0.45	0.51	0.51
LEG	0.24	0.43	0.15	0.73	0.11

For continuous traits, BW and BM performed differently. For BW, the use of FULL improved accuracy in line 1 while unexpected deterioration regardless of the use of the genomic information with much lower accuracies was found in line 2. One explanation could be a specific selection on correlated and probably antagonistic traits. A different selection strategy was previously pursued in each line on more than 20 traits, but only 3 traits were currently analyzed. For BM, the improvement of accuracies from SUB to FULL was about 0.04-0.06 and from no genomic to genomic was 0.04-0.06 (line 1) to 0.18 (line 2). For LEG, a binary trait that was analyzed as a linear, the improvement from no genomic to genomic evaluation varied from 0.06 in SUB to 0.09 in FULL for line 1 and -0.09 in SUB to 0.30 in FULL for line 2. The increase in accuracy for LEG in line 1 was more modest due to the deterioration with SUB in line 2. The low accuracies for LEG, especially in line 1 could possibly be the consequence its binary nature, a heritability > 0.3 estimated by threshold

model (results not reported), i.e. accuracy as defined for a linear trait is only an approximation of that for binary traits. Results with Bayes A were similar to SSP/SUB, with Bayes A being slightly more accurate for BM and LEG in line 1 and BM in line 2. The better performance of Bayes A in this case could be due to major genes.

Genomic selection based only on the genotyped animals appears to work well for traits with complete recording, at least moderate heritabilities, and no prior strong selection. For traits under strong selection, use of only the genotyped subset may not be useful. For traits with low hereditability and especially those preselected, an increase in accuracy is only possible if the complete data set, genotyped plus pedigree phenotypes, are used.

Conclusion

Results of genomic selection using only records of genotyped animals depend strongly on selection criteria used for genotyped animals and trait heritability. For some traits the accuracy using the subset can be higher than a BLUP evaluation using the complete population. For traits evaluated using genotyped animals only with undergoing sequential selection, the accuracies may be very low. The most accurate evaluation would involve the complete populations with multiple-trait models and all traits on which the selection was practiced. Such an evaluation is possible with a single-step methodology. A critical part of genomic selection is correct model development as flaws in the BLUP model can affect the accuracies of genomic evaluations.

Acknowledgements

The authors thank Cobb-Vantress for access to data for this study. This study was partially funded by the Holstein Association, Smithfield Premium Genetics, and by AFRI grants 2009-65205-05665 and 2010-65205-20366 from the USDA NIFA Animal Genome Program.

References

Aguilar, I., Misztal, I., Johnson, D. L. *et al.* (2010). *J. Dairy Sci.,* 93:743–752.

González -Recio, O., Gianola, D., Long, N. *et al.* (2008). *Genetics,* 178: 2305–2313.

González -Recio, O., Gianola, D., Rosa, G. J. M. *et al.* (2009). *Genet. Sel. Evol.,* 41: 3–13.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J. *et al.* (2009). *J. Dairy Sci.,* 92: 433–443.

Legarra, A., Aguilar, I., and Misztal, I. (2009). *J. Dairy Sci.,* 92: 4656–4663.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). *Genetics,* 157: 1819–1829.

Misztal, I., Legarra, A., and Aguilar, I. (2009). *J. Dairy Sci.,* 92: 4648–4655.

Misztal, I., Tsuruta, S., Strabel, T. *et al.* (2002). In *Proc 7th WCGALP*, 33: 743–744.

Pollak, E. J., van der Werf, J., Quaas, R. L. (1984). *J. Dairy Sci.,* 67: 1590–1595.

Tsuruta, S., Misztal, I., and Stranden, I. (2001). *J. Anim. Sci.,* 79:1166–1172.

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

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R. *et al.* (2009). *J. Dairy Sci.,* 92: 16–24.