

6-1-2013

Analysis of egg production in layer chickens using a random regression model with genomic relationships

Anna Wolc

Iowa State University, awolc@iastate.edu

J. Arango

Hy-Line International

P. Settar

Hy-Line International

J. E. Fulton

Hy-Line International

N. P. O'Sullivan

Hy-Line International

See next page for additional authors

Follow this and additional works at: https://lib.dr.iastate.edu/ans_pubs



Part of the [Agriculture Commons](#), [Genetics and Genomics Commons](#), [Poultry or Avian Science Commons](#), and the [Statistical Models Commons](#)

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/ans_pubs/647. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Animal Science at Iowa State University Digital Repository. It has been accepted for inclusion in Animal Science Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Analysis of egg production in layer chickens using a random regression model with genomic relationships

Abstract

Random regression models allow for analysis of longitudinal data, which together with the use of genomic information are expected to increase accuracy of selection, when compared with analyzing average or total production with pedigree information. The objective of this study was to estimate variance components for egg production over time in a commercial brown egg layer population using genomic relationship information. A random regression reduced animal model with a marker-based relationship matrix was used to estimate genomic breeding values of 3,908 genotyped animals from 6 generations. The first 5 generations were used for training, and predictions were validated in generation 6. Daily egg production up to 46 wk in lay was accumulated into 85,462 biweekly (every 2 wk) records for training, of which 17,570 were recorded on genotyped hens and the remaining on their nongenotyped progeny. The effect of adding additional egg production data of 2,167 nongenotyped sibs of selection candidates [16,037 biweekly (every 2 wk) records] to the training data was also investigated. The model included a 5th order Legendre polynomial nested within hatch-week as fixed effects and random terms for coefficients of quadratic polynomials for genetic and permanent environmental components. Residual variance was assumed heterogeneous among 2-wk periods. Models using pedigree and genomic relationships were compared. Estimates of residual variance were very similar under both models, but the model with genomic relationships resulted in a larger estimate of genetic variance. Heritability estimates increased with age up to mid production and decreased afterward, resulting in an average heritability of 0.20 and 0.33 for pedigree and genomic models. Prediction of total egg number was more accurate with the genomic than with the pedigree-based random regression model (correlation in validation 0.26 vs. 0.16). The genomic model outperformed the pedigree model in most of the 2-wk periods. Thus, results of this study show that random regression reduced animal models can be used in breeding programs using genomic information and can result in substantial improvements in the accuracy of selection for trajectory traits.

Keywords

genomic selection, random regression, egg production

Disciplines

Agriculture | Animal Sciences | Genetics and Genomics | Poultry or Avian Science | Statistical Models

Comments

This article is published as Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'sullivan, R. Preisinger, R. Fernando, D. J. Garrick, and J. C. M. Dekkers. "Analysis of egg production in layer chickens using a random regression model with genomic relationships." *Poultry science* 92, no. 6 (2013): 1486-1491. doi: [10.3382/ps.2012-02882](https://doi.org/10.3382/ps.2012-02882).

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Authors

Anna Wolc, J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, Rohan Fernando, D. J. Garrick, and Jack C. M. Dekkers

This article is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/ans_pubs/647

Analysis of egg production in layer chickens using a random regression model with genomic relationships

A. Wolc,*†‡¹ J. Arango,‡ P. Settar,‡ J. E. Fulton,‡ N. P. O'Sullivan,‡ R. Preisinger,§ R. Fernando,† D. J. Garrick,† and J. C. M. Dekkers†

**Department of Genetics and Animal Breeding, Poznan University of Life Sciences, Wolynska St. 33, 60-637 Poznan, Poland; †Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames 50011-3150; ‡Hy-Line International, Dallas Center, IA 50063; and §Lohmann Tierzucht GmbH, 27472 Cuxhaven, Germany*

ABSTRACT Random regression models allow for analysis of longitudinal data, which together with the use of genomic information are expected to increase accuracy of selection, when compared with analyzing average or total production with pedigree information. The objective of this study was to estimate variance components for egg production over time in a commercial brown egg layer population using genomic relationship information. A random regression reduced animal model with a marker-based relationship matrix was used to estimate genomic breeding values of 3,908 genotyped animals from 6 generations. The first 5 generations were used for training, and predictions were validated in generation 6. Daily egg production up to 46 wk in lay was accumulated into 85,462 biweekly (every 2 wk) records for training, of which 17,570 were recorded on genotyped hens and the remaining on their nongenotyped progeny. The effect of adding additional egg production data of 2,167 nongenotyped sibs of selection candidates [16,037 biweekly (every 2 wk) records] to the training data was also investigated. The model included a 5th order Leg-

endre polynomial nested within hatch-week as fixed effects and random terms for coefficients of quadratic polynomials for genetic and permanent environmental components. Residual variance was assumed heterogeneous among 2-wk periods. Models using pedigree and genomic relationships were compared. Estimates of residual variance were very similar under both models, but the model with genomic relationships resulted in a larger estimate of genetic variance. Heritability estimates increased with age up to mid production and decreased afterward, resulting in an average heritability of 0.20 and 0.33 for pedigree and genomic models. Prediction of total egg number was more accurate with the genomic than with the pedigree-based random regression model (correlation in validation 0.26 vs. 0.16). The genomic model outperformed the pedigree model in most of the 2-wk periods. Thus, results of this study show that random regression reduced animal models can be used in breeding programs using genomic information and can result in substantial improvements in the accuracy of selection for trajectory traits.

Key words: genomic selection, random regression, egg production

2013 Poultry Science 92:1486–1491
<http://dx.doi.org/10.3382/ps.2012-02882>

INTRODUCTION

After successful implementation of random regression models (**RRM**) in genetic evaluation of dairy cattle, the possibility of their utilization for longitudinal traits in poultry has been shown in several studies (Anang et al., 2002; Kranis et al., 2007; Luo et al., 2007; Wolc and Szwaczkowski, 2009). An important advantage of RRM is that they allow modeling of variance components (and breeding values) over time, thereby providing op-

portunities to select for persistency of egg production (Wolc et al., 2011b). Another recent advancement is the utilization of information from high-density SNP panels to increase accuracy of selection. Due to high costs, it is unlikely that all individuals will be genotyped; therefore, methods that combine information from genotyped and nongenotyped animals such as the single-step approach (Aguilar et al., 2010) and the reduced animal model (Wolc et al., 2011c) can be applied to optimize the joint use of phenotypic, genomic, and pedigree information for selection. This study estimates variance components over time and combines the above-mentioned developments, with the objective of maximizing accuracy of estimated breeding values for egg production in a purebred brown egg shell layer pop-

©2013 Poultry Science Association Inc.

Received November 5, 2012.

Accepted February 10, 2013.

¹Corresponding author: awolc@jay.up.poznan.pl

ulation, by using a random regression reduced animal model with a genomic relationship matrix compared with a pedigree-based RRM.

MATERIALS AND METHODS

Phenotypic Data

Two scenarios were evaluated: in the first scenario, for training, daily egg production up to 46 wk in lay from 5 generations of a brown egg layer line was accumulated into 85,462 biweekly (every 2 wk) records, of which 17,570 were recorded on genotyped hens (1,122 birds) and the remainder on their nongenotyped progeny (7,986 birds). In the second scenario, the egg production data of the 2,167 nongenotyped sibs of selection candidates [16,037 biweekly (every 2 wk) records] were added to the training data set. These data represented the early part of the laying season. The first scenario represents the situation of selecting individuals at hatch when no phenotypes are available for that generation, whereas the second scenario mimics selection of males for which phenotypes on female relatives become available over time but their own phenotype is not available. Validation was on 288 genotyped and phenotyped individuals from generation 6, with 5,787 biweekly (every 2 wk) records. Numbers of genotyped and phenotyped birds per generation were as described by Wolc et al. (2011a), except for a few additional missing values. All hens had records up to 10 wk of production (early evaluation), but only selected hens had records in the later periods (late evaluation). The selection was on an index combining 16 production and egg quality traits, although coselection and mating of close relatives was avoided to reduce inbreeding. Individual days for which the production of the flock dropped rapidly compared with surrounding days were excluded. The data were standardized by dividing the number of eggs recorded within a calendar-based 2-wk period by the number of days with production records in this period and multiplying by 14. If fewer than 8 d were available in a given 2-wk period, the period was excluded from analysis.

Genomic Data

For the genomic analyses, genotypes on 24,430 SNP from a custom-made Illumina SNP chip were used after quality checks (minor allele frequency > 0.025, maximum number of missing genotypes < 0.05, maximum parent-offspring mismatches < 0.05). Chromosomes 1 to 28, 2 unassigned linkage groups, and chromosome Z were represented on the SNP panel. Recorded pedigree was checked for Mendelian mismatches between parents and progeny based on available SNP genotypes. If the proportion of opposing homozygous SNP for a given parent-progeny pair was >0.1, then parents were set to unknown for the pedigree analysis.

Statistical Methods

The genomic relationship matrix was constructed for 3,908 genotyped individuals using the method of VanRaden (2008). The reduced animal model of Wolc et al. (2011c) was used to combine phenotypic information of genotyped individuals with that of their nongenotyped progeny.

The following RRM was fitted using ASReml (Gilmour et al., 2008).

$$y_{ikl} = HY_i \sum_{m=0}^{n1} b_m z_{klm} + \mathbf{Q} \sum_{m=0}^{n2} a_{km} z_{klm} + \mathbf{W} \sum_{m=0}^{n3} a1_{km} z_{klm} + \sum_{m=0}^{n4} p_{km} z_{klm} + e_{ikl},$$

where y_{ikl} is the record of hen k in period l within hatch-year i ($i = 1, 2, \dots, 16$); HY_i is the fixed effect of hatch-year i ; b_m is the m th fixed regression coefficient nested within hatch-year class; a_{km} is the m th random regression coefficient for the additive genetic effect of animal k ; $a1_{km}$ is the m th random regression coefficient for the Mendelian sampling effect of animal k ; p_{km} is the m th random regression coefficient for the permanent environmental effect of animal k ; z_{klm} is the covariate coefficient of Legendre polynomials for period l of animal k ; and e_{ikl} is the random residual effect. \mathbf{Q} is an ($N \times p$) incidence matrix where a row for a genotyped individual has zeros except for a 1 in the column corresponding to the individual, and a row for a nongenotyped individual has zeros except for 1/2 in the columns corresponding to the sire and dam of the individual. \mathbf{W} is the incidence matrix for Mendelian sampling with the element corresponding to the nongenotyped individuals equal to 1 or 0 otherwise, with variance of $a1$ constrained to be equal to half of the additive genetic variance. Including the Mendelian sampling effects in the model enabled fitting of residuals with homogeneous variance for genotyped and nongenotyped individuals within a time point. Scalars $n1$ to $n4$ are the numbers of coefficients for Legendre polynomials.

Across time, heterogeneous residual variance was assumed, with a separate value allowed for each 2-wk period.

Initially, first-order Legendre polynomials were used, as suggested by Wolc et al. (2011b) but based on preliminary results, the order of polynomials was increased to quadratic. Estimates of variance components and predictive ability were compared for alternative models using pedigree or genomic relationships to describe the variance-covariance matrices for the random genetic effects in the models (a_{km} and $a1_{km}$). Accuracy of estimated breeding values (**EBV**) was evaluated as the correlation between predicted and observed performance in generation 6 (the progeny of the last generation in training).

RESULTS

Estimates of variance components over time are plotted in Figures 1a and 1c. Estimates of residual variance were very similar under both pedigree- and genome-based models. The high residual variance in the early periods probably comes from differences in sexual maturity and reaction to the light schedule, which was hard to control when birds from different hatches were housed in the same building, as in this population. The fitting of fixed curves within hatch-weeks was undertaken to solve this problem, but it did not entirely remove the excessive residual variation. The linear model using genomic relationships resulted in higher estimates of genetic variance and lower estimates of permanent environmental variance than the model based on pedigree relationships. Estimates of genetic variance from a model with only a linear Legendre polynomial were clearly overestimated at later ages (heritability estimates were up to 0.8, Figures 1b and 1d). Thus, the order of polynomials was increased to 2 for both genetic and permanent environmental components, which resulted in significant improvement of the likelihood function ($\log L = -4,676$ with order 1 vs. $\log L = -4,153$ with order 2 polynomial), and a shape of heritability curve that was closer to what was obtained from separate, single-trait analyses of monthly records. For the model with quadratic polynomials for genetic and permanent environ-

mental effects, estimates of heritability increased with age up to mid-season (periods 8 to 14) and decreased afterward (Figures 1b and 1d), resulting in average heritability estimates of 0.20 and 0.33 for pedigree and genomic models, respectively. Further results are given only for the model with quadratic polynomials. Resulting estimates of variance components for additive genetic and permanent environmental terms are in Table 1. The estimates of genetic variance for intercept and slope were higher for the genome-based model than for the pedigree-based model but were to some extent reduced when additional information on sibs of selection candidates was added. The variance of the quadratic term for genetic effects, which allows the curve to bend back down, increased with additional data. In contrast, the intercept for the permanent environmental effect explained more variation with additional data, which suggests that there may be some level of confounding between genetic and permanent environmental effects, which leads to biased estimates when fewer data are available.

Prediction of total egg number (sum of eggs produced in 2-wk periods 1 to 23) was calculated according to analyzed models and compared with the sum of laid eggs in Table 2. Measured as correlation between predicted and realized phenotype, the prediction was more accurate when using genomic rather than pedigree relationships; the correlation of resulting EBV with phe-

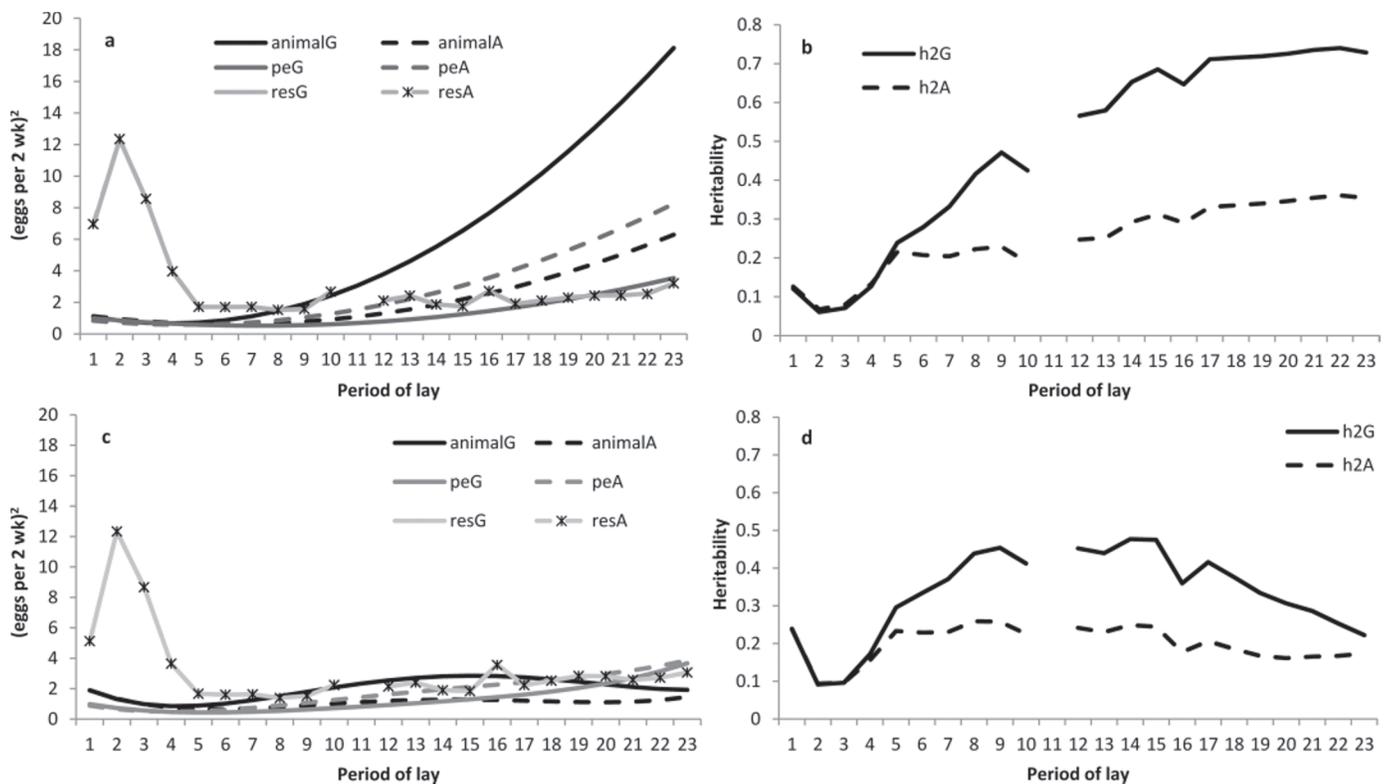


Figure 1. Estimates of variance components (a,c) and heritabilities (b,d) under random regression models (RRM) using pedigree (A) or genomic relationships (G). Linear (a,b) and quadratic (c,d) Legendre polynomials were used to model animal and permanent environmental effects and heterogeneous residual variance was assumed; animal = additive genetic variance, pe = permanent environmental variance, res = residual variance. There were no data in period 11; however, the RRM model estimates variance components as a continuous function. Thus, estimates of genetic and permanent environmental variance, but not residual variance, are available even for the period with no data.

Table 1. Estimates of additive genetic and permanent environmental variance components from random regression models using pedigree (A) or genomic (G) relationships and 5 generations of training data, with or without inclusion of production data on ungenotyped sibs of selection candidates (sibs)

Parameter	G				A			
	No sib data		With sib data		No sib data		With sib data	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Variance of intercept	2.701	0.237	2.183	0.206	1.312	0.163	1.242	0.148
Genetic effect								
Covariance between intercept and linear term	0.800	0.145	0.528	0.124	0.248	0.102	0.193	0.094
Variance of linear term	0.826	0.119	0.744	0.108	0.491	0.095	0.499	0.091
Covariance between intercept and quadratic term	-0.817	0.082	-0.643	0.075	-0.284	0.060	-0.220	0.058
Covariance between linear and quadratic term	-0.355	0.055	-0.320	0.052	-0.167	0.046	-0.194	0.045
Variance of quadratic term	0.459	0.052	0.472	0.052	0.366	0.047	0.415	0.048
Permanent environmental effect								
Variance of intercept	1.283	0.122	1.542	0.119	1.999	0.132	2.050	0.124
Covariance between intercept and linear term	0.729	0.084	0.816	0.082	1.034	0.094	1.026	0.090
Variance of linear term	1.044	0.086	1.056	0.083	1.222	0.093	1.195	0.091
Covariance between intercept and quadratic term	-0.158	0.047	-0.282	0.046	-0.441	0.045	-0.513	0.043
Covariance between linear and quadratic term	0.021	0.038	-0.018	0.037	-0.078	0.039	-0.086	0.039
Variance of quadratic term	0.188	0.036	0.249	0.037	0.212	0.039	0.249	0.039

notypes in generation 6 used for validation (progeny of generation 5) was 0.26 for the genomic relationship model vs. 0.16 for the pedigree-based model. For models with quadratic polynomial, the model with genomic relationships outperformed the pedigree-based model in selecting both top and bottom performance individuals with and without additional information on sibs, though under the linear model the top individuals from pedigree-based analysis on average performed better. Except for the first three 2-wk periods, prediction of egg records was more accurate with genomic relationships (Figure 2). Inclusion of sib information improved accuracy of prediction mostly in early lay for both pedigree- and genome-based models. This may be due to the fact that for early evaluation, all sibs have records, but sib records were not available for late evaluation.

DISCUSSION

In this study, RRM were used to predict breeding values of selection candidates using either genomic or pedigree relationship matrices. The model using genomic relationships was shown to outperform the traditional pedigree-based approach in terms of predictive ability in the next generation. The advantage of the genomic model comes from more accurate estimation of the parental average and some ability to predict Mendelian sampling terms (Wolc et al., 2011c). The use

of genomic relationships allows differentiation of full sibs without their own phenotypes or progeny records, which is especially relevant for egg production traits in males, which are traditionally evaluated based on their female relatives. In most 2-wk periods, predictions using genomic relationships without sib information were more accurate than those using pedigree relationships with sib. The greater predictive ability of the genomic over the pedigree-based model at later periods suggests that the genomic model is better for evaluation of persistency of egg production. The average accuracy from the genomic RRM utilizing longitudinal data was higher in the late periods than the accuracy reported by Wolc et al. (2011c) for this same population when analyzing combined late production (measured as egg production rate including only saleable eggs after 12 wk in lay) with a single-trait genomic model; however, the type of data used (i.e., saleable in the previous study vs. all eggs in the current study) could also have contributed to the difference.

In several other commercial layer lines, Wolc et al. (2011b) found a RRM with linear Legendre polynomials to provide a good description of the data on egg production and this was also the case for the line analyzed here when using pedigree information. However, when genomic relationships were used, the known feature of Legendre orthogonal polynomials of inflated variance components at extreme time periods became apparent.

Table 2. Correlation between the predicted and realized sum of eggs laid in 2-wk periods 1 to 23 and average total egg production of the 15 and 30 top or bottom births as ordered by pedigree- (A) or genomic- (G) based model, using linear (lin) or quadratic (quad) polynomial to model genetic variance without or with (sibs), including production records of sibs of selection candidates in training

Item	A lin	G lin	A quad	G quad	A quad sibs	G quad sibs
Correlation	0.10	0.24	0.16	0.26	0.15	0.28
Top 15	226.53	211.98	214.98	227.34	230.02	232.72
Top 30	224.11	220.14	221.83	230.82	225.66	233.09
Bottom 30	211.08	206.64	216.17	208.05	216.92	208.12
Bottom 15	215.45	200.83	214.99	208.86	212.11	198.74

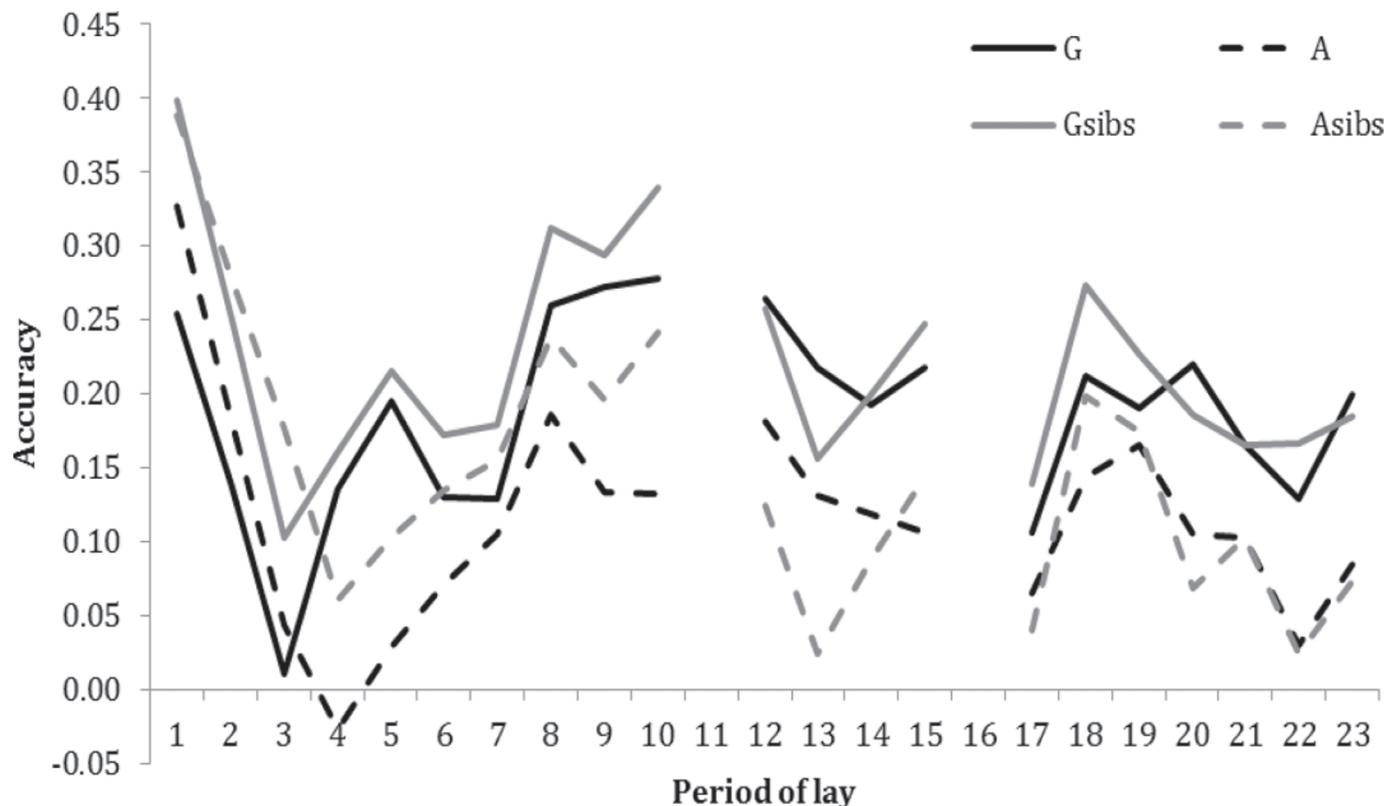


Figure 2. Correlation between predicted breeding values and phenotypes in the validation data set (progeny of animals in training). Predictions were based on a random regression model using pedigree (A) or genomic (G) relationships without or with data on sibs of selection candidates (sibs). There were no data in periods 11 and 16.

This problem was partly overcome by increasing the order of polynomials, which indicates that the order of polynomial must be carefully chosen for each population and source of information.

There are 2 main limiting factors for the practical application of genomic RRM—genotyping costs and computing time. Although genotyping costs have been greatly reduced in recent years, the use of high-density panels is still not economically justified, particularly in poultry, where genotyping cost surpasses the cost of the individual birds. Also, the use of a sire as parent is limited by fresh semen or natural mating potential because cryopreservation of semen has not been successful, and genotyping cost must be recovered from a limited number of descendants of individuals with exceptionally desirable EBV, in contrast to many other mammalian livestock species. Research on the utilization of low-density genotyping and imputation (Habier et al., 2009; Huang et al., 2012) seem to offer promise for reducing costs with little loss of accuracy.

The prolonged computing time of the genomic RRM (~2 d) could originate from both the complexity of the proposed model and the inverse of the genomic relationship matrix being dense compared with the sparse inverse of the pedigree-based numerator-relationship matrix. The complexity of the RRM is not a problem in a pedigree-based analysis because the set of mixed model equations remains sparse. This, however, is not

the case in any genome-based analysis because the mixed model equations are dense due to the dense inverse of the genomic relationship matrix. In the RRM, which has several genetic terms such as the intercept, linear and quadratic effects, the size of the dense part of the mixed model equations is larger than in a model with one genetic term. Therefore, the genomic RRM may become infeasible to implement for a large number of genotyped individuals, unless extensive computing resources are available or more efficient algorithms are developed.

In conclusion, the random regression reduced animal model can be used in breeding programs based on genomic information, leading to substantial improvements of accuracy of EBV compared with pedigree-based approaches. Practical cost-effective implementation will depend upon minimizing computing and genotyping costs, both of which are reducing at a high rate. Therefore, feasible implementation could be foreseen in the near future.

ACKNOWLEDGMENTS

This study was supported by Hy-Line International, the Erich Wesjohann Group GmbH, Visbek, Germany, and Agriculture and Food Research Initiative competitive grants 2009-35205-05100 and 2010-65205-20341

from the USDA National Institute of Food and Agriculture Animal Genome Program.

REFERENCES

- Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.
- Anang, A., N. Mielenz, and L. Schüler. 2002. Monthly model for genetic evaluation of laying hens. II. Random regression. *Br. Poult. Sci.* 43:384–390.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2008. ASReml User Guide Release 3.0. VSN Int. Ltd., Hemel Hempstead, UK.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2009. Genomic selection using low-density marker panels. *Genetics* 182:343–353.
- Huang, Y., J. M. Hickey, M. A. Cleveland, and C. Maltecca. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genet. Sel. Evol.* 44:25.
- Kranis, A., G. Su, D. Sorensen, and J. A. Woolliams. 2007. The application of random regression models in the genetic analysis of monthly egg production in turkeys and a comparison with alternative longitudinal models. *Poult. Sci.* 86:470–475.
- Luo, P. T., R. Q. Yang, and N. Yang. 2007. Estimation of genetic parameters for cumulative egg numbers in a broiler dam line by using a random regression model. *Poult. Sci.* 86:30–36.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, S. J. Lamont, and J. C. M. Dekkers. 2011a. Persistence of accuracy of estimated breeding values in layers using marker and pedigree based relationship matrices. *Genet. Sel. Evol.* 43:23.
- Wolc, A., J. Arango, P. Settar, N. O’Sullivan, and J. C. M. Dekkers. 2011b. Evaluation of egg production in layers using random regression models. *Poult. Sci.* 90:30–34.
- Wolc, A., C. Stricker, J. Arango, P. Settar, J. E. Fulton, N. O’Sullivan, D. Habier, R. Fernando, D. J. Garrick, S. J. Lamont, and J. C. M. Dekkers. 2011c. Breeding value prediction for production traits in layers using pedigree and marker based methods. *Genet. Sel. Evol.* 43:5.
- Wolc, A., and T. Szwaczkowski. 2009. Genetic evaluation of laying hens based on random regression models. *J. Appl. Genet.* 50:41–46.