Prediction Accuracy of Pedigree and Genomic Estimated Breeding Values over Generations in Layer Chickens

Z.-Q. Weng[1], A. Wolc[1,2], R. L. Fernando[1], J. C.M. Dekkers[1], J. Arango[2], P. Settar[2], J. E. Fulton[2], N. P. O'Sullivan[2] and D. J. Garrick[1].
[1]Iowa State University, Ames, IA, USA, [2]Hy-Line International, Dallas Center, IA, USA

ABSTRACT: Theoretically, the more data in the training set, the better the accuracy of the predicted breeding values. However, adding distant generations to the training data will introduce computational burden, with perhaps limited contributions to prediction. The objectives of this study were to compare the accuracy of marker-based and pedigree-based models and to evaluate the optimum number of training generations required to most accurately predict EBV in a commercial layer breeding line. On average, accuracies of EBV based on markers were higher than accuracies based on pedigree. Accuracies of all methods initially increased with successive increases in the number of generations of training data, but slightly dropped or reached an asymptote when including training generations far apart from validation. The divergence in gene frequencies in each generation, genotype by environment interactions, and selection over generations might be the causes of these decreases in accuracy.
Keywords: prediction accuracy; estimated breeding value; layer chicken

Introduction

Pedigree-derived relationships can be used to predict breeding values (EBV). Genomic estimated breeding values (GEBV) are predicted by genomic approaches which rely on genetic markers, such as single nucleotide polymorphisms (SNPs). Selection and mating strategies based on GEBV can enhance the rate of genetic improvement compared to mass selection, and control inbreeding. The accuracy of genomic prediction is influenced by marker density (Meuwissen et al. (2001)), linkage disequilibrium (LD) and linkage between markers and quantitative trait loci (QTL) (Habier et al. (2007)), distribution of QTL effects (Meuwissen et al. (2001)), size of training population (VanRaden et al. (2009)), architecture of trait (Daetwyler et al. (2010)), effective population size (Hayes et al. (2009)), structure of relationships between training and validation individuals (Habier et al. (2010)), and other factors (Goddard (2009)). The accuracy of Pedigree-based Best Linear Unbiased Prediction (PBLUP) relies on heritabilities and additive genetic relationships among individuals.

One challenge of breeding value prediction is utilizing information from non-genotyped animals with phenotypes. The two-step approach (Garrick et al. (2009)) and modifications to the reduced animal model (Wolc et al. (2011)) to account for non genotyped offspring are applicable solutions. It has typically been assumed that the more training data that are available, the better the accuracy of genomic prediction (Goddard (2009)). Further, Goddard (2009) pointed out that the selection response of genomic selection is expected to decline in comparison with phenotypic selection, if no new training information is provided over successive generations. GEBV can capture both relationships between markers and QTL, and relationships between animals, which is expected to lead to better persistence of accuracy than EBV (Habier et al. (2007)). Wolc et al. (2011) confirmed that GEBV are more persistent than pedigree-based EBV in layer chickens. The objectives of this study were to compare the accuracy of marker-based and pedigree-based prediction models, to examine the value of a family mean (FM) model which uses phenotypes from non-genotyped offspring of genotyped parents, and to quantify the increase in accuracy with additional training generations in a commercial layer breeding line.

Materials and Methods

Data comprised phenotypic records from 17,793 birds born between 2002 and 2011. Among those, 2,723 birds over 9 generations were genotyped with a 42K SNP panel (Illumina). After removing SNPs with call rate <0.95, minor allele frequency <0.025 or parentage probability <0.95, only 23,356 segregating SNPs remained across 28 autosomes. Records of 4 traits were analyzed: early and late egg color (eCO, lCO), and early and late egg weights (eEW, lEW). Early (late) measurements were taken at 26-28 (42-46) weeks. In total, there were 16,018 records for early traits, and 11,915 records for late traits. Estimates of heritability from single-trait pedigree-based animal models fitted using ASREML3.0 (Gilmour et al. (2009)) are in Table 1.

Table 1. Estimates of pedigree-based heritabilities for 4 traits from single-trait animal models

Trait[1]	h^2	Standard error
eCO	0.71	0.017
eEW	0.69	0.017
lCO	0.68	0.025
lEW	0.61	0.026

[1]early (e) and late (l) CO (egg color), and EW (average weight of 3-5 eggs).

Three models were used for predicting breeding values. 1) Animal model using pedigree relationships (PBLUP) with available phenotype records, performed using ASREML3.0 (Gilmour et al. (2009)). 2) BayesB (Meuwissen et al. (2001)) with genotyped individual records,

performed using the GenSel4.0 software (Fernando and Garrick (2013)). BayesB assumes a fraction π of markers have zero effects, and locus-specific variance for each SNP with a scaled inverse Chi-Square prior distribution for the variances. Parameter π was assumed to be 0.95. The chain length was 33,000 iterations, of which the first 3,000 were discarded for burn-in. 3) BayesB with both genotyped individual records and family means (BayesB-FM). To exploit information from non-genotyped birds, the average genotype of their genotyped parents and the non-genotyped full-sib mean phenotype were calculated. Weighting factors for the residual variance of the family mean records were $\frac{1-h^2}{(1-0.5h^2)/p}$, where p is the number of records included in the family mean (Garrick et al. (2009)). All methods accounted for fixed effects of hatch within generation. Different validation sets were represented by every generation from 2006 to 2011. EBV in validation sets were estimated based on different numbers of generations immediately preceding the validation generation. The prediction accuracy was the correlation between EBV and hatch-corrected phenotypes in the validation data, divided by the square root of the trait heritability. The expected accuracy of PBLUP was calculated as $\sqrt{1 - PEV/\sigma_g^2}$, where PEV is the prediction error variance, and σ_g^2 is genetic variance. The PEV was obtained from elements of the matrix C^{22}, representing the animal partition of the inverse coefficient matrix of the mixed model equations (Henderson (1975)). The expected accuracy of genomic prediction was calculated according to the equations as follows (Goddard et al. (2011)):

$$r_{MBV} = q\sqrt{\frac{\theta}{1+\theta}}$$

$$q^2 = \frac{M}{M + M_e}$$

$$M_e = 2N_eLk/\log(N_eL)$$

$$\theta = Tq^2h^2/M_e$$

where is M is number of markers (23,356 SNPs), M_e is the effective number of chromosome segments in the population, N_e is the effective population size, L is the average length of chromosome in Morgan (~1.07 M), k is the number of chromosomes (k=28), T is the size of training population (individuals with genotype records), and q^2h^2 is marker-based heritability. Because the N_e varied across generations, N_e =150 was used as an approximation.

Results and Discussion

Figure 1 shows prediction accuracies of EBV across different validation sets for the average of the 4 traits. Marker-based methods outperformed the pedigree-based method, which indicated that markers not only capture relationships between individuals but also QTL information through LD or linkage (Habier et al. (2007)). Adding information through family means resulted in BayesB-FM having higher accuracy than BayesB using only genotyped individual records, which implies the family mean model is a practical method to implement information from non-genotyped non parent animals. Expected and empirical accuracies of PBLUP quickly plateaued with an increasing number of training generations. However, the expected accuracy of PBLUP overestimated the empirical value, which might be caused by continuous selection over many generations. Selection induces the Bulmer effect (Bulmer (1971)), which is not taken into account in the calculation of expected accuracy in PBLUP.

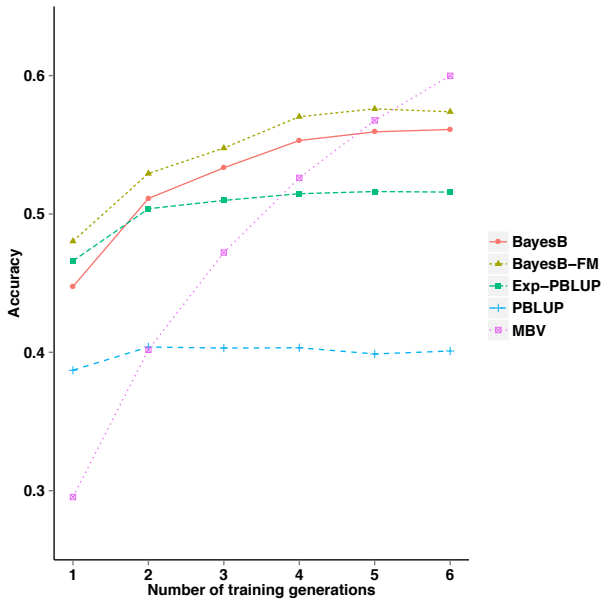

Figure 1: Average accuracies across 4 traits of predicted breeding values based on the number of training generations when using a pedigree-based model (PBLUP), BayesB with genotyped individual records (BayesB), or BayesB with genotyped individual and family mean records (BayesB-FM), the expected accuracy of PBLUP (Exp-PBLUP), and the expected accuracy of GEBV (MBV).

Empirical and expected accuracy of genomic predictions increased quickly with increasing number of training generations, but empirical accuracy slightly dropped or plateaued when adding more distant generations (Figure 1). The increase of training generations is analogous to increasing size of the training population, which is expected to lead to an increase in accuracy (Meuwissen et al. (2001)). Because the calculation of expected accuracy doesn't consider the relationship between training and validation sets, it was lower than empirical value at the beginning. Distant generations did not contribute to accurate genomic predictions, therefore empirical accuracy of genomic prediction was actually lower than the expected value when training sets exceeded 5 generations. In this population, conventional selection using pedigree and phenotypes was conducted before 2009, while genomic information has been implemented in the selection program since 2009. The transition of the selection program might introduce a reduction in

accuracy of GEBV. On the other hand, the expected accuracy of GEBV ignored the impact from population structure and generational selection which would lead to a higher accuracy than the empirical value. The reductions in prediction accuracy when all generations were used for training were not expected, though these reductions were not significant. The reductions might have been caused by divergent gene frequencies in each selected generation, the influence of environmental effects, or interactions between genotype and environment. Results differed significantly between validation sets (not shown). Different family structures, selection strategies, and sampling processes may be the causes of disparities between validation sets.

Conclusion

This study investigated accuracies of EBV for different numbers of training generations in layer chickens using pedigree or marker-based models. The prediction accuracy of EBV using markers was higher than using pedigree information. The BayesB-FM model outperformed BayesB model, because the family mean model utilized information from non-genotyped individuals. In general, the prediction accuracy increased with an increase in the number of training generations. However, it slightly decreased or became asymptotic when including distant generations in training sets. Based on this data, using 5 generations in training seems optimal for GEBV prediction, although results differed between validation sets.

Literature Cited

Bulmer, M.G. (1971). Amer. Nat. 105:201–211.

Daetwyler, H.D., Pong-Wong, R., Villanueva, B. et al. (2010). Genetics 185: 1021-1031.

Fernando, R.L., and Garrick, D. J. (2013). 237–274 in Genome-Wide Association Studies and Genomic Prediction. Humana Press, Totowa, NJ.

Garrick, D.J., Taylor, J.F., and Fernando. R.L. (2009). Genet. Sel. Evol. 41:55.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. et al. (2009). VSN Int Ltd, Hemel Hempstead, UK.

Goddard, M. (2009). Genetica. 136:245–57.

Goddard, M., Hayes, B.J., Meuwissen, T.H.E. (2011). J. Anim. Breed. Genet. 128:409–421.

Habier, D., Fernando, R.L., and Dekkers, J.C.M. (2007). Genetics. 177:2389–97.

Habier, D., Tetens, J., Seefried, F.R. et al. (2010). Genet. Sel. Evol. 42:5.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., and Goddard, M.E. (2009). J. Dairy Sci. 92: 433-443.

Henderson, C.R. (1975). Biometrics. 31:423–447.

Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. (2001). Genetics. 157:1819–29.

Quaas, R. L., and Pollak, E. J. (1980). J. Anim. Sci. 51:1277–1287.

VanRanden, P., Tassell, C.V., Wiggans, G. et al. (2009). J. Dairy Sci. 92: 16-24.

Wolc, A., Arango, J., Settar, P. et al. (2011). Genet. Sel. Evol. 43:23.