Application of low-pass sequencing to genomic prediction of egg quality in laying hens

A. Wolc[1,2*], J. Li[3], D. Lubritz[2], J. Arango[2], J. Fulton[2], P. Settar[2], K. Rowland[2] and H. Cheng[3]

[1] Department of Animal Science, Iowa State University, Ames, IA 50011, USA; [2] Hy-Line International, Dallas Center, IA 50063, USA; [3] Department of Animal Science, University of California, Davis, CA 95616, USA; [*]awolc@iastate.edu

Abstract

Accuracy of genomic prediction was compared in two layer lines using medium density SNP chip and low-pass sequencing. Three thousand individuals representing two lines were sequenced for the project. Thirty-four genomic regions were identified as genome-wide significant across 11 egg quality traits and their genotypes were added to the chip. Multi-trait single-step animal model was used for genomic prediction. A slight increase in accuracy for some of the traits was obtained by adding the significant SNPs to those present on the chip. Fitting all sequence SNPs had no advantage for predictive ability over chip despite explaining 15% more of phenotypic variance in the Genome Wide Association Study. Further optimization of the utilization of sequence data and cost reduction are required to make low-pass sequence competitive to medium density SNP chips for genomic prediction.

Introduction

Genomic selection has been applied to poultry breeding for a decade providing benefits of increased accuracy of estimated breeding values and shortening of generation intervals. With the reduction in cost, implementation evolved from a combination of low- and high-density panels to population specific medium density panels (50-60,000 SNPs). With the development in sequencing technology the costs of obtaining whole genome sequences were greatly reduced. Low pass sequencing interrogates all genomic locations in contrast to chip genotyping which only considers those SNPs included on the chip. Low pass sequencing comes at a lower cost than standard sequencing but also at a lower accuracy of genotype calls than standard sequencing. Rubinacci et al. (2021) notes that the accuracy of genotype calls from low pass sequencing can be recovered by combining the genotype calling with imputation similar to the methodology used to combine information from low- and high-density chips. If relevant reference populations are used accurate genotype calls can be obtained with 1x sequences.
Some studies suggest that low-pass sequencing can improve genome wide association studies relative to chip data (Li et al., 2021). Until now no studies using large scale sequencing were reported in the literature for laying chickens, therefore we investigated the impact of using whole genome sequence data and adding SNPs identified as significant in a Genome Wide Association Study (GWAS) to the 54k chip for genomic prediction.

Materials & Methods

Data

One hundred males from 8 major lines were selected for sequencing at 4x coverage and combined with 124 samples from publicly available sequences to provide reference for genotype calling for low-pass (1x) sequenced individuals. Genotypes for 32 million SNPs were called. Vcf files were provided by Gencove with SNPs called using their proprietary algorithm. One generation of breeding males and two generations of breeding females were sequenced at 1x coverage for a total of 3,025 sequenced individuals, including 1,540 individuals from one

White Leghorn (WL) line and 1,485 individuals from one Rhode Island Red line (RIR). Genotypes of males were used in single-step genomic prediction (they are sires of hens with phenotypes) but not in GWAS (they don't have own phenotypes). After quality control with minor allele frequency > 1%, 8.6 million markers were left in RIR and 7.6 million markers in WL line. Phenotypes are available for all of the sequenced females with more than 30,000 and 40,000 additional phenotyped relatives for RIR and WL line, respectively. Phenotypes used for the analyses include several egg quality traits such as shell strength, egg weight, yolk weight, albumen height and shell colour measured at different ages.

Validation of SNP calls

Accuracy of 1x genotype calls was validated by comparison to Axiom 54k SNP chip genotypes on the same individuals using absolute value of correlation for genotypes called as 0,1,2 and percentage of identical calls, and as percentage of identical calls for animals sequenced in duplicates.

GWAS

GWAS was performed using a mixed model in Genome Wide Complex Trait Analysis software (GCTA; Yang et al., 2011) for WL and RIR lines separately. The following statistical model was used in the GWAS:

$y_i = \mu + b\, x_i + g_i + e_i$

where y_i is the phenotype adjusted by station and hatch effect for individual i; μ is the overall mean; b is the additive effect (covariate) of the candidate SNP to be tested for association; x_i is the SNP genotype indicator of the candidate SNP for individual i coded as 0, 1 or 2; g_i is the random polygenic effect for individual i based on the relationship matrix computed from genotypes and e_i is the residual. No obvious population stratification was observed based on the principle component analysis for both lines, thus it was not considered in the GWAS. The SNPs with p-value < 0.00001 were added to the SNP panel for genomic prediction.

Genomic prediction

212 females from WL line and 228 from RIR line had phenotypes removed when conducting the GWAS and genomic prediction for the purpose of validation. The data were analysed per line with AIREMLF90 software using chip data, sequence data or chip combined with significant SNPs from the sequence for the construction of H matrix.

A series of multi-trait animal models fitting fixed effects of hatch within generation, station processing the eggs, age of birds and random effects of animal and permanent environment were applied for genomic prediction. Variance components were estimated with pedigree and only BLUP was performed using H matrix in order to reduce computational requirements. Accuracy was measured as correlation between the phenotype adjusted for fixed effects and predicted genomic breeding value divided by square root of heritability.

Results

Accuracy of genotyping with low-pass sequencing

Sequence was evaluated on 386 samples by comparing those genotypes to respective allele calls from 54k Axiom SNP Chip (currently a gold standard). Median correlation of the genotypes was 0.97 (SNPs with lower MAF tended to have lower correlation) with 98.4% calls being identical. For 96 test samples sequenced in duplicates, concordance of genotype calls around 98% was obtained compared to >99.5% for duplicates on chip.

Genomic prediction

Based on mixed model in GCTA software using only hens that had both phenotypic and sequence records sequence data explained on average 15% more variance than SNP chip. Across the egg quality traits 43 genomic regions were identified as significant at genome wide level.

In the first step, genomic prediction was performed using SNP chip genotypes. Based on the GWAS results, 7,063 and 7,216 SNPs were added to those present on the chip to conduct another set of genomic prediction. Slight improvement in accuracy was achieved for some of the traits (Table 1). The prediction accuracy based on 54k chip data was close to that based on low-pass sequencing, despite many more SNPs considered with sequencing data.

Table 1. Accuracy of GEBVs using SNP chip data (54k), SNP chip data enhanced with significant SNPs from sequence (54kPlusSig) or complete sequence (Seq) in two layer lines.

	WL line			RIR line		
	54k	54kPlusSig	Seq	54k	54kPlusSig	Seq
AH1	0.77	0.77	0.69	0.56	0.55	0.53
AH2	0.53	0.51	0.48	0.69	0.68	0.70
BS1	0.27	0.29	0.35	0.48	0.41	0.48
BSPO1	0.34	0.38	0.37	0.59	0.55	0.59
BSPO2	0.46	0.53	0.47	0.59	0.56	0.58
CO1	0.67	0.75	0.62	0.62	0.60	0.62
E3	0.65	0.65	0.65	0.50	0.47	0.43
EW1	0.45	0.48	0.45	0.31	0.31	0.30
EW2	0.47	0.48	0.50	0.25	0.24	0.24
YW1	0.51	0.52	0.45	0.3	0.31	0.34
YW2	0.53	0.54	0.52	0.35	0.34	0.32

AH = albumen height, BS = breaking strength, CO = shell color, E3,EW = egg weight, YW=yolk weight. Numbers 1 and 2 refer to data collected at early (around 26wks) and late (around 75wks) age of birds

Discussion

Previous studies (Rubinachi et al 2021, Kim et al 2021, Li et al 2021) used subsampling of high coverage sequence data to evaluate accuracy of low pass sequencing and showed that with exception of low frequency variants accurate calls can be obtained when genotype calling is combined with imputation. We used real samples that were genotyped and sequenced in duplicates under different IDs to validate low-pas sequencing and showed that the quality of information from low-pass sequence is slightly lower than from SNP chip but acceptable for GWAS and genomic prediction.

Mixed results were obtained by using sequence data for genomic prediction with several studies reporting very limited benefits (VanRaden et al., 2016; Heidaritabar et al., 2016). Yang et al. (2010) showed that the genetic variance does not depend on the number of SNPs when correcting for incomplete LD between SNPs, which explains why the genomic prediction based on SNP chip and sequencing data were similar.

Some increases in accuracy were observed when SNPs were weighted differentially based on their annotation (McLeod et al., 2016) or smaller panel was designed by combining sequence

based GWAS with annotation (Xiang et al., 2021). Therefore, the next step will be to employ differential weighting of SNPs and other computational strategies.

Conclusions

Similar or slightly higher accuracy can be obtained in chicken using low pass (1x) sequence data compared to a medium density SNP chip however further optimization is needed to extract more information from the sequence data.

References

Heidaritabar M., Calus M.P., Megens H.J., Vereijken A., Groenen M.A., et al. (2016) J Anim Breed Genet. 133(3):167-79. doi: https://doi.org/10.1111/jbg.12199

Kim, S., Shin, J. Y., Kwon, N. J., Kim, C. U., Kim, C., et al. (2021). Human genomics, 15(1), 58. https://doi.org/10.1186/s40246-021-00357-w

Li J. H., Mazur C. A., Berisa T., and Pickrell, J. K. (2021) Genome Res., 31(4), 529–537. https://doi.org/10.1101/gr.266486.120

MacLeod I.M., Bowman P.J., Vander Jagt C.J., Haile-Mariam M., Kemper K. E. et al. (2016) BMC Genomics. 17:144 https://doi.org/10.1186/s12864-016-2443-6

Rubinacci S., Ribeiro D.M., Hofmeister R. J., and Delaneau, O. (2021) Nature Genet., 53(1), 120–126. https://doi.org/10.1038/s41588-020-00756-0

VanRaden P.M., Tooker M.E., O'Connell J.R., Cole J.B., and Bickhart D.M. (2017). Genet Sel Evol., 49(1), 32. https://doi.org/10.1186/s12711-017-0307-4

Xiang R., MacLeod I.M., Daetwyler H.D. et al. (2021) Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. Nat Commun 12:860. https://doi.org/10.1038/s41467-021-21001-0

Yang J., Benyamin B., McEvoy B. P., Gordon S., Henders A. K. et al. (2010). Nature Genet. 42(7), 565–569. https://doi.org/10.1038/ng.608