Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Applications of Genomic Selection in Poultry

A. Wolc[*,†], **A. Kranis**[‡,§], **J. Arango**[†], **P. Settar**[†], **J.E. Fulton**[†], **N. O'Sullivan**[†], S. **Avendaño**[‡], **K.A. Watson**[‡], **R. Preisinger**[#], **D. Habier**[*], **S.J. Lamont**[*], **R. Fernando**[*], **D.J. Garrick**[*], **J.C.M. Dekkers**[*]

[*]Department of Animal Science, Iowa State University, Ames, USA, [†] Hy-Line International, Dallas Center, USA, [‡]Aviagen Limited, Newbridge, UK, [§] The Roslin Institute, R(D)SVS, University of Edinburgh, Scotland, [#] Lohmann Tierzucht GmbH, Cuxhaven, Germany

ABSTRACT: Here we describe the application of genomic selection in both layer and broiler breeder populations. A brown egg layer line was partitioned into two sub-lines, one used for genomic selection and the other as a control representing pedigree based selection. Generation interval in the genomic sub-line was halved and the sub-line size was reduced compared to the traditionally-selected control. The genomic sub-line outperformed pedigree-selected contemporaries in 12 of 16 traits evaluated, and genomic estimated breeding values were more accurate and persistent than pedigree-based estimates. Genome wide association studies for all available traits identified several regions associated with economically important traits. Similar improvements in prediction accuracy were observed in broilers. Estimation of the Mendelian sampling term for full sibs without own phenotypic information contributed to this gain. The development of robust imputation methods enabled the implementation of genomic selection into the routine evaluations to accelerate genetic progress.

Keywords: Genomic selection; Poultry; GWAS

Introduction

The development of tools for genome analysis (Hillier et al. (2004); Groenen et al. (2009); Groenen et al. (2011); Kranis et al. (2013)) has given poultry breeders access to genomics. This genomic information can be used to improve efficiency of selection by providing more accurate assessment of naturally occurring genetic variation between individuals and associating it with traits of economic interest. Early attempts at marker assisted selection (MAS) started in the 1960s when blood groups (Brilles (1950)) and their association with immune related traits (Hansen et al. (1967)) were discovered. Further discoveries of genetic markers such as Restriction Fragment Length Polymorphisms and Microsatellites provided the means to construct genetic maps and begin studies on localizing genomic regions associated with various traits. Initial excitement from these studies was somewhat reduced by the fact that the regions showing significant associations were very large, up to whole chromosomes, and thus included thousands of genes which could be involved in determining the traits of interest, with no tool to allow easy interpretation of the genomic information. The disappointing early MAS results were likely due to limited availability of markers, hence low genotyping resolution, and high genotyping costs, which limited the size of discovery populations. These studies were performed on crosses of animals with extreme differences in phenotypes for the traits of interest to maximize the chance of finding significant regions. Markers significant in these crosses were often already fixed by selection in commercial lines.

A new opportunity arose when in 2004 the chicken became the first livestock species sequenced (Hillier et al. (2004)). Sequencing revealed millions of single nucleotide polymorphisms (SNPs), which covered all major chromosomes. Sequence enabled development of SNP panels, which made high throughput genotyping of tens or hundreds of thousands markers possible (Avendaño et al. (2010); Groenen et al. (2011); and Kranis et al. (2013)). The cost of determining a genotype at a single locus dropped from about $1.50 in 2010 for a microsatellite marker to less than $0.0005 per SNP on a high density SNP chip (Fulton, pers. com.).

In parallel to advances in genotyping technology, statistical and computational methods were developed to capitalize on continuously increasing amounts of data. Meuwissen et al. (2001) proposed that, instead of localizing QTL regions, information from the whole genome should be used to estimate breeding values. Subsequently, several genomic prediction models were developed (see review by Gianola (2013)) and applied to simulated and real data (Habier et al. (2011)). Based on simulation studies, genomic selection in poultry was shown to have the potential to provide increased

accuracy of selection, reduced generation intervals and better control of inbreeding (Dekkers et al. (2010)). Experiences with real data brought genomics in the poultry industry from being an 'utopic objective' to envisaging deliverables (Avendaño et al. (2010)) and further into consolidation stages (Avendaño et al. (2012)). Currently, thanks to the short generation interval, the chicken is the only livestock species for which information on multiple generations of genomic selection is available. In this paper we share experiences from a multigenerational genomic selection experiment performed by Hy-Line International and Iowa State University, as well as details of implementation of genomic selection in both layer and broiler industries.

Lessons learned from a multi-generation genomic selection experiment

In order to quantify gains from genomic selection, a brown egg layer line was split into a pedigree control and a genomic sub-line, with parallel selection programs. The size of the pedigree sub-line was reduced compared to the original line (down to 60 males selected from 1,000 candidates and 360 females selected from 2,000 female candidates) but continued with traditional pedigree-based BLUP selection for an index of 16 traits (reflecting egg production and quality), in 13 month cycles, and a nested mating structure. The genomic sub-line size was reduced to 50 males and 50 females selected from totals of 300 male and 300 female candidates to reduce costs of genotyping, but without increasing the expected rate of inbreeding per year. The generation interval was reduced to 7 months, the mating structure was changed to allow each female to produce progeny with 10 different males (i.e. cross classified mating), and selection was based on the same index as in the pedigree sub-line but included genomic information in breeding value estimation. The training data for genomic selection consisted of 5 prior generations of all selected individuals in the original line. Substitution effects of SNPs were re-estimated every generation as new phenotypes on genotyped individuals became available.

Substantial variability in accuracy between generations was observed. Models that showed highest validation accuracy in the previous generation were chosen to derive genomic EBV for selection in the current generation. Different models for breeding value estimation were tested (including GBLUP, BayesB, BayesCPi and their modifications) but no single method consistently showed an advantage over other methods (Figure 1). However for egg weight traits, which were affected by a large QTL, Bayesian variable selection models tended to outperform GBLUP, whereas for other traits GBLUP had similar or in some validation sets higher accuracy. An advantage of the GBLUP model was observed in terms of the range of models that can be utilized, including random regression models for egg production (Wolc et al. (2013b)), which are not available in current software for Bayesian analysis. All models using genomic information tended to have higher accuracy than pedigree based models (Wolc et al. (2011a)). For the BayesB model, it was observed that increasing the proportion of markers fitted in the model did not negatively affect accuracy of predictions, but fitting too few markers sometimes did. For some traits, reducing the weight on, or even removing information from distantly related individuals, increased accuracy (Wolc et al. (2013)). A study on Marek's disease resistance (Wolc et al. (2013a)) showed that proper weighting of phenotypes is essential for accurate breeding value estimation, and that at least for that trait, removing markers from regions of the genome that do not show association with the trait may increase accuracy.

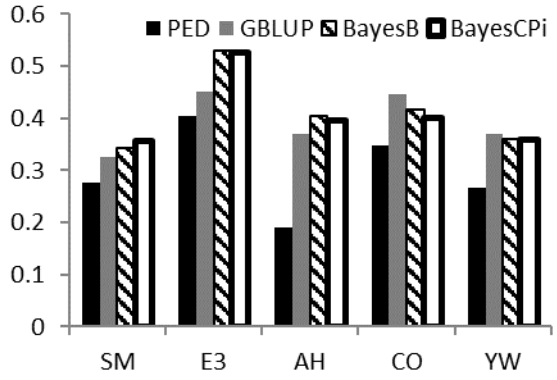

Figure 1. Validation accuracies of EBV obtained from pedigree (PED), GBLUP, BayesB, and BayesCPi methods for 5 traits: sexual maturity (SM), weight of first 3 eggs (E3), albumen height (AH), shell color (CO), and yolk weight (YW).

Persistency of accuracy over generations without retraining was higher for genomic EBV than for pedigree-based EBV, which suggests that genomic EBV captured linkage disequilibrium with QTL, in addition to relationships (Wolc et al. (2011b)). However, there was a substantial drop in accuracy from prediction of progeny to grand progeny of the last training generation, which led to the conclusion that phenotyping cannot be discontinued and constant retraining will be necessary to maintain accuracy of selection.

In the experiment, all genotyping was performed using a proprietary EW Group 42K Illumina SNP panel (Avendaño et al. (2010)). On a commercial scale, high-density genotyping all selection candidates would be economically inefficient, thus several scenarios using reduced SNP panels were evaluated. It was found that if both parents were high-density genotyped, correlations above 0.97 between true and imputed genotypes could be achieved when the low density panel contained as few as 400 of the 42,000 high-density SNPs but correlations dropped below 0.95 when the females in the pedigree were only low-density genotyped (Wolc et al. (2011c)).

Genome wide association studies (GWAS) were performed for all traits in the selection index (Wolc et al. (2012); Wolc et al. (2014)) and for feed efficiency (Wolc et al. (2013c)) and Marek's disease resistance (Wolc et al. (2013a)). Chromosomes 1, 2 and 4 were found to have the largest numbers of regions associated with the analyzed traits. Except for a large QTL on chromosome 4, which was associated with multiple traits related to egg weight, other traits showed polygenic models of inheritance. Some consistently significant regions were found for measures of egg quality at different ages (Wolc et al. (2014)).

In the final generation of the genomic selection experiment (3rd generation of pedigree sub-line, 5th generation of genomic sub-line), the two sub-lines were hatched together and raised in the same barn. For 12 out of 16 traits, the genomic sub-line significantly outperformed the pedigree sub-line, however for some of the lower heritability traits (h^2<0.3) accurate estimation of genomic breeding values remained problematic. Greater genetic progress in the genomic selection sub-line originated from shorter generation intervals and greater accuracy of selection of males (Wolc et al. (2011a)).

Implementation of genomic selection in layers

Based on the promising results from the selection experiment, genomic selection was implemented in commercial lines of layers using a newly developed 600K SNP chip (Kranis et al. (2013)) and custom designed low- and moderate-density SNP sets. With as few as 1,000 high quality SNPs and strategic high-density genotyping, a high accuracy of imputation using pedigree-based methods could be achieved. Genomic selection models have shown improvements in accuracy over pedigree-based analysis, but they don't solve the problems of low accuracy for traits with low heritability and limited number of records. The first genomic selected birds will generate descendants that will enter the market as commercial birds in 2015.

Experience with genomic selection in broilers

History

Very soon after the release of the draft chicken genome sequence (Hillier et al. (2004)), Aviagen started developing its first SNP panel. Thanks to the rapid technological advances, the chip density increased from 6K (Andreescu et al. (2007)), to 12K (Powel et al. (2011)), 42K (Wang et al. (2013)) and ultimately to 600K SNPs (Kranis et al. (2013)). Similarly, the number of available genotypes for analysis has increased from two hundred individuals per line at the initial phase of the project to more than fifty thousand birds accumulated since 2012, when routine implementation started in Aviagen.

Methodology

The increasing numbers of genotyped animals and markers per sample had a profound effect on the methods applied for utilizing genomic information in breeding programs. Originally the objective was to identify major QTL, capitalizing on historical LD detected from a GWAS, and then to implement a marker-assisted selection scheme. However, it proved difficult to detect and validate QTL explaining a large proportion of genetic variance for the main quantitative traits of economic importance for broiler breeding. As the idea of fitting all markers simultaneously to estimate genomic breeding values (GEBVs) was gaining traction, the research was directed to the development of a sustainable and cost efficient strategy to implement genomic selection for routine evaluations.

Imputation

The main challenge was the prohibitive cost of large-scale genotyping due to the large number of selection candidates and, despite a reduction in the cost per SNP, the overall price per selection candidate genotyped was relative stable since the density was increasing. The development of low-density and imputation strategies for genomic selection offered a viable solution to the problem, where most animals are genotyped with a sparse panel comprised of equally spaced markers (Habier et al. (2009)). To implement this approach, a robust imputation algorithm was required. Both a peeling algorithm employing a Gibbs sampler (Wang et al (2013)) and a heuristic approach implemented in the software program AlphaImpute (Hickey et al. (2012)) were investigated. In both cases the accuracy of imputation,

measured as the correlation between imputed and real high-density genotype, was around 0.97. AlphaImpute was also extended to accommodate the sex chromosome (Hickey et al. (2013)). Systematic monitoring (Figure 2), shows that large-scale imputation is feasible and robust, enabling the implementation of genomic selection for routine evaluations in elite broiler lines, without compromising the accuracy of GEBVs (Wang et al. (2013)).

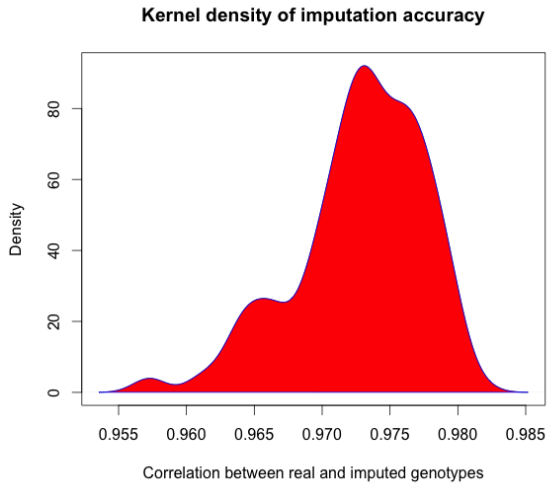

Figure 2. Density plot of the imputation accuracy measured as the correlation between true and imputed genotypes.

Accuracy

One of the main promises of incorporating genomic information is the improved prediction accuracy for traits for which phenotypes are not available for candidate animals. For such traits, in a univariate analysis, estimation of the Mendelian sampling term is impossible with pedigree-based BLUP and thus, the same EBV is assigned to full-sibs. Unless information is available on strongly correlated traits for these individuals, the ability to distinguish between full-sibs is impaired. However, with genomics, it is feasible to accurately estimate Mendelian sampling terms without necessarily requiring a multivariate model. This is illustrated in Figure 3, where parental averages from conventional BLUP are plotted against GEBVs from a ridge regression for a group of 411 individuals, which for the purposes of the analysis were assumed not to have phenotypes on body weight. Two families are highlighted in the graph. For the first one on the left side of the chart, all individuals were ranked as below average and hence, most likely none of them would be selected. Nevertheless, if the GEBVs are considered (red upper triangles in the graph), some of these full-sibs may now have qualified for selection, as they rank high. The opposite applies for the second family (depicted on the right side of the graph) for which, although the parental average is very high, some of the individuals rank quite low for GEBVs. As all of these 411 testing individuals had at least 25 offspring with body weight, it was possible to estimate adjusted progeny means, equivalent to de-regressed EBVs used in dairy breeding (depicted as green hexagons in the plot). The fact that these also exhibit the same variability is evidence that the spread of GEBVs was not an artifact and does estimate the Mendelian sampling term.

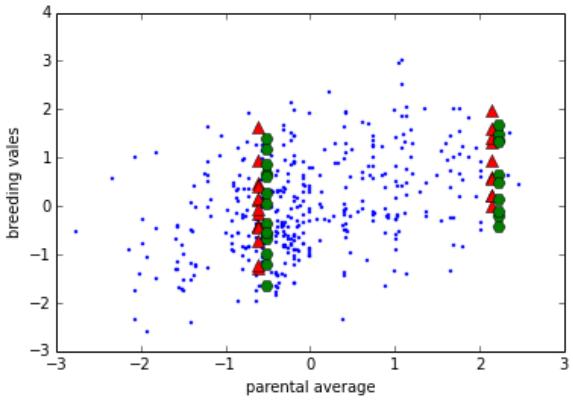

Figure 3. Relationship between breeding values accounting for the Mendelian sampling term (red triangles: GEBVs, green hexagons: adjusted progeny means) and parental average (PA) from a pedigree-based BLUP. Two families are shown (one with low PA on the left and one with high PA on the right). All sibs have the same PA, however there is significant variation in GEBVs, suggesting some sibs rank highly in the cohort. This observation was validated using adjusted progeny means for the full sibs of the two families.

Further illustration of the relative advantage in prediction accuracy of genomic selection over a traditional pedigree-based approach is shown in Figure 4. The comparison included pedigree-based estimates from multivariate BLUP and univariate GEBVs. The relative improvement from implementation of genomic selection in terms of selection accuracy, measured as the correlation between phenotype adjusted for fixed effects and pedigree/genomic EBV at the point of selection, when animals had no phenotypic records, range between 20% and 70% (Figure 4).

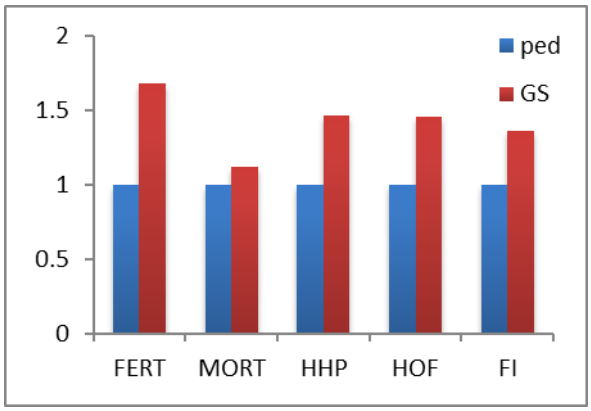

Figure 4. Relative improvement in prediction accuracy of genomic selection (GS) over pedigree-based (Ped) EBVs, measured as the correlation of EBVs with adjusted phenotype for 5 traits: fertility % (FERT), laying mortality (MORT), hen-housed egg production (HHP), hatchability % (HOF) and feed intake (FI).

With the fast accumulation of genotypes in a commercial breeding program, we already have entered the "big data" territory and thus, more sophisticated tools are required to cope with the high dimensionality of the predictions. One of the factors that increases computational requirements is the number of markers considered in the analysis. As very dense SNP panels and soon sequence data will become widely available, it appears that there is a trade-off in marginal gains in accuracy as a function of number of predictors and SNP densities beyond 100K seem to offer no additional benefits for accuracy within the same population, at least with the current methods (Abdollahi-Arpanahi et al. (2014)).

While for quantitative traits that have been selected over many generations the whole genome approach appears the best strategy, breeding objectives are enriched with more traits (Neeteson-van Nieuwenhoven et al. (2013)) that may have different genetic architecture. As the quality of the sequence assembly and annotation status improves, strategies capitalising on this prior biological knowledge may further improve prediction accuracy (Morota et al. (2014)).

Outlook

We have been able to successfully implement, and show a significant advantage for genomic selection in broiler selections using imputation from low and medium density SNP chips. However, reducing the cost of genotyping remains a constant challenge. Genotyping by sequencing (G-by-S) offers the potential to exploit the dramatic reduction in sequencing costs, as it uses low density sequence coverage to identify SNPs in selection candidates. The SNP data are then used for two benefits, firstly as a low cost genotyping method to enable imputation; and secondly to identify rare polymorphisms/QTN from large scale GWAS, which have the potential to improve training and therefore the accuracy of genomic evaluation. The utility for G-by-S in broiler breeding remains unproven; however there is significant excitement for this methodology in the plant breeding world to warrant further evaluation.

Since genomic selection is now part of routine evaluations for broilers, a natural progression is to expand into other species. Recent advances in the sequencing of turkeys (Dalloui (2010)) offer the necessary foundation to implement genomic selection to achieve faster genetic progress for turkeys.

Literature cited

Abdollahi-Arpanahi R, Pakdel A, Nejati-Javaremi A, et al. (2014). J. Anim. Breed. Genet. 15:109.
Andreescu C., Avendaño S., Brown S. R. et al. (2007). Genetics. 177:2161-2169.
Avendaño S., Watson K. A., Kranis A. (2010). Proc 9th WCGALP, 0049.
Avendaño S., Watson K.A., Kranis A. (2012). Proc 24th World Poult. Congress.
Briles, W. E. (1950). Genetics 35:656-657.
Dalloui R. A., Long J. A., Zimin A. V. et al. (2010). PLOS Biol 8
Dekkers, J.C.M, Stricker C., Fernando R.L., et al. (2010). J. Anim. Sci. 87 E Suppl. 2.
Gianola D. (2013). Genetics, 194:573-596.
Groenen, M. A. M., Wahlberg P., Foglio M. et al. (2009). Genome Res. 19:510–519.
Groenen M.A.M., Megens H.-J., Zare Y. et al. (2011). BMC Genomics, 12:274.
Habier,D., Fernando,R.L., and Dekkers,J.C. (2009) Genetics 182:343
Habier D, Fernando, R.L., Kizilkaya, K., et al. (2011). BMC Bioinformatics, 12:186.
Hansen, M. P., Van Zandt, J. N., and Law, G. R. J. (1967). Poult. Sci., 46:1268.
Hickey J.M., Kinghorn B.P., Tier B. et al (2012). Gen Sel Evol 44:9
Hickey J.M. and Kranis A. (2013). Gen Sel Evol 45:10
Hillier L.W., Miller W., Birney Eet al. (2004). Nature 432:695-716.
Kranis A., Gheyas A.A., Boschiero C., et al. (2013). BMC Genomics, 14:59.
Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E. (2001). Genet 157:1819-1829.

Morota G, Abdollahi-Arpanahi R, Kranis A, et al. (2014). J. Anim Breed 15:109
Neeteson-van Nieuwenhoven A-M, Knap P., Avendaño S. (2013). Anim. Front 3:52-57
Powell J.E., Kranis A., Floyd J., et al. (2011). Anim Gen
Wang C., Habier D., Peiris B.L. (2013). Pout Sci 92:1712-23
Wolc A., C. Stricker, J. Arango, et al., (2011a). GSE, 43:5.
Wolc A., Arango J., Settar P., et al. (2011b). GSE, 43:23.
Wolc A., Arango J., Settar P., et al. (2011c). Proc 7th European Symp. Poult. Genet.
Wolc A., Arango J., Settar P., et al. (2012). Anim. Genet. 43:87-96.
Wolc A., Arango J., Jankowski T., et al. (2013a). Avian Dis. 57:395–400.
Wolc A., Arango J., Settar P., et al. (2013b). Poult. Sci. 92:1486-1491.
Wolc A., Arango J., Jankowski T., et al. 2013c. Poult. Sci. 92:2270-2275.
Wolc A., Arango J., Settar P., et al. (2013). 21st PAG
Wolc A., Arango J., Jankowski T., et al. (2014). J. Anim. Breed. Genet (in Press)