Genomics In Poultry Breeding – From Utopias To Deliverables

S. Avendaño, K.A. Watson, A. Kranis[*]

Introduction

The main objective of a poultry breeder is to deliver products to perform competitively for a broad number of traits ranging from live performance, health, and welfare, in a wide range of production environments. Selective breeding based on combining phenotypes and pedigrees and estimation of breeding values using Best Linear Unbiased Predictions (BLUP) has been widely used in poultry breeding and there is evidence of the benefits translated to the industry (Laughlin, 2009). Genomics postulates further opportunities through extra accuracy of selection, reduction of generation intervals and exploitation of new sources of genetic variation (Dekkers, 2004). While up to 700 QTL have been reported in poultry (Abasht *et al*., 2006), there is no evidence of any implementation for routine breeding selection (Dekkers, 2004; de Koning and Hocking, 2007; Nieuwhof *et al*., 2008). Since 2004, with the release of the first draft of the chicken genome sequence (International Chicken Genome Sequencing Consortium, 2004) and the availability of between 3 million SNPs (International Chicken Polymorphism Map Consortium, 2004) and up to 7 million SNPs (Rubin *et al*., 2010) in the public domain, there has been a notorious expansion in SNP panel development, combining public and proprietary information. This has been coupled with an increasing availability of high throughput genotyping platforms and genome sequencing services and the development of a wide array of theoretical methods for incorporating genomics information in routine breeding programs. Nevertheless, the question of whether the postulated benefits of the use of genomics information in poultry breeding will translate, in a cost effective way, in product performance advantages is still open. This paper will discuss the current state of genomics technology in poultry and present the approaches taken by global layer and broiler breeding companies within the EW Group to incorporate genomics approaches in breeding programs.

SNP panel development

The release of the draft sequence of the chicken genome (International Chicken Genome Sequencing Consortium, 2004) provided the key to allow the poultry breeding industry to actively pursue the implementation of genomic selection (GS). Indeed, poultry breeders have led the development of high density (HD) SNP panels with a focus on describing the molecular basis of quantitative variation in commercial populations. In contrast with other livestock populations, to date, no commercial avian SNP panel is available.

[*] Aviagen Limited, Newbridge, EH28 8SZ, Midlothian, Scotland, UK.

Aviagen Ltd developed its first high density SNP in 2005, soon after the release of the draft chicken genome sequence and since then other poutry breeding organisations have made their genomics projects public (O'Keefe, 2009). In the last five years, Aviagen has been expanding and refining SNP panels to include the results of extensive whole genome associations on a range of economically important traits (Ye *et al.*, 2006; Powell *et al.*, 2008) and the structure of linkage disequilibrium (LD) within and between broiler and layer populations (Andreescu *et al.*, 2007; Abasht *et al.*, 2006). In 2009, the EW Group (including Aviagen Ltd, Hy-Line International and Lohmann LTZ) developed an optimized proprietary 42K Illumina iSelect BeadChip to meet ongoing SNP genotyping requirements. The 42K BeadChip was developed without public funding and has been made available to a number of academic partners (Iowa State University, University of Wisconsin, The Roslin Insitute). Also, there has been an explosion in the scope for SNP genotyping of commercial populations across livestock populations moving the frontier of SNP panels from tens of thousands to hundreds of thousands of SNPs. For instance, the release of a commercially available high density HD Illumina BeadChips with the BovineHD with more than 500K SNPs is expected in 2010 (www.illumina.com/agriculture). Affymetrix is also actively entering the agrigenomics market with the development of a Bovine GeneChip based on the GeneTitan platform. The EW Group is currently working towards SNP panels with >500K SNP content based on whole genome sequence information across both broilers and layers of maximum phylogenetic diversity.

In parallel to the improvements in SNP genotyping platforms, swift advances in sequencing technologies have been seen. In a recent paper by Rubin *et al.* (2010), whole genome resequencing of multiple chicken genotypes has identified >7 Million SNPs. Identification of new SNP variation provides a tremendous resource for new SNP panel development and improvement of the chicken genome build alike. A novel source of polymorphism is Copy Number Variation (CNV), which has proved useful in explaining the genetic basis of variation and can be identified by resequencing (Wright *et al.*, 2009) or comparative genome hybridization arrays (Crooijmans *et al.*, 2010). On the other hand, the full utility of CNV in genomic selection applications has yet to be demonstrated.

Whole Genome Methods and Statistical tools

The application of marker-assisted selection (MAS) requires the existence of markers that are in strong LD with QTLs that influence economically important traits. If prior information exists on the physiological function of a gene or a group of genes, then a candidate-gene approach can be followed (Ye *et al.*, 2006). However, a genome scan using LD-mapping can be more effective in discovering novel markers. For this purpose, a custom analysis workflow was developed to allow to efficiently run whole genome associations for a large number of traits within minutes, while addresssing statistical issues affecting power (Hassen *et al.*, 2009). However, in broilers the extent of LD in broiler chickens was found to be smaller than in layers (Andreescu *et al.*, 2007) and this had implications for validation of the association studies. The use of haplotypes instead of single or triplets of SNPs offered higher accuracy, particularly with medium-density SNP panels (Powell *et al.*, 2008) and when the localized genomic architecture is considered (Powell, 2009). Nevertheless, the identification of robust SNPs validated across generations is a challenge and evidence from extensive

human studies highlighted similar problems (Jakobsdottir *et al.*, 2009; Manolio *et al.*, 2009). In addition case-control studies, aiming to detect associations for traits postulated to be regulated by a smaller number of loci, are a promising and cost-efficient strategy, particularly when based on pooled samples (Peiris *et al.*, 2010).

For the purposes of genetic evaluation, whole-genome breeding values (GEBVs) are appealing. The first application in animal breeding of whole genome evaluation was using a machine learning to identify a set of SNPs across the genome as classifiers for dichotomous or continuous traits (Long *et al.*, 2007; 2008; 2009). The development of a novel framework, in which non-parametrically derived kernels were embedded in a linear model, was shown to achieve higher predictive ability in cross-validation studies (Gonzalez-Recio *et al.*, 2008; 2009). These approaches can be extended to accommodate non-additive components for more accurate prediction of future phenotypes (Gianola and de los Campos, 2008). The application of genomic selection (Meuwissen *et al.*, 2001) provides a parametric methodology that has been shown to be effective in dairy cattle (VanRaden *et al.*, 2009). . Markers can also be used to calculate a DNA-based relationship matrix and predict breeding values under BLUP, namely GBLUP (Misztal *et al.*, 2009; Hayes *et al.*, 2009).

For a large-scale commercial application of GS, the cost of genotyping might be prohibitive due to the large number of selection candidates in poultry. However, it has been shown that subsets of high density assays can have high predictive ability when only SNPs with high predictive ability are included (Weigel *et al.*, 2009). Alternatively, the use of genotype imputation methods can alleviate the cost through the use of equally spaced low density SNP panels (ELD) and achieve similar accuracies as GEBV computed from high-density panels. Under a low density genomic selection approach, a generic evenly spaced SNP panel can achieve comparable accuracy with GS for any number of traits over consecutive generations.

Implementation - opportunities and challenges

Genomic selection offers great opportunities to maximize selection accuracy, particularly in traits for which there is no much information to estimate the Mendelian Sampling Term at the time of selection (e.g., breeder performance traits). On the other hand the current cost of genotyping selection candidates with high density panels for routine evaluation in poultry breeding schemes remains a challenge. Genotyping costs currently range between $150 and $250 USD per selection candidate. In consequence, strategies to make the use of GS a cost effective alternative are required. These can range from modifications in the breeding program through reducing the scale of the selected population while keeping the same effective population size and/or genotyping a sub-set of selection candidates, to strategies that use low-density SNP panels that either (i) use a sub-set of SNPs with the greatest predictive ability (Gonzalez-Recio *et al*., 2009; Long *et al*., 2007; Weigel *et al*., 2009), or (ii) use a combination of equally spaced low-density and HD SNP panels (Habier *et al*., 2009). Novel approaches towards implementing GS in routine layer and broiler breeding programs within the EW group, Hy-Line and Aviagen Ltd, respectively, are summarised below.

Layers: Hy-Line International GS experiment.

A GS based breeding program was set-up aimed to substantially increase response per year compared to a traditional program by halving the generation interval from 12 to 6 months. To make HD genotyping of all selection candidates feasible, the GS population size and structure was optimized to maximize the response with no increments in the rate of inbreeding per year compared to the traditional scheme. Table 1 shows the layout of the traditional and the GS program, with predicted accuracies, and rates of response and inbreeding obtained from simulation. All selection candidates in the GS population are genotyped using a proprietary optimized 42,000 SNP panel. GEBVs were predicted for a range of commercially important traits including egg production, egg weight, egg shell quality and sexual maturity using a Bayesian mixed model implementation (Bayes-C-π) and GBLUP (Misztal *et al*., 2009; Hayes *et al*., 2009), as described in Wolc et al. (2010). For Bayes-C-π a training set including a total of 293 sires and 913 females across four generations was used. Accuracy of (G)EBV ranges at an early selection age for early production traits when phenotypes were not available, for pedigree based EBV (PEBV), Bayes- C-π and GBLUP are shown in Table 2.

Table 1: Traditional and a Genomic Selection (GS) based layer breeding program

	Traditional		Genomic	
Parameters	♂	♀	♂	♀
Candidates	1000	3000	250	250
Phenotypes	0	3000	0	0
Selected	60	360	50	50
GI-months	12	12	6	6
Accuracy[1]	0.44	0.62	0.6	0.6
ΔG^{1} (σ_P)	0.48		0.80	
ΔF^{1} (%)	1.4		1.14	

[1] Deterministic predictions using pseudo-BLUP indices and the approach described by Dekkers (2008)

Table 2. Selection accuracies of pedigree based EBV (PEBV) and Genomic Selection using Bayes-C-π and GBLUP for early production traits (based on Wolc et al. 2010).

	PBLUP	Bayes-C-π	GBLUP
Minimum	0.180	0.310	0.310
Average	0.345	0.451	0.424
Maximum	0.450	0.640	0.570

These results show the potential of whole genome based genetic evaluation methods over classical pedigree based prediction of EBV when no phenotypes are available at the time of selection. In addition, implementation of HD-GEBVs can indeed be made feasible through re-designing the breeding program.

Broilers: Aviagen Ltd GS initiative

This strategy focuses on capitalizing from the benefits of GS through maximizing the accuracy of GEBVs in a full scale breeding program and with no changes in the generation interval. With between 100,000 and 150,000 typically placed individuals per year in an elite broiler line, full HD genotyping of all selection candidates is a rather onerous option. In

order to rationalize the genotyping cost through using evenly spaced low-density (ELD-GS) SNP panels on selection candidates, two alternative approaches were used: (i) Evaluate predicting ability of sub-sets of ELD SNP selected using LASSO (least absolute shrinkage operator) as described by Weigel *et al*. (2009), and (ii) Combine HD SNP panels with equally spaced low density panels (ELD) that accurately track HD SNP allelic effects by imputing HD haplotypes through within family co-segregation (Habier *et al*., 2009).

The LASSO experience

While more markers may enhance predictive ability, the number of markers required for accurate prediction may vary, depending on training sample size, genetic structure of the population and genetic architecture of the target trait. To evaluate this, the Bayesian LASSO (Park and Casella, 2008) was fitted to egg-production data with varying numbers of evenly spaced markers: 20, 40, 50, 80, 120, 150, 300, 400, 800, 1500 and 21,985 SNPs. Records were adjusted offspring means and own performance for males and females, respectively. The dataset was randomly partitioned into training (TRN, consisting of 80% of data points) and testing (TST, 20% of observations); predictive ability was evaluated via the cross-validation correlation (CV-COR) between predictions and realizations in TST. This statistic has uncertainty due to sampling of TRN and TST sets. To evaluate this, 50 random partitions of the data were generated, each of which yielded an estimate of CV-COR for each of the set sizes. Figure 1 gives CV-COR versus set size. Each point corresponds to one random partition of the data into TRN and TST set and the solid line gives the average correlations for each set-size.

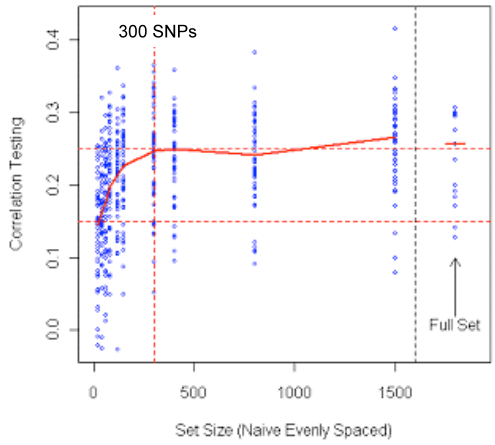

Figure 1. Predictive ability (correlation between observed and predicted values) for sub-sets of SNPs of different sizes.

Results indicate that measures of predictive ability are subject to large uncertainty. This has been partially neglected in the animal breeding literature, and our results suggest that careful consideration of uncertainty of measures of predictive ability is needed for model comparison. The average CV-COR reached a plateau at around 400 SNPs. This is in contrast

with result from other populations, such as those in Weigel *et al*. (2009). Differences in the genetic structure of the populations and the extent of LD may explain why different number of markers may be needed to reach a plateau in the CV-COR vs set size curve.

The HD-ELD SNP panel experience

The strategy of combining HD-ELD SNP panels for GS was tested by simulating genotypes for a real broiler pedigree of 13 generations. A genome of 8 chromosomes of 75cM containing 800 QTL was simulated. The HD panel included 8,000 SNPs (0.075cM intervals), while the HD panel included 75 SNPs (8cM intervals). A Bayes-B implementation was used to estimate HD-SNP effects and predict GEBV using one generation (1500 males and 1500 females) of HD individuals as the training set. HD haplotypes were imputed sequentially for ELD genotyped candidates from generations 2, 3 and 4 following Habier *et al*. (Habier *et al*., 2009). The accuracy of HD-GEBV, ELD-GEBV and traditional BLUP was compared across four successive generations under four scenarios: (i) ELD-offspring, ELD-sires, ELD-dams; (ii) ELD-offspring, HD-Sires, ELD-dams; (iii) ELD-offspring, ELD-sires, HD-dams; and (iv) ELD-offspring, HD-sires, HD-dams (Figure 2).

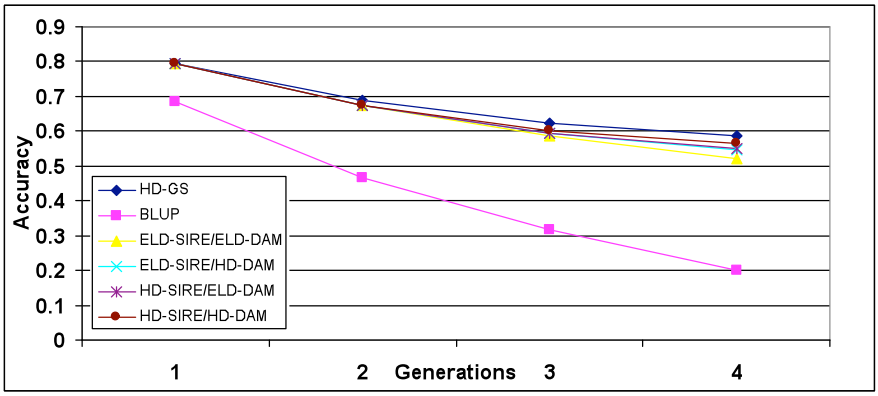

Figure 2: Accuracy of GS using HD panels of a combination of ELD (8cM spacing) and HD on sires and dams.

These results suggest that the combination of ELD and HD SNP panels could be an attractive strategy for rationalizing the cost of genotyping. The percentage loss of accuracies of the ELD panel strategy is relatively minor, holding up to 88.8% compared to that of the HD panel at generation 4. Validation of these results with real data is ongoing using over 1200 high density genotyped (42K SNPs) individuals across three generations for training and 1100 progeny individuals genotyped using an ELD panel of 384 SNPs.

Final considerations

The continuous expansion of SNP panels, the increasing capabilities of high throughput genotyping by genotyping providers and the consolidation of statistical tools for predicting breeding values using genomics information suggest that the incorporation of genomics in routine breeding is potentially within reach by commercial poultry breeding companies. On the other hand, with poultry breeding being such highly competitive market, there is no margin of error. In this context, exhaustive validation of SNP or haplotype effects with independent data is essential to avoid fitting spurious effects and introducing sources of bias in routine genetic evaluations. Specific implementation of genomics for routine breeding will require a huge deal of flexibility in terms of accommodating structural changes in the breeding program structure and to be able to incorporate one or a combination of new methods and statistical approaches. Like any other research and development strategy, implementation of genomics will have to be cost effective, and the corresponding tools repeatable, tractable and capable of dealing with the high-throughput nature of poultry breeding.

References

Abasht,B., Pitel,F., Lagarrigue,S. *et al.* (2006) *Genet Sel Evol.* 38 (3), 297-311.

Andreescu,C., Avendano,S., Brown,S.R. *et al.* (2007) *Genetics.* 177 (4), 2161-2169.

Crooijmans,R.P.M.A., Fife,M., Fizgerald,T. *et al.* (2010) *Plant & Animal Genomes XVIII Conference January 9-13, 2010* .

de Koning, D. J. and Hocking, P. M. (2007) in "*Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish*". Food and Agriculture Organization of the United Nations (FAO)

Dekkers,J.C. (2004) *J Anim Sci.* 82 E-Suppl:E313-328., E313-E328.

Dekkers,J.C (2008) JABG

Dekkers,J.C., Sticker,C., Fernando,R.L. *et al.* (2009) *ADSA CSAS ASAS Montreal, Canada, July 12-16* .

Gianola,D. and de los Campos,G. (2008) *Genet Res.* 90 (6), 525-540.

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

Gonzalez-Recio,O., Gianola,D., Long,N. *et al.* (2008) *Genetics* 178 (4), 2305-2313.

Habier,D., Fernando,R.L., and Dekkers,J.C. (2009) *Genetics.* 182 (1), 343-353.

Hassen,A., Avendano,S., Hill,W.G. *et al.* (2009) *J Anim Sci.* 87 (3), 868-875.

Hayes,B.J., Bowman,P.J., Chamberlain,A.J. *et al.* (2009) *J Dairy Sci.* 92 (2), 433-443.

International Chicken Genome Sequencing Consortium. (2004) *Nature.* 432 (7018), 695-716.

International Chicken Polymorphism Map Consortium. (2004) *Nature.* 432 (7018), 717-722.

Jakobsdottir,J., Gorin,M.B., Conley,Y.P. *et al.* (2009) *PLoS Genet.* 5 (2), e1000337.

Laughlin,K. (2009) *Temperton Fellowship* Report No.15, Harper Adams University College, Newport, Shropshire, UK.

Long,N., Gianola,D., Rosa,G.J. *et al.* (2007) *J.Anim Breed.Genet.* 124 (6), 377-389.

Long,N., Gianola,D., Rosa,G.J. *et al.* (2008) *Dev.Biol (Basel).* 132:373-6., 373-376.

Long,N., Gianola,D., Rosa,G.J. *et al.* (2009) *Genet Sel Evol.* 41:18., 18.

Manolio,T.A., Collins,F.S., Cox,N.J. *et al.* (2009) *Nature.* 461 (7265), 747-753.

Meuwissen,T.H.E., Hayes,B.J., and Goddard,M.E. (2001) *Genetics* 157 (4), 1819-1829.
Misztal,I., Legarra,A., and Aguilar,I. (2009) *J Dairy Sci.* 92 (9), 4648-4655.
Nieuwhof, G. J., Winters, M., and Avendano, S (2008) in "*Animal Production and Animal Science Worldwide, WAAP book of the year 2007*". Wageningen Academic Publishers
O'Keefe,T. (2009) *WATT Poultry, USA* .
Park,T. and Casella,G. (2008) *J.Amer.Statistical Assoc.* 103 (482), 681-686.
Peiris,B.L., Ralph,J., Lamont,S.J. *et al.* (2010) *Midwest ADSA/ASAS Meeting, Des Moine* .
Powell,J., Kranis,A., and Haley,C.S. (2008) *Proceeding of XXIII World's Poultry Congress, Brisbane, Australia.*
Powell,J.E. (2009) *Edinburgh University PhD Thesis* .
Rubin,C.-J., Zody,M.C., Eriksson,J. *et al.* (2010) *Nature* (in press).
VanRaden,P.M., Van Tassell,C.P., Wiggans,G.R. *et al.* (2009) *J Dairy Sci.* 92 (1), 16-24.
Weigel,K.A., de los Campos,G., Gonzalez-Recio,O. *et al.* (2009) *J Dairy Sci.* 92 (10), 5248-5257.
Wolc, A., Stricker, C., Arango, J., Settar, P., Fulton, J., O'Sullivan, N., Habier, D., Fernando, R., Garrick, D., Lamont, S.J., and Dekkers, J.C.M. (2010). 9th WCGALP.
Wright,D., Boije,H., Meadows,J.R.S. *et al.* (2009) *PLoS Genet* 5 (6), e1000512.
Ye,X., Avendano,S., Dekkers,J.C. *et al.* (2006) *Poult.Sci.* 85 (9), 1555-1569.