182. DNA pooling is a cost effective method of including commercial crossbred data in selection of purebreds

M.N. Aldridge[1*], J. Marjanovic[1], J.M. Henshall[2], B. de Klerk[2], K. Peeters[3] and Y. de Haas[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Cobb Vantress B.V. Koorstraat 2, 5831 GH Boxmeer, the Netherlands; [3]Hendrix Genetics B.V., P.O. Box 114, 5830 AC Boxmeer, the Netherlands; michael.aldridge@wur.nl

Abstract

It is generally economically unviable to individually genotype large quantities of commercial crossbred broilers to predict purebred breeding values for crossbred performance. Pooling DNA of crossbreds with similar phenotypes, and genotyping the pool, could be a cost-effective alternative. To test this hypothesis, we used a dataset from a broiler three-way cross experiment consisting of ~9,000 individual genotypes and constructed pools of four different sizes (5, 10, 25, and 50 individuals). Estimated SNP effects and predicted sire breeding values from pooled genotypes were compared to results from individual genotypes, where number of individual genotypes was equal to number of pools. The pool size of 50 and 58 pools had a reasonable accuracy and bias (0.45 and 0.76, respectively) compared to 58 genotyped individuals (0.14 and 0.24). Our results indicate that DNA pooling may be used as a cost-effective means to obtain information on commercial crossbreds for selection in purebreds.

Introduction

Production, health, and welfare traits are routinely recorded on commercial crossbred broilers. Due to cost and logistics, it is not viable to link information from individual commercial crossbred animals to purebred selection candidates. Nevertheless, it is desirable to include these traits in selection of purebreds for crossbred performance. Particularly traits measurable at point of slaughter, observable only on animals raised in a commercial environment, or traits for which genetic correlation between purebred and crossbred performance is low. A potential solution is DNA pooling, where individuals with a similar phenotype are grouped together, and then genotyped as one sample. DNA is extracted from each individual, and amplified to a known concentration. A pool is formed by combining equal contributions of DNA, and then the pool is genotyped. In our previous study we showed that the genotypes of DNA pools are a good representation of individuals within the pool (Marjanovic *et al.*, 2020). DNA pooling can be used to estimate sire breeding values (GEBVs) by building a pseudo genomic relationship matrix (Bell *et al.*, 2017). However, this approach would be difficult to implement with complex crossbreeding structures, as relationships between pools and each sire would be relatively equal. Alternatively, SNP effects could be estimated, from which GEBVs of purebred selection candidates can be predicted. Our objective was to determine if DNA pooling of crossbred genotypes and estimation of SNP effects could be used to accurately and unbiasedly predict GEBVs for crossbred performance in purebred animals.

Materials & methods

Available data. Cobb Vantress provided a dataset from a broiler three-way cross experiment. From a single generation and across five batches, 156 genotyped purebred sires (A) were mated with 1,027 crossbred dams (BC). In total there were 9,262 crossbred progeny (A(BC)) that had a body weight recorded at 35 days after hatching (BW35) and each crossbred animal was genotyped (50,961 SNPs). The crossbred progeny were raised in an environment with 'commercial-like' properties, and each batch was divided across five pens.

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_182, © M.N. Aldridge *et al.* 2022

Sampling scenarios. In total, 40 contemporary groups were formed based on the fixed effects batch×pen×sex×age, with between 108 and 408 individuals in each. The majority of a sire's offspring within a batch were housed in the same pen, and each pen represented multiple sires, this means sires were not equally distributed across all contemporary groups. Each contemporary group was divided into a high end and a low end (top and bottom 25% performing animals based on BW35, respectively). From the high end of one contemporary group, five individuals were randomly sampled, forming one high pool. Each individual could only appear in one pool, and a pool would only be formed if it could be filled completely (each pool must have 5 individuals). This sampling was repeated for the low end and all contemporary groups. In total, 890 pools were formed (445 high pools and 445 low pools). Pool sizes of 10, 25, and 50 were tested with the same sampling method, which resulted in 426, 158, and 58 pools respectively. For each pool a pseudo genotype was created by calculating the allele frequency of individual genotypes in that pool (pools were not actually genotyped). The phenotype for the corresponding pool was calculated as the mean BW35 for the individuals within the pool. To compare pooling and individual genotyping, three sampling methods were tested. Individuals were selected randomly from the full population (Full), randomly from animals in the high end (High only), or randomly from animals in either the high or low end (High/Low). The number of individuals genotyped was equal to the total number of pools. For the larger individual sampling sizes (890, 426, 158) each contemporary group was represented. In addition, one analysis used all 9,262 individual crossbred genotypes as the golden standard.

Estimating SNP effects. Previously parameter estimates for BW35 were estimated with the same dataset (Duenk *et al.* 2019). They were used as input for the program MiXBLUP with the 'hpblup solver', which was used to estimate SNP effects (Ten Napel *et al.*, 2017). For scenarios that included pools, the pseudo genotypes were treated as if they were imputed genotypes. For scenarios that used the individual genotypes (no pooling), the actual genotypes were used. For the 156 line A sires, the SNP genotypes were used to predict GEBVs following Meuwissen (2001). For each scenario with pooling (5, 10, 25, and 50) and individual genotyping (Full, High only, High/Low), the sampling of individuals and estimation of SNP effects and GEBVs were repeated ten times. In addition, GEBVs were estimated using SNP effects for the golden standard.

Accuracy, bias, and cost. Accuracy and bias were calculated for each scenario and presented as means of the ten replicates. Accuracy was estimated as a weighted correlation between the predicted sire GEBV and the mean offspring performance (Calculated from the full dataset for each sire). Bias was calculated using R (Team R Core, 2013), as a weighted regression of the mean offspring performance on the sire GEBVs, and multiplied by two (for an expectation of 1.00, rather than 0.5 of just the sire). Weighted correlations and regressions were used because some of the sires had significantly more crossbred offspring than others. The weights used, were reliabilities of corrected offspring performance, estimated following Cameron (1997), as:

$$(1/4nh^2) / (1+1/4(n\text{-}1)h^2), \quad (1)$$

where n is the number of offspring and h^2 is the estimated heritability.

The cost for each sampling method, was set at € 28 per sample to be genotyped and an additional 20 cents per DNA extraction.

Results

When the full population of crossbred progeny (9,262) were genotyped, the highest accuracy (0.88) and no bias (1.00) were achieved, but at the highest cost (€ 259,336). With less than 10% of those costs, a reasonable accuracy (0.62) and bias (0.98) was achieved by genotyping 890 individuals (Table 1), but only if genotyped animals were sampled from both the high and low performing end (High/Low). Sampling from the Full population or just the High end, reduced accuracy (Full: 0.49 and High: 0.29) and increased bias (Full: 0.85 and High: 0.51). The High/Low sampling was optimal regardless of the number of individuals genotyped. Therefore, the comparison between pooling and individual sampling is made with High/Low.

With a pool size of 5 (4,450 individuals in 890 pools), the accuracy and bias (0.51 and 0.87) was lower than High/Low sampling. The pool size of 10 (4,260 individuals in 426 pools) outperformed individual High/Low sampling for both accuracy (Pool size 10: 0.49 and High/Low 426: 0.42) and bias (Pool size 10: 0.83 and High/Low 426: 0.74). Due to the cost of extracting additional DNA samples, the cost of 426 pools (€ 12,780) is slightly higher than individual genotyping (€ 11,928). With large pool sizes (25 and 50) and small number of pools (158 and 58), pooled sampling significantly outperformed accuracy and bias of the equivalent number of individual genotypes, at a lower cost (€5,214 and €2,204). With pools of 25, the accuracy was higher (Pool size 25: 0.48 and High/Low: 0.24) and bias lower (Pool size 25: 0.82 and High/Low: 0.42) than 158 individual genotypes. The extreme pooling of 50 individuals in 58 pools had a higher accuracy than 58 individual genotypes (Pool size 50: 0.45 and High/Low 58: 0.14) and less bias (Pool size 50: 0.76 and High/Low 58: 0.24).

Discussion

Pooling individuals based on phenotypic performance within contemporary groups, and then estimating SNP effects from the DNA pools, is an effective method of estimating GEBVs for a sire line. This is in line with the results of Bell *et al.* (2017) with a different approach where a pseudo G-matrix is formed instead. If there is sufficient budget, individual genotyping is preferred. However, to maximise accuracy and reduce bias, animals from both the low and high performing end should be included. Pool size and the number of pools had limited impact on accuracy and bias. The purpose of DNA pooling is cost saving, when the budget is limited, maximising pool size and minimizing number of pools is preferential.

Table 1. Estimated accuracy, bias, and economic cost of individual or pooled genotyping.

Sampling method	Number of individuals	Accuracy[1]	Bias[2]	Cost per sample genotyped	Number of samples genotyped	Total cost
Population	9,262	0.88	1.00	€ 28	9,262	€ 259,336
Full 890	890	0.49	0.85	€ 28	890	€ 24,920
High 890	890	0.29	0.51	€ 28	890	€ 24,920
High/Low 890	890	0.62	0.98	€ 28	890	€ 24,920
High/Low 426	426	0.42	0.74	€ 28	426	€ 11,928
High/Low 158	158	0.24	0.42	€ 28	158	€ 4,424
High/Low 58	58	0.14	0.24	€ 28	58	€ 1,624
Pool size 5	4,450	0.51	0.87	€ 29	890	€ 25,810
Pool size 10	4,260	0.49	0.83	€ 30	426	€12,780
Pool size 25	3,950	0.48	0.82	€ 33	158	€ 5,214
Pool size 50	2,900	0.45	0.76	€ 38	58	€ 2,204

[1] SE for individual genotyping methods ranged between 0.01 and 0.03, and for pooling were all <0.01.

[2] SE for individual genotyping methods ranged between 0.01 and 0.04, and for pooling were all <0.01.

It is worth noting that we formed pseudo genotypes as allele frequencies, rather than actually genotyping pools as it would be required in practice. Previously we showed small differences depending on how the pseudo genotype is calculated and some differences compared to real genotyping of pools (Marjanovic *et al.*, 2020). Therefore, our method of obtaining pseudo genotypes may have influenced our findings. Furthermore, our cost analysis only considered the additional cost of extracting DNA, while there would be other additional costs. Finally, we acknowledge that the accuracies are likely to be spurious (important for EBVs of the high performing end), as the relationship between sires and crossbred progeny are likely to be different across strategies. A cross-validation will be performed to confirm our findings.

Regardless of limitations, in the most extreme scenario (58 pools with 50 individuals in each pool), the accuracy and bias outperformed most individual genotyping scenarios. The accuracy and bias from 58 pools suggests substantial genetic progress could be made using predicted GEBVs in the sire line. On the other hand, selection based on information from 58 individual genotypes would result in poor outcomes due to low accuracy. DNA pooling and estimating SNP effects is a promising method for situations with limited resources but there is a desire to include traits in selection, which is often the case with commercial crossbreds.

Funding

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri and Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

References

Bell A.M., Henshall J.M., Portro-Neto L.R., Dominik S., McCulloch R., *et al.* (2017) Genet. Sel. Evol. 49(28):1-7. https://doi.org/10.1186/s12711-017-0303-8

Cameron N.D. (1997) Selection indices and prediction of genetic merit in animal breeding. Wallingford: CAB International; 1997. pp. ix + 203.

Duenk P., Calus M.P.L., Wientjes Y.C.J., Breen V.P., Henshall J.M., *et al.* (2019) Genet. Sel. Evol. 51(6):1-11. https://doi.org/10.1186/s12711-019-0447-9

Marjanovic J., de Klerk B., de Haas Y., Dibbits B.W., and Aldridge M.N. (2020) Proc. of the 71st Annual Meeting of the European Federation of Animal Science, Online conference.

Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001). Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Team R Core. (2013). R: A language and environment for statistical computing. http://r.meteo.uni.wroc.pl/

Ten Napel J., Vandenplas J., Lidauer M., Stranden I., Taskinen M., *et al.* (2017) MiXBLUP, user-friendly software for large genetic evaluation systems – Manual Wageningen, the Netherlands V2.1-2017-08. http://www.mixblup.eu/documents/Manual%20MiXBLUP%202.1_June%202017_V2.pdf