A COMPARISON BETWEEN THE USE OF PEDIGREE OR GENOMIC RELATIONSHIPS TO CONTROL INBREEDING IN OPTIMUM-CONTRIBUTION SELECTION

M. Sharif-Islam[1], M. Henryon[2], J.H.J. van der Werf[3], A.C. Sørensen[2], T.T. Chu[4], B. J. Wood[5] and S. Hermesch[1]

[1]Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia
[2]Danish Pig Research Centre, Danish Agriculture and Food Council, Axeltorv 3, 1609 Copenhagen V, Denmark
[3]School of Environmental and Rural Science, University of New England, Armidale, NSW, 2351 Australia
[4]Center for Quantitative Genetics and Genomics, Aarhus University, Denmark
[5]School of Veterinary Science, University of Queensland, Brisbane, QLD 4072, Australia

SUMMARY

Stochastic simulation was used to test the hypothesis that optimum-contribution selection with genomic relationships using marker loci with low minor allele frequency (MAF) below a predefined threshold (referred as TGOCS) to control inbreeding maintained more genetic variation than pedigree relationships (POCS) at the same rate of true genetic gain (ΔG_{true}). Criteria to measure genetic variation were the number of segregating QTL loci (quantitative trait loci) and the average number of founder alleles per locus. Marker alleles having a MAF below 0.025 were used in forming the genomic relationships in TGOCS strategy. For centering in establishing genomic relationships, when the allele frequency of marker loci with low MAF set to 0.5 the TGOCS strategy maintained 66% fewer founder alleles than POCS and there were 30% fewer QTL segregating. This TGOCS strategy maintained 61% fewer founder alleles than GOCS and 28% fewer segregating QTL loci. When the allele frequency of marker loci with low MAF was set to observed allele frequency these figures were 8%, 2%, 5% and 2%, respectively. Using marker loci with low MAF in the TGOCS strategy was inferior to both GOCS and POCS. Both TGOCS and GOCS were affected by the same constraint that is LD (linkage disequilibrium) between markers and QTL. Therefore, POCS is a more efficient method to maintain genetic variation in the population until a better way to use genomic information in optimum-contribution selection is identified.

INTRODUCTION

Optimum-contribution selection (OCS) can use either pedigree or genomic relationships to control inbreeding. Simulation studies showed that using pedigree relationships to control inbreeding in OCS realise more true genetic gain (ΔG) than genomic relationships at the same rate of true inbreeding (ΔF), where the true inbreeding coefficient of an individual is the observed proportion of marker loci in its genome with alleles that are identical-by-descent (IBD) (Sonesson *et al*. 2012, Henryon *et al*. 2019). Using pedigree relationships to control inbreeding in OCS (referred to as POCS) realises more ΔG than using genomic relationships based on all markers (GOCS) because POCS manages expected genetic drift without restricting selection at QTL (Henryon *et al*. 2019). By contrast, GOCS penalises changes in allele frequencies at marker loci generated by genetic drift or selection. Because these marker alleles are in linkage disequilibrium with QTL alleles, it restricts changes in favourable QTL alleles. This implies that GOCS in its current form is unlikely to realise more genetic gain than POCS at the same rate of true inbreeding.

* A joint venture of NSW Department of Primary Industries and University of New England

An alternative strategy is needed for using genomic information to control inbreeding. One possible strategy is to carry out GOCS by establishing genomic relationships using only marker loci with low minor-allele frequencies (MAF) in the generation under selection. All other marker loci are excluded in this form of GOCS. Focusing on marker alleles with low MAF is of interest because it is these marker alleles that are particularly susceptible to being lost through selection or genetic drift. While most changes in allele frequencies due to drift or selection will be reversible, the extinction of a particular allele constitutes an irreversible loss of variation. Using marker loci in establishing genomic relationships, the genotypes of animals at each locus are centred around a pre-defined allele frequency; it could reduce the loss of alleles by promoting a shift in allele frequencies towards the pre-defined frequencies. Assuming allele frequency of 0.5 for all marker loci in establishing genomic relationships, more weight will be given on the heterozygotes by moving allele frequency towards 0.5 (Meuwissen *et al*. 2020). In case of using calculated allele frequency other than 0.5, rare alleles will be given more weight than common alleles (Forni *et al*. 2011). Preventing the loss of minor alleles may maintain more segregating loci that contribute to genetic variation in the population. Since markers are in LD (linkage disequilibrium) with QTL, both the number of segregating QTL and the number of founder alleles maintained in the population can be used as criteria to assess loss of genetic variation, at least in simulation studies. This reasoning leads to the hypothesis that GOCS using marker loci with MAF below a predefined threshold in the generation under selection – hereafter referred to as TGOCS- maintains more segregating QTL and founder alleles than POCS at the same ΔG.

MATERIALS AND METHODS

Procedure. Stochastic simulation was used to estimate the number of segregating QTL and founder alleles (the average number of founders that contributed alleles to each locus, averaged over all founder loci) realised in the last generation after applying TGOCS, GOCS or POCS and making comparisons at the same ΔG_{true}. TGOCS included all low frequency alleles at the marker loci when MAF at the marker was below 0.025 in the first generation under selection. Marker allele frequencies were calculated in the OCS candidates but the allele frequencies used for the centering of genotypes was set to either 0.5 (Scenario, TGOCS_0.5) or to the allele frequency found in the base population (scenario, TGOCS_base). GOCS used genomic relationships calculated from all markers having a MAF above 0. The criterion for selection was the true breeding value (TBV) of a single trait with a genetic variance of 1. Each breeding scheme was run for ten discrete generations. Each replicate was initiated by sampling a unique base population from the founder population. Selection candidates were genotyped before selection.

Breeding scheme. A total of 25 matings were made from 250 selection candidates by OCS in each generation. Animals were selected randomly in generations 1 to 3. Selection based on TBV was introduced in generations 4 to 10. Males that we selected were allocated up to 25 matings. All male candidates were considered potential parents by OCS. The top 25 females were allocated a single mating each. The 25 sire and 25 dam matings were paired randomly. Each dam produced ten offspring resulting in 25 full-sib families and 250 offspring. Offspring were assigned as males or females with a probability of 0.5.

Genetic models. The founder population was established using a Fisher-Wright inheritance model to generate LD between QTL and markers following the study of Henryon *et al*. (2019). The genome was 30 Morgan and consisted of 18 pairs of autosomal chromosomes; each chromosome was 167 cM long. A total of 7,702 QTLs and 54,218 markers were located across the genome and were all segregating in generation $t = -1$. An additive effect of every mutant allele at each QTL followed an exponential distribution. No major QTL was simulated. Markers were distinct from QTL and were used to form genomic relationships in TGOCS and GOCS. A total of 6,012 founder loci were placed evenly across the genome in the base population (generation=0). These founder

loci were not used in establishing genomic relationships.

Optimum-contribution selection. POCS was carried out by maximising $\boldsymbol{U_t(c) = c'a - \omega\, c'A\, c}$, where **c** is a vector of genetic contributions to the next generation, **a** is a vector of TBV, ω is a penalty applied to the average estimated relationship of the next generation, and **A** is a pedigree relationship matrix (Henryon *et al*. 2019). The penalty, ω, was constant across generations. It was calibrated to realise approximately 1.00 ΔG_{true} in all scenarios. GOCS was carried out by replacing **A** with a genomic-relationship matrix, **G,** which was calculated as $\frac{ZZ'}{\sqrt{2p'(1-p)}}$, where $\mathbf{Z} = \boldsymbol{M - 1(2p)'}$ and **M** is a matrix of count of mutant alleles with element M_{ij}=0, 1 or 2 for each animal at each marker. Allele frequency **p**, was calculated using all OCS candidates in the generation under selection for GOCS_base and TGOCS_base while the **p** was set to 0.5 for centering in TGOCS_0.5 and GOCS_0.5.

Data analysis. The number of founder alleles, segregating QTL and number of markers below the threshold maintained in the last generation for each of the five scenarios were calculated for each replicate. ΔG_{true} was calculated as the linear regression of G_t on t, where G_t is the average TBV of animals born at generations, t=4...10 for each replicate. All results were expressed as the mean of 300 replicates.

Software. The breeding program was simulated using the software package ADAM (Pedersen *et al*. 2009) then OCS was carried out using EVA software (Berg *et al*. 2006).

RESULTS AND DISCUSSION

The results did not support the premise that TGOCS maintains more segregating QTL or founder alleles than POCS at the same rate of ΔG_{true}. Results showed that POCS maintained more QTL alleles and IBD alleles than TGOCS (Table 1). This makes POCS a robust method to use in animal breeding. Similar to GOCS, using marker information in TGOCS does not help to maintain more alleles in the population. In addition, TGOCS_0.5 maintained significantly fewer (66% and 8%) founder alleles and (30% and 2%) segregating QTL than TGOCS_base and POCS (Table 1). To the best of our knowledge, the proposed method TGOCS has not been investigated while GOCS has been investigated previously (e.g. Sonesson *et al*. 2012; Henryon *et al*. 2019). Results show that TGOCS maintained significantly fewer founder alleles and segregating QTL than GOCS (Table 1). Consequently, TGOCS was inferior to both GOCS and POCS. GOCS was also inferior to POCS in this study, which is supported by the results found in the study of Sonesson *et al*. (2012) and Henryon *et al*. (2019). Therefore, POCS remains the worthy method to maintain more QTL alleles and founder alleles in the population.

Table 1. Numbers (N) of founder, QTL or markers alleles maintained in the last generation (standard errors) realised by scenarios of alternative optimum-contribution selection (OCS) at the same rate of true genetic gain

OCS scenarios	N founder alleles	N favourable QTL alleles	N marker alleles with MAF<0.025	N marker alleles with MAF>0.025
POCS	20.19 (0.03)	2617.17 (2.01)	3221.15 (10.48)	32840.08 (17.08)
TGOCS_0.5	6.78 (0.04)	1825.11 (5.24)	2129.75 (20.82)	24039.93 (57.37)
TGOCS_base	18.54 (0.13)	2557.70 (5.72)	4580.87 (36.17)	30779.85 (106.63)
GOCS_0.5	17.25 (0.04)	2541.87 (2.12)	2581.27 (10.78)	32509.65 (17.75)
GOCS_base	19.59 (0.09)	2596.90 (2.50)	3266.68 (10.69)	32565.85 (20.69)

Abbreviations: POCS: Optimal contribution selection (OCS) based on pedigree relationships; GOCS: OCS with genomic relationships using all marker loci; TOCS: OCS with GOCS using marker loci with low minor allele frequency (MAF) below a predefined threshold (MAF<0.025). Allele frequencies were set either to 0.5 (TGOCS_0.5/GOCS_0.5) or base population allele frequency (TGOCS_base/GOCS_base).

TGOCS_base did not maintain more IBD and favourable QTL alleles than GOCS_base. The reason could be that we simulated very small populations and LD between markers and QTL. Even if we used only a subset of markers having MAF below 0.025, still there are enough markers. Therefore, TGOCS_base could not overcome LD between markers and QTL. If we would simulate more matings, we believe that TGOCS_base and GOCS_base would produce similar results. TGOCS_0.5 also could not maintain more minor alleles than GOCS by attempting to increase allele frequency towards 0.5 at markers with low MAF. A possible reason could be that only 25 matings were simulated in this study. It had less flexibility to move allele frequency of all markers towards 0.5. However, simulating more matings might not help because allele frequency towards 0.5 is suboptimal when genetic gain is concerned. So, by giving more weight to markers with low MAF, TGOCS_0.5 ultimately lost more markers which consequently lost more QTL alleles. Therefore, TGOCS_0.5 maintained fewer favourable segregating QTL alleles than TGOCS_base. It indicates that it is difficult to maintain more genetic variation by using genomic information in its current form in OCS because of LD between markers and QTL.

By contrast, POCS can manage the expected genetic drift without restricting selection at QTLs (Henryon *et al*., 2019). Since POCS does not depend on the markers, POCS can increase the allele frequency at some favourable QTL without much affecting allele frequency at other QTLs. Thus, POCS could maintain more favourable QTL alleles in the population than TGOCS. Since QTL and founder alleles are in LD, POCS maintained more founder alleles than TGOCS at the same ΔG_{true}. So, genomic information used in TGOCS in its current form could not help maintain more QTL and founder alleles than POCS at the same ΔG_{true}. This study gives more insight into the underlying mechanisms of why use of pedigree relationships in OCS is superior to using genomic relationships in OCS to maintain genetic variation in the population. However, this study assessed genetic variation across the whole genome but controlling genetic diversity in specific regions of genome might also be of interest. Research should be conducted how genomic relationships can be used to control genetic diversity in different regions of the genomes while maintaining rate of true genetic gain in the population.

ACKNOWLEDGEMENTS

M. Sharif-Islam is supported by an Animal Genetics and Breeding Unit Scholarship at the University of New England, Australia, and a PhD top-up scholarship from the Australasian Pork Research Institute Ltd (APRIL).

REFERENCES

Henryon M., Liu H., Berg P., Su G., Nielsen H.M., Gebregiwergis G.T. and Sorensen A.C. (2019) *Genet. Sel. Evol.* **51:** 39.

Sonesson A.K., Woolliams J.A. and Meuwissen T.H.E. (2012) *Genet. Sel. Evol.* **44:** 27.

Pedersen L.D., Sorensen A.C., Henryon M., Ansari-Mahyari S. and Berg P (2009) *Livestock Sci.* **121:** 343.

Berg P., Nielsen J. and Sorensen M.K., (2006) *WCGALP* 27-09.

Forni S., Aguilar I. and Misztal I. (2011) *Genet. Sel. Evol.* **43:** 1

Meuwissen T.H.E., Sonesson A.K., Gebreyohans G. and Woollliams J.A. (2020). *Front. Genet*. **11**: 880.