AN OPTIMAL CONTRIBUTION SELECTION TECHNIQUE THAT UTILISES NON-ADDITIVE GENETIC COMPONENTS

Z. Loh, J. H. J. van der Werf, S. Clark

School of Environmental & Rural Science, University of New England, Armidale, NSW, 2351
Australia

SUMMARY

This study proposes an optimal contribution selection method (OCS) that utilizes both additive and non-additive genetic components. Using a genetic algorithm, the contribution of sires toward a cohort of dams, along with their mate allocation, were optimized under a constraint of a 1% increment of inbreeding per generation. The inclusion of dominance into the OCS increases the total genetic gain in offspring initially by 30.5% improvement from +4.02 to +5.27 units compared to using additive genetic component alone for a trait with a 15% dominance-to-additive variance ratio. Compared to additive-only OCS, optimization of the dominance component resulted in one-off additional gains, with no additional merit thereafter, despite continued optimization. In conclusion, this inclusion of dominance in mate allocation can give a significant genetic lift in total genetic merit.

INTRODUCTION

While optimal contribution selection (OCS) has successfully optimized the additive genetic gain in livestock breeding systems within a constraint of inbreeding, it has only focused on estimated breeding values (EBVs) and generally not focused on optimizing the non-additive genetic component, such as dominance. Dominance could explain a significant proportion of the genetic variance for some traits, but it has been difficult to exploit due to its dependency on sire-dam mating configuration and the difficulty of predicting these specific effects. The advent of genomic information, however, allows direct prediction of the expected offspring heterozygosity, which could be used to predict dominance effects for mate allocation.

The aim of this study was to develop an OCS that could optimize both additive and non-additive genetic components, using information easily available to a breeding program. It is anticipated this OCS can be used in improving both additive and non-additive effects in a trait.

LAYOUT OF THE OPTIMAL CONTRIBUTION SELECTION METHOD

The OCS requires several inputs: sire and dam genotype arrays of size $N_m \times M$ and $N_f \times M$ respectively, with N_m, N_f and M be number of sires, dams and markers respectively; sire and dam phenotypic vector of length N_m and N_f respectively, and narrow sense heritability. The genotype, phenotype and heritability were used to calculate the sire EBVs ($\hat{\boldsymbol{\beta}}_m$) and sire GRM ($\boldsymbol{G}$) using method by VanRaden (2008). A targeted level of increment of consanguinity (ΔF_t) were also needed for the OCS.

This OCS has three phases: the first phase optimized additive and inbreeding components; the second phase optimized the non-additive genetic components only, and the final phase combined the results from both phases. Such a design was needed to improve the feasibility of the method from the significantly increased sample space when optimizing the dominance genetic components.

To initialize the GA, 1500 candidate solutions of length N_f, denoted as $\boldsymbol{s}$, that contain the indices of sires that paired with each dam, were generated, with the i-th entry of $\boldsymbol{s}$ contains which sire that would be paired with i-th dam. This formatting was required due to the mate-specific nature of the

dominance component, which depends on the exact permutations of the sires. This set of s vectors were compiled into a sire index matrix of size $1500 \times N_f$, denoted as S_1.

The first phase of this OCS optimized the additive and inbreeding coefficients, which were initialized by translating S_1 into its corresponding sire proportion matrix X_1, defined as a matrix of size $1500 \times N_m$ with its i-th row and j-column representing the proportion of j-th sire that would contribute into the next generation for the i-th solution. The objective function for this phase was defined as follows:

$$f_{obj}(X_1)_{AI} = X_1 \widehat{\boldsymbol{\beta}}_m{}' - \lambda_i * diag(X_1 A X_1') \quad [1]$$

where λ_i denoted the scalar weightage for the inbreeding component for this phase of OCS.

From this objective function, the top two ss in term of $f_{obj}(X_1)_{AI}$ were chosen, which were propagated into a new S_1. This new S_1 was subjected to five genetic operators: mutation, where sires in S_1 were replaced with new sires; vertical and horizontal recombination, where the part of S_1 were exchanged, column-wise and row-wise, respectively, and vertical and horizontal inversions, where the orders of sires in S_1 were reversed, column-wise and row-wise respectively. The hyper-parameters values for these operators were based on Srinivas and Patnaik (1994).

This phase was then iterated with the new S_1. For each iteration, the λ_i was adjusted with the amount $100(average(diag(X_1 A X_1')) - \Delta F_t)$. The mutation, recombination and inversion rate were also adjusted adaptively based on the method by Srinivas and Patnaik (1994). This process continued until convergence, defined as the point where the slope of the curve of $f_{obj}(X_1)_{AI}$ is less than 1×10^{-3} across the last 50 iterations. To reduce the chance of premature convergence for subsequent phases, this phase was repeated eight times, with the converged solutions from each repeat pooled into a new sire index array, S_2. From each repeat, the average of the λ_i at the point of convergence, denoted by λ_{avg}, was also recorded.

The S_2 was then used for Phase 2 optimization, which maximizes the offspring dominance component. From S_2, 3000 solutions were resampled and altered using genetic operators. Only vertical recombination and horizontal inversion were used on S_2, as they only affect the permutations of the sires within the ss, thus with no effects on its additive and inbreeding scores, thus not affecting their Phase 1 optimality. The performance of each solution in S_2 was tested, with the objective function for k-th solution defined as follows:

$$f_{obj}(S_2)_D = \sum_{i=1}^{N_f} H_{S_2(k,i),i} \quad [2]$$

where $H_{S_2(k,i),i}$ is defined as the expected heterozygosity for $S_2(k,i)$-th sire and i-th dam, which $S_2(k,i)$ is the k-th row and i-th column of S_2. The top two ss in terms of $f_{obj}(S_2)_D$ were extracted from S_2 and used to generate a new S_2 array, subjected to vertical recombination and horizontal inversion. This phase was iterated until convergence, defined as the point where the slope of the curve of $f_{obj}(S_2)_D$ is less than 1×10^{-4} across the last 200 iterations. To increase the chance of finding the global maximum, Phase 2 was repeated eight times, and the solutions pooled into S_3.

In the final phase, the S_3 was translated into its corresponding sire proportion array X_3. The performance of each solutions was evaluated as follows:

$$f_{obj}(S_3, X_3)_{ADI} = X_3 \widehat{\boldsymbol{\beta}}_m{}' + \sum_{i=1}^{N_f} H_{S_3(k,i),i} - \lambda_{avg} * (average(diag(X_3 A X_3')) - \Delta I_t) \quad [3]$$

Equation [3] served as the final objective function for the OCS. The top s in terms of $f_{obj}(S_3, X_3)_{ADI}$ were deemed as the optimized solution, and were the final output of the OCS.

TESTING THE OPTIMAL CONTRIBUTION SELECTION METHOD

The OCS was tested with simulated genotypic arrays generated using QMSim (Sargolzaei and Schenkel 2009). For the ancestral population, 5,000 animals and 20,000 loci across 10 chromosomes of 100 cM each were simulated. This population was gene-dropped for 1,000 generations, with the population size increasing up to 10,000 in the final generation. Either 500 or 1000 sires and dams were then randomly chosen for genotyping and these were selection candidates (Table 1).

From all loci, 500 of them were assigned as QTL, with both additive and dominance effects. Using these effect sizes, the phenotypes were calculated as follows:

$$\boldsymbol{y} = \boldsymbol{Z_a\beta} + \boldsymbol{Z_h\delta} + \boldsymbol{e} \qquad [4]$$

where $\boldsymbol{y}$ is the phenotype vector; $\boldsymbol{Z_a}$ is the additive genotypic array encoded in the format of $\{0,1,2\}$; $\boldsymbol{Z_h}$ is the heterozygosity array with a value of 1 for heterozygotes and 0 otherwise; $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are vectors with additive and dominance effect sizes for each QTL, respectively, and $\boldsymbol{e}$ is a vector with the residual component of the phenotypes. Both $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ were generated using a gamma distribution, with shape parameters set at 0.3 and scale parameters provided in Table 1. The vector $\boldsymbol{e}$ was generated using a normal distribution, with mean zero and variance $\frac{(1-h^2)var(\boldsymbol{G\beta})}{h^2}$, where h^2 is the narrow sense heritability. The h^2 was set at 0.3 for all simulations.

These genotypes and phenotypes were used in a four-generation selection program. Three selection regimes were tested: truncation genomic selection (denoted as TS), OCS with additive component (OCSA); and OCS with both additive and dominance components (OCSAD). The ΔF_t is set at 1% per generation for OCSA and OCSAD. To ensure validity of comparison for TS, the proportion of sires selected was determined by the number of selected top sires that would produce the same ΔF_t. A non-selected population (NSEL) was used to establish the offspring baseline performance. For each generation, the additive, dominance and total genetic merits (TGM) from each selection regime were recorded.

The parameters and values tested in this study were provided in Table 1. When a parameter was under study, default values were used for other parameters. When neither the additive and dominance genetic variances were under study, the default scale parameters of the effect size distributions were chosen such that the dominance genetic variance is 15% of the additive genetic variance. For each set of parameter values and selection regimes, 20 replicates were conducted. To test the performance between selection regimes, a two-sample Welch's t-test was used, with the performance deemed significantly different if the $logpval = -\log_{10}(p - value) \geq 3$.

Table 1. Parameters and values tested in this study

Parameters	Default values	Alternative values
Number of Sires and Dams	500	1000
Additive Effect Size Scale Parameter	1.0	3.0
Dominance Effect Size Scale Parameter	0.5	1.5

RESULTS

The additive, dominance and TGM across four generations for the different selection regimes were provided in Figure 1. The first-generation total genetic merit under different parameter values and selection regimes were provided in Table 2.

Compared to TS, both OCS methods significantly improved the additive genetic component of the offspring across all parameter values tested. The OCSAD method significantly improved the dominance component compared to OCSA from +0.17 to +1.51 ($logpval = 22.54$), and this led to a 30.5% additional improvement in TGM from +4.02 to +5.27 ($logpval = 8.48$) in the first generation of selection under the default parameter values. The additional gain from the dominance

component is a one-off genetic lift, however, with no further additional increments in dominance genetic merit despite its continued optimization in the subsequent generations (Figure 1b).

The improvement in TGM in OCSAD compared to OCSA was observed for all parameter values tested, although these parameters affect the significance of improvement. For example, by increasing the scale parameter for additive QTL effect sizes from 1.0 to 3.0, which increases the additive genetic variance, the increment in TGM becomes less significant ($logpval = 1.59$). While this change of parameter value has decreased the dominance-to-additive variance ratio to 2.1%, the TGM for OCSAD is still 11.4% higher than OCSA after the first generation of selection, indicating the potential merit of mate allocation in exploiting dominance variation.
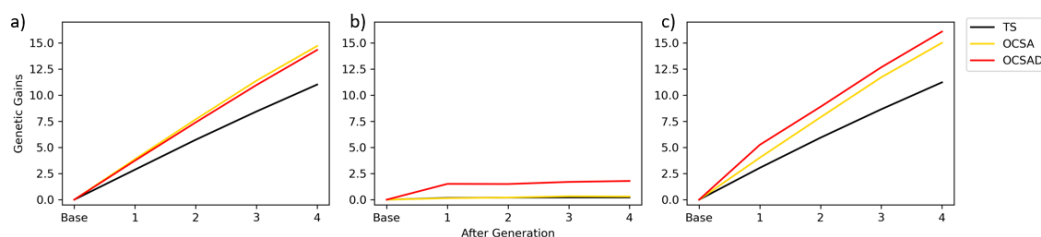

Figure 1. Plots for the base scenario showing (a) additive, (b) dominance and (c) total genetic merit of the offspring under truncation selection (TS), additive-inbreeding OCS (OCSA) and additive-dominance-inbreeding OCS (OCSAD) across four generations

Table 2. First generation total genetic merit with truncation selection (TS), additive-inbreeding OCS (OCSA) and additive-dominance-inbreeding OCS (OCSAD) under varying parameter values and selection regimes. Superscripts with different letters (row wise) denote significant differences between selection regimes

Parameter values	Value tested	Total genetic merit		
		TS	OCSA	OCSAD
Number of sires and dams (default)	500	3.045[a]	4.019[b]	5.247[c]
(alternative)	1000	4.010[a]	4.527[a]	5.827[b]
Additive effect size scale parameter	3.0	8.845[a]	11.328[b]	12.616[b]
Dominance effect size scale parameter	1.5	3.052[a]	3.953[b]	7.855[c]

DISCUSSION AND CONCLUSION

In this study, an OCS method that optimized the additive and dominance component was proposed. Using heterozygosity for all loci as a proxy for the optimization of dominance, with a 15% dominance-to-additive variance ratio, this method improved the initial TGM by 30.5% compared to only optimizing the additive component. The one-off lift from the dominance component optimization means that after the first generation both OCS would have the same rate of genetic gain despite the continued optimization of dominance. Some computational aspects of the proposed method could be further optimised.

In conclusion, an OCS that optimizes additive and dominance effects was proposed in this study, and gave a significant lift in total genetic merit of a selected trait. The method can be used to improve the within-population genetic merit for economically important traits in livestock.

REFERENCES

Sargolzaei M and Schenkel F.S. (2009). *Bioinformatics*. **25:** 680.
Srinivas M. and Patnaik L.M. (1994). *IEEE Sys. Man. Cybern.* **32:** 115.
VanRaden P.M. (2008). *J. Dairy Sci.* **75:** 3136.