# Optimum contribution selection combined with weighting rare favourable alleles increases long-term genetic gain

**H. Liu**[1]**, A.C. Sørensen**[1]**, P.Berg**[1,2]

[1]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University,
[2]Nordic Genetic Resource Center, Norway

**ABSTRACT:**

This simulation study was conducted according to the hypothesis: the selection strategy where optimum contribution selection (OCS) are combined with genomic estimated breeding values with additional weights on rare favorable alleles (wGEBV) will boost long-term genetic gain, while at the same time effectively controlling inbreeding. Six breeding schemes were simulated by combining GEBV or wGEBV and truncation selection (TS) or OCS. Optimum contribution selection was further categorized into OCSA and OCSG depending on whether pedigree ($A$) or genomic data ($G$) was used for penalizing average co-ancestry. The selection was performed for 40 generations. It was shown that OCSG combined with wGEBV resulted in up to 38.8% higher cumulative genetic gain than TS based on GEBV, although the benefit from using wGEBV was less pronounced for OCSG than OCSA. Furthermore, the control of genome-wide inbreeding in GS can be more successfully achieved using OCSG than OCSA irrespective of selection criteria.

**Keywords:** Genomic selection; Optimum contribution selection; Rare alleles

## Introduction

The selection scheme based on genomic estimated breeding values (GEBV), which is referred to as genomic selection (GS), provides a higher accuracy of selection and thus a higher short-term genetic gain compared to conventional selection methods. However, GEBV might not be the optimal selection criterion if the accumulated changes are desired to be maximized over a longer time horizon (Dekkers and van Arendonk 1998).

An alternative selection criterion, where additional weight was put on rare favorable alleles for GEBV (wGEBV), was first proposed by Goddard (2009) and further developed by Jannink (2010). The results from Jannink (2010) showed that truncation selection (TS) based on this index increased the long-term genetic gain though it slightly decreased short-term genetic gain compared to TS based on GEBV.

An alternative to TS is optimum contribution selection (OCS), which offers opportunities to maximize the long-term genetic gain (Meuwissen 1997). It attempts to achieve higher genetic gain by penalizing the average co-ancestry among selected parents. The co-ancestry can be realized by either additive genetic relationships based on pedigree data ($A$) or genomic data ($G$). Sonesson et al. (Sonesson et al. 2012) indicated that, when selection criterion is GEBV, a genomic relationship matrix should be used for penalizing the average co-ancestry, in order to obtain an effective control of genome-wide inbreeding.

To our knowledge, the approach combining OCS and wGEBV has so far not been considered. We hypothesize that by combining them, diversities in families are expected to be widened so that more genetic variance will be maintained and thus more long-term genetic gain will be achieved. This hypothesis was tested by a stochastic simulation study. Besides the genetic gain, pedigree, genomic and true inbreeding resulting from each selection scheme was also investigated to understand the mechanisms causing differences in cumulative genetic gain.

## Materials and Methods

Two selection criteria, GEBV and wGEBV and two selection rules, TS and OCS, were combined for generating breeding schemes. Optimum contribution selection was further categorized into OCSA using pedigree data and OCSG using genomic data (Table 1). All simulations were replicated 100 times.

**Selection criteria:** To obtain GEBV, Bayesian Lasso (BL) was used. Details regarding the model are described in (Liu et al. (2014)). To obtain wGEBV, the method of weighting markers in Jannink (2010) was modified by additionally accounting for the time horizon.

For wGEBV, the estimation of effect of each marker was the same as for GEBV. The selection criterion was

$$g_i = \sum_{j=1}^{p} x_{ij} \beta_j w_D,$$

where $g_i$ is wGEBV of individual $i$, $x_{ij}$ indicates the genotype at marker $j$ in individual $i$, and is coded as 0, 1 or 2 depending on the number of copies of allele "1", $\beta_j$ is the estimated marker effect and $w_D$ is the weight for each marker.

**Table 1. Overview of selection schemes.**

| criterion \ rule | Truncation selection | Optimum contribution selection | |
|---|---|---|---|
| | | $A$ [A] | $G$ |
| **GEBV** | *GEBV* | *OCSA* | *OCSG* |
| **wGEBV** | *wGEBV* | *wOCSA* | *wOCSG* |

[A] The relationship matrix used to constrain the co-ancestry between animals ($A$=numerator relationship matrix and $G$=genomic relationship matrix)

To obtain $w_D$, we used the probability density function of the beta distribution $\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$. The shape parameters $\beta$ and α were set to 1 and 0.2, given previously empirical results. Then, the weight used for marker $j$ was $w_{Dj} \sim f(p_j; \alpha + t * \frac{(1-\alpha)}{N}, 1)$, where $p_j$ is the frequency of favorable allele of marker $j$, $t$ is the generation in which selection is done and N is the generation for which the accumulated genetic gain is to be maximized (40 in the current study). Whether a marker allele is favorable or not was determined by the sign of the estimated marker effect. As $\alpha + t * \frac{(1-\alpha)}{N}$ is approaching 1 (t -> N), $w_{Dj}$ approaches 1 for each allele, which means that selection is to maximize the genetic response in the next generation (short-term gain).

**Selection rules:** With TS, the best 5 males and 25 females among all selection candidates in each generation were selected according to the selection criterion. With OCSA, the average co-ancestry was computed from pedigree data. The contribution of animals in generation t ($G_t$) to average relationship in $G_{t+1}$ was presented as a function of current genetic contributions $c$, $c'Ac$, where $A$ is the numerator relationship matrix constructed based on the full pedigree. This represents the average relationship of parents weighted by their contributions. The index $H_A$ to be maximized was:

$$H_A = w_{merit} * c' * \hat{\alpha} + w_{rel} * c' * A * c,$$

where $w_{merit}$ is the weight on the genetic merit, $\hat{\alpha}$ is a vector of estimated GEBV using BL, and $w_{rel}$ is the weight on the average additive genetic relationship. Optimum genetic contributions were computed by maximizing $H_A$ using an evolutionary algorithm, as implemented in the program EVA (Berg et al. 2007). The values of $w_{merit}$ and $w_{rel}$ were set to 1 and -5 respectively, which was determined by the decision space spanned by different values of $w_{rel}$ (results not shown). The solution of the optimization was given as a number of matings for each selected candidate. The number of males resulted from using OCSA was somewhat higher than 5. An alternative

index, referred to as OCSG, was similar to $H_A$, except that the relationship among selection candidates was represented by genomic relationship $G$ matrix (Yang et al. 2010). The $G$ matrix in $G_t$ was constructed based on genotypes of animals in $G_{t-1}$ and $G_t$.

**Genome structure and trait simulation:** Details regarding the historical population and processes of assigning markers to chromosomes are described in (Liu et al. 2014). To analyze the hitch-hiking effect, a single quantitative trait locus (QTL) was positioned at 50 cM on chromosome 1. For the remaining four chromosomes, 50 QTL (minor allele frequency>0.05) were randomly assigned to the non-marker positions. Additive effect of each QTL ($a_j$) was sampled from a gamma distribution $\Gamma(1.48,11.2)$ and scaled to ensure a heritability ($h^2$) of 0.1. In addition, and independent of markers, a set of 100 tags were evenly placed across each chromosome of base animals, i.e. two alleles at any locus for animal $i$ were tagged as $i, -i$ to infer IBD status directly from homozygosity. The environmental terms were drawn from a random normal distribution $N$ (0, $1 - h^2$) and were added to the TBV to obtain the phenotypic record of each animal.

**Data Analysis:** Genetic gain, genetic variance and inbreeding were recorded every generation. The differences between schemes with respect to those records were tested using ANOVA and HSD Tukey test.

## Results and Discussion

**Cumulative genetic gain:** Among all selection schemes, *wOCSG* led to the highest cumulative genetic gain over 40 generations (Table 2). In general, OCS schemes always resulted in higher cumulative genetic gain compared to TS schemes at the expense of slightly lower short-term gain. Specifically, compared to *GEBV*, *OCSG* increased the cumulative genetic gain by 35%, followed by *OCSA* which increased the gain by 25%. Compared to *wGEBV*, *wOCSG* increased the cumulative genetic gain by 19%, which was lower than the increase from *GEBV* to *OCSG*.

**Table 2. Genetic response and genetic variance in the last generation**

| Selection scheme | Gain_1 [A] | Gain_40 [B] | Genetic Variance [C] |
|---|---|---|---|
| **Truncation selection** | | | |
| *GEBV* | 0.0351[ab] | 0.435[e] | 0.00001[c] |
| *wGEBV* | 0.0322[b] | 0.495[d] | 0.0002[b] |
| **Optimum contribution selection** | | | |
| *OCSA* | 0.0377[ab] | 0.542[c] | 0.0001[b] |
| *wOCSA* | 0.0398[a] | 0.590[b] | 0.0005[b] |
| *OCSG* | 0.0398[a] | 0.587[b] | 0.0024[ab] |
| *wOCSG* | 0.0383[ab] | 0.604[ab] | 0.0046[a] |

Different superscripts in the same column show significant differences (p<0.05)
[A] Genetic gain in $G_1$, representing short-term genetic gain
[B] Cumulative genetic gain $G_{40}$, representing long-term genetic gain
[C] Genetic variance in $G_{40}$

Effects of using different selection criteria and selection rules appeared non-additive, as shown in Table 2. The interaction was lower in OCSG schemes than OCSA schemes, as it benefits less from using wGEBV, probably because both OCSG and wGEBV were based on the same data. The animals carrying the rare alleles were less related to others, which can be realized by $G$ matrix. But OCSG itself had no consideration on whether the rare allele is the good one or not, whereas wGEBV takes it into account. Then the families selected based on GEBV or wGEBV would be different. Furthermore, the rare segments contribute to $G$ matrix can be diluted by the family share, so the emphasis on rare alleles in $OCSG$ is not as effective as $wOCSG$. Therefore, we still observed a higher genetic variance and genetic gain in $wOCSG$ than in $OCSG$.

**Hitch-hiking:** In general, the effect of hitch-hiking led to an asymmetric pattern where a distinct peak was shown at the position of the QTL in chromosome 1 (Figure 1). Among schemes using GEBV as criterion, $GEBV$ showed highest $F_{IBD}$ across chromosome 1, followed by $OCSA$ and $OCSG$. The footprint of selection for $OCSA$ was broad, which indicates that the use of OCSA did not restrict inbreeding in the region surrounding the QTL, in line with Sonesson et al. (2012). In contrast, $OCSG$ managed to select animals carrying favorable alleles and meanwhile maintaining genetic variation for the remaining segments of the genome. This pattern is expected, because the selection on the favorable alleles would not be allowed to affect the allele frequencies of other regions (e.g. deleterious genes) due to hitch-hiking. Using wGEBV further decreased proportion of IBD across the genome relative to GEBV, which was consistent with the increased variance observed in Table 1.
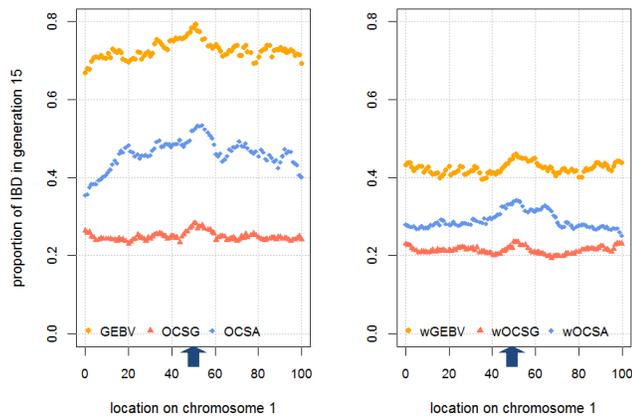


**Figure 1. The hitch-hiking of QTL along chromosome 1 in generation 15.** The QTL was located in the middle of the chromosome, which is marked by an arrow. The type and the color of lines in the two plots were the same for the pair of strategies that share the same selection rule but being with different selection criteria (e.g. $OCSA$ vs $wOCSA$)

**Rate of pedigree, genomic and true inbreeding:** In all selection schemes, rate of true inbreeding at QTL ($\Delta F_{qtl}$) was higher than at tags ($\Delta F_{true}$). The difference was more pronounced with $OCSA$, which was in line with the pattern of hitch-hiking on chromosome 1 (Figure 1). Genomic inbreeding tended to overestimate true inbreeding in all selection schemes to some extent. In $OCSA$ and $wOCSA$, penalizing co-ancestry by $A$ was only able to constrain the pedigree inbreeding, but due to substantial hitch-hiking, the true inbreeding was consequently much higher than expected. In contrast, penalizing co-ancestry by $G$ constrained genomic inbreeding successfully constrained true inbreeding. This was because $G$ led to accurate estimates of true allele sharing, and thereby selection tends to act on the most different animals within the family (results not shown).

### Conclusion

The strategy combining wGEBV with OCS was very promising, as it provided higher gain and/or lower true inbreeding than using each of them alone in genomic breeding programs. It is recommended to use wOCSG, as not only did it boost the cumulative genetic gain, but also it restricted increase in true inbreeding across the genome.

**Table 3. The mean rate of inbreeding over first 15 generations measured by pedigree information, genomic markers and IBD at tags and QTL.**

| Selection scheme | $\Delta F_{ped}$ | $\Delta F_{gen}$ | $\Delta F_{true}$ | $\Delta F_{qtl}$ |
|---|---|---|---|---|
| **Truncation selection** | | | | |
| GEBV | 0.068[a] | 0.116[a] | 0.098[a] | 0.106[a] |
| wGEBV | 0.054[b] | 0.056[c] | 0.052[b] | 0.057[c] |
| **Optimum contribution selection** | | | | |
| OCSA | 0.028[c] | 0.078[b] | 0.055[b] | 0.063[b] |
| wOCSA | 0.026[c] | 0.043[d] | 0.031[c] | 0.034[d] |
| OCSG | 0.026[c] | 0.030[e] | 0.026[d] | 0.029[ef] |
| wOCSG | 0.025[c] | 0.024[e] | 0.021[e] | 0.023[f] |

Different superscripts in the same column show significant differences (p<0.05)
The numbers with grey shading mean the category of inbreeding that is constrained by OCS

### Literature Cited

Berg, P., Nielsen, J., and Sorensen, M. K. (2006). *WCGALP*, CD-ROM communication no. 27-09

Dekkers, J. C. M., and van-Arendonk, J. A. M. (1998). *Genet. Res.,* 71: 257-275.

Goddard, M. (2009). *Genetica*, 136: 245-257.

Jannink, J. L. (2010). Dynamics of long-term genomic selection. *Genet. Sel. Evol.,* 42:35.

Liu, H. , Sorensen, A. C., Meuwissen, T. H. et al. (2014) *Genet. Sel. Evol.,* 46:8.

Meuwissen, T. H. (1997). *J. Anim. Sci.,* 75: 934-940.

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

Yang J., Benyamin, B., McEvoy, B. P. Et al. (2010). *Nat. genet.,* 42: 565-569.