Optimum contribution selection in a dairy cattle population with different relationship matrices

E. Gautason[1*], G. Sahana[1], B. Guldbrandtsen[2] and P. Berg[1,3]

[1] Aarhus University, Center for Quantitative Genetics and Genomics, Blichers Allé 20, 8830 Tjele, Denmark; [2] University of Copenhagen, Dept. of Veterinary and Animal Sciences, Grønnegårdsvej 3, 1870 Frederiksberg C, Denmark; [3] Norwegian University of Life Sciences, Department of Animal and Aquacultural Sciences, Oluf Thesens vei 6, 1433 Ås, Norway; [*]egill@au.dk

Abstract

We compared the effects of using three genomic relationship matrices and a pedigree based relationship matrix for optimum contribution selection (OCS) with a constraint on number of sires selected, on genetic gain, kinship and additive genetic variance. This was done by simulating a genomically selected small dairy cattle population. The cattle population underwent genomic selection for 11 generations. In total, 6,000 females and 2,500 males were genotyped per generation. We compared a pedigree-based matrix as well as VanRaden's method 1 to construct genomic relationship matrices using either allele frequencies of the base population, of the current generation, or allele frequencies among all genotyped animals. The pedigree-based OCS resulted in higher kinship and more loss of genetic variance than using genomic OCS based on base population frequencies. Thus, genomic OCS with genomic relationship matrices using base population frequencies are preferable compared with pedigree-based OCS.

Introduction

Genomic selection has increased genetic gain in dairy cattle in recent years, but in some cases at the cost of higher inbreeding rates (Makanjuola *et al*. 2020). Small local breeds that are not genomically selected have become increasingly less competitive to the transboundary breeds. However, simulations have shown that small dairy cattle populations can also benefit from genomic selection (Thomasen *et al*. 2014; Obšteter *et al*. 2019), but research is lacking on ways to balance genetic gain and inbreeding in small populations. To achieve genetic gains through selective breeding in future generations, genetic diversity must be preserved. Inbreeding has to be managed to avoid inbreeding depression. This is not least the case in small populations. Optimum contribution selection (OCS; Meuwissen 1997) maximizes genetic gain while restraining inbreeding. OCS has been implemented using pedigree relationships (POCS), but genomic OCS (GOCS) can be an alternative by using a genomic relationship matrix (GRM). A common way to form GRMs is VanRaden's method one (VanRaden 2008). This method scales the cross-product of centered genotype scores by $\sum 2p(1\text{-}p)$, where p is the marker allele frequency. The choice of reference allele frequencies to use is an important parameter to consider when constructing GRMs. VanRaden (2008) recommended the use of base population frequencies. Berg et al. (subm.) found important differences in results depending on the type of reference used for GOCS. The aim of this study was to compare four scenarios, three different allele frequencies to construct GRMs and the pedigree based relationship matrix, on genetic gain, kinship and additive genetic variance in a small dairy cattle population undergoing genomic selection.

Materials & Methods

The simulations were modelled to mimic the breeding program of the Icelandic Cattle population (Sigurdsson and Jonmundsson 2011). QMSim (Sargolzaei and Schenkel 2009) was used for simulating a base population. The R package Modular Breeding Program Simulator (MoBPS; Pook *et al*. 2020) was used for breeding program simulations. We used GMATRIX (Su and Madsen 2012) to construct GRMs. EVA (Berg *et al*. 2006) was used to optimize genetic contributions of selection candidates. DMU (Madsen and Jensen 2013) was used to estimate breeding values.

Historical population. We simulated a genome of 29 chromosomes each with 2,000 evenly spaced loci. The mutation rate was 2.5×10^{-5} per locus per generation. For each replicate of scenarios, a new historical population was simulated. The population size was 1,000 individuals for 2,000 generations and increased over 200 generations to 12,000. Genotype data were converted into Plink *ped* format (Purcell *et al*. 2007) and read into MoBPS. We randomly sampled 2,000 segregating loci to be quantitative trait loci (QTL), which were not used for construction of GRM. We assigned effects to the QTL by drawing from a gamma distribution with shape 0.4 and scale parameter 1.66. We discarded very low frequency marker loci so that the genotyping markers reflected the allele frequency distribution of commercial SNP chips. The number of polymorphic genotyping marker loci were around 39,000 at the start of the genomic selection, with some variation across replicates. We simulated a sex-based trait with a constant heritability of 0.4. We used three replicates for each scenario.

Breeding program structure and breeding value prediction. We simulated four discrete generations of pedigree-based best linear unbiased prediction (PBLUP) selection before genomic selection started. We then simulated three generations of genomic prediction using single-step genomic BLUP (ssGBLUP), and subsequently eight generations of GBLUP. Each generation contained 6,000 males and 6,000 females. Each cow had exactly two calves. In each generation of genomic selection, 2,500 male and 6,000 female selection candidates were selected for genotyping based on their parent average and their genomic breeding values (GEBVs) were estimated. Selection was according to optimum contribution selection, only on the sire side, with additional constraints described below. Dams of selection candidates had phenotypes at the stage of selection. The model for genetic evaluation included the additive genetic animal effect, an intercept and a residual: $\mathbf{y=Xb+Za+e}$, where $\mathbf{y}$ was a vector of animal phenotypes, $\mathbf{b}$ was a vector of fixed effects, $\mathbf{a}$ was a vector of (genomic) estimated breeding values where $\mathbf{a}$ followed $N(\mathbf{0}, \mathbf{A}\sigma_g^2)$, $N(\mathbf{0}, \mathbf{H}\sigma_g^2)$ and $N(\mathbf{0}, \mathbf{G}\sigma_g^2)$ for PBLUP, ssGBLUP and GBLUP, where $\mathbf{A}$ was the numerator relationship matrix, $\mathbf{H}$ was a combined pedigree and genomic relationship matrix (Christensen and Lund 2010) and $\mathbf{G}$ was a GRM computed using VanRaden's method one. $\mathbf{X}$ and $\mathbf{Z}$ were design matrices and $\mathbf{e}$ was a vector of random residuals following $N(\mathbf{0}, \mathbf{I}\sigma_e^2)$. We used the true values of variance components.

Optimum contribution selection. To constrain the number of sires selected, we constructed a 'pseudo-female' out of all female selection candidates. The pseudo-female had a relationship with itself that was equal to the mean relationship among female selection candidates, and a relationship with each bull that was equal to the mean relationship of all female candidates to that bull. The pseudo-female was assigned 30 matings, and each bull was allowed one mating. The 30 selected bulls were then randomly mated to females and used equally. The genetic contributions were optimized to achieve the same target rate of inbreeding, while maximizing genetic gain, and selecting a fixed number of bulls, and using them equally. Genetic contributions c were optimized in each generation to maximize the genetic level in the offspring generation, G: $\mathbf{G=c'\hat{a}}$ where $\hat{\mathbf{a}}$ was a vector of GEBVs, subject to the constraint: $\mathbf{c'Rc=C}$ where

R was a relationship matrix, and the target rate of average coancestry was set to 0.005, and the constraint that the sum contributions of males and females each equaled ½ (Meuwissen 1997; Woolliams *et al*. 2015).
We used the first method of VanRaden (2008) to compute GRMs: $\mathbf{G} = \mathbf{ZZ}^{\mathbf{T}}/(2\sum p_j(1-p_j))$, where $\mathbf{Z} = \mathbf{M} - \mathbf{P}$ where **M** is the genotypic matrix for genotyped animals for which rows are genotypes with values 2 and 0 for the two homozygotes, and 1 for heterozygotes; **P** is a matrix where all elements in the jth column are $2p_j$, where p_j is the frequency of the allele at locus j that was counted in **M**. We used three estimates of p_j: *Base* used the allele frequencies in the base population, *Recent* used genotyped animals in the current generation, and *All* used all genotyped animals to estimate p_j. Additionally, we used a pedigree-based relationship matrix (*Pedigree*). Thus the study composed of four scenarios of OCS: *All*, *Base*, *Recent* and *Pedigree*.

Statistical analysis. We computed the mean true breeding value and estimated mean identical-by-descent (IBD) based genomic kinship for the selection candidates in each generation. Kinship 'represents the value of chromosome segments of two different individuals to be IBD at a random position when randomly drawing one of the two haplotypes of each individual' (T. Pook, pers. comm.). We used the following linear model for generations six to 16:

$$y = Replicate + Scenario + Scenario*Generation + error \quad (1)$$

where y was breeding value, kinship or additive genetic variance. We used the standard errors of the linear regression coefficients for significance testing. The interaction term of scenario and generation gave a regression coefficient nested within scenario.

Results

Figure 1 shows the average results of the simulation from generation 1 to 16. Genetic gain was 0.827, 0.846, 0.818 and 0.835 genetic standard deviations; kinship rate was 0.81%, 0.84%, 0.73% and 0.98%; and loss of additive genetic variance was -3.0%, -3.0%, -2.9% and -3.6% per generation, for *All*, *Recent*, *Base* and *Pedigree*, respectively, according to the linear model. Pedigree-based OCS resulted in significantly higher rate of increase in kinship than GOCS ($P<0.01$), and *Base* resulted in significantly lower kinship increase than *Recent* and *Pedigree* ($P<0.05$). *Pedigree* resulted in significantly more loss of genetic diversity than genomic OCS ($P<0.05$). The difference in genetic gain among scenarios was not statistically significant.

Discussion

The results suggest that genomic OCS can outperform pedigree-based OCS, contrary to the conclusion of Henryon *et al*. (2019), but agreeing with the recommendation of Sonesson *et al*. (2012). Henryon *et al*. (2019) assessed inbreeding rate at a set of IBD loci not used for prediction, while we assessed true kinship using recombination points and founder chromosome segments. The faster build-up of kinship, and more rapid loss of genetic diversity observed with POCS, suggests that genomic OCS is preferable to pedigree-based OCS under the conditions in this simulation. Our results agree with results of Berg *et al*. (*subm.*). They found that distant reference frequencies for GOCS resulted in less build-up of kinship than using recent frequencies. We simulated a population with number of markers and animals resembling a feasible breeding scheme for the Icelandic dairy cattle population. This included simulating a breeding scheme that used a fixed number of bulls and equal use of these bulls. While this is a deviation from the theory of optimum contribution selection, we believe that this resembles a practical and cost-effective way for genomic breeding of a small dairy cattle population, since number of AI bulls is a limiting factor in a small closed population. Our results suggest that in such settings, GOCS should be conducted, using old animals to compute reference allele frequencies for the GRM.

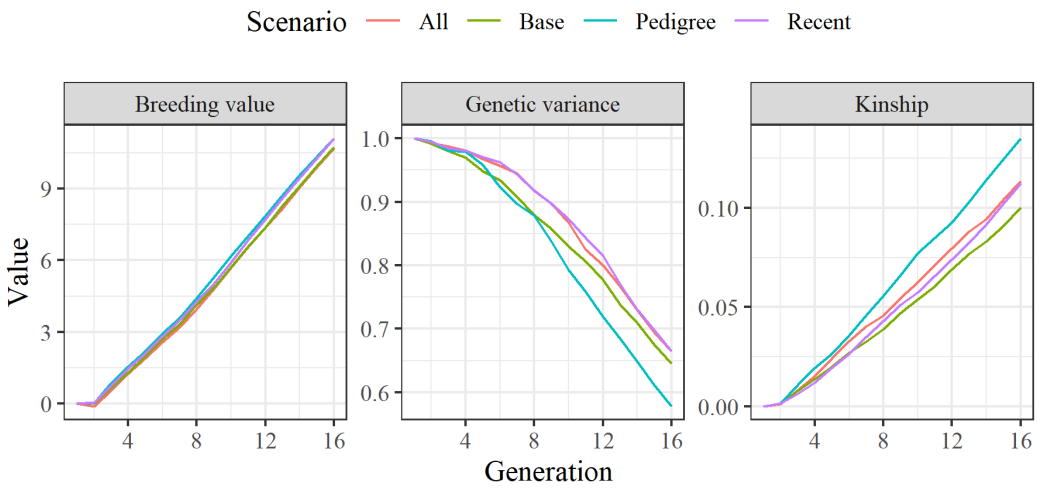

Figure 1. Mean breeding value, additive genetic variance and kinship. Genomic selection started in generation six. Breeding value is in units of additive genetic standard deviation, genetic variance is the proportion of variance in the base population and kinship is based on genomic identity-by-descent.

References

Berg P., Nielsen J., and Sørensen M.K. (2006) Proc. of the 8th WCGALP. Belo Horizonte, Minas Gerais, Brazil.

Christensen, O.F., and Lund M.S. (2010) Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2

Henryon M., Liu H., Berg P., Su G. et al. (2019) Genet. Sel. Evol. 51(1):39. https://doi.org/10.1186/s12711-019-0475-5

Madsen P., and Jensen J. (2013) DMU: A Package for Analysing Multivariate Mixed Models.

Makanjuola B.O., Miglior F., Abdalla E.A., Maltecca C. et al. (2020) J. Dairy Sci. 103(6):5183–5199. https://doi.org/10.3168/jds.2019-18013

Meuwissen T.H.E. (1997) J. Anim. Sci. 75(4):934–940. https://doi.org/10.2527/1997.754934x

Obšteter J., Jenko J., Hickey J.M., and Gorjanc G. (2019) J. Dairy Sci. 102(11):9971–9982. https://doi.org/10.3168/jds.2019-16853

Pook T., Schlather M., and Simianer H. (2020) G3: Genes Genom. Genet. 10(6):1915–1918. https://doi.org/10.1534/g3.120.401193

Purcell S., Neale B., Todd-Brown K., Thomas L. et al. (2007) Am. J. Hum. Genet. 81(3):559–575. https://doi.org/10.1086/519795

Sargolzaei M., and Schenkel F.S. (2009) Bioinformatics. 25(5):680–1. https://doi.org/10.1093/bioinformatics/btp045

Sigurdsson A., and Jonmundsson J.V. (2011) Icel. Agric. Sci. 24(2011):55–64.

Sonesson A.K., Woolliams J.A., and Meuwissen T.H. (2012) Genet. Sel. Evol. 44:27. https://doi.org/10.1186/1297-9686-44-27

Su G., and Madsen P. (2012) User's guide for GMATRIX Ver2.

Thomasen J.R., Sørensen A.C., Lund M.S., and Guldbrandtsen B. (2014) J. Dairy Sci. 97(9):5822–5832. https://doi.org/10.3168/jds.2014-7906

VanRaden P.M. (2008) J. Dairy Sci. 91(11):4414–23. https://doi.org/10.3168/jds.2007-0980

Woolliams J.A., Berg P., Dagnachew B.S., and Meuwissen T.H.E. (2015) J. Anim. Breed. Genet. 132:89–99. https://doi.org/10.1111/jbg.12148