Maximising Genetic Gain Whilst Controlling Rates Of Genomic Inbreeding Using Genomic Optimum Contribution Selection

A.K. Sonesson[*], J.A. Woolliams[†‡] and T.H.E Meuwissen[§]

Introduction

For genomic breeding values, the effects of dense genetic markers are first estimated in a test population and later used to predict breeding values of selection candidates (Meuwissen et al, 2001). Genomic breeding values generally have higher accuracy than conventional BLUP breeding values (e.g. Goddard, 2009), because the genetic markers provide a more accurate relationship matrix than the pedigree based matrix. For example, the relationship between two full-sibs is 0.5 when based on pedigree, but markers may show that the true relationship is between 0 and 1.

Optimum contribution selection (Meuwissen, 1997, Grundy et al., 1998) is a selection method that maximises genetic gain while restricting the rates of inbreeding of the progeny by restricting the relationship of the parents. Until now, the pedigree based relationship matrix has been used to restrict the rates of inbreeding, which constrains the inbreeding at a neutral locus that is not linked to any QTL. It may be questioned whether such a locus exists, especially since genomic selection results suggest that there is a large number of QTL in the genome (Luan et al., 2009). Thus, using genomic relationships may yield a more precise control of the genomic inbreeding.

The aim of this study is to compare genetic gain, pedigree and genomic based rates of inbreeding when using pedigree based or genomic optimum contribution selection (GOCS) and where in both cases selection is for genome wide EBVs (GWEBV). GOCS constrained the average genomic Identity-by-Descent (IBD), but local rates of inbreeding were also studied. The breeding design resembled a sib-test design as commonly used for aquaculture species.

[*] Nofima Marin, Post Box 5010, 1430 Ås, Norway
[†] The Roslin Institute & R(D)SVS, University of Edinburgh, Roslin, Midlothian EH25 9PS, United Kingdom.
[‡] University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway

Material and methods

Simulation of populations. Briefly, a base population with effective population size of 1000 was simulated for 4000 generations. Details are described in Sonesson and Meuwissen (2009). The genome consisted of 10 pairs of chromosomes, each of size 1M. Randomly 100 SNPs were selected as QTL, if MAF>.05, and the 100 SNPs with highest MAF were selected as markers. QTL-effects were sampled from a gamma-distribution with a shape parameter of 0.4 and a scale parameter of 1.66 (Hayes and Goddard, 2001). Total genetic variance was standardized to 10. In addition, 100 artificial IBD markers were positioned at equal distances at each chromosome. These IBD markers were not involved in the selection in any way, but were assigned unique founder alleles in generation G0, in order to monitor the increase of the genomic IBD at these positions. 100 sires and 100 dams from generation 4000 were randomly selected to create generation G0, consisting of 3000 or 6000 selection candidates (Ncand), which were genotyped, and 3000 or 6000 test-full-sibs (Ntest), which were phenotyped and genotyped. In later generations, selection was by GOCS. GWEBV were predicted by BLUP (Meuwissen et al., 2001).

Genomic optimum contribution selection. The optimum contribution selection algorithm of Meuwissen (1997) was used, ie. genetic level of next generation of animals, $\mathbf{G_{t+1}} = \mathbf{c_t'GWEBV_t}$, was maximised, where $\mathbf{c_t}$ is a vector of genetic contributions of the selection candidates to generation t+1. Rates of inbreeding are restricted by constraining the average coancestry of the selection candidates to $\overline{C}_{t+1} = \mathbf{c_t'A_tc_t}/2$, where $\mathbf{A_t}$ is a (n x n) relationship matrix among the selection candidates, $\overline{C}_{t+1} = 1-(1-\Delta F_d)^t$, and ΔF_d is the desired rates of inbreeding (Grundy et al., 1998) being 0.005 or 0.010 per generation. The relationship matrix is here either based on pedigree data or on genomic data, in which case it equals $\mathbf{XX'}$ (Goddard, 2009), where $\mathbf{X}$ is a matrix of marker genotypes (X_{ij} is $-2p/\sqrt{[p(1-p)]}$, $1-2p/\sqrt{[p(1-p)]}$, or $2(1-p)/\sqrt{[p(1-p)]}$ for '0 0', '1 0' or '1 1' gentypes). Parents of candidates (and 1 or 2 testsibs) were sampled with replacement according to probabilities $\mathbf{c_t}$.

Results and discussion

Table 1 shows that ΔF_d was achieved on the scale it was constrained (pedigree or genomic). However, when ΔF_{ped} was constrained $\Delta F_{genomic}$ was 3 times too high. Furthermore, ΔG was 37-60% higher than when $\Delta F_{genomic}$ was constrained. When ΔF_{ped} was constrained, ΔG was higher (+~60% when ΔF_d =0.005 and +37% when ΔF_d =0.010), because $\Delta F_{genomic}$ was increased.

Table 1: Genetic gain (ΔG), rate of inbreeding based on pedigree (ΔF_{ped}) and on genomic IBD ($\Delta F_{genomic}$)[1] relationship matrices at generation G10 when either ΔF_{ped} or $\Delta F_{genomic}$ was constrained. Results are based on average of 100 replicates.

Ncand	Ntest	ΔF_d	ΔG (se)	ΔF_{ped} (se)	$\Delta F_{genomic}$ (se)
				ΔF_{ped} constrained	
3000	**3000**	**0.005**	3.08 (0.035)	0.0050 (0.0001)	0.0151 (0.001)
3000	**6000**	**0.005**	3.10 (0.035)	0.0048 (0.0001)	0.0162 (0.001)
3000	**6000**	**0.010**	3.31 (0.037)	0.0098 (0.0003)	0.0333 (0.002)
				$\Delta F_{genomic}$ constrained	
3000	**3000**	**0.005**	1.91 (0.033)	0.0041 (0.0001)	0.0051 (0.0002)
3000	**6000**	**0.005**	1.95 (0.024)	0.0039 (0.0001)	0.0052 (0.002)
3000	**6000**	**0.010**	2.41 (0.028)	0.0071 (0.0002)	0.0100 (0.0004)

[1] ΔF_{genome} is based on the increase of the IBD as monitored by the IBD markers.

Figure 1 shows that schemes that constrained ΔF_{ped}, genomic IBD is substantially more variable across the genome, than when $\Delta F_{genomic}$ was constrained. It seems that when constraining ΔF_{ped} the optimum contribution algorithm tries to find ways to increase the frequency of the largest QTL as quickly as possible, and uses deficiencies in the pedigree based relationship to do so. Furthermore, it seems that constraining $\Delta F_{genomic}$, results in a quite evenly distributed increase of the IBD across the genome, such that putting extra IBD constraints in regions with large QTL is not needed, i.e. constraining the average genomic IBD across the genome seems to suffice.

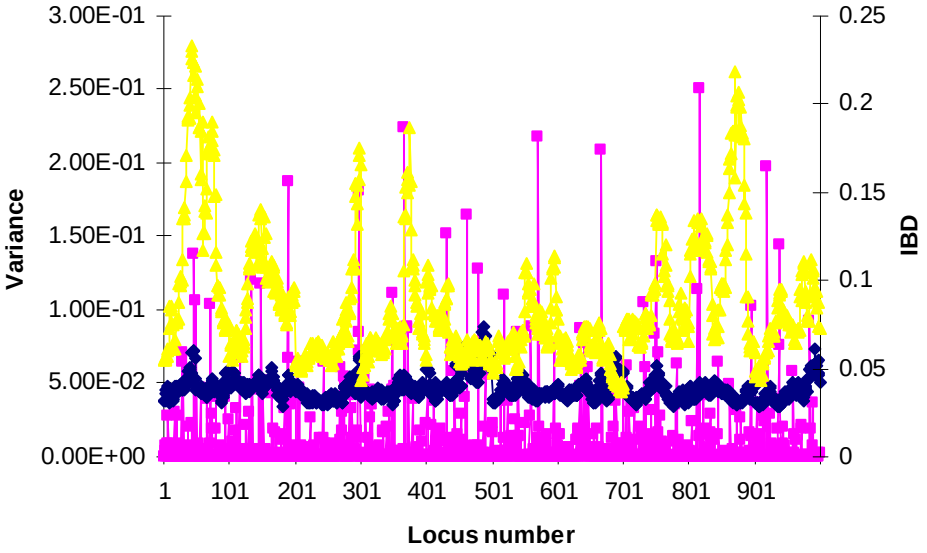

Figure 1: Variance explained by QTL in generation G0 (pink) and IBD in generation G10 when ΔF_{ped} (yellow) or $\Delta F_{genomic\ (blue)}$ was constrained. Results are from one replicate with Ncand=3000, Ntest=3000 and ΔF_d=0.005. IBD across the genome was monitored by the IBD markers.

Conclusion

The results show that when genomic methods are used with traits that have many QTL contributing to the variance the concept of inbreeding at a neutral locus is no longer tenable. Therefore this requires a re-consideration on what is an appropriate rate of loss of diversity when directly measured in the genome. Previous selection methods could have promoted similar rates of loss of diversity on the genomic scale, but the control measures applied were on a scale based on pedigree and neutral loci, and resulted in a much more variable loss of diversity compared to constraining the $\Delta F_{genomic}$.

References

Grundy, B., Villanueva, B and Woolliams, J.A. (1998) *Genet. Res.* 72: 159-168

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

Hayes, B.J. and Goddard, M.E. (2001) *Genet. Sel. Evol.* 33: 209-229

Luan, T., Woolliams, J.A. Lien, S, Kent, M., Svendsen, M. and Meuwissen, T.H.E. (2009) *Genetics* 183: 1119-1126

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001). *Genetics* 157: 1819-1829

Sonesson, A.K. and Meuwissen, T.H.E. (2009). *Genet. Sel. Evol*. 41:37