Controlling Inbreeding with Genomic Selection

Jack Dekkers and Sam Clark

Inbreeding can be managed by incorporating genomic information into Optimal Contribution selection strategies.

One simple way of using genomic information to manage inbreeding within the breeding program is to constrain mating's based on the genomic relationships between individuals.

There are many ways to build the genomic relationship matrix (GRM) here we will show two common methods.

The first was proposed by VanRaden (2008).
It uses an incidence matrix **M**, which specifies which alleles each individual inherited, coded -1, 0, 1. A second matrix **P** is defined such that it contains the allele frequencies expressed as a difference from 0.5 and multiplied by 2, such that column i of **P** is 2(pi - 0.5) where the frequency of the second allele at locus i is pi. The subtraction of **P** from M gives Z, which sets the expected value of u to 0. Subtraction of **P** gives more credit to rare alleles than to common alleles when calculating genomic relationships. Therefore **G** = **ZZ'**/[2∑pi(1 - pi)]. The division by 2∑pi (1 - pi) makes **G** similar to the numerator relationship matrix (**A**).

Other genomic matrices have been proposed, such as the one used by Yang et al. (2010). They combined A_{ijk} for all of the SNPs using a similar weighting scheme to that proposed by VanRaden. The main differences are that they weight the off-diagonal and diagonal elements differently, when j≠k then (off-diagonals);

$$A_{jk} = \frac{1}{N}\sum\nolimits_i A_{ijk} = \frac{1}{N}\sum\nolimits_i \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1-p_i)}$$

When j is equal to k (diagonals), then;

$$A_{jk} = \frac{1}{N}\sum\nolimits_i A_{ijk} = 1 + \frac{1}{N}\sum\nolimits_i \frac{x_{ij}^2 - (1-2p_i)x_{ij} + 2p_i^2}{2p_i(1-p_i)}$$

These estimates of relationship are all relative to a base population in which the average relationship between individual is zero (all individuals are completely unrelated). Although there are slight differences between the two methods they result in very similar predictions. The genomic relationship matrix can be incorporated into an optimal contribution selection strategy to manage inbreeding at the genomic level.

Observations based on Sonesson, Woolliams, and Meuwissen. 2012. Genomic selection requires genomic control of inbreeding. Genetics Selection Evolution 44:27

Measures of inbreeding:

F_{ped} = level of inbreeding based on pedigree

F_{IBD} = level of inbreeding at IBD markers (based on average homozygosity of unique founder alleles across genome)

Truncation selection on pedigree-based BLUP (TBLUP), GEBV from GBLUP or BayesB.

Table 1 Truncation selection on breeding values estimated using TBLUP or GBLUP

Breeding value estimation	ΔG (se)	ΔF_{ped} (se)	ΔF_{IBD} (se)
TBLUP	2.49 (0.035)	0.0156 (0.0001)	0.0235 (0.0009)
GBLUP	2.77 (0.026)	0.0053 (0.0002)	0.0209 (0.0005)
BayesB	2.73 (0.027)	0.0053 (0.0001)	0.0235 (0.0005)

Genetic gain (ΔG) and rate of inbreeding based on pedigree (ΔF_A) or on genomic IBD (ΔF_G) relationship matrices at generation *G10* with truncation selection and either GBLUP or BayesB breeding value estimates[a].
[a]Number of test sibs = 3000; number of selection candidates = 3000.

With selection: $\Delta F_{ped} < \Delta F_{IBD}$

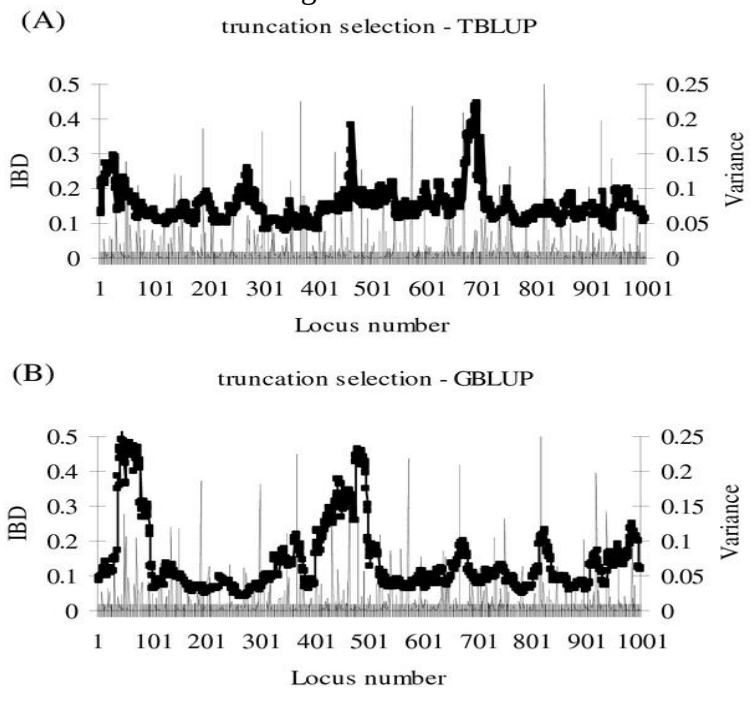
ΔF_{ped} much lower for GBLUP than TBLUP

ΔF_{IBD} only slightly lower for GBLUP than TBLUP

With GS, $\Delta F_{IBD} >> \Delta F_{ped}$

Higher genome-wide IBD profile with TBLUP than GBLUP (Fig.1)

Figure 1. IBD for one replicate of truncation selection with 3000 selection candidates and 3000 test sibs**.**
Thin lines = Variance (σ_g^2-unit) explained by QTL in *G0*
Thick lines = IBD in generation *G10*



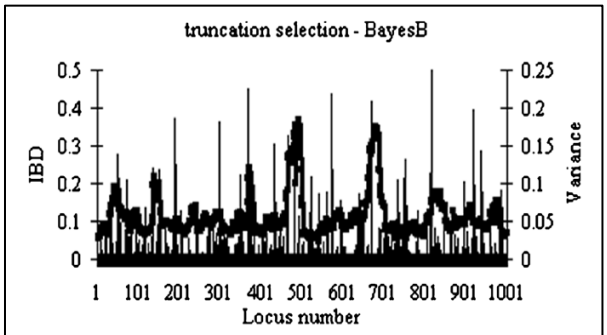


Figure 4 Identity-by-Descent for one replicate of truncation selection on breeding values estimated by BayesB. Variance (σ_g-unit^2) explained by QTL in generation *G0* (—) and IBD (—) in generation *G10* with truncation selection and BayesB estimated breeding values; results are from one replicate with 3000 selection candidates and 3000 test sibs.

Should inbreeding be controlled using pedigree or genomic relationships?

Pedigree inbreeding evaluates inbreeding at a neutral locus that is not linked to any QTL.

- Does such a locus exists? All loci are linked to QTL?
- Control inbreeding by controlling genome-based inbreeding?
  - Also as a way to manage footprints of selection?

Optimal contribution selection based on pedigree (Meuwissen, 1997):

$$\text{Max } \bar{g}_{t+1} = \mathbf{c}_t' \hat{\mathbf{g}}_t \qquad \text{Subject to } \mathbf{Q}'\mathbf{c}_t = \frac{\mathbf{1}}{\mathbf{2}}$$

$$\tfrac{1}{2}\mathbf{c}_t' \mathbf{A}_t \mathbf{c}_t = \bar{C}_{t+1}$$

$\bar{g}_{t+1}$ = mean BV in the next generation

$\hat{\mathbf{g}}_t$ = the vector of BLUP EBV of candidates for selection in generation t

$\mathbf{c}_t$ = a vector of contributions of selection candidates to the next generation

$\mathbf{Q}$ = a known incidence matrix for sex (the first column contains one for male candidates and the second column one for female candidates)

$\frac{\mathbf{1}}{\mathbf{2}} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$ and ensures that contributions of males and of all females sum to ½.

$\mathbf{A}_t$ = **pedigree-based** relationship matrix among selection candidates in generation t.

$\bar{C}_{t+1}$ = average **pedigree-based** coancestry among all progeny in generation t+1

= ½ weighted average genetic relationship among selected parents = $\tfrac{1}{2}\mathbf{c}_t' \mathbf{A}_t \mathbf{c}_t$.

= set equal to $\Delta F(t+1)$ when objective is to restrict rate of inbreeding per generation to ΔF and generation 0 is non-inbred

To control genomic inbreeding, replace pedigree-based relationship matrix, $\mathbf{A}_t$, by genomic relationship matrix, $\mathbf{G}_t$.

Scenarios: ΔF constraint set based on pedigree (ΔF_A constraint) or genomics (ΔF_G constraint)
Selection on GBLUP or TBLUP

Table 2 Optimum contribution selection on breeding values estimated using TBLUP or GBLUP

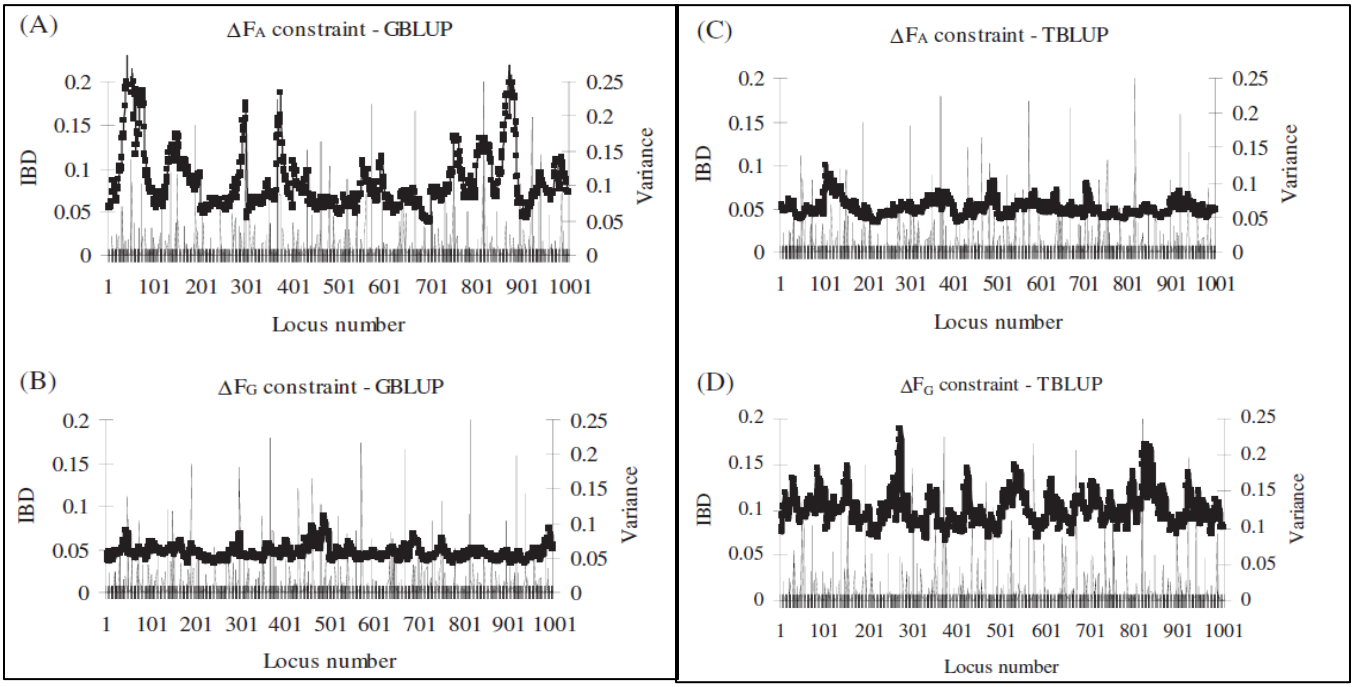
Ntest	ΔF_d	ΔG (se)	ΔF_{ped} (se)	ΔF_{IBD} (se)
		ΔF_A constraint – GBLUP		
3000	0.005	3.08 (0.035)	0.0050 (0.0001)	0.0211 (0.0004)
6000	0.005	3.10 (0.035)	0.0048 (0.0001)	0.0226 (0.0004)
6000	0.010	3.31 (0.037)	0.0098 (0.0003)	0.0422 (0.0008)
		ΔF_G constraint – GBLUP		
3000	0.005	1.91 (0.026)	0.0041 (0.0001)	0.0051 (0.0001)
6000	0.005	1.95 (0.024)	0.0039 (0.0001)	0.0053 (0.0001)
6000	0.010	2.41 (0.028)	0.0071 (0.0002)	0.0102 (0.0002)
		ΔF_A constraint – TBLUP		
3000	0.005	2.26 (0.003)	0.0050 (0.0001)	0.0068 (0.0001)
6000	0.005	2.50 (0.003)	0.0049 (0.0001)	0.0074 (0.0001)
6000	0.010	2.63 (0.003)	0.0102 (0.0002)	0.0151 (0.0003)
		ΔF_G constraint – TBLUP		
3000	0.005	1.41 (0.041)	0.0193 (0.0004)	0.0121 (0.0002)
6000	0.005	1.44 (0.039)	0.0185 (0.0004)	0.0122 (0.0002)
6000	0.010	1.48 (0.046)	0.0300 (0.0008)	0.0183 (0.0003)

Genetic gain (ΔG), rate of inbreeding based on pedigree (ΔF_{ped}) and on genomic IBD (ΔF_{IBD}) relationship matrices at generation *G10* when the constraint on relationship was either pedigree-based (ΔF_A) or marker-based (ΔF_G) with TBLUP or GBLUP breeding value estimates[a].

[a]*Ntest* = number of test sibs; ΔF_d = desired rates of inbreeding; number of selection candidates = 3000.

- **Constraining pedigree inbreeding** (ΔF_A constraint) constrains pedigree inbreeding for both TBLUP and GBLUP but does not constrain genomic inbreeding. $\Delta F_{ped} < \Delta F_{IBD}$
- **Constraining genomic inbreeding** (ΔF_G constraint) constrains ΔF_{IBD} only for GBLUP
  - → pedigree-inbreeding < desired with selection on GBLUP
  - → pedigree- and genomic inbreeding > desired with selection on TBLUP
- **Selection on TBLUP**
  - ⇨ Flat IBD profile when constraining on pedigree inbreeding (ΔF_A constraint)
  - ⇨ Variable IBD profile when constraining on genomic inbreeding (ΔF_G constraint)
- **Selection on GBLUP**
  - ⇨ Variable IBD profile when constraining on pedigree inbreeding (ΔF_A constraint)
  - ⇨ Flat IBD profile when constraining on genomic inbreeding (ΔF_G constraint)

Figure 2. IBD for one replicate of optimum contribution selection



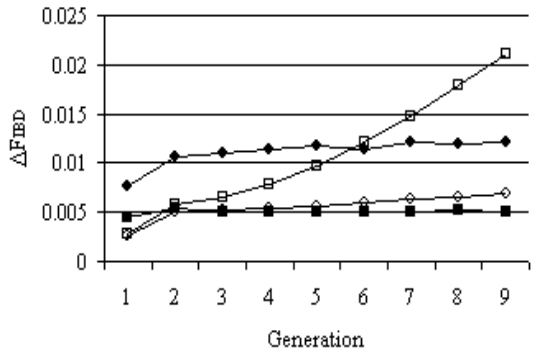


Figure 3. Rates of genomic inbreeding over generations

□ = ΔF_A constraint and GBLUP - unstable

• = ΔF_G constraint and TBLUP – stable but too high

◦ = ΔF_A constraint and TBLUP - stable

▪ = ΔF_G constraint and GBLUP - stable

3000 selection candidates and 3000 test sibs and a desired rate of genomic inbreeding of 0.005.

Ideally: increase IBD at QTL but with limited footprint of selection
→ sharp IBD peaks at QTL

ΔF_G constraint with GBLUP → very flat profile → spread-out selection pressure

Observations based on Clark, Hickey, Kinghorn and van der Werf. 2013. The effect of genomic information on optimal contribution selection in livestock breeding programs. Genetics Selection Evolution 45:44

Why is inbreeding lower when truncation selection is performed on different breeding values?
What selection space is available?
How does this effect within and between family selection?

If truncation selection is performed the correlation between the breeding values of relatives is important. Animals with high BV's tend to be related and therefore tend to be selected.

Table 1- Intra class correlation between breeding values from the half sib and full sib populations

Breeding value type	Half sib correlation	Full Sib correlation	Accuracy
PA EBV	0.55	1.0	0.45
GEBV	0.50	0.85	0.57
TBV	0.26	0.53	1.0

Table 2- The proportion of variation in breeding value explained by between family (Sire and Dam) and within family (MS) information (some real data).

	LIC					*ADHIS*			
BV	**Sire**	**Dam**	**MS+e**	***Prop. of PT***	**BV**	**Sire**	**Dam**	**MS+e**	***Prop. of PT***
PA EBV	0.56	0.44	0.001	0.001	**PA EBV**	0.44	0.52	0.04	0.05
GEBV	0.43	0.26	0.31	0.56	**GEBV**	0.33	0.37	0.30	0.36
PT	0.21	0.31	0.48	1.0	**PT**	0.16	0.32	0.52	1.0

In the Tables above all full siblings would have the same breeding value estimated using pedigree (PA EBV). More selection space is available when using GEBV for selection between full siblings (in this case a lower correlation is better). Variation in GEBV is greater because they utilize some information other than that from direct relatives. Utilizing this variation, which is likely to be due to Mendelian sampling (Daetwyler et al. 2007), allows for lower inbreeding because the breeding values of relatives become more different when the amount of MS variance explained becomes higher. Consequently as breeding values become more accurate it is easier to select between close relatives.

Therefore genomic information helps to manage inbreeding in two ways:

1. Using genomic relationships helps to restrict genomic inbreeding. This enables greater response to selection especially when selecting from large full sib families.

2. Using GEBV's allow for great selection space because they contain information about Mendelian sampling.

References

1. Sonesson, Woolliams, and Meuwissen. 2012. **Genomic selection requires genomic control of inbreeding.** Genetics Selection Evolution 44:27
2. Yang, Benyamin, McEvoy et al. (2010) **Common SNPs explain a large proportion of the heritability for human height.** Nature Genetics 42,565-571.
3. VanRaden. 2008. **Efficient methods to compute genomic predictions**. Journal of Dairy Science 91, 4414-4423.
4. Daetwyler, Villanueva, Bijma and Woolliams. 2007. **Inbreeding in genome-wide selection.** Journal of Animal Breeding and Genetics 124, 369–376
5. Meuwissen. 1997. **Maximizing the response of selection with a predetermined rate of inbreeding**. Journal of Animal Science. 75, 934–940
6. Clark, Hickey, Kinghorn and van der Werf. 2013. **The effect of genomic information on optimal contribution selection in livestock breeding programs.** Genetics Selection Evolution 45:44