Sonesson et al. Genetics Selection Evolution 2012, 44:27
http://www.gsejournal.org/content/44/1/27

GSE Genetics Selection Evolution

RESEARCH Open Access

Genomic selection requires genomic control of inbreeding

Anna K Sonesson[1*], John A Woolliams[2,3] and Theo HE Meuwissen[2]

Abstract

Background: In the past, pedigree relationships were used to control and monitor inbreeding because genomic relationships among selection candidates were not available until recently. The aim of this study was to understand the consequences for genetic variability across the genome when genomic information is used to estimate breeding values and in managing the inbreeding generated in the course of selection on genome-enhanced estimated breeding values.

Methods: These consequences were measured by genetic gain, pedigree- and genome-based rates of inbreeding, and local inbreeding across the genome. Breeding schemes were compared by simulating truncation selection or optimum contribution selection with a restriction on pedigree- or genome-based inbreeding, and with selection using estimated breeding values based on genome- or pedigree-based BLUP. Trait information was recorded on full-sibs of the candidates.

Results: When the information used to estimate breeding values and to constrain rates of inbreeding were either both pedigree-based or both genome-based, rates of genomic inbreeding were close to the desired values and the identical-by-descent profiles were reasonably uniform across the genome. However, with a pedigree-based inbreeding constraint and genome-based estimated breeding values, genomic rates of inbreeding were much higher than expected. With pedigree-instead of genome-based estimated breeding values, the impact of the largest QTL on the breeding values was much smaller, resulting in a more uniform genome-wide identical-by-descent profile but genomic rates of inbreeding were still higher than expected based on pedigree relationships, because they measure the inbreeding at a neutral locus not linked to any QTL. Neutral loci did not exist here, where there were 100 QTL on each chromosome. With a pedigree-based inbreeding constraint and genome-based estimated breeding values, genomic rates of inbreeding substantially exceeded the value of its constraint. In contrast, with a genome-based inbreeding constraint and genome-based estimated breeding values, marker frequencies changed, but this change was limited by the inbreeding constraint at the marker position.

Conclusions: To control inbreeding, it is necessary to account for it on the same basis as what is used to estimate breeding values, i.e. pedigree-based inbreeding control with traditional pedigree-based BLUP estimated breeding values and genome-based inbreeding control with genome-based estimated breeding values.

Background

Traditional pedigree-based BLUP (Best Linear Unbiased Prediction) estimated breeding values (EBV) [1] are based on pedigree information and recordings of selection candidates and relatives, e.g. sibs of candidates, as in aquaculture breeding schemes, where many traits (e.g. disease resistance and fillet quality) cannot be measured on the candidates. 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 [2]. Genome-based EBV, i.e. EBV based on high-density marker data across the genome, generally have higher accuracy than pedigree-based BLUP EBV, because genetic markers provide a more accurate relationship matrix than pedigree [3], which is based on expected genetic relationships. For example, the expected relationship between two full-sibs is 0.5 but markers show that the true relationship deviates from 0.5 [4] and varies among pairs of sibs, depending on the

* Correspondence: anna.sonesson@nofima.no
[1]Nofima AS, 1431 Ås, Norway
Full list of author information is available at the end of the article

segregation of the parental chromosomes. However, the increased accuracy of genome-based EBV can differ between methods used to estimate them and, e.g., the number of genes affecting the trait. The genomic BLUP methodology has shown highest accuracy for traits without large quantitative trait loci (QTL) but the BayesB method has shown highest accuracy for traits with known large QTL, because it puts higher weight on genetic markers with large effects [5,6].

Optimum contribution selection [7,8] is a selection method that maximises genetic gain while restricting the rates of inbreeding in the progeny by restricting relationships between selected parents. Until now, pedigree-based relationship matrices have been used to control inbreeding rates, which constrain inbreeding rates at a neutral locus that is not linked to any QTL. It may be questioned whether such a locus exists, especially since genomic selection and other studies suggest that most traits are affected by a large number of QTL across the genome [9-12]. Thus, using genomic relationships may help to better control genome-based inbreeding, and may provide a tool for breeders to manage footprints of selection [13-16].

The aim of this study was to understand the consequences for genetic variability across the genome when genomic information is used to estimate EBV and in managing the inbreeding generated in the course of selection on genome-enhanced EBV. The consequences are measured by genetic gain and the pedigree- and genome-based rates of inbreeding and local inbreeding across the genome. Breeding schemes are compared by simulating truncation selection or optimum contribution selection with a restriction on pedigree- or genome-based inbreeding, and with selection on genome- or pedigree-based BLUP EBV. The trait under selection is a trait for which information on selection candidates comes from full-sibs, which provides a challenging test for developing theory, because the use of genome-based EBV is most beneficial in this situation [3]. In addition, these so-called sib-tests are commonly applied in practical breeding schemes.

Methods

Simulation of populations

A base population with an effective size of 1000 was simulated for 4000 generations. Details are described in [17]. One hundred sires and 100 dams from generation 4000 were randomly selected to create generation *G0*, consisting of 3000 selection candidates (*Ncand*) and 3000 or 6000 test sibs (*Ntest*). In later generations (*G1-G10*), selection was done by truncation or optimum contribution selection. Inbreeding coefficients based on pedigree (F_{ped}) and rates of inbreeding based on pedigree (ΔF_{ped}) assumed that the *G0* individuals were unrelated base individuals.

Simulation of the genomes has been described elsewhere [17]. The genome consisted of 10 pairs of chromosomes (1Morgan each). All polymorphisms were generated during the 4000 generations of the Fisher-Wright population model [18,19]. The infinite sites mutation model [20] was used to create new bi-allelic single nucleotide polymorphisms (SNP), using a mutation rate of 10^{-8} per nucleotide and assuming 10^6 nucleotides per cM. Inheritance of the SNP followed Mendel's law and the Haldane mapping function [21] was used to simulate recombination. One hundred SNP per chromosome were sampled randomly without replacement from SNP with a minor allele frequency (MAF) > 0.05 and used as QTL, i.e. the total number of QTL was 1000. From the remaining SNP, 1000 SNP with the highest MAF over all chromosomes were chosen as genetic markers. In addition, 100 artificial identical-by-descent (IBD) markers were positioned at equal distances on each chromosome. These IBD markers were not involved in selection, but were assigned unique founder alleles in generation *G0*, in order to monitor the increase of the local genomic IBD at these positions.

Additive effects of the QTL alleles were sampled from a gamma distribution with a shape parameter of 0.4 and a scale parameter of 1.66 [9]. The QTL effects were assumed to be either positive or negative with a probability of 0.5 because the gamma distribution only gives positive values. After sampling, these QTL effects were standardized so that the total genetic variance was 1.

Calculation of phenotypic values and true and estimated breeding values

The true breeding value (*TBV*) of an individual was calculated as:

$$TBV_i = \sum_{j=1}^{1000} \left(x_{ij1} g_{j1} + {}_{xij} g_{j2} \right)$$

where x_{ijk} is the number of copies of the k^{th} allele that individual i has at the j^{th} QTL position and g_{jk} is the effect of the k^{th} allele at the j^{th} position. The phenotypic values, y_i, of individuals in the sib-test were simulated by:

$$y_i = TBV_i + \varepsilon_i$$

where ε_i is an error term for animal i, which was normally distributed with mean zero and variance σ_e^2, which was adjusted so the heritability was 0.4.

Marker effects, $\hat{a}_j$, were predicted using the genome-based BLUP method described in [2], named GBLUP hereafter. The statistical model used was:

$$y_i = \mu + \sum_j^n X_{ij} a_j \; + \; e_i$$

where y_i is the record of test sib i, μ is the overall mean, n is the total number of markers, X_{ij} denotes the

standardised marker genotype, a_j is the random effect of the j^{th} marker and Var(a_j) is assumed $1/n$ since the total genetic variance was standardised to 1, e_i is a random residual. X_{ij} was standardised to a mean of 0 and a variance of 1:

$$X_{ij} = -2p_j/\sqrt{H_j}$$

denotes that the individual is homozygous for the first allele; $(1\text{-}2p_j)/\sqrt{H_j}$ denotes that it is heterozygous; and $(2\text{-}2p_j)/\sqrt{H_j}$ denotes that it is homozygous for the second allele, where H_j is the marker heterozygosity and p_j is the frequency of the second allele. Division by $\sqrt{H_j}$ results in every SNP explaining an equal amount of variance a priori (independent of the frequency of the SNP). The genetic variance explained by a SNP is $R^{2*}V_{QTL}$, where R^2 is the linkage disequilibrium between the SNP and the nearest QTL, and V_{QTL} is the variance due to this QTL. Division by $\sqrt{H_j}$, avoids making the assumption that SNP with high H_j have proportionally higher R^2, which would especially be questionable in a situation where most QTL have a low frequency, as is the case when the QTL alleles are in mutation-drift equilibrium and even more so if the QTL are under selection. However, QTL with low allele frequencies are expected to have lower V_{QTL}. Here, none of the markers used had low H_j, thus correction of X_{ij} by $\sqrt{H_j}$ is expected to affect the results only marginally.

Genome-based BLUP EBV (*GEBV*) were estimated by summing across the estimated effects of the markers:

$$GEBV_j = \sum_j^n X_{ij}\hat{a}_j$$

In addition, *GEBV* were calculated with method BayesB, as described in [2]. BayesB uses the same statistical model as GBLUP but attempts to reduce the weight of SNP that are estimated to have no association with QTL. It also assumes a priori that many SNP have no effect and few SNP (1000 here) have t-distributed effects.

Traditional pedigree-based BLUP EBV (*TEBV*) were estimated with the method described in [1], named TBLUP hereafter, in which genetic relationships are based on pedigree information. Pedigree recording started in generation *G0*.

Optimum contribution selection and mating

The optimum contribution selection algorithm of [7] was used, i.e. the genetic level of the next-generation animals, $g_{t+1} = \mathbf{c_t}'\mathbf{EBV_t}$ ($\mathbf{EBV_t}$ contains either *GEBV* or *TEBV*), was maximised, where $\mathbf{c_t}$ is a vector of genetic contributions of the selection candidates to generation $t+1$. Rates of inbreeding were restricted by constraining the average relationship of the selection candidates to $\bar{C}_{t+1} = \mathbf{c_t}'\mathbf{A_t}\mathbf{c_t}/2$, where $\mathbf{A_t}$ was a relationship matrix among the selection candidates, $\bar{C}_{t+1} = 1 - (1 - \Delta F_d)^t$, and ΔF_d was the desired rate of inbreeding [8], i.e. 0.005 or 0.010 per generation. The relationship matrix was either based on pedigree or genomic data. For the latter, it equalled $\mathbf{G_t} = \mathbf{X_t}\mathbf{X_t}'/n$ [3].

Having calculated the optimum contribution vector $\mathbf{c_t}$, the next generation of offspring were produced by sampling a male and a female parent with replacement, according to the probabilities given by $2\mathbf{c_t}$, which resulted in random mating. One hundred full-sib families were created each generation, from *G1* to *G10*. Each family was split into 30 selection candidates and 30 or 60 test sibs. The test sibs were recorded for the trait.

Recording

For the schemes using *TEBV*, test sibs were only phenotyped, while for those using *GEBV*, test sibs were phenotyped and genotyped to estimate the SNP effects and selection candidates were genotyped. This sib-test provided a challenging test for the management of genetic variation but it is also very relevant as it in such circumstances that the use of genomic data in breeding value estimation is of greatest value.

Truncation selection and mating

For a simple comparison of TBLUP and GBLUP, truncation selection was used instead of optimum contribution selection. Each generation, 100 sires and 100 dams were selected from 3000 selection candidates on their breeding values estimated either from TBLUP or GBLUP. These sires and dams were pair-wise mated to produce 100 full-sib families for the next generation, using sampling without replacement.

Calculation of genomic identity-by-descent

Genomic IBD was obtained by calculating the allele frequencies of the founder alleles at the IBD markers, i.e. f_{ij} for founder allele j at IBD marker i. Their homozygosity, i.e. probability of IBD was then calculated for IBD marker i as $\Sigma_j {f_{ij}}^2$. To calculate ΔF_{IBD}, this IBD probability was averaged over all IBD markers to evaluate the overall IBD over the genome in generations*G0* to *G10*. IBD profiles differed between replicates because the position and size of the QTL differed.

Statistics

TBLUP and GBLUP schemes were compared using either truncation selection or optimum contribution selection. In the latter case, the constraint was based on relationships derived either from pedigree (ΔF_A) or from markers (ΔF_G). The schemes were run for ten generations (*G1*-*G10*) and summary statistics for each of the

schemes were based on 100 replicated simulations. The breeding schemes were compared for rates of inbreeding per generation (ΔF) and genetic gain (ΔG), expressed in genetic standard deviation units of generation *G0* (σ_a) in generation *G10*. The rates of inbreeding per generation were calculated in two ways, either from the pedigree ΔF_{ped} or using the IBD loci ΔF_{IBD}. Here, the symbol ΔF_A denotes a constraint on pedigree-based relationships and thus on inbreeding, and ΔF_{ped} denotes the rate of inbreeding as calculated from the pedigree that results from the breeding scheme. Inbreeding coefficients at each IBD locus were also stored to analyse inbreeding rates as a function of locus and generation.

Results

Truncation selection with TBLUP and GBLUP breeding values

The truncation selection schemes evaluated constitute classical selection with comparisons made at constant selection intensity and scheme size. For these schemes, genetic gain was 11% higher for GBLUP than for TBLUP and, although ΔF_{ped} was much lower for GBLUP than TBLUP, ΔF_{IBD} was only slightly lower for GBLUP than TBLUP (Table 1). Rate of inbreeding measured by ΔF_{IBD} was 51% greater than inbreeding measured by ΔF_{ped} for TBLUP schemes but 292% greater for GBLUP schemes (Table 1). Thus, with GBLUP schemes the increase in genomic inbreeding was well above the increase in pedigree inbreeding. A higher genome-wide IBD profile was obtained with TBLUP than with GBLUP (Figure 1). These results show the importance of considering the basis for constraining rates of inbreeding.

Optimum contribution selection with TBLUP and GBLUPEBV

Table 2 shows that, with the current practice of explicitly constraining ΔF_A, the desired rate of inbreeding, ΔF_d, was observed in ΔF_{ped} but not when the constraint was implemented based on ΔF_G. Table 2 also shows that when selection was on TBLUP, the observed ΔF_{IBD} substantially exceeded ΔF_d.

When ΔF_A was constrained, ΔG was substantially greater with GBLUP than with TBLUP, by ~ 35% when *Ntest* = 3000 and by ~25% when *Ntest* = 6000. Due to the inadequacy of constraining ΔF_A, ΔF_{IBD} increased above ΔF_d even more strongly with GBLUP than with TBLUP (by ~320-360% with GBLUP and ~35-50% with TBLUP). The greatest increase in $\Delta FIBD$ was observed when *Ntest* = 6000.

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)

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

When ΔF_G was constrained, ΔG was again greater with GBLUP than with TBLUP, however in this case, the magnitudes of the increase depended on ΔF_d, i.e. by ~35% when ΔF_d was set at 0.005 and *Ntest* = 3000 and by ~65% when ΔF_d was set at 0.010. When applying the constraint using ΔF_G with TBLUP, ΔF_{IBD} substantially exceeded the desired ΔF_d. The observed ΔF_{ped} was even more extreme, and was ~50% greater than ΔF_{IBD}. However, when applying the constraint using ΔF_G with GBLUP, ΔF_{ped} was ~70-80% of the desired ΔF_d.

With GBLUP, schemes that constrained ΔF_A showed a substantially more variable IBD profile across the genome (Figure 2A), than those that constrained ΔF_G (Figure 2B). In contrast, with TBLUP, schemes that constrained ΔF_A showed little variation in the genome-wide IBD profile (Figure 2C), while those that constrained ΔF_G showed a very erratic profile (Figure 2D).

Schemes that constrained ΔF_G showed a constant ΔF_{IBD} over generations but at a higher level than the constraint when selection was on TBLUP (Figure 3). Schemes that constrained ΔF_A showed an increase in ΔF_{IBD} over generations, in particular when selection was on GBLUP. This increase in ΔF_{IBD} over generations is probably due to fixation of favourable alleles, which occurs faster with GBLUP. This increased ΔF_{IBD} over time also suggests that the constraint on ΔF_A becomes less restrictive over time when selection uses *GEBV*.

Use of *GEBV* derived using BayesB showed very similar results as using *GEBV* from GBLUP in terms of ΔG and accuracy of selection (Table 3). BayesB had significantly higher ΔF_{IBD} (0.0235 compared to 0.0209), which can be explained by a larger focus on some SNP, which increased local IBD values but also the overall ΔF_{IBD}. These results are in concordance with [5], who found that BayesB is advantageous for traits with a few large QTL and many smaller QTL. Here, 1000 QTL were simulated, which disadvantaged BayesB compared to GBLUP. With BayesB, the IBD profile had a few IBD peaks but was generally quite uniform (Figure 4).

Discussion

Several methods for the management of ΔF_A have been suggested in the literature [7,8,13,14,16]. With the development of genomic selection, genomic relationships have become available in addition to the traditional pedigree-based relationships. An immediate question then is whether the constraint should be based on pedigree (ΔF_A) or genomic (ΔF_G) relationships, when

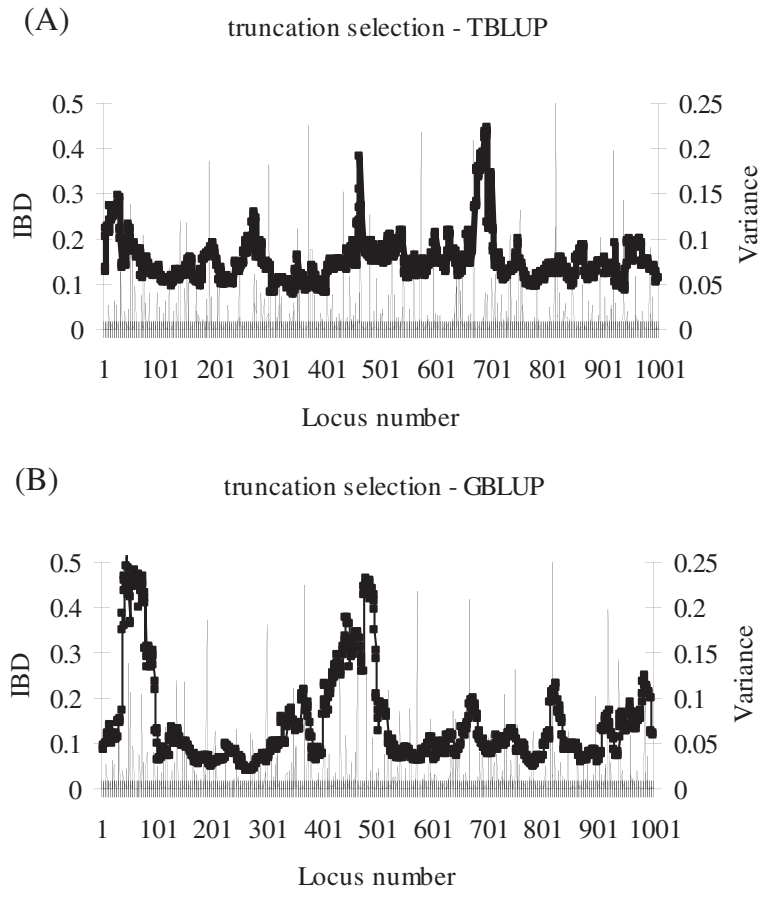

Figure 1 Identity-by-Descent for one replicate of truncation selection on breeding values estimated by TBLUP or GBLUP. Variance (σ^2_g-unit) explained by QTL in generation *G0* (—) and Identity-by-Descent (IBD; ▬) in generation *G10* with truncation selection and (**A**) TBLUP (BLUP based on pedigree relationships) or (**B**) GBLUP (BLUP based on genomic relationships); results are from one replicate with 3000 selection candidates and 3000 test sibs.

combined with breeding value estimation based on pedigree (TBLUP) or genomic (GBLUP) information. Our results address this question by showing that if the information sources used to estimate breeding values and to constrain ΔF are identical, i.e. either both pedigree-based or both genomic-based, then the resulting rates of genomic inbreeding correspond to the desired values and the rates are reasonably uniform across the genome. However, if different information is used to calculate breeding values and to constrain ΔF, e.g., the EBV are based on genomic information and the ΔF constraint is on pedigree information, the resulting rate of inbreeding, based on ΔF_{IBD}, is much higher than expected.

It is important to recognise the distinction between the three relationships measures between two individuals u and v that are considered in this study: a_{uv}, the numerator relationship derived from pedigree; g_{uv}, the identity by state relationship obtained from the markers accumulated over time and used to estimate *GEBV*; and t_{uv}, the relationship between the identity-by-descent markers defined in the base population used for reference. EBV were estimated either using a_{uv} (TBLUP) or g_{uv} (GBLUP) and rates of inbreeding were controlled using either a_{uv} or g_{uv}, since these values are accessible and known. However, the examination of the development of relationships over time is based on a_{uv} or t_{uv} since these measure IBD with respect to the base generation *G0*, with the objective of controlling the drift in an unknown locus from the start of selection. Using a_{uv} for this control will be predictive and unbiased for a neutral locus that is unlinked to variants with an effect, whereas t_{uv} is an empirical value that is not restricted by neutrality or position of the locus. Using g_{uv} to control inbreeding has the advantage of acting upon relationships that already exist in generation *G0*, whereas both a_{uv} and t_{uv} assume that *G0* animals are unrelated.

Interpretation of the consequences of these differences in breeding value estimation, control of inbreeding and assessment of IBD may be helped by considering the impacts when u and v are full-sib candidates. For the

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.

simulated sib-tested trait, the distinctions between EBV estimated using TBLUP versus GBLUP are clear, since u and v will have the same EBV with TBLUP but information on Mendelian sampling terms to differentiate the sibs is available with GBLUP. For the same reason, full-sibs u, v and w will be treated as having equal pair-wise relationships when measured using a_{uv} but different pair-wise relationships when measured using g_{uv}.

The incentive to include genomic data into breeding schemes comes from its use in breeding value estimations. With GBLUP, the results of Table 2 are interpretable from the framework outlined in the previous paragraph. When ΔF_A is constrained, the relationship of two selected individuals u and v will be greater than their pedigree relationship since their high merit implies that on average they are more likely to share QTL, as well as flanking segments that, over time, will tend to become more homozygous along with the QTL. Thus the genome will contain segments where segregation is not free from the influence of selection and where there is more homozygosity than predicted by a_{uv} and so $E[t_{uv}|a_{uv}] > a_{uv}$, i.e. when we condition on or constrain a_{uv}, we expect t_{uv} to exceed a_{uv}. Hence, constraining ΔF_A will result in underestimating ΔF_{IBD}. The extent of this underestimation will depend on the density of the QTL across the genome and the linkage disequilibrium between the QTL. In a similar way, when ΔF_G is constrained, $E[a_{uv}|g_{uv}] < g_{uv}$, and so ΔF_{ped} will be less than the target value, whereas $E[t_{uv}|g_{uv}] \sim g_{uv}$ in this context.

If breeding value estimation is based on TBLUP then truncation selection will select whole full-sib families but not with selection with optimized contributions. When inbreeding management is based on ΔF_A, then $E[t_{uv}|a_{uv}] > a_{uv}$ after selection, for the same reasons as before, i.e. large segments are not free of the influence of selection, and ΔF_{IBD} is greater than the desired rate of inbreeding. With TBLUP, the impact on ΔF_{IBD} is not as large as when using GBLUP since the QTL are not identified as accurately and response to selection is less. The most challenging outcomes occur when ΔF_G is constrained. In this case, two full-sibs might be selected if they appear less related based on markers than expected based on pedigree. On the one hand, $E[t_{uv}|g_{uv}]$ is greater than g_{uv} since the prediction errors of t_{uv} from g_{uv} will be positive because variation not explained by g_{uv} is more likely to reflect the full-sib pedigree relationship. On the other hand, $E[a_{uv}|g_{uv}]$ will be much greater than g_{uv} since the genomic relationships of those sibs that are selected will be below average and thus lower than a_{uv}. Hence, ΔF_{ped} will be greater than ΔF_{IBD}. It is notable that the combination of TBLUP with ΔF_G delivers the least gain for close to the highest rates of inbreeding, by either measure (see Table 2). Although this combination may not be practical, it is instructive.

The dynamics of ΔF_{IBD} over time can also be explained within this framework. The only combination without a stable trend in inbreeding was when ΔF_A was used as the constraint with GBLUP; all other

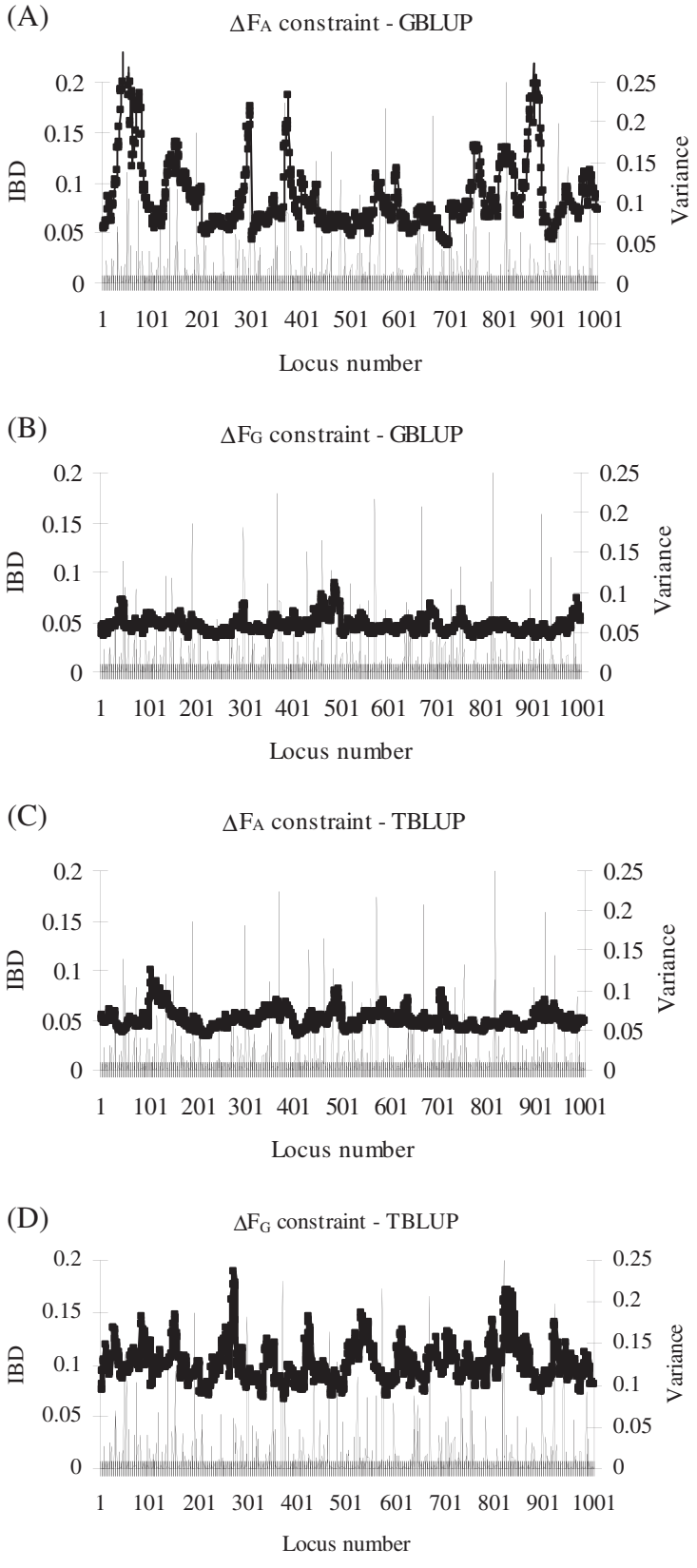

Figure 2 (See legend on next page.)

(See figure on previous page.)
Figure 2 Identity-by-Descent for one replicate of optimum contribution selection on breeding values estimated by TBLUP or GBLUP. Variance (σ^2_g-unit) explained by QTL in generation *G0* (—) and Identical-by-descent (IBD; —) in generation *G10* with ΔF_A constraint based on pedigree relationships and GBLUP (BLUP based on genomic relationships) (**A**), ΔF_G constraint and GBLUP (**B**), ΔF_A constraint and TBLUP (BLUP based on pedigree relationships) (**C**) or ΔF_G constraint based on marker relationships and TBLUP (**D**); results are from one replicate with 3000 selection candidates and 3000 test sibs and a desired rate of inbreeding of 0.005.

combinations showed a stable ΔF_{IBD} (Figure 3). With GBLUP and ΔF_A, ΔF_{IBD} continuously increased during generations *G1* to *G10* (i.e. the rate of inbreeding increased), probably because the estimates of marker effects are more persistent over generations than the pedigree relationships, implying that changes across many generations of the frequencies of (selected) chromosome segments are not picked up by pedigree relationships. As a result, GEBV favour specific chromosome segments generation after generation, thereby increasing ΔF_{IBD} but not ΔF_{ped}. Thus, especially in the longer term, discrepancies in the information used to estimate the breeding values and to control inbreeding will hamper the control of inbreeding.

Some of the outcomes observed here depend on the number and distribution of effects of the QTL: we simulated 1000 QTL on a genome of 10 M. With a much smaller number of QTL, i.e. when the genome is predominantly composed of neutral loci that are weakly linked to QTL, ΔF_{IBD} is expected to be closer to ΔF_{ped}. However, increasing evidence from well-studied traits such as human height indicates that many traits are composed of many QTL each explaining a small part of the variance [12]. Trait heritability was rather high here (0.4). With a lower trait heritability, the number of sib-tested animals needs to be increased in order to obtain similar accuracies as found in the present study [22].

Inbreeding at QTL positions is desirable in breeding schemes because this increases the frequency of the positive alleles towards homozygosity. However, narrowing the genomic IBD peak at the QTL positions is also desired, so that the remaining genome is as little as possible affected by selection at individual QTL, i.e. the footprint of selection should be as small as possible. Based on the hitchhiking effect [23], a broad genomic IBD peak around the largest QTL would have been expected. However, when GBLUP was used and ΔF_G was constrained, the genome-wide IBD profile was rather flat (Figure 1B), which suggests that this selection method spreads the selection pressure quite evenly over many loci in order to control ΔF_{IBD}. This implies that, in practical breeding schemes, it is not necessary to implement additional constraints on the genomic inbreeding at positions surrounding large QTL in order to avoid excessive rates of genomic inbreeding in QTL regions.

A reduction of the footprint of selection may also be achieved by increasing the frequencies of a broad spectrum of QTL alleles slowly, instead of heavily selecting on the biggest QTL with the danger of a large selection footprint. The latter seems to have occurred with genomic optimum contribution selection with estimation based on GBLUP, since the increase of the IBD was rather flat across the genome for that scenario (Figure 1B). The GBLUP method assumes that all SNP explain the same genetic variance ($1/n$, i.e. 5000 here). Use of BayesB resulted in an IBD profile with more peaks (Figure 4) but not to the extent that local restrictions of inbreeding would be required. Further research is needed to investigate whether in situations with a few large QTL and using BayesB, directed measures would be required to reduce the footprint of selection in the regions of large QTL, and how such measures should be implemented. In case the SNP effects are estimated accurately and do not change over time, a model that

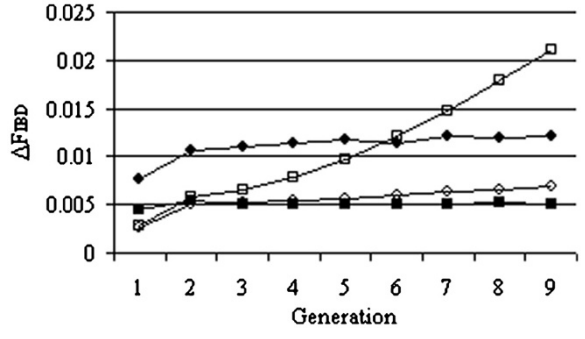

Figure 3 Rates of genomic inbreeding over generations. Rates of inbreeding based on genomic IBD (ΔF_{IBD}) over generations for schemes with ΔF_A constraint based on pedigree relationships and GBLUP (BLUP based on genomic relationships)(□), ΔF_G constraint based on genomic relationships and GBLUP (■), ΔF_A constraint and TBLUP (BLUP based on pedigree relationships) (◦) or ΔF_G constraint and TBLUP (•); results are from the scheme with 3000 selection candidates and 3000 test sibs and a desired rate of inbreeding of 0.005.

Table 3 Truncation selection on breeding values estimated using GBLUP or BayesB

Breeding value estimation	ΔG (se)	ΔF_{ped} (se)	ΔF_{IBD} (se)
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.

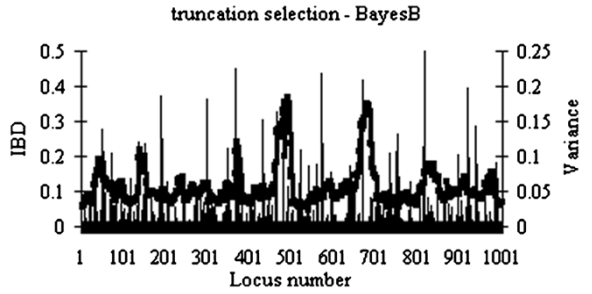

Figure 4 Identity-by-Descent for one replicate of truncation selection on breeding values estimated by BayesB. Variance (σ_g-unit^2) explained by QTL in generation *G0* (—) and IBD (—) in generation *G10* with truncation selection and BayesB estimated breeding values; results are from one replicate with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

maximises the genetic gain over a specified time horizon can also be used [3], since this will spread the selection intensity optimally across the genomic regions in order to maximise long-term genetic gain.

In our study, allelic effects were assumed purely additive, whereas dominance and higher order interactions may occur. Dominance interactions can be partly accommodated by including a regression on (genomic) inbreeding in the model used to estimate breeding values. This will correct the (G)EBV so that they are valid for matings that result in non-inbred offspring. In practice, animals are expected to be mated to related selection candidates and thus their (G)EBV should be corrected for the inbreeding depression times their expected future inbreeding, which is their average coancestry with the selection candidates [24].

The design of the breeding scheme studied here resembles that of aquaculture breeding schemes, which rely heavily on sib-testing. In this design, traditional selection relies only on family information and either selects entire full-sib families or rejects them. The use of genomic selection makes it possible to estimate within-family deviations, and thus to distinguish between family members. This sib-testing design may have exaggerated the differences between genomic and traditional selection, and their effects on genome structure, because in most practical breeding schemes traditional selection also yields an estimate of the within-family deviation. However, in such schemes, genomic selection will estimate within-family deviations more accurately and thus the general outcomes of our study will still hold. In addition, genomic selection will be applied mainly in situations in which traditional selection yields little or no information on the within-family genetic component.

This study shows that serious interactions can occur between the methods used to estimate breeding values and the types of relationships used to control inbreeding. Results showed that outcomes are most stable and predictable, i.e. the final outcome reflects the constraint, when the same information is used to estimate breeding values and to control inbreeding. Thus, genomic selection has to be combined with genomic control of inbreeding in order to effectively manage ΔF_{IBD}.

Conclusions

Desired control of inbreeding was only achieved when it was managed using the same information as is used to estimate breeding values, i.e. pedigree-based inbreeding control with pedigree-based estimation of breeding values and genome-based inbreeding control with genome-based estimation of breeding values. In addition, the genome-based estimation of breeding values allows management of changes in genomic inbreeding, and thus changes in pedigree-based inbreeding are probably no longer relevant.

Competing interests

The authors declare that they have no competing interest.

Authors' contributions

AKS wrote the main computer program, ran computer programs and drafted the manuscript. JAW contributed in setting up the study, interpreting results and writing the manuscript. THEM wrote computer modules for genome-wide breeding value estimation and for Fisher-Wright populations. All authors have approved the final manuscript.

Acknowledgements

Helpful comments from two reviewers are gratefully acknowledged. All authors acknowledge financial support from Project 190442/S40 by Norwegian Research Council. JAW and THEM gratefully acknowledge that the research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 222664 ("Quantomics"). However, the publication reflects only the author's views and the European Community is not liable for any use that may be made of the information contained herein.

Author details

[1]Nofima AS, 1431 Ås, Norway. [2]University of Life Sciences, Department of Animal and Aquacultural Sciences, Nofima 1432 Ås, Norway. [3]The Roslin Institute and R(D) SVS, University of Edinburgh, Roslin, Midlothian EH25 9PS, UK.

Received: 30 January 2012 Accepted: 26 July 2012
Published: 16 August 2012

References

1. Henderson C: *Applications of Linear Models in Animal Breeding*. Canada: Guelph University Press; 1984.
2. Meuwissen THE, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157:**1819–1829.
3. Goddard ME: **Genomic selection: prediction of accuracy and maximisation of long term response.** *Genetica* 2009, **136:**245–257.
4. Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG: **Assumption-free estimation of heritability from genome-wide identy-by-descent sharing between full siblings.** *PLoS Genet* 2006, **2:**e41.
5. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA: **The impact of genetic architecture on genome-wide evaluation methods.** *Genetics* 2010, **185:**1021–1031.
6. Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME: **Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits.** *PLoS Genet* 2010, **6:**e1001139.

7. Meuwissen THE: **Maximizing the response of selection with a predefined rate of inbreeding.** *J Anim Sci* 1997, **75:**934–940.
8. Grundy B, Villanueva B, Woolliams JA: **Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development.** *Genet Res* 1998, **72:**159–168.
9. Hayes B, Goddard ME: **The distribution of the effects of genes affecting quantitative traits in livestock.** *Genet Sel Evol* 2001, **33:**209–229.
10. Chamberlain AJ, McPartlan HC, Goddard ME: **The number of loci that affect milk production traits in dairy cattle.** *Genetics* 2007, **177:**1117–1123.
11. Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen THE: **The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation.** *Genetics* 2009, **183:**1119–1126.
12. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, de Andrade M, Feenstra B, Feingold E, Hayes MG, Hill WG, Landi MT, Alonso A, Lettre G, Lin P, Ling H, Lowe W, Mathias RA, Melbye M, Pugh E, Cornelis MC, Weir BS, Goddard ME, Visscher PM: **Genome partitioning of genetic variation for complex traits using common SNPs.** *Nat Genet* 2011, **43:**519–525.
13. Fernández B, Santiago E, Toro MA, Caballero A: **Effect of linkage on the control of inbreeding in selection programmes.** *Genet Sel Evol* 2000, **32:**249–264.
14. Villanueva B, Pong-Wong R, Fernández J, Toro MA: **Benefits from marker-assisted selection under an additive polygenic genetic model.** *J Anim Sci* 2005, **83:**1747–1752.
15. Roughsedge T, Pong-Wong R, Woolliams JA, Villanueva B: **Restricting coancestry and inbreeding at a specific position on the genome by using optimized selection.** *Genet Res* 2008, **90:**199–208.
16. Pedersen LD, Sørensen AC, Berg P: **Marker-assisted selection reduces expected inbreeding but can result in large effects of hitchhiking.** *J Anim Breed Genet* 2010, **127:**189–198.
17. Sonesson AK, Meuwissen THE: **Testing strategies for genomic selection in aquaculture breeding programs.** *Genet Sel Evol* 2009, **41:**37.
18. Fisher RA: *The Genetical Theory of Natural Selection*. Oxford: Clarendon; 1930.
19. Wright S: **Evolution in Mendelian populations.** *Genetics* 1931, **16:**97–159.
20. Kimura M: **The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations.** *Genetics* 1969, **61:**893–903.
21. Haldane JBS: **The combination of linkage values, and the calculation of distances between the loci of linked factors.** *J Genet* 1919, **8:**299–309.
22. Daetwyler HD, Villanueva B, Woolliams JA: **Accuracy of predicting the genetic risk of disease using a genome-wide approach.** *PLoS One* 2008, **3:**3395.
23. Smith MJ, Haigh J: **The hitch-hiking effect of a favourable gene.** *Genet Res* 1974, **23:**23–35.
24. VanRaden PM: **Inbreeding adjustments and effect on genetic trend estimates.** *Interbull Bull* 2005, **33:**81–84.

doi:10.1186/1297-9686-44-27
Cite this article as: Sonesson *et al.*: Genomic selection requires genomic control of inbreeding. *Genetics Selection Evolution* 2012 **44**:27.