

# Restricting coancestry and inbreeding at a specific position on the genome by using optimized selection

T. ROUGHSEGE<sup>1</sup>\*, R. PONG-WONG<sup>2</sup>, J. A. WOOLLIAMS<sup>2</sup> AND B. VILLANUEVA<sup>1</sup>

<sup>1</sup>Sustainable Livestock Systems, SAC, West Mains Road, Edinburgh EH9 3JG, UK

<sup>2</sup>Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, UK

(Received 20 August 2007 and in revised form 27 November 2007)

## Summary

Over recent years, selection methodologies have been developed to allow the maximization of genetic gain whilst constraining the rate of inbreeding. The desired rate of inbreeding is achieved by constraining the group coancestry using the numerator relationship matrix computed from pedigree. It is shown that when the method is applied to mixed inheritance models, where a QTL is segregating together with polygenes, the rate of inbreeding achieved in the region around a QTL is greater than the desired level. The constraint on group coancestry at specific positions around the QTL is achieved by using a relationship matrix computed from pedigree and genetic markers. However, the rate of inbreeding realized at the position of constraint is lower than that expected given the assumed relationship between group coancestry and the subsequent rate of inbreeding. The use of markers in the calculation of the relationship matrix allows the selection of candidates with very low or zero relationships because they are homozygous for alternative alleles, which results in a heterozygosity amongst their offspring higher than would be expected given their allele frequencies. A generation of random selection restored the expected relationship between group coancestry and inbreeding.

## 1. Introduction

The need to restrict the increase in the rate of inbreeding ( $\Delta F$ ) in genetic improvement programmes has been widely recognized. Restricting  $\Delta F$  allows restriction of the decrease in genetic variability and, most importantly, the reduction in fitness-related traits. Over recent years, selection methodologies ('optimized selection') have been developed to maximize genetic gain whilst constraining  $\Delta F$  by optimizing the contributions made by parents to the next generation (Meuwissen, 1997; Grundy *et al.*, 1998). Although these methodologies were developed assuming the infinitesimal genetic model, they are also effective for mixed inheritance models where the identified QTL are known to be segregating with the polygenes (Villanueva *et al.*, 2002).

A key component of optimized selection is the numerator relationship matrix, as the selection method takes into account all genetic relationships between

selection candidates together with their estimated breeding values. The use of the numerator relationship matrix conditional on pedigree ( $\mathbf{A}_p$ ) to restrict global coancestry in the optimization tool ensures maintenance of  $\Delta F$  and genetic variation at the desired level for loci unlinked to those under selection but not for selected loci or loci linked to them (Villanueva *et al.*, 2005).

It may be desirable to maintain genetic variation in specific genome regions close to loci under selection. An example would be the MHC region where loci affecting traits selected by breeders (e.g., performance traits) might be located. Bot *et al.* (2004) suggested that loci affecting clean fleece weight are segregating in the MHC region in sheep. In pigs, the SLA region on chromosome 7 has possible associations with many production and reproduction traits (Vaiman *et al.*, 1988). QTL detection studies have reported back fat thickness effects between  $-3$  and  $+3.7$  mm for specific SLA haplotypes (de Koning *et al.*, 1999). A restriction on  $\Delta F$  at a specific region in the genome

\* Corresponding author. e-mail: tim.roughsedge@sac.ac.uk

could be achieved by applying optimized selection using a numerator relationship matrix constructed using pedigree and molecular markers located in and around that region.

The study aims to assess (i) the effect of optimized selection (using a numerator relationship matrix based on pedigree only) on genetic gain and inbreeding at a specific region of the genome where a QTL affecting the selected trait is located; (ii) the comparative effectiveness of using marker information combined with pedigree for restricting inbreeding at that position; and (iii) to explore how controlling inbreeding with markers fits into existing theory.

## 2. Materials and methods

### (i) Genetic and population models

A trait under the genetic control of a single known additive bi-allelic QTL and polygenes was simulated. The total genetic value for the  $i$ th individual was  $G_i = v_i + u_i$ , where  $v_i$  is the QTL effect and  $u_i$  is the polygenic effect. The genotypic values due to the QTL for individuals with genotypes BB, Bb and bb were respectively  $a$ , 0 and  $-a$ , where  $a$  is defined as half the difference between the two homozygotes (Falconer & MacKay, 1996). The additive genetic variance contributed by the QTL in the base generation ( $t=0$ ) was thus  $\sigma_q^2 = 2p(1-p)a^2$ , where  $p$  is the initial frequency of the favourable allele B (Falconer & MacKay, 1996), which was equal to 0.15.

The QTL was located 50 cM from one end of a 250 cM long continuous segment of the genome. Across the segment,  $m$  markers (markers  $M_1$ ), spaced evenly at 10 cM intervals from positions 0 to 250 cM were simulated. At  $t=0$ , the number of alleles per marker was 10 for all  $M_1$  markers. Marker alleles for a particular individual born at  $t=0$  were sampled from a uniform distribution with an equal probability given to all allelic variants. The  $M_1$  markers were used to compute the numerator relationship matrices conditional on pedigree and markers to be used in the optimized selection. In addition, another set of markers (markers  $M_2$ ) was simulated at 1 cM intervals and was used to estimate the observed inbreeding coefficients calculated as the homozygosity by descent, at each position along the genome segment. For each of these  $M_2$  markers, all  $N$  unrelated individuals ( $N/2$  males and  $N/2$  females) at  $t=0$  were sequentially allocated two unique alleles, i.e.,  $2N$  alleles in total were allocated to each  $M_2$  marker. Information from these markers was not utilized in the optimized selection procedure. All  $M_1$  markers were simulated in linkage equilibrium with each other and with the QTL.

Selection was carried out for eight discrete generations with  $N$  offspring produced in each generation

providing  $N/2$  male and  $N/2$  female candidates for selection as parents of the next generation. At  $t=1$ , the polygenic effect of each individual was sampled from a normal distribution with mean zero and variance  $\sigma_u^2$ . Alleles at the QTL were randomly allocated according to the starting QTL allelic frequencies. The QTL and polygenes were in linkage equilibrium at  $t=0$ . The phenotype of individual  $j$  ( $P_j$ ) was obtained by adding the total genetic value ( $G_j$ ) to a normally distributed environmental component with mean zero and variance  $\sigma_e^2$ . In subsequent generations, the polygenic effect was calculated as the average polygenic effect of the parents plus a random Mendelian term sampled from a normal distribution with mean zero and variance  $(\sigma_u^2/2)[1 - (F_s + F_d)/2]$ , where  $F_s$  and  $F_d$  are the pedigree relationship matrix inbreeding coefficients of the sire and dam, respectively. Marker and QTL alleles were inherited from the parent chromosome segments in classical Mendelian fashion, allowing for recombination assuming the Haldane mapping function. It was assumed that all individuals were genotyped for all  $M_1$  markers and for the QTL.

### (ii) Estimation of breeding values

The effect of the QTL was assumed to be known without error. The total estimated breeding value was the polygenic estimated breeding value plus the assumed known QTL breeding value. Polygenic estimated breeding values were obtained from best linear unbiased prediction (BLUP) using phenotypic data corrected for the QTL effect ( $P_i - v_i$ ), and the polygenic heritability. The breeding value for the QTL was calculated as  $2(1-p)a$ ,  $(1-2p)a$  and  $-2pa$  for individuals with genotype BB, Bb and bb, respectively (Falconer & MacKay, 1996), and  $p$  was updated at each generation.

### (iii) Selection and mating

The optimization algorithm, described by Meuwissen (1997), for obtaining maximum genetic gain whilst constraining  $\Delta F$  to a specified level was used (Appendix A). The **A** matrix utilized in the optimization to constrain  $\Delta F$  was either the numerator relationship matrix conditional on pedigree (**Ap**) or the numerator relationship matrix conditional on pedigree and markers (**Am**). The **Am** matrix was calculated at the position of the genome where there was an intention to restrict  $\Delta F$  for a particular scenario. The deterministic method of Pong-Wong *et al.* (2001) was used to compute **Am**. Matrix **Am** becomes the same as **Ap** when the flanking markers around a genome position become uninformative. Mating of parents resulting from the optimized selection procedure was at random.

(iv) *Genetic gain and inbreeding*

Average polygenic, QTL and total true breeding values and inbreeding coefficient of individuals born at each generation were computed. The average inbreeding coefficient ( $F$ ) was calculated as:

$$F = \frac{1}{N} \left( \sum_{i=1}^N A_{ii} - 1 \right), \quad (1)$$

where  $A_{ii}$  refers to the diagonal element of the numerator relationship matrix. The rate of inbreeding ( $\Delta F$ ) at generation  $t$  was calculated as  $(F_t - F_{t-1}) / (1 - F_{t-1})$ . Both  $F$  and  $\Delta F$  were obtained using three different relationship matrices: **Ap** ( $F_P$ ), **Am** ( $F_{M_1}$ ) and **An** ( $F_{M_2}$ ), the numerator relationship matrix conditional on  $M_2$  markers. Matrix **An** was calculated at all positions on the genome where  $M_2$  markers were simulated. Elements of **An** for a particular position were obtained using:

$$A_{ij} = \frac{1}{2} (h_{r(i),r(j)} + h_{r(i),m(j)} + h_{m(i),r(j)} + h_{m(i),m(j)}),$$

where  $h_{x(i),y(j)}$  is 1 if allele  $x$  of individual  $i$  is identical to allele  $y$  of individual  $j$  and 0 otherwise. Alleles  $x$  and  $y$  are either paternally ( $r$ ) or maternally ( $m$ ) inherited. The  $F_{M_2}$  and  $\Delta F_{M_2}$  calculated using the **An** matrix are considered the observed inbreeding coefficient and rate of inbreeding determined by homozygosity by descent for that position on the genome. Each founder received a unique pair of alleles at each locus and therefore identical  $M_2$  alleles can only be identical by descent.

(v) *Coancestry*

Different coancestry parameters were calculated in order to gain insight into the mechanism by which the optimization tool restricts  $\Delta F$  (Appendix B). These included the calculation of the group coancestry weighted by the candidate contribution solutions of the optimized selection algorithm, the group coancestry based on equal contributions of candidates and the average pairwise coancestry between the males and females weighted by the contribution solutions of the optimized selection algorithm.

(vi) *Scenarios investigated*

Six scenarios were simulated. In all scenarios,  $\sigma_u^2$  was 0.2,  $\sigma_e^2$  was 0.8 and  $N$  was 120. Scenario 1 was used to investigate the effect that optimized selection using **Ap** has on the inbreeding in a genome segment where there are no QTL affecting the trait under selection. In scenario 2, **Ap** was again used but a QTL with effect  $a=0.1$  on the trait was placed on the genome segment. Scenarios 3–5 were used to investigate the

Table 1. Total ( $G_T$ ), polygenic ( $G_P$ ) and QTL ( $G_Q$ ) genetic means and frequency of the favourable QTL allele ( $p$ ) at generations four and eight under different scenarios with QTL effect ( $a$ ) and varying the numerator relationship matrix used in the optimization procedure (**A**) to constrain the rate of inbreeding. The QTL was at position 50 cM

Scenario	$a$	<b>A</b>	$t$	$G_P^a$	$G_Q^a$	$G_T^a$	$p$
2	0.1	<b>Ap</b>	4	1.03	0.04	1.07	0.60
			8	1.95	0.09	2.04	0.94
3	0.1	<b>Am</b> at 55 cM	4	1.08	0.01	1.09	0.43
			8	2.06	0.02	2.08	0.61
4	0.1	<b>Am</b> at 105 cM	4	1.08	0.04	1.12	0.61
			8	2.04	0.09	2.13	0.95
5	0.1	<b>Am</b> at 155 cM	4	1.08	0.05	1.13	0.60
			8	2.01	0.09	2.10	0.96

<sup>a</sup> Measured in genetic standard deviations. Standard error: from 0.005 to 0.015.

effectiveness of **Am** constructed 5, 55 and 105 cM away from the QTL locus for constraining inbreeding in the region around the QTL that has effect  $a=0.1$ . Scenario 6 was the same as scenario 3 except that the QTL effect was set to zero in order to examine the effect of the restriction achieved in the absence of a QTL under selection. The constraint applied to the desired rate of inbreeding was to 0.02 across all scenarios.

## 3. Results

When using **Am** in the optimization, approximately half of the replicates failed to give a solution for all eight generations of selection due to the constraint  $C$  not being met ( $\lambda_0$  was negative). Only those replicates where a solution was successfully achieved (positive  $\lambda_0$ ) in all generations contributed to the results presented. Results for the successful generations of those replicates that did not achieve all eight generations were the same as those observed in replicates that successfully achieved eight generations (results not shown).

(i) *Genetic gain and QTL allele frequency*

Polygenic ( $G_P$ ), QTL ( $G_Q$ ) and total ( $G_T$ ) genetic gains and frequency of the favourable QTL allele ( $p$ ) for scenarios 2–5 are shown in Table 1. Results for scenario 1 are not comparable with those from the other scenarios (the QTL effect was zero in scenario 1, making the overall genetic variation lower for this scenario).

When the QTL had an effect on the selected trait and **Ap** was used in the optimization (scenario 2),  $p$  increased from 0.15 at  $t=0$  to 0.94 at  $t=8$ . In contrast, when **Am** calculated at 5 cM from the QTL position was used (scenario 3),  $p$  rose to only 0.61 in generation 8. However,  $G_T$  was higher in scenario 3 compared with scenario 2, with lower  $G_Q$  being over-compensated by a higher  $G_P$ .

Results in terms of  $p$  and  $G_Q$  from scenarios where the **Am** used in the optimization was calculated at 55 and 105 cM from the QTL (scenarios 4 and 5) were very similar to those when using **Ap**, where no constraint was specifically applied to the QTL region (scenario 2). However,  $G_P$  was higher in scenarios 4 and 5 than in scenario 2, resulting in a 4 and 3% greater  $G_T$  in scenarios 4 and 5, respectively.

### (ii) Rate of inbreeding

It can be seen that for scenario 2 (Fig. 1b) the use of **Ap** to restrict the rate of inbreeding across the genome did not restrict inbreeding in the region around the QTL under positive selection. This is a demonstration of the hitch-hiking effect (Maynard-Smith & Haigh, 1974), which results in a signature selective sweep in which neutral loci close to a QTL under selection change in frequency, resulting in reduced variation in that region. When **Am** constructed 5 cM from the QTL was used to restrict the rate of inbreeding, the effect of the selective sweep was eliminated (Fig. 1c). However, when **Am** constructed 55 cM from the QTL position was used in the restriction (Fig. 1d), the effect of the selective sweep led to an asymmetric pattern of restriction with the inbreeding coefficients rising more rapidly at positions away from the restriction position adjacent to the QTL. The observed average rate of inbreeding ( $\Delta F_{M2}$ ) achieved in scenarios 1–6 is shown in Table 2. The rate of inbreeding computed from the pedigree ( $\Delta F_P$ ) is also shown for comparison. In scenario 1, no QTL effect was simulated and **Ap** was used in the optimization and, as expected, both  $\Delta F_{M2}$  and  $\Delta F_P$  were restricted to the desired value (0.02). However, in scenario 2, when the QTL had an effect on the selected trait, **Ap** no longer constrained  $\Delta F_{M2}$  at the QTL position to the desired value and  $\Delta F_{M2}$  at that position rose by 50% relative to scenario 1.

In scenario 3, where the **Am** calculated 5 cM from the QTL position was used in the optimization,  $\Delta F_P$  rose to 0.039, nearly 2.5 times greater than the desired value (Table 2). However, the observed  $\Delta F_{M2}$  at the QTL position and at the position where the **Am** imposed the restriction were approximately half the desired value. At positions further from the position of restriction, the observed  $F_{M2}$  in this scenario rose and reached the value of the  $F_P$  approximately 80 cM from the restriction position (results not shown). The asymptotic value of  $F_P$  and  $F_{M2}$  (observed at positions

away from the QTL) was higher than those seen in scenario 2.

In scenarios 4 and 5, the  $\Delta F_P$  was similar to that observed in scenario 3 and the observed  $\Delta F_{M2}$  at the position of restriction in scenarios 4 and 5 (55 and 105 cM from the QTL, respectively) was also approximately half the desired value as in scenario 3 (Table 2). At the QTL position, the observed  $\Delta F_{M2}$  was higher than that seen in scenario 3 and also higher than that seen in scenario 2 where **Ap** was used and no specific restriction was imposed at the QTL region.

### (iii) Coancestry

The previous section showed a clear discrepancy between the observed and expected  $\Delta F_{M2}$  values at the position of restriction. In order to investigate further the lower than expected observed  $\Delta F_{M2}$  when **Am** was used in the optimization, additional parameters were calculated: (i) the average group coancestry weighted by the optimal contributions ( $f_o$ ); (ii) the average group coancestry assuming that all candidates have equal contributions ( $f_e$ ); and (iii) the average pairwise coancestry between the male and female selected parents weighted by the optimal contributions, i.e., excluding self and within-sex coancestries ( $f_o^-$ ) at a given generation. Results are shown for these parameters in Table 3 for two scenarios: scenario 1, where the QTL had no effect on the trait and the matrix used in the optimization was the **Ap** matrix, and scenario 6, where the QTL had no effect and the matrix used in the optimization was **Am** calculated at 55 cM.

As expected, when **Ap** was used in the optimization (scenario 1), the constraint on  $\Delta F_P$  was achieved at each generation and the values of observed coancestry and inbreeding coefficients at position 55 cM computed from **An**, and at all other positions along the genome segment computed from **An** (results not shown), were very similar to those computed from the pedigree. The specific constraint applied was achieved ( $f_o = C$ ) at each generation and  $f_o$  at  $t$  was equal to  $f_e$  at  $t+1$ . The parents selected from the group of candidates had  $f_o^-$  equal to  $f_e$  and their matings resulted in a group of offspring with  $F$  equal to  $f_o^-$  of their parents (i.e.,  $F$  at  $t$  was equal to  $f_o^-$  at  $t-1$ ). This clearly shows that when **Ap** is used in the optimization in a situation where there is no QTL, then the desired constraint is achieved across the genome.

When **Am** calculated at 55 cM was used in the optimization (scenario 6), the observed  $f_o$  at the restriction position achieved the desired value in each generation ( $f_o = C$ ). However, the pedigree  $f_o$ , which represents the optimally weighted group coancestry across all neutral loci in an infinitesimal model, had an increased value, which relates to the higher  $\Delta F_P$  observed in scenarios 3–6. When **Am** was used in the

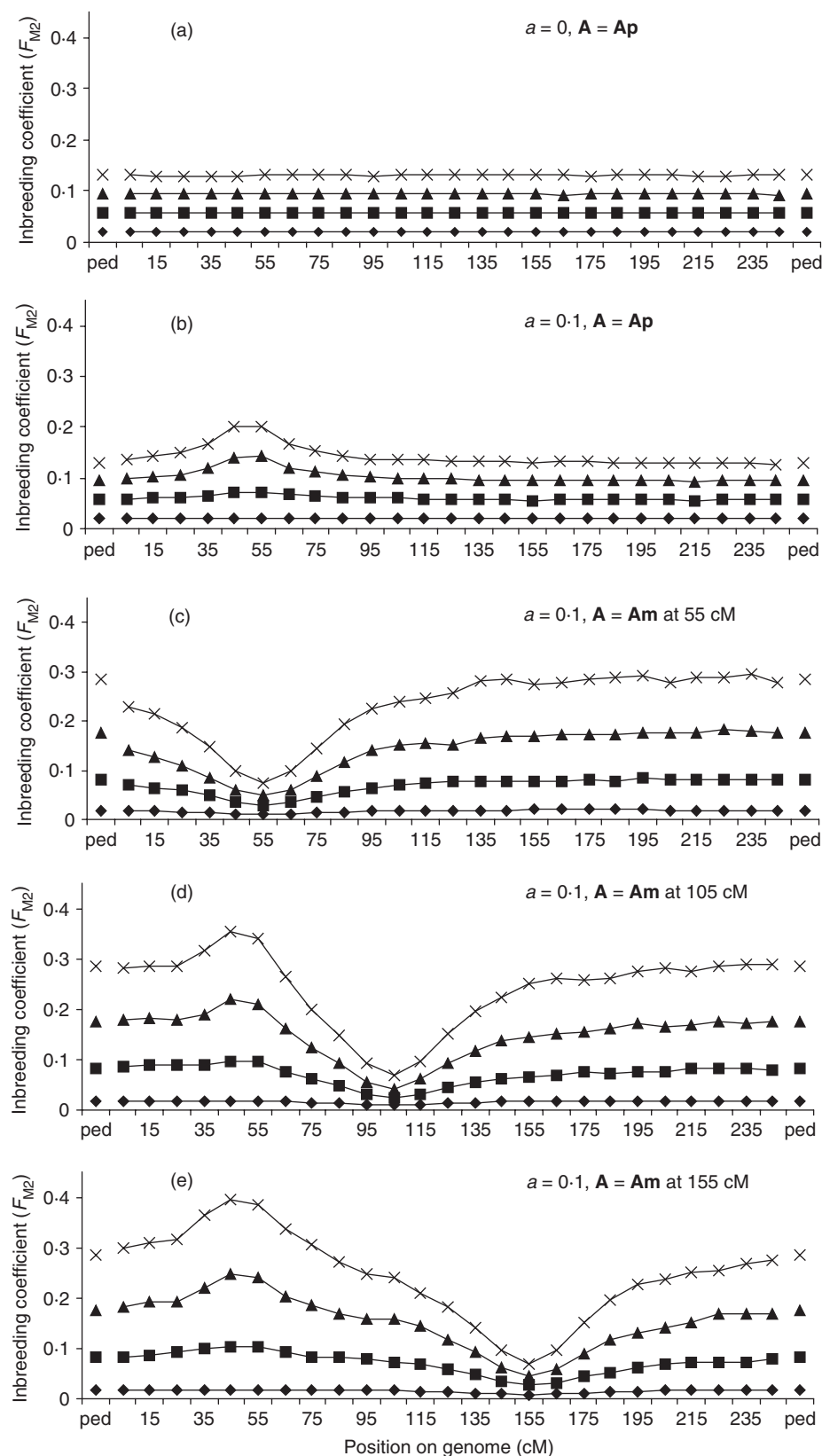


Fig. 1. Observed inbreeding coefficient ( $F_{M2}$ ) computed from  $A_n$  calculated as homozygosity by descent, at different positions along the genome segment at generations 2 ( $\blacklozenge$ ), 4 ( $\blacksquare$ ), 6 ( $\blacktriangle$ ) and 8 ( $\times$ ) when optimized selection was practised under different scenarios varying in the QTL effect ( $a$ ) and the relationship matrix used in the optimization procedure ( $A$ ). The average inbreeding coefficient calculated using  $A_p$  (ped) is also shown for comparison.

Table 2. Observed rate of inbreeding ( $\Delta F_{M_2}$ ) calculated as homozygosity by descent using **An** at different positions, and pedigree rate of inbreeding computed from **Ap** ( $\Delta F_P$ ), when different numerator relationship matrices (**A**) were used in the optimization to constrain the rate of inbreeding. The optimization aimed to restrict  $\Delta F$  to 0.02. Values in boldface indicate  $\Delta F_{M_2}$  at the position where the constraint was targeted. The QTL was at position 50 cM with additive effect (a)

Scenario	a	<b>A</b>	$\Delta F_{M_2}$ at position (in cM <sup>a</sup> )				$\Delta F_P$
			50	55	105	155	
1	0.0	<b>Ap</b>	0.019	0.019	0.019	0.019	<b>0.019</b>
2	0.1	<b>Ap</b>	0.033	0.029	0.019	0.019	<b>0.019</b>
3	0.1	<b>Am</b> at 55 cM	0.009	<b>0.009</b>	0.032	0.038	0.039
4	0.1	<b>Am</b> at 105 cM	0.055	0.048	<b>0.009</b>	0.034	0.039
5	0.1	<b>Am</b> at 155 cM	0.061	0.054	0.033	<b>0.009</b>	0.039
6	0.0	<b>Am</b> at 55 cM	0.010	<b>0.010</b>	0.035	0.039	0.041

<sup>a</sup> Averaged across generations 3–8. Standard error for scenarios 1 and 2:  $\sim 0.001$ . Standard error for scenarios 3–6:  $\sim 0.002$ .

optimization, the magnitude of the observed pairwise coancestry between the selected males and females ( $f_{\sigma}^-$ ) calculated using **An** was less than that seen when **Ap** was used in the optimization (scenarios 1 and 2). Lower observed  $f_{\sigma}^-$  when **Am** was used in the optimization compared to when **Ap** was used resulted in a lower observed  $F_{M_2}$  in the offspring at the position of restriction, leading to a lower observed  $\Delta F_{M_2}$  than that used to calculate the desired constraint ( $C$ ). The same effect was seen when **Am** was used in the coancestry calculation. The coancestries and inbreeding coefficients calculated using **Ap** in scenario 6, though higher than the observed values in the QTL region using **An**, followed the expectation, i.e.,  $f_{\sigma}^-$  was of a similar magnitude to  $f_e$ , the group coancestry, and the average inbreeding coefficient of the offspring was that expected from  $f_{\sigma}^-$ .

#### 4. Discussion

This study has quantified inbreeding in the region around a QTL under selection when the  $\Delta F$  was constrained using a numerator relationship matrix either conditional on pedigree (**Ap**) or conditional on pedigree and markers (**Am**) constructed at different positions on the genome. The constraint on  $\Delta F$  was achieved using optimized selection that constrained the level of group coancestry amongst the candidates

selected as parents weighted by the contributions each makes to the next generation.

It has been shown that **Am** constructed at a specific position on the genome can be used to constrain the level of weighted group coancestry ( $f_{\sigma}$ ) at that position within the optimization to the desired level ( $C$ ). This resulted in the desired rate of increase in unweighted group coancestry ( $\Delta f_e$ ) at that position on the genome. However, the consequent rate of inbreeding measured using the inbreeding coefficients, at the same position, did not follow our expectation as determined by  $C$ . The  $\Delta F_{M_2}$  measured by the observed inbreeding coefficients ( $F_{M_2}$ ) at the restriction position was consistently below our expectation, our expectation being that the rate of inbreeding would follow the rate of coancestry across generations. At positions away from the restriction, the observed  $\Delta F_{M_2}$  was seen to rise and reach a maximum equal to the unconstrained rate of inbreeding observed in the rate of pedigree inbreeding ( $\Delta F_P$ ). The quantitative results of this study were specific to a single QTL allele starting frequency and assumed marker information content; however, qualitatively we have no reason to believe why they should not exemplify more general phenomena.

In the simulation of the base population, the QTL and  $M_1$  markers were simulated in linkage equilibrium. However, this should not have a direct impact on the schemes simulated where the QTL is assumed to be known and the marker information is only related to the calculation of **Am**. By definition, all animals at the base population are assumed to be unrelated regardless of whether they share the same marker allele or not. During the calculation of **Am**, the marker information is used to determine the inheritance pattern from the parents to offspring, not to determine any further degree of relationship arising before the base generation. However, linkage disequilibrium would have an indirect impact by increasing the need to utilize more distant markers for inferring IBD (Identity By Descent). Hence, the assumption of linkage equilibrium introduces no bias but a small increase in the precision of IBD estimation that could otherwise be achieved by including additional markers.

Under random mating in the absence of self-fertilization, we expect  $f_{\sigma(t)} \cong F_{t+2}$  (Caballero & Toro, 2000), which holds with our expectation from eqn (7) where  $f_e$  is equivalent to  $f_{\sigma}$  as no constraint on group coancestry was assumed. Under pedigree relationship, the use of **Ap** in optimized selection constrains  $f_{\sigma}$  to a desired level, which is then expected to produce an equal restriction on  $\Delta F$ . This relationship between  $f_{\sigma}$  and  $\Delta F$  was shown to hold true in this and previous studies (e.g., Grundy *et al.*, 1998).

When **Am** is constructed at a single position on the genome and used in the optimization, as a mechanism

Table 3. Coancestry, inbreeding coefficients<sup>a</sup> and rate of inbreeding computed from **Ap** (pedigree) or **An** (observed) over generations (*t*) when **Ap** was used in the optimization (scenario 1) and when **An** constructed at 55 cM was used in the optimization (scenario 6). The QTL had no effect on the trait. Last column shows desired constraint imposed on coancestry (*C*) in optimization in each generation

Scenario	<i>t</i>	Computed from <b>Ap</b> <sup>a</sup>					Computed from <b>An</b> at 55 cM <sup>a</sup>					<i>C</i>
		$f_o$	$f_e$	$f_o^-$	<i>F</i>	$\Delta F$	$f_o$	$f_e$	$f_o^-$	<i>F</i>	$\Delta F$	
1	1	0.020	0.002	0.000	0.000	0.000	0.020	0.002	0.000	0.000	0.000	0.020
	2	0.040	0.020	0.020	0.000	0.000	0.040	0.020	0.020	0.000	0.000	0.040
	3	0.059	0.040	0.039	0.020	0.020	0.059	0.039	0.039	0.020	0.020	0.059
	4	0.078	0.059	0.057	0.039	0.019	0.077	0.058	0.057	0.039	0.019	0.078
	8	0.149	0.132	0.132	0.114	0.021	0.148	0.130	0.130	0.111	0.018	0.149
6	1	0.020	0.002	0.000	0.000	0.000	0.020	0.002	0.000	0.000	0.000	0.020
	2	0.048	0.020	0.019	0.000	0.000	0.040	0.020	0.009	0.000	0.000	0.040
	3	0.088	0.048	0.047	0.019	0.019	0.059	0.040	0.017	0.009	0.009	0.059
	4	0.135	0.088	0.086	0.048	0.030	0.078	0.059	0.023	0.016	0.007	0.078
	8	0.355	0.297	0.295	0.234	0.066	0.152	0.133	0.067	0.050	0.008	0.149

<sup>a</sup>  $f_o$ : average group coancestry weighted by the optimal contributions from eqn (B1);  $f_e$ : average group coancestry assuming that all candidates have equal contributions from eqn (B2);  $f_o^-$ : average pairwise coancestry between selected male and female parents from eqn (B3); *F*: average inbreeding coefficient from eqn (1);  $\Delta F$  is the rate of inbreeding.

to restrict  $\Delta F$  at that position to a predefined level, the observed  $\Delta F_{M_2}$  was found to be lower than expected. The theory behind the expectation that  $f_{o(t)} = F_{t+2}$  is dependent on the average frequency of the alleles in the selected male parents being the same as those in the selected female parents. If due to sampling or other effects the gene frequencies of the two sexes of parents selected were different, then we would expect an excess of heterozygous individuals in comparison to our expectation under Hardy–Weinberg equilibrium (Robertson, 1965). If our hypothesis is that such a situation arose when **An** was used in the optimization, then we would expect that  $f_o > f_o^-$  and a  $\Delta F$  lower than expected. This was tested using a single replicate of the scenario 3 simulation. In the simulation, **An** calculated at 5 cM away from the QTL was used to constrain the group coancestry in the optimization. The gene frequencies of the  $M_2$  markers at the position of restriction for the male and female parents weighted by their relative contributions and the expected excess of heterozygotes were calculated following Robertson (1965) as shown in Appendix C. The results of this exercise demonstrated that the difference between  $f_o$  and  $f_o^-$  was equal to the expected deviation from heterozygosity (Table 4). This explains why  $f_o$  was not subsequently realized in *F*, as  $F_t = f_{o(t-1)}$ .

In the previous illustration, an excess of heterozygotes due to variation in sample gene frequency between the parent sexes was demonstrated, which in turn resulted in *F* being lower than we would expect from  $f_o$ . There is, however, no reason to expect this pattern of variation in the sampling of alleles between

Table 4. Demonstration that the deviation of observed average pairwise coancestry between selected male and female parents ( $f_o^-$ ) from the observed average group coancestry weighted by the optimal contributions ( $f_o$ ) calculated using **An** at 55 cM, can be attributed to the excess of heterozygotes created due to the non-random segregation of alleles between sexes. Results are from eight generations (*t*) of a single replicate of the scenario 3, where **An** at 55 cM was used in the optimization

<i>t</i>	<i>C</i> <sup>a</sup>	$f_o^b$	$f_o^-^b$	$f_o - f_o^-$	Excess heterozygote deviation <sup>c</sup>
1	0.020	0.020	0.000	0.020	0.020
2	0.040	0.040	0.006	0.034	0.034
3	0.059	0.057	0.015	0.042	0.042
4	0.078	0.074	0.010	0.064	0.064
5	0.096	0.094	0.056	0.038	0.038
6	0.114	0.099	0.036	0.063	0.063
7	0.132	0.128	0.051	0.077	0.077
8	0.149	0.152	0.091	0.061	0.061

<sup>a</sup> Desired constraint used in optimization.

<sup>b</sup>  $f_o$ : eqn (B1);  $f_o^-$ : eqn (B3).

<sup>c</sup> Predicted using eqn (C1) in Appendix C.

the two sexes of selected candidates. In fact, the sex of the candidates was not a factor in the non-random sampling of alleles. This was tested by randomizing the sex of the selected candidates following selection subject to there being equal numbers of gametes contributed from male and female candidates. Following

randomization, the same outcome was obtained as seen in Table 4, with the expected deviation from heterozygosity again being equal to the difference between  $f_o$  and  $f_o^-$ . This demonstrates that sex does not have a bearing on the non-random sampling of alleles; rather, there is a between-selected-individual non-random sampling of alleles, which results in an excess of heterozygous offspring.

We can explain the repeated excess of heterozygous offspring observed in the following way. The optimization maximizes gain at a given rate of inbreeding. In the scenarios where **Am** was used in the restriction rather than **Ap**, fewer individuals were selected with higher average individual contributions, potentially increasing intensity. This potential increase in intensity is facilitated by the nature of **Am**, which has more variable diagonals and many more small off-diagonal elements than **Ap**, which estimates average neutral relationship derived from probabilities. The  $N(N-1)$  off-diagonal values drive the constraint in the optimization and it is therefore possible to select individuals who have a high proportion of very low or zero relationships with other candidates, and consequently a greater proportion of the group coancestry may arise from self-coancestries, allowing fewer but greater contributions. Such solutions are attractive since it allows greater contributions from individuals with greater breeding values. Therefore, in each generation, the more dispersed values of **Am** lead to selection of individuals that have low or zero relationships because they are homozygous for alternative alleles, resulting in a higher heterozygosity amongst the offspring than would be expected from the array of candidate genotypes. For the reason that more loci are averaged to obtain the **Am** used, the deviation from the expectations of **Ap** diminish as the dispersion among the elements of **Am** diminish (T. Roughsedge, unpublished results).

Having established that the discrepancy between  $f_o$  and  $f_o^-$  was due to non-random inheritance of alleles at the position of restriction in the selected candidates, we can hypothesize that removing the constraint at  $t-1$  and allowing random selection to take place would restore the expected relationship between group coancestry and average inbreeding at  $t$ . In order to test this hypothesis, a further simulation was run. The scenario 3 simulation parameters were again used, but for generation 7 random selection rather than optimized selection was practised, with 20 of the candidates of each sex randomly selected as equally contributing parents. The simulation was run for 1000 replicates. The optimization makes six generations of selection with the constraint on the rate of inbreeding being applied using **Am**. If the discrepancy between  $f_o$  and the subsequently realized  $F_{M2}$  was due to the selection of parents with non-random samples of alleles, then the introduction of random selection

Table 5. Effect on coancestry and inbreeding calculated using **Am** at 55 cM over generations ( $t$ ), when introducing random selection at  $t=7$  and  $t=8$  following optimized selection using **Am** (55 cM) at  $t=1-6$

$t$	Observed coefficients at 55 cM <sup>a</sup>				
	$f_o$	$f_e$	$f_o^-$	$F$	$C^b$
1	0.020	0.000	0.000	0.000	0.020
2	0.040	0.018	0.009	0.000	0.040
3	0.059	0.038	0.017	0.009	0.059
4	0.078	0.057	0.025	0.016	0.078
5	0.097	0.077	0.034	0.025	0.096
6	0.116	0.096	0.046	0.035	0.114
7	0.128	0.115	0.115	0.044	na
8	0.140	0.127	0.126	0.115	na

<sup>a</sup>  $f_o$ : average group coancestry weighted by the optimal contributions from eqn (B1);  $f_e$ : average group coancestry assuming that all candidates have equal contributions from eqn (B2);  $f_o^-$ : average pairwise coancestry between selected male and female parents from eqn (B3);  $F$ : average inbreeding coefficient from eqn (1).

<sup>b</sup> Desired constraint used in optimization. Applied at  $t=1-6$  in which optimized selection is undertaken.

should select a sample of parents with a random sample of the alleles from the previous generations parents, which were selected under the optimization. In the simulation, the  $F_{M2}$  realized in  $t=8$  was approximately equal to the  $f_o$  at  $t=6$  (Table 5). This further demonstrates that the optimized selection method constrains group coancestry to the desired level, but non-random segregation of allele frequency at the restriction position occurs in selected parents. The discrepancy between  $f_o$  and the realized  $\Delta F_{M2}$  can be attributed to this non-random segregation of alleles between the selected parents.

This raises a question about which rate of inbreeding it is appropriate to constrain when undertaking selection. In the optimized selection procedure, we actually constrain the rate of group coancestry:

$$\Delta f_e = \frac{f_e(t) - f_e(t-1)}{1 - f_e(t-1)}$$

By constraining the rate of group coancestry, the loss of genetic variation is managed but no absolute control is placed on the inbreeding coefficients. As we have seen in this study, when relationships are constructed at a single locus and the group coancestry at this locus is constrained, the relationship between group coancestry at  $t$  and average inbreeding coefficients at  $t+2$  breaks down. The loss of genetic variation is still managed but  $\Delta F$  is less than we would expect. This highlights why we should treat the



management of genetic variation and inbreeding coefficients as separate issues.

A technique to simultaneously constrain both pedigree group coancestry and also group coancestry around a QTL under selection may provide a solution to simultaneously being able to constrain the  $\Delta F_{M2}$  on the segment of the genome around the QTL and the pedigree  $\Delta F_P$ . This would require an approach that allowed two or more constraints to be introduced to the quadratic optimization such as semi-definite programming (Pong-Wong & Woolliams, 2007). Such an approach could be useful when a QTL is segregating in a region where a loss in genetic variability is not desirable. Here, a constraint on both the group coancestry in the region around the QTL and also the pedigree group coancestry could avoid a loss in genetic variation in the region and also the average genetic variation. This study has shown that it is possible to erode the relationship between the rate of coancestry and the rate of inbreeding under random mating. Hence, we have reinforced the concept that in order to manage the loss of long-term genetic variation, group coancestry should be managed rather than the inbreeding coefficient. If, however, there is interest in the effects of inbreeding depression, then the management of group coancestry based on a numerator relationship matrix conditional on pedigree and marker information does not provide a reliable estimate of inbreeding coefficients. In order to manage inbreeding depression, it is advisable to take a two-stage approach with the first stage managing the loss of variation through the constraint of group coancestry, followed by mate allocation to manage inbreeding depression. This message has a clear bearing on both selection and conservation breeding programmes.

**Appendix A. Optimized selection**

The optimization algorithm, described by Meuwissen (1997), maximizes the objective function:

$$f(\mathbf{c}_t) = \mathbf{c}_t^T \mathbf{g}_t - \lambda_0 \left( \frac{\mathbf{c}_t^T \mathbf{A}_t \mathbf{c}_t}{2} - C_t \right) - (\mathbf{c}_t^T \mathbf{Q} - \frac{1}{2}) \boldsymbol{\lambda},$$

where  $\mathbf{c}_t$  is the solution vector (of dimension  $N$ ) of optimal contributions of candidates at generation  $t$ ,  $\mathbf{g}$  is the vector of total estimated breeding values,  $\mathbf{A}$  is the numerator relationship matrix  $N \times N$  of candidates,  $\mathbf{Q}$  is a known incidence matrix  $N \times 2$  with ones for males and zeros for females in the first column and ones for females and zeros for males in the second column,  $C$  is the constraint on the rate of inbreeding,  $\frac{1}{2}$  is a vector of halves of dimension 2, and  $\lambda_0$  and  $\boldsymbol{\lambda}$  (a vector of dimension 2) are Lagrangian multipliers. The Lagrangian multiplier  $\lambda_0$  is chosen such that  $\mathbf{c}_t^T \mathbf{A}_t \mathbf{c}_t / 2$  is constrained to  $1 - (1 - \Delta F)^t$

(Grundy *et al.*, 1998), where  $\Delta F$  is the desired rate of inbreeding.

**Appendix B. Coancestry formulae**

All three relationship matrices ( $\mathbf{A}_p$ ,  $\mathbf{A}_m$ ,  $\mathbf{A}_n$ ) can be decomposed into four sub-matrices,  $\mathbf{A}_{mm}$ ,  $\mathbf{A}_{mf}$ ,  $\mathbf{A}_{fm}$  and  $\mathbf{A}_{ff}$ , each of dimensions  $N/2 \times N/2$ , relating to the male by male, male by female, female by male and female by female numerator relationships, respectively. The solution vector of optimal parental contributions ( $\mathbf{c}$ ) can also be decomposed into two sub-vectors, one for male ( $\mathbf{c}_m$ ) and one for female ( $\mathbf{c}_f$ ) parent contributions, each of dimensions  $N/2$ . The  $\mathbf{c}$  vector sums to 1 and each sub-vector ( $\mathbf{c}_m$  and  $\mathbf{c}_f$ ) sums to 0.5. The average group coancestry of candidates ( $f_o$ ) calculated using their optimal contributions is the average pairwise coancestry of a given group of individuals including within-sex and self-coancestries weighted by their contributions, and is obtained from the optimization procedure as:

$$f_o = \frac{\mathbf{c}^T \mathbf{A} \mathbf{c}}{2} = \frac{1}{2} \cdot \begin{bmatrix} \mathbf{c}_m \\ \mathbf{c}_f \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{A}_{mm} & \mathbf{A}_{mf} \\ \mathbf{A}_{fm} & \mathbf{A}_{ff} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{c}_m \\ \mathbf{c}_f \end{bmatrix}. \tag{B1}$$

This differs from the group coancestry ( $f_e$ ) defined by Cockerham (1967) in which all candidates are assumed to make an equal contribution. Let  $\mathbf{c}_e$  be a vector of dimension  $N$  with all elements equal to  $1/N$  (i.e., all candidates make an equal contribution). Then, the equal contributions group coancestry ( $f_e$ ) can be calculated as

$$f_e = \frac{\mathbf{c}_e^T \mathbf{A} \mathbf{c}_e}{2}. \tag{B2}$$

Finally, the average pairwise coancestry between the male and female individuals selected as parents weighted by their contributions ( $f_o^-$ ) is:

$$f_o^- = 2 \cdot \begin{bmatrix} \mathbf{c}_m \\ \mathbf{0} \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{A}_{mm} & \mathbf{A}_{mf} \\ \mathbf{A}_{fm} & \mathbf{A}_{ff} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{0} \\ \mathbf{c}_f \end{bmatrix}, \tag{B3}$$

where  $\mathbf{0}$  is a vector of dimension  $N/2$  with all elements set to zero.

**Appendix C. Excess of heterozygotes when allele frequency differs between parents**

Let  $q_1$  be the frequency of allele A of a bi-allelic gene in a group of male parents and  $q_2$  be the frequency in a group of female parents. Then, following random mating, the expected proportion of heterozygotes in the offspring is  $q_1(1 - q_2) + q_2(1 - q_1)$ . However, the mean gene frequency of the parents is  $(q_1 + q_2)/2$ , and from the Hardy–Weinberg equilibrium we would expect the frequency of heterozygotes in the offspring to be  $(q_1 + q_2)(1 - (q_1 + q_2)/2)$ . Therefore the excess of

heterozygotes when  $q_1 \neq q_2$  is  $\frac{1}{2}(q_1 - q_2)$ . This theory was extended beyond the bi-allelic case by Robertson (1965):

$$H = \frac{1}{2} \sum_{i=1}^L \sum_{j=i+1}^L (q_{i,1} - q_{i,2})(q_{j,1} - q_{j,2}), \quad (\text{C1})$$

where  $H$  is the excess of heterozygotes,  $q_{x,y}$  is the frequency of allele  $x$  in sex  $y$  and  $L$  is the total number of possible alleles at the locus considered.

The deviation of the observed  $f_o^-$  from the observed  $f_o$  at  $t$  calculated using **An** at the restriction position when **Am** is used in the optimization (e.g., Table 3, scenario 6) is accounted for by the deviation in expected heterozygotes calculated in eqn (A4). This results in the following equality:

$$\begin{aligned} & \frac{1}{2} \cdot \begin{bmatrix} \mathbf{c}_m \\ \mathbf{c}_f \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{An}_{mm} & \mathbf{An}_{mf} \\ \mathbf{An}_{fm} & \mathbf{An}_{ff} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{c}_m \\ \mathbf{c}_f \end{bmatrix} \\ &= 2 \cdot \begin{bmatrix} \mathbf{c}_m \\ \mathbf{0} \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{An}_{mm} & \mathbf{An}_{mf} \\ \mathbf{An}_{fm} & \mathbf{An}_{ff} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{0} \\ \mathbf{c}_f \end{bmatrix} \\ &+ \frac{1}{2} \sum_{i=1}^{2N} \sum_{j=i+1}^{2N} (q_{i,1} - q_{i,2})(q_{j,1} - q_{j,2}), \end{aligned}$$

where  $\mathbf{An}_{xy}$  is the sub-matrix between sexes  $x$  and  $y$  of the numerator genetic relationship matrix calculated at the same position as **Am** was constructed;  $\mathbf{c}_x$  is the contributions vector for sex  $x$ ;  $q_{i,1}$  is the frequency of allele  $i$  in sex 1,  $q_{i,2}$  is the frequency of allele  $j$  in sex 2,  $q_{j,1}$  is the frequency of allele  $j$  in sex 1 and  $q_{j,2}$  is the frequency of allele  $j$  in sex 2,  $2N$  is the total number of  $M_2$  alleles.

We are grateful to the SEERAD for funding this work. R.P.-W. and J.A.W. acknowledge funding from the BBSRC.

## References

- Bot, J., Karlsson, L. J. E., Greef, J. & Witt, C. (2004). Association of the MHC with production traits in Merino ewes. *Livestock Production Science* **86**, 85–91.
- Caballero, A. & Toro, M. A. (2000). Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research* **75**, 331–343.
- Cockerham, C. C. (1967). Group inbreeding and coancestry. *Genetics* **56**, 89–104.
- de Koning, D. J., Janss, L. L. G., Tattink, A. P., van Oers, P. A. M., de Vries, B. J., Groenen, M. A. M., van der Poel, J. J., de Groot, P. N., Brascamp, E. W. & van Arendonk, J. A. M. (1999). Detection of quantitative trait loci for backfat thickness and intramuscular fat content in pigs (*Sus scrofa*). *Genetics* **152**, 1679–1690.
- Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th edn. Harlow, UK: Addison-Wesley Longman.
- Grundy, B., Villanueva, B. & Woolliams, J. A. (1998). Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genetical Research* **72**, 159–168.
- Maynard-Smith, J. & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23–35.
- Meuwissen, T. H. E. (1997). Maximising the response to selection with a predefined rate of inbreeding. *Journal of Animal Science* **75**, 934–940.
- Pong-Wong, R., George, A. W., Woolliams, J. A. & Haley, C. S. (2001). A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genetics, Selection, Evolution* **33**, 453–471.
- Pong-Wong, R. & Woolliams, J. A. (2007). Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semi-definite programming. *Genetics Selection Evolution* **39**, 3–25.
- Robertson, A. (1965). The interpretation of genotypic ratios in domestic animal populations. *Animal Production* **7**, 319–324.
- Vaiman, M., Renard, C. & Borgeaux, N. (1988). SLA, the major histocompatibility complex in swine; its influence on physiological and pathological traits. In *The Molecular Biology of the Major Histocompatibility Complex of Domestic Animal Species* (ed. C. M. Warner, M. F. Rothschild & S. J. Lamont), pp. 23–38. Ames, IA: Iowa State University Press.
- Villanueva, B., Pong-Wong, R., Fernández, J. & Toro, M. A. (2005). Benefits from marker-assisted selection under an additive polygenic genetic model. *Journal of Animal Science* **83**, 1747–1752.
- Villanueva, B., Pong-Wong, R. & Woolliams, J. A. (2002). Marker assisted selection with optimised contributions of the candidates to selection. *Genetics, Selection, Evolution* **34**, 679–703.