

# Potential gain from including major gene information in breeding value estimation

C Larzul<sup>1</sup>, E Manfredi<sup>2</sup>, JM Elsen<sup>2</sup>

<sup>1</sup> *Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas cedex;*

<sup>2</sup> *Institut national de la recherche agronomique, Station d'amélioration génétique des animaux, 31320 Castanet-Tolosan, France*

(Received 18 March 1996; accepted 11 December 1996)

**Summary** – Two indexes were compared for the selection of a quantitative trait in the case of a mixed inheritance. The first index did not consider the major genotype information (standard method) whereas the second index took this information into account (modified method). Two types of selection scheme were considered: individual selection and selection based on a progeny test. The model for the estimation of genetic progress and evolution of allele frequencies takes overlapping generations into account. All of the effects studied suggested a large number of interactions. However, it can be concluded that information about the major gene should be put into the selection indexes when the heritability is low, the major gene effect high and its initial frequency small, in particular for a recessive major gene. The selection pressure has little influence on the results. In the short term, the modified method is of more value in the case of individual selection than in the case of selection based on a progeny test. On the whole, the extra genetic gain of the modified method is limited and considering the major genotypes in the selection indexes without any change of the selection scheme is probably not the best way to use this information.

**selection / genetic gain / major gene**

**Résumé** – **Intérêt de l'inclusion de l'information au locus majeur dans l'indice de sélection.** *Le but de l'étude est de comparer l'application de deux indices dans le cas d'une sélection sur un caractère quantitatif soumis à l'effet d'un gène majeur. Dans le premier cas, l'indice ne prend pas en compte l'information sur le génotype au locus majeur (méthode standard) alors que le deuxième indice prend en compte cette information (méthode modifiée). Deux types de schémas sont considérés : sélection individuelle et sélection sur descendance. Le calcul du progrès génétique et de l'évolution des fréquences alléliques est réalisé pas à pas en considérant des générations chevauchantes. Tous les effets étudiés sur la supériorité de la méthode modifiée sur la méthode standard suggèrent de nombreuses interactions. Cependant, il ressort que la prise en compte de l'information sur le gène majeur dans l'indexation est avantageuse dans les cas de faible héritabilité, de fort effet du gène majeur et de faible proportion initiale de l'allèle favorable surtout lorsque cet allèle*

*est récessif. Le taux de sélection n'a que peu d'influence sur les résultats. Enfin, l'intérêt de la méthode modifiée est plus visible et plus rapide dans la sélection individuelle que dans la sélection sur descendance. Il n'en demeure pas moins qu'en dehors des conditions extrêmes précédemment citées, l'intérêt de la méthode modifiée sur la méthode standard reste pour le moins limité et la prise en compte de l'information sur les génotypes au locus majeur dans l'indice de sélection, sans modification du schéma de sélection, ne constitue sûrement pas le meilleur outil de valorisation de cette information pour la sélection.*

**sélection / gain génétique / gène majeur**

## INTRODUCTION

Most of quantitative genetics theory and its application to animal breeding is based on the assumption that a trait is controlled by a very large number of small independent genes. Nevertheless, evidence of genes with a large effect on quantitative traits is increasingly being found in livestock: double muscling in pigs (Ollivier, 1980), cattle (Hanset and Michaux, 1985), Callipyge in sheep (Cockett et al, 1994), dwarfism in poultry (Mérat and Ricard, 1974), hyperovulation in sheep (Booroola gene: Piper and Bindon, 1982; Inverdale gene: Davis et al, 1991), high milk protein content in goats (Grosclaude et al, 1987), low technological yield for the cooking of ham in pigs (Le Roy et al, 1990), high milk flow in goats (Ricordeau et al, 1990). In order to take greater advantage of this genetic variability for animal improvement, specific genetic evaluation methods and selection schemes should be applied (Smith, 1967; Soller, 1978; Smith and Webb, 1981; Smith, 1982; Hoeschele, 1990; Gibson, 1994). Alternatively, organisation of matings including genotypic information may be proposed for a more efficient fixation of recessive favourable alleles (eg, Caballero et al, 1991).

In this paper, genotypes at the major locus were perfectly identified, an infrequent situation at the present time (eg, milk protein content in goats, halothane in pigs) but which should become more frequent in the future thanks to progress made in molecular genetics. The usefulness of including the major genotype information in breeding value estimation was evaluated by comparing it with the standard situation where this information is not considered. This comparison was performed in the framework of selection schemes for a trait measured on young animals from both sexes, eg, growth rate (scheme I) and for a trait measured on females only with a progeny test of sires, eg, milk production, (scheme II). Various populations with different genetic contexts (heritability, major gene effect, initial allele frequencies) and organisation (selection pressure, number of generations selected) were studied. Standard and modified situations were compared based on the genetic progress they were expected to produce. The selection schemes considered were very simplified, only the main features of the situations studied were kept. This paper considers, as did Gibson (1994), a dynamic model where the evolution of allele frequencies and genetic means are described step by step, using a model matching the proposition made by Hill (1974) and Elsen and Mocquot (1974). This is a generalization of the Smith (1982) model.

## METHODS

### *Description of the selection schemes*

The generations were overlapping and in demographic equilibrium within an infinite population. The age structure of the population was constant for both sexes. A constant selection pressure of 80% was assumed for the dam–daughter path. The three other paths (sire–son, sire–daughter, dam–son) were selected with the same selection pressure  $q$ . The situations studied, even if somewhat arbitrary, were expected to reflect an average situation for performance test and progeny test selection schemes.

Scheme I is a model of a selection plan organized for instance in a meat sheep or cattle breed, with the trait measured in both sexes at the same time, when animals are between 0 and 1 year old. The generation interval is about 2 years. Only one selection step was considered before the first reproduction for each of the two sexes. The proportions of available breeding animals per age class are given in table I.

Scheme II is a model of a selection plan organized in a dairy species. The generation interval is about 3 years in the present study. The trait was measured only in females. Males were selected after a progeny test on 40 daughters whereas females were selected on their own performance after their first reproduction. In this scheme, a constant 30% of the daughters was supposed to be born from young progeny tested males. The result of the progeny test was available when the young males were 2 years old. The first reproduction of females was not used for replacement. The proportions available per age class are given in table I.

**Table I.** Proportions of available reproducers per age class and sex in schemes I and II.

Age (years)	Scheme I		Scheme II	
	males	females	males	females
0-1	0.40	0.40	0.30	0.0
1-2	0.25	0.25	0.0	0.40
2-3	0.20	0.20	0.35	0.30
3-4	0.10	0.10	0.25	0.20
4-5	0.05	0.05	0.10	0.10

### *Genetic model*

The principles of the model were those of Smith (1982). The whole population was divided into classes defined by the major genotype  $i$  at a single major locus ( $i = AA, AB$  or  $BB$ ,  $A$  being the favourable allele), the age  $j$  and the sex  $k$ . At a given time  $t$ , the components of the classes were their relative size  $\alpha_{ijkt}$ , their major locus genotypic mean value  $C_i$  and their polygenic mean  $\mu_{ijkt}$ . Time 0 ( $t = 0$ ) determined the situation before the selection process began, thus the whole population was considered homogeneous for allele frequencies and polygenic value. At  $t = 1$ , the first generation after selection was applied was born. The

$\alpha_{.jkt} = \sum_i \alpha_{ijkt}$  have been given above. They were constrained to  $\sum_j \alpha_{.jkt} = 1$ . The evolution of the population was described through the evolution of the components  $\mu_{ijkt}$  and  $\alpha_{ijkt}$ , assuming the within class variances to be constant during the whole selection process.

The model included three types of relations as described below.

### Ageing without selection

Between two successive classes of ages  $j - 1$  and  $j$  at time  $t$  and  $t + 1$  without selection, two equalities occurred

$$\begin{aligned}\mu_{ijkt+1} &= \mu_{ij-1kt} \\ \alpha_{ijkt+1} &= \alpha_{ij-1kt} \frac{\alpha_{.jkt}}{\alpha_{.j-1kt}} = \alpha_{ijkt}\end{aligned}$$

### Ageing with selection

When selection was carried out between the ages  $j - 1$  and  $j$ , the previous relations became

$$\mu_{ijkt+1} = \mu_{ij-1kt} + \Delta_{ij-1kt}$$

where  $\Delta_{ij-1kt}$  is the mean polygenic superiority of selected individuals in the class  $ij - 1kt$  at time  $t$ . In practice, there is only one selection step for reproducers, so that only one age class was considered for ageing with selection:  $j = 1$  for both sexes in scheme I;  $j = 2$  for females and  $j = 3$  for males in scheme II

$$\alpha_{ijkt+1} = \alpha_{ij-1kt} \frac{\alpha_{.jkt}}{\alpha_{.j-1kt}} \frac{q_{ijkt}}{q_{jk}}$$

where  $q_{ijkt}$  is the selection pressure for class  $ijk$  at time  $t$  and  $q_{jk}$  is the selection pressure, which is supposed to be constant, for the set of individuals of age  $j$  and sex  $k$ .

### Replacement

The components of the newborn individuals depended on the components of their parents ( $k = s$  for sire,  $k = d$  for dam)

$$\begin{aligned}\mu_{i0kt+1} &= \sum_{i_s j_s} \alpha_{i_s j_s st} \sum_{i_d j_d} \alpha_{i_d j_d dt} \tau_{i_s i_d i} \frac{1}{2} (\mu_{i_s j_s st} + \mu_{i_d j_d dt}) \\ \alpha_{i0kt+1} &= \alpha_{.i0kt+1} \sum_{i_s j_s} \alpha_{i_s j_s st} \sum_{i_d j_d} \alpha_{i_d j_d dt} \tau_{i_s i_d i}\end{aligned}$$

with  $\tau_{i_s i_d i}$  being the probability that an individual has genotype  $i$  given its parents genotypes  $i_s$  and  $i_d$ .

### ***Estimation of the selection pressures and selection differentials***

Since the algebra used is similar for male and female selection and since the selection is performed in only one step, neither the index  $k$  nor the index  $j$  are specified. In order to simplify the algebra, the index  $t$  is also suppressed.

A reproducer  $r$  is characterized by its global genetic value  $\Gamma_r$  which includes its polygenic value  $g_r$  and its major locus genotypic value  $G_r$ . The parental value  $H_r$  of a reproducer was defined as the expected progeny performance  $X_p$ , ie, half the breeding value defined by Falconer (1989). It was estimated by the selection index  $I = \hat{H}_r$  corresponding to the expectation of  $X_p$  dependent on various types of information according to the case: own performance  $X_r$  (scheme I and females in scheme II) or offspring performances  $X_o$  (males in scheme II) and with the genotypic information at the major locus,  $G_r$  and  $G_o$ . In the standard method, the selection is made on an index supposed to be an expectation of the parental value when ignoring the existence of the major locus: the index  $I$  is defined as a simple regression on the own performance value  $X_r$  (scheme I) or offspring performances  $X_o$  (males in scheme II). The evolution of genetic value of selected reproducers, applying either index, has to be calculated as well as changes of allele frequencies and polygenic mean of each genotype.

The joint probability density of the genetic value  $\Gamma_r$  of the reproducer  $r$  and of its index  $I$  is  $f(\Gamma_r, I)$ . This density is a mixture of subdensities  $\phi_{i_r}$  corresponding to genotypes  $i_r$ :

$$f(\Gamma_r, I) = \sum_{G_r=i_r} \alpha_{i_r} \phi_{i_r}(\Gamma_r, I)$$

with  $\alpha_{i_r}$  being the  $i_r$  class frequency within the considered group of reproducers. In practice: in scheme I,  $\alpha_{i_r} = \frac{\alpha_{i0st}}{\sum_i \alpha_{i0st}}$  for males and  $\alpha_{i_r} = \frac{\alpha_{i0dt}}{\sum_i \alpha_{i0dt}}$  for females and,

in scheme II,  $\alpha_{i_r} = \frac{\alpha_{i2st}}{\sum_i \alpha_{i2st}}$  for males and  $\alpha_{i_r} = \frac{\alpha_{i1dt}}{\sum_i \alpha_{i1dt}}$  for females. The within

subclass distributions,  $\phi_{i_r}(\Gamma_r, I)$ , were assumed to be multi-normal distributed with the moments  $E_{i_r}$  and  $V_{i_r}$  depending on the particular case considered. The components of these moments are

$C_{i_r}$  :  $i_r^{th}$  genotypic mean value

$\mu_{i_r}$  : polygenic mean of the  $i_r^{th}$  major genotype class

$\sigma_g^2$  : within genotype additive polygenic variance

$\sigma_p^2$  : within genotype phenotypic variance

$h^2$  : within major genotype polygenic heritability  $h^2 = \frac{\sigma_g^2}{\sigma_p^2}$

The within genotype variances,  $\sigma_g^2$  and  $\sigma_p^2$ , were independent of both genotype  $i_r$  and time  $t$ .

The within genotype polygenic mean superiority of the selected individuals is given by

$$\Delta_{i_r} = \int_{\Gamma_r=-\infty}^{+\infty} \Gamma_r \int_{I=T}^{+\infty} \phi_{i_r}(\Gamma_r, I) dI d\Gamma_r / q_{i_r}$$

where  $T$  is the selection threshold (the  $I$  value above which the candidates are selected) and  $q$  the selection pressure corresponding to  $T$ :

$$q = \sum_{G_r=i_r} \alpha_{i_r} q_{i_r}$$

with

$$q_{i_r} = \int_{\Gamma_r=-\infty}^{+\infty} \int_{I=T}^{+\infty} \phi_{i_r}(\Gamma_r, I) dI d\Gamma_r$$

Application of these principles to the different cases studied is described in the *Appendix*.

In all cases studied, the threshold is found iteratively, as described by Ducrocq and Quaas (1988). However, contrary to the standard situation, the breeding value evaluation taking the major locus genotype into account was obtained after a two level iterative process: since the parental value  $H_r$  has been defined as the progeny mean, it depends on the genotypic structure of the selected mate (ms) population (the  $\alpha_{i_{ms}}$  and  $\mu_{i_{ms}}$ ) which itself depends on the  $\alpha_{i_{rs}}$  and  $\mu_{i_{rs}}$  of the selected reproducers (rs). Taking as a starting point the genotypic structure of the mate population before selection, the solution was obtained iteratively with a given selection pressure  $q$ . In order to simplify the algebra of the young male indexes  $I$ , it was assumed that the characteristics (mean polygenic values and major genotype frequencies) of the female population (when selecting males) could replace those of their future mates.

### **Comparison criteria**

The value of including the genotype information in the parental value estimation was measured by the extra genetic gain as compared with the standard method. Starting from an initial point where all within major genotype classes were assumed to have equal polygenic means ( $\mu_{i_{jkt}} = \mu \forall i, j, k$ ), the nonlinear change of the  $\alpha_{i_{jkt}}$  and  $\mu_{i_{jkt}}$  over time differed between the two parental value estimation methods. The evolution of the 0-1-year old females ( $y_t = \sum_i \alpha_{i0t}(\mu_{i0t} + C_i) / \alpha_{i0t}$ ) was used as a measure of genetic progress, but our primary criterion was:

$$C(t_f) = \sum_{t=0}^{t=t_f} \delta y_t$$

with  $t_f$  being the number of years considered and  $\delta y_t$  the difference between both methods for year  $t$ .

This criterion was preferred to the final deviation  $\delta y_{t_f}$  which gives only a partial description of the differences between both methods. Preliminary analyses showed that comparisons between the methods were hardly influenced by the inclusion of

a discounting factor in the  $t$ -summations, and the comparisons were finally limited to a nondiscounted criterion. The methods were also compared according to the evolution of the allele frequencies.

### Cases studied

The selection methods were compared for various combinations of the following parameters:

*Genetic parameters:* the within major genotype heritability coefficient ( $h^2$ ) was given values between 0.1 and 0.5 and the 'major gene effect' defined here as  $\Delta C = C_{AA} - C_{BB}$  between 1 and 3 within genotype phenotypic standard deviations. Allele  $A$  was dominant ( $AA = AB = \Delta C, BB = 0$ ), additive ( $AA = 2AB = \Delta C, BB = 0$ ) or recessive ( $AA = \Delta C, AB = BB = 0$ ) over the allele  $B$ . Initial frequency  $p$  for allele  $A$  was tested between 0.1 and 0.9.

$$\text{The global heritability } (H^2 = \frac{\sigma_g^2 + \sum_{G_r=i_r} frq(G_r)(C_{G_r}^2) - [\sum_{G_r=i_r} frq(G_r)(C_{G_r})]^2}{\sigma_p^2 + \sum_{G_r=i_r} frq(G_r)(C_{G_r}^2) - [\sum_{G_r=i_r} frq(G_r)(C_{G_r})]^2},$$

with  $frq(G_r)$  the frequency of genotype  $G_r$ , which includes both polygenes and major genes, depends on polygenic heritability, major gene effect (both constant) and allele frequencies (variable with time). Initial  $H^2$  is between 0.11 and 0.81 (table II).

*Population structure:* the selection pressure  $q$  was given values of 5, 10 and 20%.

**Table II.** Global heritability  $H^2$  (and proportion of overall phenotypic variance due to major gene) in relation to the  $A$  allele initial frequency  $p(A)$ , polygenic heritability  $h^2$ , degree of dominance and major gene effect  $\Delta C$ .

$\Delta C$	$h^2$	$p(A) = 0.1$				$p(A) = 0.5$					
		A recessive		A additive	A dominant		A recessive	A additive			
1	0.1	0.11	(1.0%)	0.14	(4.3%)	0.22	(13.3%)	0.24	(15.8%)	0.20	(11.1%)
	0.2	0.21		0.23		0.31		0.33		0.29	
	0.3	0.31		0.33		0.39		0.41		0.38	
	0.4	0.41		0.43		0.48		0.49		0.47	
	0.5	0.50		0.52		0.57		0.58		0.56	
2	0.1	0.13	(3.8%)	0.24	(15.3%)	0.44	(38.1%)	0.49	(42.9%)	0.40	(33.3%)
	0.2	0.23		0.32		0.50		0.54		0.47	
	0.3	0.33		0.41		0.57		0.60		0.53	
	0.4	0.42		0.49		0.63		0.66		0.60	
	0.5	0.52		0.58		0.69		0.71		0.67	
3	0.1	0.17	(8.2%)	0.36	(18.8%)	0.62	(58.1%)	0.67	(62.8%)	0.58	(52.9%)
	0.2	0.27		0.43		0.66		0.70		0.62	
	0.3	0.36		0.50		0.71		0.74		0.67	
	0.4	0.45		0.57		0.75		0.78		0.72	
	0.5	0.54		0.64		0.79		0.81		0.76	

## RESULTS AND DISCUSSION

### *Evolution of mean genetic and polygenic values*

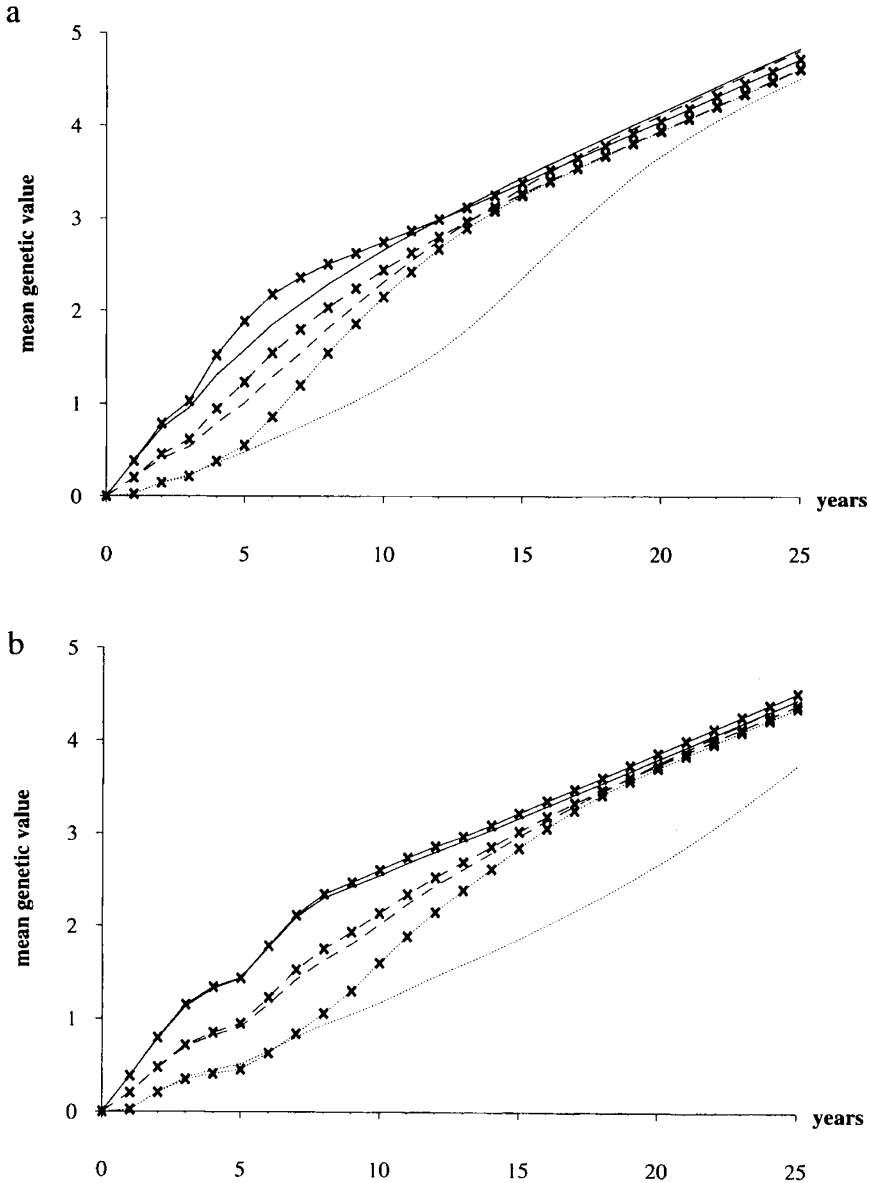
The evolution of the mean genetic values of young females is illustrated in figure 1 for the case of  $A$  dominant, additive and recessive with  $h^2 = 0.3$ ,  $p = 0.1$ ,  $q = 0.1$  and  $\Delta C = 2$ . In scheme I (fig 1a), when  $A$  is dominant or additive, the difference is nil at the beginning of the process. In the medium term, the modified method shows a higher increase of mean genetic value, essentially owing to the faster fixation of the favourable allele. In the long term, the standard method appears more efficient when comparing the final mean genetic value. When allele  $A$  is recessive, the modified method is slightly less efficient in the short term ( $-0.02\sigma_p$ ), but from year 3, this method becomes and remains more efficient than the standard selection ( $+0.08\sigma_p$ ). The reduced efficiency of the modified method within the very first years is observed for the large major gene effect ( $\Delta C = 2$  or  $\Delta C = 3$ ), but not for  $\Delta C = 1$ . In scheme II (fig 1b), with the same parameters, the maximal difference between both methods is lower than that observed in scheme I. When  $A$  is dominant, mean genetic value is always higher when applying the modified method, with a nil difference at the beginning that vanishes in the long run ( $+0.06\sigma_p$ ). For  $A$  additive, the modified method becomes less efficient than the standard method within the first 25 years of selection (year 17). In the long term (not shown), the modified method becomes less efficient for  $A$  recessive but not for  $A$  dominant. Lower mean genetic values are observed for the case of  $A$  recessive in the first five generations for the modified selection ( $-0.05\sigma_p$ ).

The lower efficiency in the long term of marker assisted selection or combined selection when taking into account a major gene, when effects of alleles are additive, is now established (Gibson, 1994; Woolliams and Pong-Wong, 1995). The recessive case is not mentioned in these studies. The relative superiority of one method compared to the other is dependent on the rate of fixation of the favourable allele, but also on polygenic value evolution till fixation. An example is given in figure 2a and b in the case of  $A$  recessive and additive with  $\Delta C = 2$ ,  $h^2 = 0.3$ ,  $p = 0.1$  and  $q = 0.1$ . The polygenic mean increases more rapidly when the standard indexes are applied. This phenomenon is observed for both selection schemes, with a stronger effect in the case of scheme II during the early years. In the case of individual selection and  $A$  recessive, this tendency changes after fixation of the favourable effect in the modified method (year 15) giving a faster increase of polygenic values in this modified method as compared to the standard one. When the favourable allele is fixed in the standard scheme, the evolution patterns become parallel. In the case of scheme II, these phenomena do not appear during the first 25 years of selection.

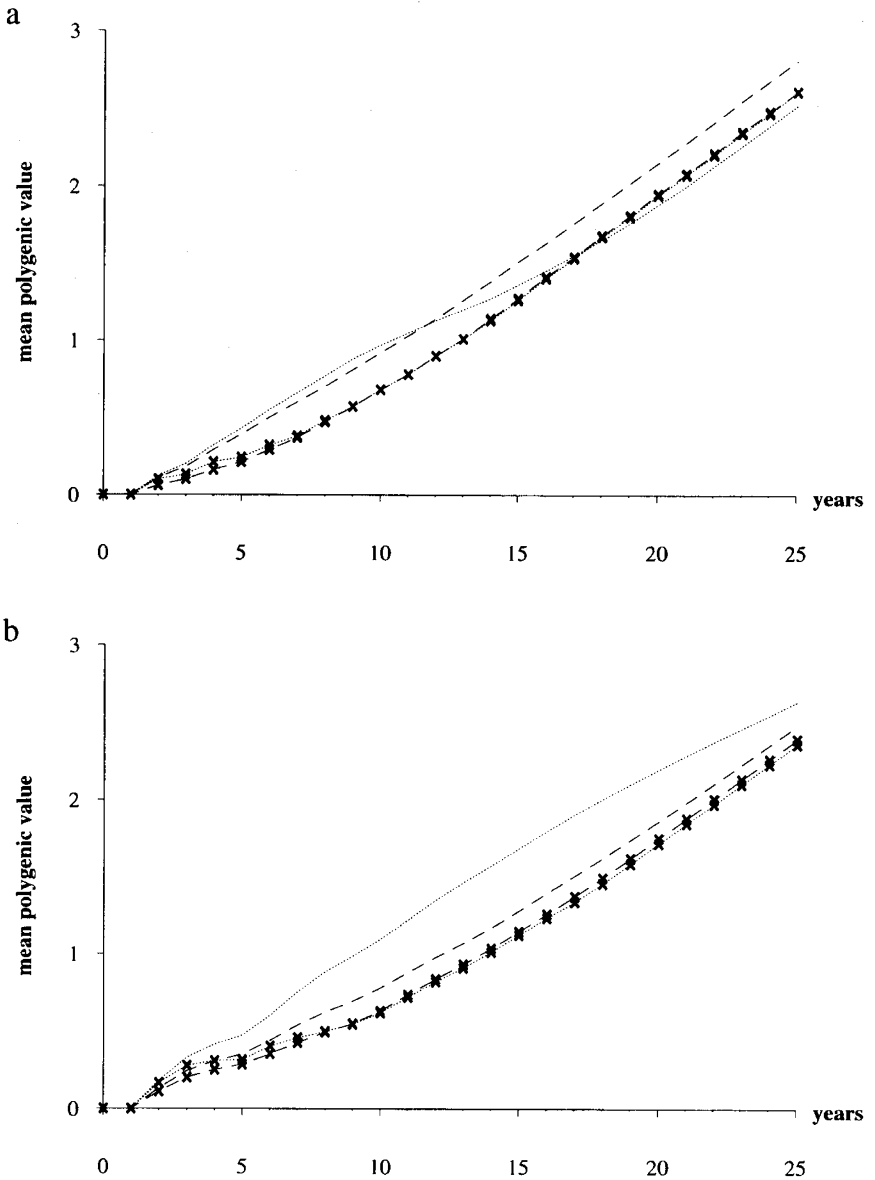
### *Choice of period length $t_f$*

Our criterion is a measure of the weighted surface between both mean genetic value curves, truncated at the final time  $t_f$ . The criterion  $C(t_f)$  reaches its maximum value for intermediate  $t_f$ , as illustrated in figure 3 for  $h^2 = 0.3$ ,  $\Delta C = 2$ ,  $p = 0.1$  and  $q = 0.10$  for  $A$  recessive. In this situation, the maximum is achieved at year 12





**Fig 1.** Evolution of the mean genetic values of young females in phenotypic standard deviation (polygenic heritability  $h^2 = 0.3$ , allele  $A$  initial frequency  $p = 0.1$ , selection pressure  $q = 0.1$ , major gene effect  $\Delta C = 2$ ). (a) Selection scheme on individual performance (scheme I). (b) Selection scheme on progeny test performances (scheme II). Standard method: index without genotypic information: —,  $A$  dominant; ---,  $A$  additive; - · - ·,  $A$  recessive. Modified method: index with genotypic information: —×—,  $A$  dominant; - × - ,  $A$  additive; - · - · × - · - ·,  $A$  recessive.

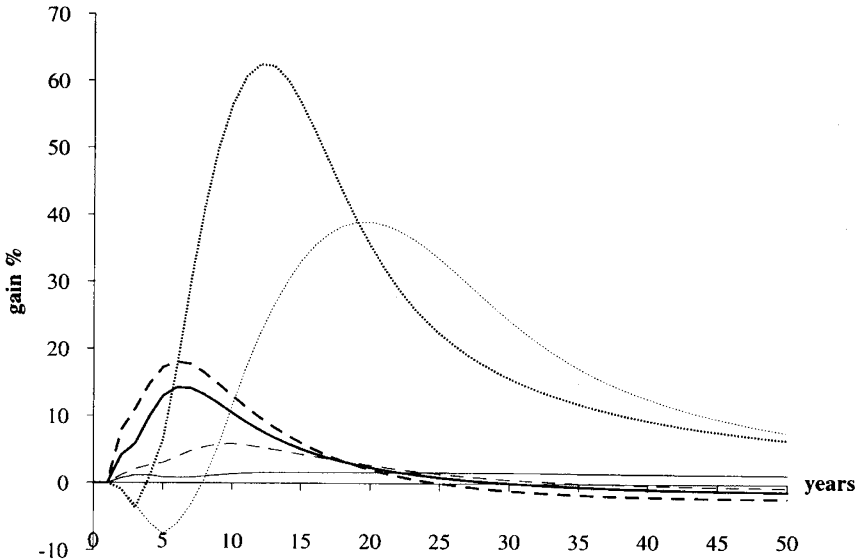


**Fig 2.** Evolution of the mean polygenic values of young females in phenotypic standard deviation (polygenic heritability  $h^2 = 0.3$ , allele  $A$  initial frequency  $p = 0.1$ , selection pressure  $q = 0.1$ , major gene effect  $\Delta C = 2$ ). (a) Selection scheme on individual performance (scheme I). (b) Selection scheme on progeny test performances (scheme II). Standard method: index without genotypic information: ———,  $A$  additive; - - - - -,  $A$  recessive. Modified method: index with genotypic information: - × - -,  $A$  additive; - - - × - - -,  $A$  recessive.

in the case of scheme I, and at year 22 in the case of scheme II. For  $A$  dominant and additive, the maximum is lower and achieved earlier.

Figure 3 indicates that including the major gene information in the selection criterion gives a slightly negative result in the very first few years, only in the case of a recessive favourable allele. This is probably due to the nonoptimality of our criterion when considering, in the evaluation of the breeding value  $I$  of the future reproducer, the genotypic structure of the contemporary mate population before selection as fully representative of the whole genotypic structure of the dams. An optimal index should take into account the whole future mate population structure. In fact, this negative result in the very first few years appears when the initial frequency of allele  $A$  is lower than or equal to 0.1 (not shown) and when allele  $A$  is recessive. In this case, the modified method permits selection of  $AB$  genotypes instead of  $BB$  genotypes, even if their polygenic values are lower. The proportion of  $AB$  in mates is not high enough to increase the proportion of  $AA$  in the progeny greatly, thus the lower polygenic gain is probably not counterbalanced by the increase of  $AA$  genotypes.

In the following discussion, unless otherwise mentioned, the results are given for  $t_f = 10$ , a period length for which differences between both methods are maximal.



**Fig 3.** Effect of the period length of selection on the evolution of the extra gain (polygenic heritability  $h^2 = 0.3$ , allele  $A$  initial frequency  $p = 0.1$ , selection pressure  $q = 0.1$ , major gene effect  $\Delta C = 2$ ). Scheme I: selection scheme on individual performances: ———,  $A$  dominant; - - - - ,  $A$  additive; ·····,  $A$  recessive. Scheme II: selection scheme on progeny performances: ———,  $A$  dominant; - - - - ,  $A$  additive; ·····,  $A$  recessive.

### ***Major gene and polygenic effects***

The influence of heritability and major gene effect parameters on genetic progress is described in figure 4a and b, considering an initial allele  $A$  frequency  $p = 0.10$  and a selection pressure  $q = 0.10$ .

The gain  $C(t_f)$  decreases when the heritability increases: the greater the extent to which the genetic variation may be explained by the major gene, the more it becomes worthwhile to include the corresponding information in the breeding evaluation. This result was already observed by Smith (1967) who compared selection based on (1) individual performance, (2) known genetic loci and (3) a selection index of (1) and (2) on the basis of their short term responses. Marker assisted selection is also most useful when the heritability of the trait is low (Lande and Thompson, 1990; Ruane and Colleau, 1995; Meuwissen and Goddard, 1996), at least in the short term.

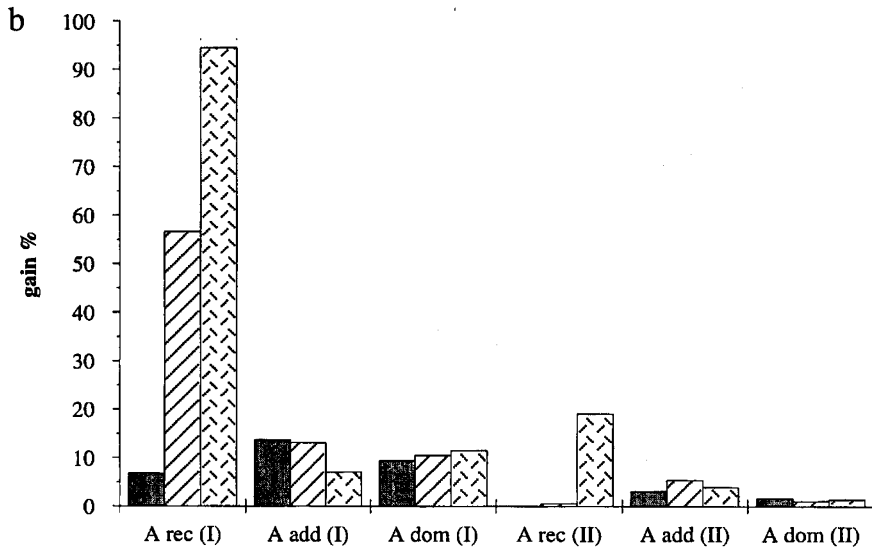
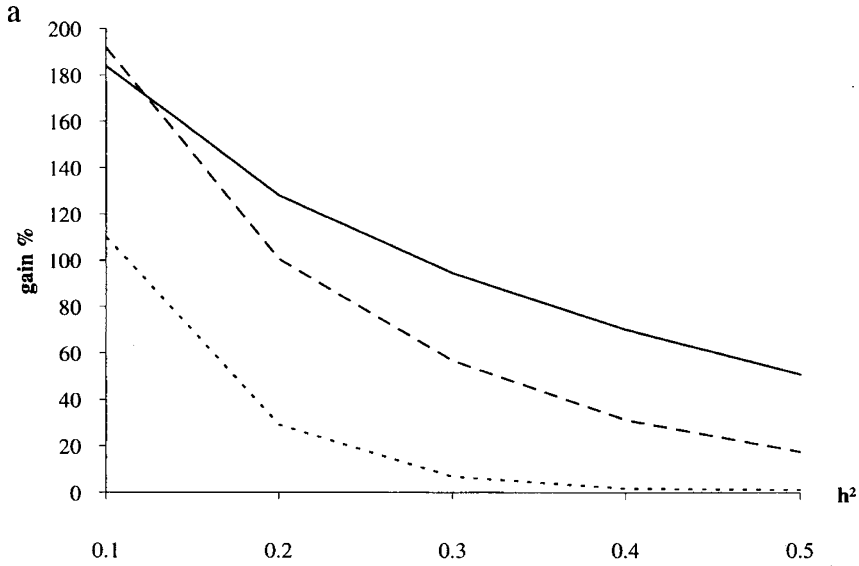
The effect of the deviation between  $AA$  and  $BB$  depends on the degree of dominance: the gain  $C(t_f)$  is higher with increasing major gene effect when  $A$  is recessive, and lower in other situations. The main value of including the genotypic information in the parental value estimation is the possibility of selecting carriers which do not show their superiority when only their phenotypes are considered: this is the case when  $A$  is recessive or when  $A$  is codominant or dominant but with a small effect.

This gain  $C(t_f)$  may be quite important when the favourable allele  $A$  is recessive (up to 200% in scheme I) but decreases when its dominance over  $B$  increases. It becomes nearly nil for full dominance in scheme II. These results, which confirm the previous hypothesis, could be explained by the following arguments. When  $A$  is recessive and infrequent, the standard selection has poor efficiency for increase of allele  $A$  frequency, for  $AB$  and  $BB$  have the same value. Thus, they have nearly the same chance of being selected if the number of reproducers to be retained is higher than the number of  $AA$  in the candidates. On the contrary, the modified selection distinguishes  $AB$  and  $BB$  candidates, and thus is more efficient to increase  $A$  allele frequency in the short term. That is not the case when allele  $A$  is dominant or when its frequency is high enough. This difference is also reduced in scheme II because  $AB$  and  $BB$  genotypes of the reproducers are more distinct with progeny testing.

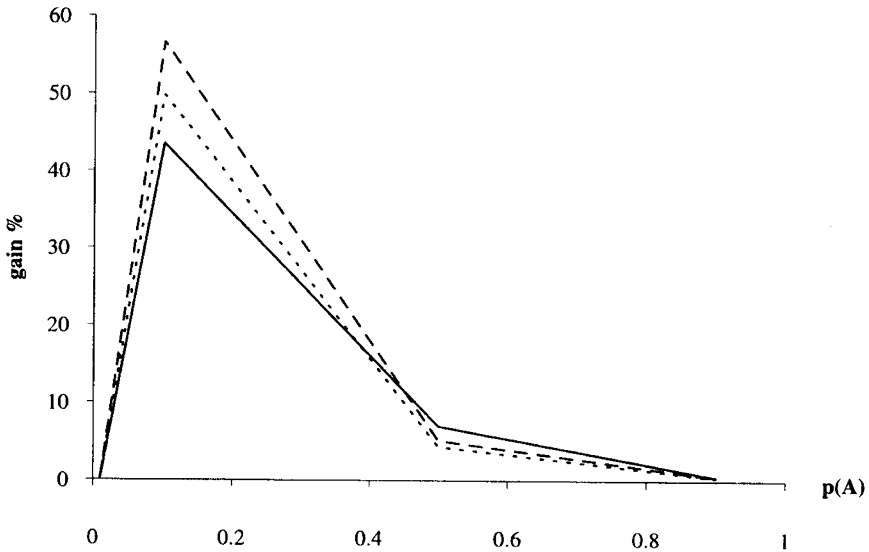
### ***Allele frequencies***

An illustration of the influence of the initial allele  $A$  frequency on the difference in genetic progress between methods is given in figure 5 for scheme I,  $A$  recessive, considering an heritability  $h^2 = 0.3$  and a major gene effect  $\Delta C = 2$ , reflecting the general findings.

In scheme II, the gain  $C(t_f)$  is very low and the differences owing to the initial frequency  $p$  are negligible. In scheme I, the gain reaches a maximum for small  $p$  values, with the exception of the recessive allele  $A$  case where a maximum is obtained for intermediate values (0.10), while no gain is obtained with a very small initial  $p$ . This result is due to the curvilinearity of allele  $A$  frequency evolution with selection. This is illustrated in figure 6a and b where  $A$  is recessive and  $q = 0.10$ : for an intermediate period length  $t_f$  of 10 years, the difference between the standard



**Fig 4.** Effect of the polygenic heritability  $h^2$  and the major gene effect  $\Delta C$  on the genetic gain (allele  $A$  initial frequency  $p = 0.1$ , selection pressure  $q = 0.1$ ). (a) Scheme I, favourable allele  $A$  recessive: - · - · - ·,  $\Delta C = 1$ ; - - - -,  $\Delta C = 2$ ; ———,  $\Delta C = 3$ . (b) Schemes I and II, effect of major gene dominance and major gene effect  $\Delta C$  (polygenic heritability  $h^2 = 0.3$ ): ■,  $\Delta C = 1$ ; ▨,  $\Delta C = 2$ ; ▩,  $\Delta C = 3$ .



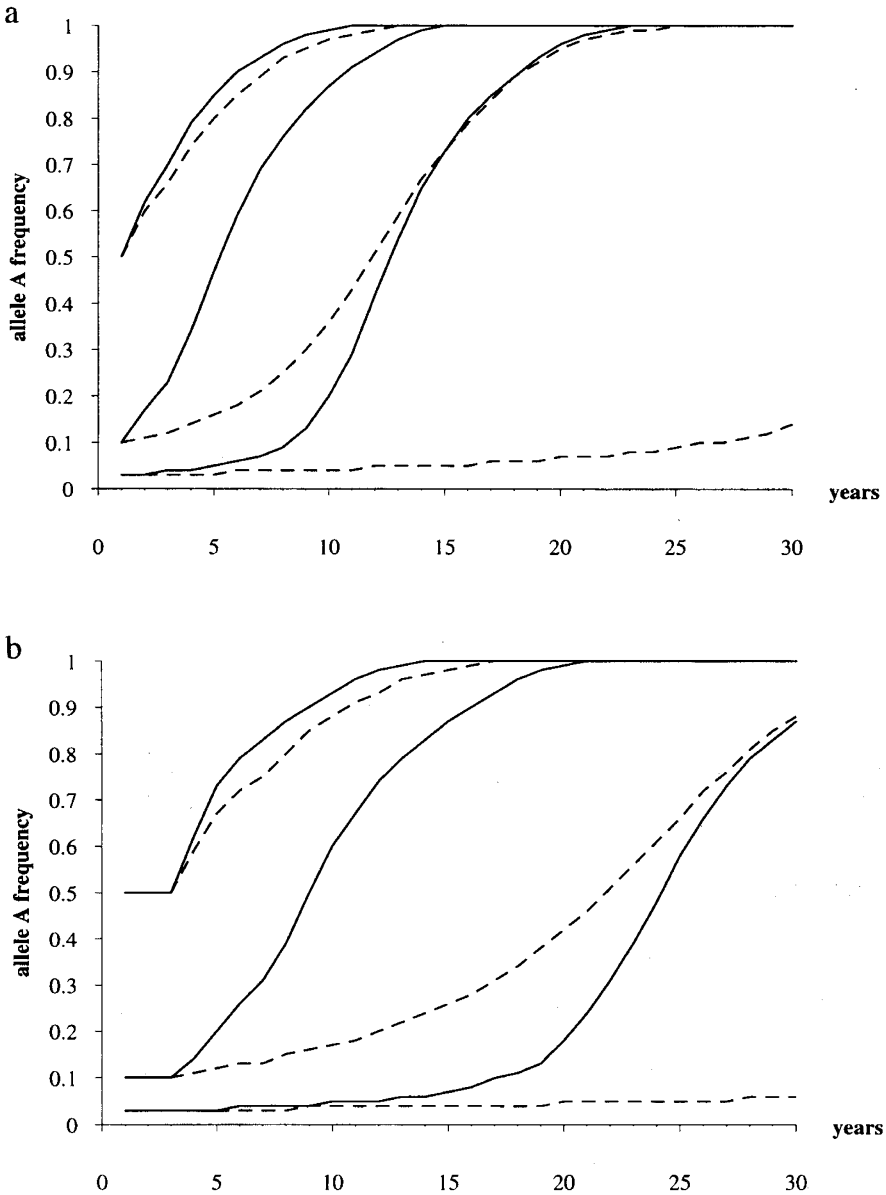
**Fig 5.** Effect of the allele  $A$  initial frequency and the selection pressure on the genetic gain (scheme I, favourable allele  $A$  recessive, polygenic heritability  $h^2 = 0.3$ , major gene effect  $\Delta C = 2$ ): - - - -,  $q = 0.05$ ; - · - · -,  $q = 0.10$ ; —,  $q = 0.20$ .

and modified methods is maximum when  $p = 0.10$ . This is not true for a longer period, when the allele  $A$  frequencies may differ between the two methods. These results clearly show that the value of putting genotypic information in the parental values is directly related to the acceleration of allele  $A$  fixation that it permits, which in turn depends on the starting point  $p$ .

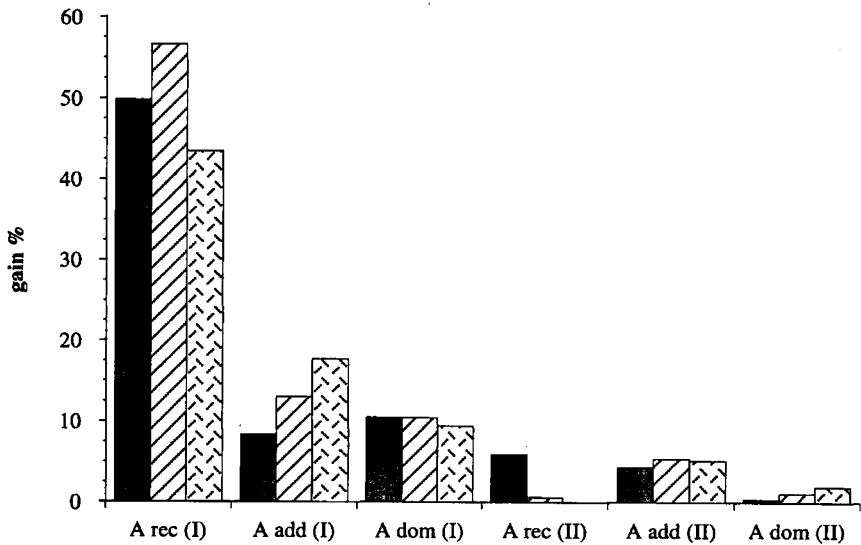
### **Selection scheme and selection pressure**

Figures 3 and 4b were drawn for schemes I and II. In all the cases studied, the maximum gain  $\mathcal{C}(t_f)$  was much higher for scheme I. There might be two reasons for this difference: (1) more complete information about the whole genetic value of reproducers was available from the progeny test than from the performance test, thus diminishing the value of including major gene data and (2) the longer time taken by scheme II to take into account the extra information (ie, to increase allele  $A$  frequency) on the major gene in parental value evaluation. A comparison based on a longer length period  $t_f$  should give a higher  $\mathcal{C}(t_f)$  in scheme II and a lower one in scheme I (cf fig 3).

The effect of selection pressure  $q$  depends on the degree of dominance, scheme (fig 7) and initial allele frequencies (fig 5) but in general, it seems to have a very limited influence on the gain  $\mathcal{C}(t_f)$ .



**Fig 6.** Effect of the allele *A* initial frequency on the evolution of this frequency (favourable allele *A* recessive, polygenic heritability  $h^2 = 0.3$ , major gene effect  $\Delta C = 2$ , selection pressure  $q = 0.10$ ). (a) Scheme I and (b) scheme II. - - - -, index without genotypic information; —, index with genotypic information.



**Fig 7.** Effect of the selection pressure on the genetic gain (polygenic heritability  $h^2 = 0.3$ , allele  $A$  initial frequency  $p = 0.1$ , major gene effect  $\Delta C = 2$ ): ■,  $q = 0.05$ ; ▨,  $q = 0.10$ ; ▩,  $q = 0.20$ .

## GENERAL DISCUSSION

We found that, in comparison with the traditional breeding value estimation, which assumes polygenic inheritance, the inclusion of information about the genotype at a major locus is valuable in limited circumstances, which could roughly be defined as those cases in which the standard methods are less effective at fixing the favourable allele (very low  $A$  initial frequency, recessivity of  $A$ ) or when most of the gain comes from the major gene itself (low heritability, short term results). The value of including the major gene information in the selection indexes may be very high in the most favourable cases (200% increase of the genetic gain), but it is more often low or slightly negative.

The situation of additive QTL was studied by others with divergent results. Negative long term results were obtained by Gibson (1994) and Woolliams and Pong Wong (1995). On the contrary, Zhang and Smith (1992), Gimelfarb and Lande (1994) found positive extra rates of genetic responses with marker-assisted selection based on the use of linkage disequilibrium, a situation much less favourable than ours owing to the progressive disappearance of marker-QTL associations. However, in all these studies, a diminution of the superiority of the modified methods when considering more generations is constant. The divergences between results may come from the characteristics of the genes studied, number of generations simulated as well as type of modelling used. The higher efficiency of methods accounting for major gene information over standard methods with lower global heritability and



higher initial favourable gene frequency was already shown by Smith (1967) for an additive major gene and by many others for additive marker QTL (eg, Lande and Thompson, 1990; Gimelfarb and Lande, 1994; Edwards and Page, 1994; Ruane and Colleau, 1995, 1996).

Contrary to others, our criterion for evaluating the efficiency of alternative methods did not consider only the mean genetic level or gene frequency at a given time but included the dynamics of the evolution due to selection. We emphasized that most of the selection schemes are able to fix favourable alleles and the differences between schemes are to be appreciated in the way they reach this state. This modelling is classical for the comparison of selection plans and needed for their economic evaluation.

Our model assumed an infinite number of loci and population size and considered only the evolution of major genotype frequencies and mean polygenic values with selection. Linkage disequilibrium between major gene and the polygenes was automatically accounted for in the model, but not the Bulmer effect within major genotype. The corresponding reduction in polygenic variance should occur in both standard and modified selection schemes. Whether this reduction is higher in the modified or standard forms is far from obvious for three reasons. First, as compared to the standard situation, the modified method induced a weaker selection pressure on polygenes within favourable genotypes, and a stronger one within unfavourable genotypes. Second, the individuals of a given genotype may have progeny of different genotypes (*BB* giving for instance *AB* offspring, and even *AA* grand-offspring) with a corresponding redistribution of polygenic variation. Third, the evolutions of allele frequencies at a polygene on the one hand, and of linkage disequilibrium between polygene loci on the other depend on their location relative to the major locus. Thus a full model should describe not only this possible variance reduction but also should deal with the linkage between the major locus and some of the minor loci controlling the trait. In a simulation of standard and modified selection schemes of type I describing the polygenic value as the sum of identified QTL (defined by their location on the genome and allele effect), Fournet et al (1995, 1997) did not find any significant difference in polygenic variance reduction between selection schemes. Finally, the Bulmer effect is most important at the beginning of a selection scheme while modification of a selection scheme to account for the segregation of a major gene should occur in an already running scheme, minimizing this effect.

This study only dealt with a possible change in breeding value estimations without any modification of the selection plan. The information given by genotypes at a major locus may be used to change the organization of the selection scheme itself. The most effective for schemes type II would probably be a preselection of young males based on their own genotype before (or by replacing) their progeny test: Smith (1967), Soller (1978), Gomez-Raya and Gibson (1993). This kind of preselection was studied when QTLs, indirectly detected through the use of marker information, were known (eg, Soller and Beckmann, 1983; Kashi et al, 1990; Meuwissen and Van Arendonk, 1992; Brascamp et al, 1993). A dynamic model similar to the model used in this paper could yield more information on the efficiency of these new plans.

## REFERENCES

- Brascamp EW, Van Arendonk JAM, Groen AF (1993) Economic appraisal of the utilization of genetic markers in dairy cattle breeding. *J Dairy Sci* 76, 1204–1213
- Caballero A, Keightley PD, Hill WG (1991) Strategies for increasing fixation probabilities of recessive mutations. *Genet Res Camb* 58, 129–138
- Cockett NE, Jackson SP, Shay T, Nielsen D, Moore SS, Steele MR, Barendse W, Green RD, Georges M (1994) Chromosomal localisation of the callipyge gene in sheep (*Ovis aries*) using bovine DNA markers. *Proc Natl Acad Sci, USA*, 91
- Davis GH, McEwan JC, Fenessy PF, Dodds KG, Farquhar PA (1991) Evidence for the presence of a major gene influencing ovulation rate on the X chromosome of sheep. *Biol. Reprod.* 44, 620–624
- Ducrocq V, Quaas RL (1988) Prediction of genetic response to truncation across generations. *J Dairy Sci* 71, 2543–2553
- Edwards MD, Page NJ (1994) Evaluation of marker-assisted selection through computer simulation. *Theor Appl Genet* 88, 676–382
- Elsen JM, Mocquot JC (1974) Recherches pour une rationalisation technique des schémas de sélection des bovins et des ovins. *Bulletin Technique du Département de Génétique Animale* 17, INRA, Paris
- Falconer DS (1989) *Introduction to Quantitative Genetics*. Longman, London, 3rd ed
- Fournet F, Hospital F, Elsen JM (1995) A FORTRAN program to simulate the evolution of genetic variability in a small population. *Comput Applic Biosci* 11, 469–475
- Fournet F, Elsen JM, Barbieri ME, Manfredi E (1997) Effect of including major gene information in mass selection: a stochastic simulation in a small population. *Genet Sel Evol*, 29, 35–56
- Gibson JP (1994) Short term gain at the expense of long term response with selection on identified loci. In: *Proc 5th World Cong Genet Appl Livest Prod* 21, 201–204
- Gimelfarb A, Lande R (1994) Simulation of marker assisted selection in hybrid populations. *Genet Res Camb* 63, 39–47
- Gomez-Raya L, Gibson JP (1993) Within-family selection at an otherwise unselected locus in dairy cattle. *Genome* 36, 433–439
- Grosclaude F, Mahé MF, Brignon G, Distasio L, Jeunet R (1987) A Mendelian polymorphism underlying quantitative variations of goat  $\alpha$ -S1 casein. *Genet Sel Evol* 19, 399–412
- Hanset R, Michaux C (1985) On the genetic determinism of muscular hypertrophy in the Belgian White and Blue cattle breed. I. Experimental data. *Genet Sel Evol* 17, 359–368
- Hill WG (1974) Prediction and evaluation of response to selection with overlapping generations. *Anim Prod* 18, 117–140
- Hoeschele I (1990) Potential gain from the insertion of major genes into dairy cattle. *J Dairy Sci* 73, 2601–2618
- Kashi Y, Hallerman E, Soller M (1990) Marker-assisted selection of candidate bulls for progeny testing programmes. *Anim Prod* 51, 63–74
- Lande R, Thompson R (1990) Efficiency of marker assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756
- Le Roy P, Naveau J, Elsen JM, Sellier P (1990) Evidence for a new major gene influencing meat quality in pigs. *Genet Res Camb* 55, 33–40
- Meuwissen THE, Van Arendonk JAM (1992) Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. *J Dairy Sci* 75, 1651–1659
- Meuwissen THE, Goddard ME (1996) The use of marker haplotypes in animal breeding schemes. *Genet Sel Evol* 28, 161–176

- Mérat P, Ricard FH (1974) Etude d'un gène de nanisme lié au sexe chez la poule: importance de l'état d'engraissement et gain de poids chez l'adulte. *Ann Génét Sél Anim* 6, 211–217
- Ollivier L (1980) Le déterminisme génétique de l'hypertrophie musculaire chez le Porc. *Ann Génét Sél Anim* 12, 383–394
- Piper LR, Bindon BM (1982) The *Booroola Merino* and the performance of medium non-peppin crosses at Armidale. In: *The Booroola Merino* (LR Piper, BM Bindon, RD Nethery, eds), CSIRO, Melbourne, 9–20
- Ricordeau G, Bouillon J, Le Roy P, Elsen JM (1990) Déterminisme génétique du débit de lait au cours de la traite des chèvres. *INRA Prod Anim* 3, 121–126
- Ruane J, Colleau JJ (1995) Marker assisted selection for genetic improvement of animal populations when a single QTL is marked. *Genet Res Camb* 66, 71–83
- Ruane J, Colleau JJ (1996) Marker assisted selection for a sex-limited character in a nucleus breeding population. *J Dairy Sci* 79, 1666–1678
- Smith C (1967) Improvement of metric traits through specific genetic loci. *Anim Prod* 9, 349–358
- Smith C (1982) Estimates of trends in the Halothane gene in pig stocks with selection. *Z Tierzüchtg Züchtgsbiol* 99, 232–240
- Smith C, Webb AJ (1981) Effects of major genes on animal breeding strategies. *Z Tierzüchtg Züchtgsbiol* 98, 161–169
- Soller M (1978) The use of loci associated with quantitative effects in dairy cattle improvement. *Anim Prod* 27, 133–139
- Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. *Theor Appl Genet* 67, 25–33
- Woolliams JA, Pong-Wong R (1995) Short-term versus long-term responses in breeding schemes. In: *46th Annual Meeting of the EAAP, Prague, 4–7 September 1995*
- Zhang W, Smith C (1992) Simulation of marker-assisted selection utilizing linkage disequilibrium. *Theor Appl Genet* 83, 813–820

## APPENDIX

### Individual selection

#### Without genotypic information

The selection index  $I$  is

$$I = \frac{h^2}{2}(X_r - \mu) \quad [A1]$$

with  $\mu$ , the general mean and  $h^2$ , the polygenic heritability (note that the ranks of the candidates are not affected by the value of this coefficient which could be replaced by the global heritability with no effect on the genetic progress).

The joint distribution of the genetic value  $\Gamma_r$  of the reproducer  $r$  and its index  $I$  is  $f(\Gamma_r, I)$ , a mixture of subdensities  $\phi_{i_r}$  corresponding to genotypes  $i_r$ :

$$f(\Gamma_r, I) = \sum_{G_r=i_r} \alpha_{i_r} \phi_{i_r}(\Gamma_r, I)$$

with  $\alpha_{i_r}$  being the  $i_r$  class frequency for the considered group of reproducers.

The moments  $E_{i_r}$ , the within genotype expected means of  $\Gamma_r$  and  $I$ , and  $V_{i_r}$ , the variance-covariance matrix of  $\Gamma_r$  and  $I$ , are

$$E_{i_r} = \begin{pmatrix} C_{i_r} + \mu_{i_r} \\ \frac{h^2}{2}(C_{i_r} + \mu_{i_r} - \mu) \end{pmatrix} \quad [A2]$$

$$V_{i_r} = \begin{pmatrix} \sigma_g^2 & \frac{h^2}{2}\sigma_g^2 \\ \frac{h^2}{2}\sigma_g^2 & \frac{h^4}{4}\sigma_p^2 \end{pmatrix} \quad [A3]$$

Thus the marginal density  $\phi_{i_r}(I)$  of  $I$  is the normal variate  $\mathcal{N}$  of expectation  $(\frac{h^2}{2}(C_{i_r} + \mu_{i_r} - \mu))$  and variance  $(\frac{h^4}{4}\sigma_p^2)$ , and the density  $\phi_{i_r}(\Gamma_r|I)$  of  $\Gamma_r$  given  $I$  is the normal  $\mathcal{N}(C_{i_r} + \mu_{i_r} + 2(I - \frac{h^2}{2}(C_{i_r} + \mu_{i_r} - \mu)), \sigma_g^2(1 - h^2))$ .

Following Ducrocq and Quaas (1988), from [A2] and [A3], the global selection pressure is given by:

$$q = \sum_{i_r} \alpha_{i_r} \left[ 1 - \Phi \left( \frac{T - \frac{h^2}{2}(C_{i_r} + \mu_{i_r} - \mu)}{\sqrt{\sigma_p^2 h^4 / 4}} \right) \right] = \sum_{i_r} \alpha_{i_r} q_{i_r} \quad [A4]$$

where  $\Phi$  is a cumulative normal function.

From [A2] and [A4], the deviation  $\Delta_{i_r}$  between selected and candidate reproducers of the  $i_r^{th}$  subclass is given by:

$$\Delta_{i_r} = \frac{1}{q_{i_r}} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{T - \frac{h^2}{2}(C_{i_r} + \mu_{i_r} - \mu)}{\sigma_p h^2 / 2} \right)^2 \right\} h^2 \sigma_p \quad [A5]$$

### With genotypic information

The selection index  $I$  must consider the possible mates with genotypes ( $G_m$ ) and ages ( $a_m$ )

$$\begin{aligned} I &= E(X_p | X_r, G_r) \\ &= \sum_{i_m} \sum_{j_m} \sum_{i_p} p(G_m = i_m, a_m = j_m) p(G_p = i_p | G_r, G_m) E(X_p | G_p, X_r, G_r, G_m, a_m) \end{aligned} \quad [A6]$$

The probability  $p(G_m = i_m, a_m = j_m)$  will be noted  $\beta_{i_m, j_m}$ .

The probability  $p(G_p = i_p | G_r, G_m)$  is given by the Mendelian rules.

The conditional expectation

$E(X_p|G_p, X_r, G_r, G_m, a_m)$  is given by  $C_{i_p} + \frac{\mu_{i_r} + \mu_{i_m j_m}}{2} + \frac{h^2}{2}(X_r - C_{i_r} - \mu_{i_r})$

It follows that:

$$I = \sum_{i_p} p(G_p = i_p | G_r = i_r) C_{i_p} + \frac{1}{2} \mu_{i_r} + \frac{1}{2} \sum_{j_m} \sum_{i_m} \beta_{i_m, j_m} \mu_{i_m j_m} + \frac{h^2}{2} (X_r - C_{i_r} - \mu_{i_r}) \quad [A7]$$

The first term ( $\sum_{i_p} p(G_p = i_p | G_r = i_r) C_{i_p}$ ), mean major locus value of future progeny of the sires with genotype  $G_r = i_r$  depends on the genotypic frequencies  $p(G_m)$  of their mates. It will be noted  $C_{i_p|i_r}$ .

The second term ( $\frac{1}{2} \mu_{i_r}$ ) is half the polygenic value of the sires with genotype  $G_r = i_r$ .

The third term ( $\sum_{j_m} \sum_{i_m} \beta_{i_m, j_m} \mu_{i_m j_m}$ ), the polygenic value of their mates, depends on their genotypic frequencies. It will be noted  $\mu_m$ .

The moments  $E_{i_r}$  and  $V_{i_r}$  are:

$$E_{i_r} = \begin{pmatrix} C_{i_r} + \mu_{i_r} \\ C_{i_p|i_r} + \frac{1}{2} \mu_{i_r} + \frac{1}{2} \mu_m \end{pmatrix} \quad [A8]$$

$$V_{i_r} = \begin{pmatrix} \sigma_g^2 & \frac{h^2}{2} \sigma_g^2 \\ \frac{h^2}{2} \sigma_g^2 & \frac{h^4}{4} \sigma_p^2 \end{pmatrix} \quad [A9]$$

The selection pressure  $q$  is given in this case by:

$$q = \sum_{i_r} \alpha_{i_r} \left[ 1 - \Phi \left( \frac{T - C_{i_p|i_r} - \frac{1}{2} \mu_{i_r} - \frac{1}{2} \mu_m}{\sqrt{\sigma_p^2 h^4 / 4}} \right) \right] = \sum_{i_r} \alpha_{i_r} q_{i_r} \quad [A10]$$

The deviation  $\Delta_{i_r}$  between selected and candidate reproducers of the  $i_r^{th}$  subclass is given by:

$$\Delta_{i_r} = \frac{1}{q_{i_r}} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{T - C_{i_p|i_r} - \frac{1}{2} \mu_{i_r} - \frac{1}{2} \mu_m}{\sqrt{\sigma_p h^2}} \right)^2 \right\} h^2 \sigma_p \quad [A11]$$

### Progeny test selection

#### Without genotypic information

The selection index  $I$  is

$$I = \tau \sum_{n=1}^N \frac{X_{o_n} - \mu}{N} \quad [\text{A12}]$$

with  $\mu$ , the general mean of offspring,  $N$  the number of tested offsprings and  $\tau = \frac{1/4Nh^2}{1 + 1/4(N-1)h^2}$ .

The equation [A1] becomes in this situation:

$$\begin{aligned} f(\Gamma_r, I) &= \sum_{i_r} \alpha_{i_r} \phi_{i_r}(\Gamma_r, I) \\ &= \sum_{i_r} \alpha_{i_r} \phi_{i_r}(\Gamma_r, I) \\ &= \sum_{i_r} \alpha_{i_r} \sum_{\gamma} p(\gamma) \phi_{i_r, \gamma}(\Gamma_r, I) \end{aligned} \quad [\text{A13}]$$

where  $\gamma$  is a combination giving for each offspring  $n$ , the genotypes and ages of the individual and its dam and  $p(\gamma)$  its probability.

The moments  $E_{i_r, \gamma}$  and  $V_{i_r, \gamma}$  are

$$E_{i_r, \gamma} = \left( \begin{array}{c} C_{i_r} + \mu_{i_r} \\ \frac{\tau}{N} \sum_n (C_{i_{o_n}} + \mu_{i_{o_n} j_{o_n}} - \mu) \end{array} \right) \quad [\text{A14}]$$

$$V_{i_r, \gamma} = \left( \begin{array}{cc} \sigma_g^2 & \frac{\tau}{2} \sigma_g^2 \\ \frac{\tau}{2} \sigma_g^2 & \frac{\tau}{4} \sigma_g^2 \end{array} \right) \quad [\text{A15}]$$

The density  $\phi_{i_r, \gamma}(I|\Gamma_r)$  of  $I$  given  $\Gamma_r$  is the normal variate  $\mathcal{N}(M_{i_r, \gamma}, \sigma_{i_r, \gamma}^2)$ , with

$$M_{i_r, \gamma} = \frac{\tau}{N} m_{i_r, \gamma} + \frac{\tau}{2} (\Gamma_r - C_{i_r} - \mu_{i_r}) \quad [\text{A16}]$$

with  $m_{i_r, \gamma} = \sum_n \frac{C_{i_{o_n}} + \mu_{i_{o_n} j_{o_n}} - \mu}{N}$ , and

$$\sigma_{i_r, \gamma}^2 = \frac{1}{4} \tau (1 - \tau) \sigma_g^2 \quad [\text{A17}]$$

It is assumed that  $\sum_{\gamma} p(\gamma) \phi_{i_r, \gamma}(I|\Gamma_r)$  is approximately normal  $\mathcal{N}(M_{i_r}, \sigma_{i_r}^2)$ , with

$$M_{i_r} = \tau \sum_{\gamma} p(\gamma) m_{i_r, \gamma} + \frac{\tau}{2} (\Gamma_r - C_{i_r} - \mu_{i_r}) \quad [\text{A18}]$$

and

$$\sigma_{i_r}^2 = \frac{1}{4}\tau(1-\tau)\sigma_g^2 + r_{i_r}^2 \quad [\text{A19}]$$

with  $r_{i_r}^2 = \sum_{\gamma} p(\gamma)M_{i_r\gamma}^2 - M_{i_r}^2$ .

Thus the marginal density  $\phi_{i_r}(I)$  of  $I$  is the normal variate

$$\mathcal{N}\left(\tau \sum_{\gamma} p(\gamma)m_{i_r\gamma}, \frac{\tau}{4}\sigma_g^2 + r_{i_r}^2\right)$$

and the density  $\phi_{i_r}(\Gamma_r|I)$  of  $\Gamma_r$  given  $I$  is the normal

$$\mathcal{N}\left(C_{i_r} + \mu_{i_r} + \frac{\tau\sigma_g^2/2}{1/4\tau\sigma_g^2 + r_{i_r}^2}(I - M_{i_r}), \sigma_g^2 - \frac{\tau^2\sigma_g^4/4}{1/4\tau\sigma_g^2 + r_{i_r}^2}\right)$$

This gives:

$$q = \sum_{i_r} \alpha_{i_r} \left[ 1 - \Phi\left(\frac{T - \tau m_{i_r}}{\sqrt{1/4\tau\sigma_g^2 + r_{i_r}^2}}\right) \right] = \sum_{i_r} \alpha_{i_r} q_{i_r} \quad [\text{A20}]$$

With  $\rho_{i_r} = \frac{\tau\sigma_g^2/2}{1/4\tau\sigma_g^2 + r_{i_r}^2}$ , the deviation  $\Delta_{i_r}$  between selected and candidate reproducers of the  $i_r^{\text{th}}$  subclass is given by:

$$\Delta_{i_r} = \frac{1}{q_{i_r}} \left[ \rho_{i_r} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{T - \tau m_{i_r}}{\sqrt{1/4\sigma_g^2\tau + r_{i_r}^2}}\right)^2\right\} \sqrt{1/4\sigma_g^2\tau + r_{i_r}^2} \right] \quad [\text{A21}]$$

### With genotypic information

The selection index  $I$  is defined similarly to the individual selection index with genotype informations:

$$\begin{aligned} I &= E(X_p|\mathbf{X}_o, G_r, \mathbf{G}_o) \\ &= \sum_{i_m} \sum_{j_m} \sum_{i_p} p(G_m = i_m, a_m = j_m) p(G_p = i_p|G_r, G_m) E(X_p|\mathbf{X}_o, G_r, \mathbf{G}_o, G_m, a_m) \end{aligned} \quad [\text{A22}]$$

The conditional expectation  $E(X_p|\mathbf{X}_o, G_r, \mathbf{G}_o, G_m, a_m)$  is given by:

$$C_{i_p} + \frac{\mu_{i_r} + \mu_{i_m j_m}}{2} + \tau \sum_{n=1}^N \frac{X_{o_n} - C_{i_{o_n}} - \mu_{i_{o_n}}}{N}$$

It follows that:

$$I = C_{i_p|i_r} + 1/2\mu_{i_r} + \mu_m + \tau \sum_{n=1}^N \frac{X_{o_n} - C_{i_{on}} - \mu_{i_{on}}}{N} \quad [\text{A23}]$$

As in the previous case,

$$f(\Gamma_r, I) = \sum_{i_r} \alpha_{i_r} \sum_{\gamma} p(\gamma) \phi_{i_r\gamma}(\Gamma_r, I) \quad [\text{A24}]$$

The moments  $E_{i_r\gamma}$  and  $V_{i_r\gamma}$  are in this case:

$$E_{i_r\gamma} = \begin{pmatrix} C_{i_r} + \mu_{i_r} \\ C_{i_p|i_r} + \frac{1}{2}\mu_{i_r} + \frac{1}{2}\mu_m \end{pmatrix} \quad [\text{A25}]$$

$$V_{i_r\gamma} = \begin{pmatrix} \sigma_g^2 & \frac{\tau}{2}\sigma_g^2 \\ \frac{\tau}{2}\sigma_g^2 & \frac{\tau}{4}\sigma_g^2 \end{pmatrix} \quad [\text{A26}]$$

Contrary to the previous situation,  $E_{i_r\gamma}$  does not depend on  $\gamma$ . Thus  $f(\Gamma_r, I) = \sum_{i_r} \alpha_{i_r} \phi_{i_r}(\Gamma_r, I)$  is normally distributed (without approximation).

The selection pressure  $q$ , and deviation  $\Delta$  are given in this case by:

$$q = \sum_{i_r} \alpha_{i_r} \left[ 1 - \Phi \left( \frac{T - (C_{i_p|i_r} + \frac{1}{2}\mu_{i_r} + \frac{1}{2}\mu_m)}{\sqrt{\tau\sigma_g^2/4}} \right) \right] \quad [\text{A27}]$$

and

$$\Delta_{i_r} = \frac{1}{q_{i_r}} \left[ \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{T - (C_{i_p|i_r} + \frac{1}{2}\mu_{i_r} + \frac{1}{2}\mu_m)}{\sqrt{1/4\tau\sigma_g^2}} \right)^2 \right\} \sqrt{1/4\tau\sigma_g^2} \right] \quad [\text{A28}]$$