

# Estimation of the average effects of specific alleles detected by the pseudo-testcross QTL mapping strategy

C Plomion, CÉ Durel

*Laboratoire de génétique et amélioration des arbres forestiers,  
Institut national de la recherche agronomique, BP45, 33610 Cestas, France*

(Received 30 August 1995; accepted 21 March 1996)

**Summary** – In a full-sib progeny of a mating between heterozygous parents, the analysis of marker-trait associations using dominant RAPD (random amplified polymorphic DNA) markers allows the detection of specific quantitative trait loci (QTL) for each parent of the cross. Here we develop a method aimed at estimating the breeding value of such QTL alleles in the population, for the incorporation of marker-assisted selection in operational forest tree-breeding programs. The proposed methodology expands on the pseudo-testcross QTL mapping strategy which is based on the selection of single dose markers present in one parent and absent in the other. It specifically exploits the fact that one of the parents of the full-sib family is double null for the RAPD markers bracketing the QTL so that, by looking at its half-sib family, ‘band present’ allele frequencies of the two markers can be obtained at the population level. The half-sib family of the other parent, which is doubly heterozygous, is then used to estimate the average effect (ie, the additive effect) of the two QTL alleles.

**QTL / breeding value / full-sib / half-sib / pseudo-testcross**

**Résumé** – Détermination des effets moyens des allèles à un QTL spécifique détecté par la stratégie du pseudo-testcross. Dans une famille de pleins-frères issue du croisement entre deux individus hétérozygotes, l’analyse des associations entre des marqueurs moléculaires dominants (RAPD, polymorphisme de l’ADN amplifié au hasard) et des caractères quantitatifs permet de détecter des locus impliqués dans l’expression des variations phénotypiques de caractères quantitatifs (QTL). Ces QTL sont spécifiques de chacun des parents du croisement. Nous développons ici un modèle permettant de déterminer la valeur générale des allèles au QTL au sein de la population, en vue d’une utilisation dans les programmes de sélection chez les arbres forestiers. La méthode proposée prolonge la localisation de QTL spécifiques par la stratégie du pseudo-testcross, basée sur la sélection de marqueurs à simple dose présents chez un parent et absents chez l’autre. Elle exploite le fait qu’un des deux parents de la famille de plein-frères est homozygote nul pour les deux marqueurs RAPD bordant le QTL. Ainsi, en observant l’une de ses descendance de demi-frères, il est possible d’estimer les fréquences dans la population des allèles codant pour la présence de bande RAPD. Une descendance de demi-frères issue de l’autre parent

*double hétérozygote aux marqueurs bordant le QTL est alors utilisée pour déterminer les effets moyens (additifs) des allèles au QTL.*

**QTL / valeur additive / plein-frère / demi-frère / pseudo-testcross**

## INTRODUCTION

The use of molecular markers as a complementary tool for breeding is based on linkage disequilibria between molecular markers and QTLs (quantitative trait loci) involved in the control of quantitative traits. Marker-assisted selection (MAS) in crop plants has been investigated in essentially two directions: (i) for genotype construction (eg, Young and Tanksley, 1989), and (ii) for predicting the breeding value of an individual progenitor (Lande and Thompson, 1990).

From population genetic studies, it is known that wild allogamous species such as forest trees are often in linkage equilibrium. It is likely, therefore, that the alleles at QTLs and the alleles at marker loci are randomly associated in different individuals (Avery and Hill, 1979; Muona, 1982; Beckman and Soller, 1983; Beckman and Soller, 1986; Hasting 1989; Lande and Thompson, 1990). Consequently, it would be difficult to establish significant marker-QTL associations at the population level in large mating populations. This absence of linkage disequilibria has been raised against the feasibility of MAS in forest trees (Strauss et al, 1992). With the considerable decrease in cost and increased automation of the RAPD (random amplified polymorphic DNA) technique (Williams et al, 1990), however, it is now possible to construct a single-tree map for every individual in an elite breeding population: an extreme alternative to deal the linkage equilibrium that was presented and discussed by Grattapaglia and Sederoff (1994). Recently, Grattapaglia et al (1995) extended their approach to QTL mapping within a full-sib family. QTLs are then defined in a narrow genetic background and could be used to identify the best performers for vegetative propagation. However, there is no evidence that such specific QTL will be significant in broader genetic backgrounds. Experimental results in crop plants have demonstrated the inconsistency of QTL expression across populations (Tanksley and Hewitt, 1988; Graef et al, 1989; Beavis et al, 1991). Conversely, QTLs defined against a wide genetic background could be more useful in breeding than QTLs defined in a specific background.

The aim of this paper is to show that the general value of a 'specific' QTL detected in a full-sib family following the method of Grattapaglia et al (1995) can be easily evaluated, provided that both parents of the full-sib family are involved in maternal half-sib (open-pollinated or polycross) families. Such two-generation pedigrees are widely available in most forest tree-breeding programs that involve the simultaneous estimation of specific and general combining abilities of selected trees. We actually expect that some QTLs controlling economically important traits exist at the population level and can be detected in a broad genetic background. The proposed strategy assumes that the QTL mapping has already been performed in a single full-sib family and therefore concentrates on the estimation of the effects of the QTL alleles.

## MODELS AND METHODS

### *Two-way pseudo-testcross mapping strategy using dominant RAPD markers*

This strategy essentially exploits the high levels of heterozygosity of outbred individuals and the efficiency of the RAPD assay in uncovering large numbers of genetic markers in an informative configuration (Grattapaglia and Sederoff, 1994). Single-dose RAPD markers are screened in such a way that they are in a heterozygous state in one parent and absent in the other, or vice versa, and therefore segregate 1:1 ratio in the F1 progeny following a testcross configuration (Carlson et al, 1991; Cai et al, 1994; Echt et al, 1994; Weeden et al, 1994; Kubisiak et al, 1995). Two separate sets of linkage data are therefore obtained and a specific genetic map is constructed for each parent. A quantitative trait dissection analysis is carried out independently for both parents of the cross under the conventional backcross model and leads to the detection of individual 'specific' QTLs. In the QTL analysis, the null hypothesis tests an allelic substitution averaged over the alleles inherited from the other parent (Leonard-Schippers et al, 1994; Grattapaglia et al, 1995).

### *Genetic model*

The model described in this paper is based on a full-sib (FS) progeny between two heterozygous parents  $P_X$  and  $P_Y$ , and on half-sib (HS) progenies of these parents. We suppose that the linkage analysis has already been carried out in the FS family and that a QTL has been detected in  $P_X$ . The same analysis could be applied for a QTL detected in  $P_Y$ . In that case, the analysis would be performed for markers heterozygous in  $P_Y$  and homozygous null in  $P_X$ .

Let  $Q$  denote a quantitative trait locus lying between linked dominant molecular marker loci  $A$  and  $B$ . Let  $r_1$  and  $r_2$  denote recombination frequencies between  $A$  and  $Q$  and  $B$  and  $Q$ , respectively. As the QTL detection has already been performed, the values of  $r_1$  and  $r_2$  are known. Besides the recombination rates, the haplotypes of  $P_X$  are also known from the mapping experiment. In the pseudo-testcross configuration, the genotypes are:  $A_1A_2Q_1Q_2B_1B_2$  for  $P_X$  and  $A_2A_2Q_3Q_4B_2B_2$  for  $P_Y$ , where  $A_1$  and  $B_1$  are the alleles coding for the presence of a particular RAPD fragments,  $A_2$  and  $B_2$  the alleles coding for the absence of the same fragments,  $Q_1Q_2$  the QTL genotype of  $P_X$  and  $Q_3Q_4$  the QTL genotype of  $P_Y$ . Since QTL alleles are unknown,  $Q_3$  and  $Q_4$  may be identical to  $Q_1$  and  $Q_2$ . For  $P_X$  the expected frequencies of the gametes  $A_1Q_1B_1$ ,  $A_2Q_2B_2$ ,  $A_1Q_1B_2$ ,  $A_2Q_2B_1$ ,  $A_1Q_2B_2$ ,  $A_2Q_1B_1$ ,  $A_1Q_2B_1$  and  $A_2Q_1B_2$  are:  $(1 - r_1)(1 - r_2)/2$ ,  $(1 - r_1)(1 - r_2)/2$ ,  $(1 - r_1)r_2/2$ ,  $(1 - r_1)r_2/2$ ,  $r_1(1 - r_2)/2$ ,  $r_1(1 - r_2)/2$ ,  $r_1r_2/2$  and  $r_1r_2/2$ , respectively.

Let  $Q_p$  denote any quantitative trait allele (QTA), in the population at the studied locus, and  $f_{A_1}$ ,  $f_{B_1}$  and  $f_{Q_p}$  the allele frequencies of  $A_1$ ,  $B_1$  and  $Q_p$  in the pollen pool. For a two-allele marker model, frequencies of the alternative alleles  $A_2$  and  $B_2$  are  $(1 - f_{A_1})$  and  $(1 - f_{B_1})$ , respectively. Assuming linkage equilibrium in the population, the expected frequencies for gametes  $A_1Q_pB_1$ ,  $A_2Q_pB_2$ ,  $A_1Q_pB_2$ , and  $A_2Q_pB_1$  in the pollen are:  $f_{A_1}f_{B_1}f_{Q_p}$ ,  $(1 - f_{A_1})(1 - f_{B_1})f_{Q_p}$ ,  $f_{A_1}(1 - f_{B_1})f_{Q_p}$ , and  $(1 - f_{A_1})f_{B_1}f_{Q_p}$ , respectively.

Let  $\mu_{qp}$  denote the trait mean of QTL genotype  $Q_qQ_p$  ( $q = 1, 2$ ) and  $\Theta_r$  the expected value of the mean of marker phenotype  $r$  weighted by the frequency of the different combinations of marker-QTL genotypes included in  $r$ . According to the Fisher model (Kempthorne, 1957), the trait mean of QTL genotype  $Q_kQ_l$  is  $\mu_{kl} = a_k + a_l + d_{kl}$ , where  $a_k = \sum_{p=1}^n f_{Qp} \mu_{kp}$  and  $a_l = \sum_{p=1}^n f_{Qp} \mu_{lp}$  are the average effects of QTA  $k$  and  $l$ ,  $d_{kl}$  denotes the dominance effect and  $n$  is the number of QTA in the population.

The expected values are derived using double-crossover gamete frequencies. Coupling-linkage phase of the marker presence alleles of the mother tree is assumed, but the same analysis could be carried out for the repulsion phase.

## RESULTS

### *Determination of the marker allelic frequencies $f_{A1}$ and $f_{B1}$ in the pollen pool*

The HS progeny produced with  $P_Y$  as the female parent was used to determine the allele frequencies of markers A and B in the population (pollen pool of the polycross). Since  $P_Y$  was double null for the bracketing RAPD markers, the maximum likelihood estimates for  $f_{A1}$  and  $f_{B1}$  are simply the observed frequencies ie,  $f_{A1} = n_{A1}/N$  and  $f_{B1} = n_{B1}/N$ , with  $n_{A1}$  and  $n_{B1}$  the number of HS individuals carrying the 'RAPD band present' alleles for markers A and B, respectively.

### *Evaluation of the average effect of $Q_1$ and $Q_2$*

The HS progeny produced with  $P_X$  as the female parent was used to determine the average effect of  $Q_1$  and  $Q_2$  mapped in  $P_X$ .  $P_X$  is heterozygous for the two bracketing RAPD markers, which makes it possible to carry out genotype discrimination within the limits of the fact that RAPDs are dominant. The different marker-QTL genotypes are presented in table I. They belong to only four marker phenotype classes:  $[A_1B_1]$ ,  $[A_1B_2]$ ,  $[A_2B_1]$ , and  $[A_2B_2]$ . Each marker phenotype class contains the two QTAs in variable frequencies, according to the gamete frequencies in the pollen pool and the maternal gamete frequencies (table I). Therefore, the frequency of each QTA within each marker phenotype class and the expected values of marker phenotype means  $\Theta_r$  could be calculated as described in table II.

We assume normal distributions of the genotypic values of  $Q_1Q_p$  and  $Q_2Q_p$  with means  $a_1$  and  $a_2$  respectively, and variance  $\sigma^2$ . Thus, the observed values of traits  $Y_i$ ,  $Y_i/Q_1$  and  $Y_i/Q_2$  are normally distributed with means  $a_1$  and  $a_2$ , respectively, and variance  $\sigma^2$ .

Maximum likelihood estimates of the three parameters:  $a_1 = \sum_{p=1}^n f_{Qp} \mu_{1p}$ ,  $a_2 = \sum_{p=1}^n f_{Qp} \mu_{2p}$ , and  $\sigma^2$  can be derived from the four equations related to the different marker locus phenotypes (table II). Using flanking markers, the likelihood of

**Table I.** Marker-QTL genotypes from a mating between female parent  $P_X$  and a random male parent; the cell numbers ([1] to [32]) are used in table II.

Female parent genotype	Random male parent genotype			
	$A_1Q_pB_1$ ( $f_{A1}f_{Qp}f_{B1}$ )	$A_1Q_pB_2$ ( $f_{A1}f_{Qp}(1-f_{B1})$ )	$A_2Q_pB_1$ ( $(1-f_{A1})f_{Qp}f_{B1}$ )	$A_2Q_pB_2$ ( $((1-f_{A1})f_{Qp}(1-f_{B1}))$ )
$A_1Q_1B_1$	$A_1A_1Q_1Q_pB_1B_1$ <sup>[1]</sup>	$A_1A_1Q_1Q_pB_1B_2$ <sup>[2]</sup>	$A_1A_2Q_1Q_pB_1B_1$ <sup>[3]</sup>	$A_1A_2Q_1Q_pB_1B_2$ <sup>[4]</sup>
$A_1Q_1B_2$	$A_1A_1Q_1Q_pB_2B_1$ <sup>[5]</sup>	$A_1A_1Q_1Q_pB_2B_2$ <sup>[6]</sup>	$A_1A_2Q_1Q_pB_2B_1$ <sup>[7]</sup>	$A_1A_2Q_1Q_pB_2B_2$ <sup>[8]</sup>
$A_1Q_2B_2$	$A_1A_1Q_2Q_pB_2B_1$ <sup>[9]</sup>	$A_1A_1Q_2Q_pB_2B_2$ <sup>[10]</sup>	$A_1A_2Q_2Q_pB_2B_1$ <sup>[11]</sup>	$A_1A_2Q_2Q_pB_2B_2$ <sup>[12]</sup>
$A_2Q_2B_2$	$A_2A_1Q_2Q_pB_2B_1$ <sup>[13]</sup>	$A_2A_1Q_2Q_pB_2B_2$ <sup>[14]</sup>	$A_2A_2Q_2Q_pB_2B_1$ <sup>[15]</sup>	$A_2A_2Q_2Q_pB_2B_2$ <sup>[16]</sup>
$A_2Q_2B_1$	$A_2A_1Q_2Q_pB_1B_1$ <sup>[17]</sup>	$A_2A_1Q_2Q_pB_1B_2$ <sup>[18]</sup>	$A_2A_2Q_2Q_pB_1B_1$ <sup>[19]</sup>	$A_2A_2Q_2Q_pB_1B_2$ <sup>[20]</sup>
$A_2Q_1B_1$	$A_2A_1Q_2Q_pB_1B_1$ <sup>[21]</sup>	$A_2A_1Q_2Q_pB_1B_2$ <sup>[22]</sup>	$A_2A_2Q_1Q_pB_1B_1$ <sup>[23]</sup>	$A_2A_2Q_1Q_pB_1B_2$ <sup>[24]</sup>
$A_1Q_2B_1$	$A_1A_1Q_2Q_pB_1B_1$ <sup>[25]</sup>	$A_1A_1Q_2Q_pB_1B_2$ <sup>[26]</sup>	$A_1A_2Q_2Q_pB_1B_1$ <sup>[27]</sup>	$A_1A_2Q_2Q_pB_1B_2$ <sup>[28]</sup>
$A_2Q_1B_2$	$A_2A_1Q_2Q_pB_2B_1$ <sup>[29]</sup>	$A_2A_1Q_1Q_pB_2B_2$ <sup>[30]</sup>	$A_2A_2Q_1Q_pB_2B_1$ <sup>[31]</sup>	$A_2A_2Q_1Q_pB_2B_2$ <sup>[32]</sup>

**Table II.** Marker locus genotypes and phenotypes, and expected values of marker phenotype means weighted with the frequency in which the different combinations of marker-QTL genotypes occur; cell numbers refer to table I.

Marker genotype	Marker phenotype (see table I)	Cell number	Expected value of marker-QTL phenotype mean	Incidence vector $\mathbf{X}$ of the linear model (3)
$A_1A_1B_1B_1$	$[A_1B_1]$	[1], [25]	$\Theta_1 = E(Y_i/[A_1B_1]) = [A/(A+B)]a_1 + [B/(A+B)]a_2$ , with	$\mathbf{X} \begin{pmatrix} X_{11} \dots X_{1n_1} = (A-B)/(A+B) \\ X_{21} \dots X_{2n_2} = (C-D)/(C+D) \\ X_{31} \dots X_{3n_3} = (E-F)/(E+F) \\ X_{41} \dots X_{4n_4} = (G-H)/(G+H) \end{pmatrix}$
$A_1A_1B_1B_2$	$[A_1B_1]$	[2], [5], [9], [26]	$A = \{(1-\tau_1)(1-r_2) + (1-\tau_1)r_2f_{B1} + \tau_1(1-r_2)f_{A1} + \tau_1r_2f_{A1}f_{B1}\}/2$	
$A_1A_2B_1B_1$	$[A_1B_1]$	[3], [17], [21], [27]	$B = \{\tau_1r_2 + (1-\tau_1)r_2f_{A1} + \tau_1(1-r_2)f_{B1} + (1-\tau_1)(1-r_2)f_{A1}f_{B1}\}/2$	
$A_1A_2B_1B_2$	$[A_1B_1]$	[4], [7], [11], [13], [18], [22], [28], [29]		
$A_1A_2B_2B_2$	$[A_2B_2]$	[8], [12], [14], [30]	$\Theta_2 = E(Y_i/[A_2B_2]) = [C/(C+D)]a_1 + [D/(C+D)]a_2$ , with	$\mathbf{X} \begin{pmatrix} X_{11} \dots X_{1n_1} = (A-B)/(A+B) \\ X_{21} \dots X_{2n_2} = (C-D)/(C+D) \\ X_{31} \dots X_{3n_3} = (E-F)/(E+F) \\ X_{41} \dots X_{4n_4} = (G-H)/(G+H) \end{pmatrix}$
$A_1A_1B_2B_2$	$[A_2B_2]$	[6], [10]	$C = \{(1-\tau_1)r_2(1-f_{B1}) + \tau_1r_2f_{A1}(1-f_{B1})\}/2$	
$A_2A_2B_1B_2$	$[A_2B_2]$	[15], [20], [24], [31]	$D = \{\tau_1(1-r_2)(1-f_{B1}) + (1-\tau_1)(1-r_2)f_{A1}(1-f_{B1})\}/2$	
$A_2A_2B_1B_1$	$[A_2B_2]$	[19], [23]	$\Theta_3 = E(Y_i/[A_2B_2]) = [E/(E+F)]a_1 + [F/(E+F)]a_2$ , with	
$A_2A_2B_1B_2$	$[A_2B_2]$	[16], [32]	$E = \{\tau_1(1-r_2)(1-f_{A1}) + \tau_1r_2(1-f_{A1})f_{B1}\}/2$	$\mathbf{X} \begin{pmatrix} X_{11} \dots X_{1n_1} = (A-B)/(A+B) \\ X_{21} \dots X_{2n_2} = (C-D)/(C+D) \\ X_{31} \dots X_{3n_3} = (E-F)/(E+F) \\ X_{41} \dots X_{4n_4} = (G-H)/(G+H) \end{pmatrix}$
$A_1A_2B_1B_1$	$[A_2B_2]$	[16], [32]	$F = \{(1-\tau_1)r_2(1-f_{A1}) + (1-\tau_1)(1-r_2)(1-f_{A1})f_{B1}\}/2$	
$A_2A_2B_1B_1$	$[A_2B_2]$	[16], [32]	$\Theta_4 = E(Y_i/[A_2B_2]) = [G/(G+H)]a_1 + [H/(G+H)]a_2$ , with	
$A_2A_2B_1B_2$	$[A_2B_2]$	[16], [32]	$G = \{\tau_1r_2(1-f_{A1})(1-f_{B1})\}/2$	
$A_2A_2B_1B_1$	$[A_2B_2]$	[16], [32]	$H = \{(1-\tau_1)(1-r_2)(1-f_{A1})(1-f_{B1})\}/2$	

In the incidence vector,  $n_1, n_2, n_3$  and  $n_4$  represent the total number of individuals within the marker phenotype classes  $[A_1B_1], [A_1B_2], [A_2B_1]$ , and  $[A_2B_2]$  respectively.

occurrence of the observed performance of all HS individuals could be expressed as follows:

$$L = \prod_{r=1}^4 \prod_{i=1}^{n_r} \phi_r(y_{ri})$$

where  $y_{ri}$  is the observed value for the quantitative trait of individual  $i$  belonging to marker phenotype classes  $r$  (ie,  $[A_1B_1]$ ,  $[A_1B_2]$ ,  $[A_2B_1]$ , or  $[A_2B_2]$ ), and  $\phi(y_{ri})$  is the probability density of  $y_{ri}$ . Each marker phenotype class is a mixture of two populations of QTL genotypes including with certainty  $Q_1$  or  $Q_2$  respectively. Therefore, the probability density functions of the four marker phenotype classes can be written as a linear combination of the two component normal distributions weighted according to their proportions, eg, for the marker phenotype class  $[A_1B_1]$ :

$$\phi_{[A_1B_1]}(y_i) = [A/(A+B)]\phi_{Q_1}(y_i) + [B/(A+B)]\phi_{Q_2}(y_i)$$

where  $\phi_{Q_1}(y_i)$  and  $\phi_{Q_2}(y_i)$  were the probability density functions of genotypes  $Q_pQ_1$  and  $Q_pQ_2$  ( $A$  and  $B$  are defined in table II).

There is no analytical solution of the problem, but maximum likelihood estimates of the four parameters could be found using the EM algorithm suggested by Lander and Botstein (1989) or the quasi-Newton algorithm used by Knott and Haley (1992). Alternatively, approximations of maximum likelihood estimators at the first order of probability (Rebaï et al, 1995) could be determined for both QTA effects,  $a_1$  and  $a_2$ , using the following model:

$$Y_{ri} = \mu + \sum_{r=1}^4 (g_r \Theta_r) + \varepsilon_{ri} \quad [1]$$

where  $Y_{ri}$  is the phenotypic value of individual  $i$  belonging to the marker phenotype  $r$ ;  $\mu$  is the general mean;  $g_r$  is an indicator variable indexing the  $r$ th ( $r = 1, \dots, 4$ ) marker phenotype class, taking values of 1 for individual  $i$  of the  $r$ th phenotypic class and 0 otherwise;  $\Theta_r = E(Y_i/MP_r) = E(Y_i/Q_1) P_r(Q_1/MP_r) + E(Y_i/Q_2) P_r(Q_2/MP_r)$ , with  $MP_r$  the marker phenotype  $r$  ( $[A_1B_1]$ ,  $[A_1B_2]$ ,  $[A_2B_1]$ ,  $[A_2B_2]$ );  $E(Y_i/Q_1)$  and  $E(Y_i/Q_2)$  are the additive effects of  $Q_1$  and  $Q_2$  respectively;  $P_r(Q_1/MP_r)$  and  $P_r(Q_2/MP_r)$  are the probabilities of genotype  $Q_1Q_p$  and  $Q_2Q_p$  respectively, given the marker phenotype  $r$  (see table II); and  $\varepsilon_{ri}$  is a random normal variable associated with each record with mean 0 and variance  $\sigma'^2$ . This variance is assumed to be identical for each class  $r$ .

This linear model [1] could be written as

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\alpha}a_1 + \boldsymbol{\beta}a_2 + \boldsymbol{\varepsilon} \quad [2]$$

where  $\mathbf{Y}$  is the column vector of the phenotypic values; and  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are the column vectors of respectively the  $a_1$  and  $a_2$  probability coefficients in  $\Theta$ . Because of the linear dependency between  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  ( $\boldsymbol{\alpha} = \mathbf{1} - \boldsymbol{\beta}$ ), equation [2] can be written as:

$$\mathbf{Y} = \boldsymbol{\mu} + (2\boldsymbol{\alpha} - \mathbf{1})a + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \mathbf{X}_a a + \boldsymbol{\varepsilon} \quad [3]$$

with the constraint  $a_1 + a_2 = 0$  and assuming  $a_1 = -a_2 = a$ . The least squares estimators for  $a$  and  $\text{var}(a)$  are:  $\hat{a} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , and  $\text{var}(\hat{a}) = \sigma'^2(\mathbf{X}'\mathbf{X})^{-1}$  with

$\mathbf{X}$  the incidence vector of the model ( $\mathbf{X} = 2\boldsymbol{\alpha} - \mathbf{1}$ ). An expression of  $\mathbf{X}$  is shown in table II.

## DISCUSSION

The construction of a single-tree maps was first investigated using the megagametophyte of conifers. The megagametophyte is a haploid tissue derived from the same megaspore that gives rise to the maternal gametes, and can be easily obtained from germinating seeds. This mapping strategy has been widely reported (Conkle, 1980; Tulsieram et al, 1992; Nelson et al, 1993, 1994; Binelli and Bucci, 1994; Plomion et al, 1995) and is being used to dissect quantitative traits in loblolly pine (O'Malley and McKeand, 1994). The half-sib analysis depends on the ability to uniquely identify the marker alleles derived from the common progenitor, and detects exclusively the effect of QTLs in a heterozygous state. The identification of the maternal genetic contribution can be easily obtained with the analysis of the megagametophyte because it carries exactly the same genetic constitution as the fertilized ovule that gives rise to the embryo scored for the quantitative trait (Grattapaglia et al, 1992). Assuming a random genetic contribution from the male parents in the pollen pool, each segregating QTL detected in the mother tree can be characterized as differences in the average allele effect (Falconer, 1989). QTL defined in this way will be of great interest, because they will be major components of breeding value. A major drawback of this half-sib QTL mapping strategy is, however, that it requires the development of specific QTL mapping populations and cannot be used to analyze QTL in existing plantations. The megagametophyte is, in fact, a temporary tissue that can only be collected from germinants. Another limitation is that the megagametophyte is a special feature of conifers, and is not biologically available in most tree species. In order to implement the marker information to tree improvement, Liu et al (1993) developed a general strategy for genetic mapping and QTL analysis, based on diploid HS progenies. Such two-generation pedigrees are widely available in forest tree-breeding programs and could be used for immediate application of marker-assisted selection. Their approach, however, might prove to be misleading because it is based on the strong hypothesis that informative dominant or codominant markers exist at a low frequency in the pollen pool. In addition, a large population size would be required in the initial QTL detection step (van der Beek et al, 1995). For instance, Weller et al (1990) estimated that in a half-sib experiment, in order to detect a QTL with an effect of  $0.3\sigma$ , 1 000–2 000 offspring will be required. The relatively small number of offspring per HS family generally available in existing forest tree progeny tests (of the order of 50–200 individuals) is insufficient for QTL detection when considering this strategy. Conversely, in an FS progeny, Strauss et al (1992) estimated that just 200 individuals would permit the detection of half of the additive genetic variance at  $\alpha = 0.01$  when the within-family trait heritability is 0.5 and five QTLs control the trait. Therefore, the two-way pseudo-testcross mapping strategy applied to FS families could be the method of choice for the application of MAS in conventional tree-breeding programs. Detecting QTLs, however, is expensive and laborious, and if one runs an experiment for the mapping of QTLs with FS progeny, one should determine whether the mapped

QTLs are also expressed in a wide range of genetic backgrounds. This could result in greater efficiency later on for use in marker-assisted selection.

We demonstrated that this can be done by studying the two maternal HS progenies involving both parents of an FS family. As we noticed before, in the pseudo-testcross strategy, the QTL analysis tests the difference between the average trait value of ( $Q_1Q_3 + Q_1Q_4$ ) versus ( $Q_2Q_3 + Q_2Q_4$ ) in parent  $P_X$ ; this is equivalent to the following contrast:  $a_1 - a_2 + 1/2(d_{13} + d_{14} - d_{23} - d_{24})$ . With the method we developed in this paper, the QTL analysis led to the measurement of a purely additive contrast ( $a_1 - a_2$ ).

The proposed approach relies on the previous identification of specific QTLs in a single FS progeny and therefore presents some limitations. As pointed out by Grattapaglia et al (1995), the only QTLs that can be detected in the FS are those that are heterozygous in the parents of the cross and where the differential effect between the alternative QTA is relatively large. In principle, the access of average breeding values of most QTL alleles of the population could be achieved by analyzing progeny tests involving several half-sib or full-sib families. In domestic animals, Weller et al (1990) have proposed a 'granddaughter' design where the marker genotype is determined on the sons of heterozygous sires and the quantitative trait value measured on the daughters of the sons. Such a design involves three generations and is of particular interest because many half-sib families of relatively small sizes can be obtained from a single sire. Van der Beek et al (1995) have also demonstrated the value of three-generation pedigrees for mapping QTL in out-bred populations. Conversely three-generation pedigrees are still rare in most forest tree-breeding programs, and two-generation pedigrees involving large HS families obtained by polymix cross or open pollination and connected FS families can be easily produced. QTL mapping strategies have also been developed for cases where several unrelated FS families with few individuals are available (Knott and Haley, 1992). When family sizes are large, it is possible to find marker QTL associations within a single pedigree without the need to accumulate data on individual markers across pedigrees. However, when family sizes are small an increase of QTL detection power can be obtained with a large number of FS families because the linkage phase can be accurately determined. At this point it must be pointed out that in farm animals, marker-assisted selection models have often been developed for multiallelic codominant molecular markers. Indeed, restriction fragment length polymorphisms (RFLP), simple sequence repeats (SSR) and minisatellites markers have been developed for linkage map construction (Hillel et al, 1990; Barendse et al, 1994; Bishop et al, 1994; Broad and Hill, 1994; Crawford et al, 1994; Ellegren et al, 1994; Rohrer et al, 1994; Archibald et al, 1995) and quantitative trait dissection analysis (Andersson et al, 1994; Georges et al, 1995) for the major animal species. For such types of marker loci, most sires will be heterozygous at most markers and for most offspring it will be possible to uniquely identify which parental allele is transmitted. The development of such powerful marker techniques is still limited in forestry. To our knowledge, only two species have been mapped with RFLPs: *Pinus taeda* (Devey et al, 1994) and *Populus trichocarpa*  $\times$  *P. deltoides* (Bradshaw et al, 1994). Conversely, dominant RAPD markers have been intensively used for the development of single tree maps and for QTL analysis using the two-way pseudo-testcross strategy. RAPDs provide a fast, efficient and reliable way to gather marker data.

Other advantages with this method are the requirement of only a small amount of DNA for PCR reactions, the rapidity with which polymorphisms can be screened and the potential automation of the technique. Since primers can be arbitrarily chosen, any individual can be mapped with the same set of primers. A limitation of the RAPD fragments, however, is their questionable locus specificity when assessments are made on different individuals. On the one hand, although the same set of primers can be used, this does not mean that the same fragments corresponding to the same loci will always be amplified. On the other hand, the same fragment size observed in two different genotypes does not necessarily correspond to sequence homology. Finally, their diallelic nature is the major limitation for the unequivocal determination of parental marker alleles transmitted to each offspring, unless one of the alleles is rare (Soller, 1978).

Another important point is that when large QTL effects are confirmed in HS progeny, it is likely that the favorable QTAs exist at a low frequency in the breeding population. Such rare and favorable alleles would be of great value to increase the average value of the breeding population through MAS. Conversely, QTAs detected in FS families may turn out to be unimportant at the population level. This could very well happen if other alleles at the same loci with larger average effects exist at relatively high frequencies in the pollen pool. Indeed, a QTL effect (eg,  $Q_1$  as favorable allele and  $Q_2$  as unfavorable allele) could be observed in a specific unfavorable background ( $Q_3$  and  $Q_4$  unfavorable QTL alleles in  $P_Y$ ), whereas this effect would be smaller in a wide genetic background (pollen pool) where QTA  $Q_p$  would be frequently favorable.

The proposed strategy assumes that a specific QTL has already been detected in an FS family and that only those segments containing QTLs of interest are tracked in the two HS progenies. Therefore, the additional genotyping costs required for the estimation of the average QTA-effects may be lower than in the mapping experiment. Indeed, the costs will mainly depend on the size of the HS family necessary for this estimation. In order to be certain that the estimated substitution effects will be representative for the whole population, the number of half-sib progeny will have to be large if the number of QTAs is also large. Evidence for multiple QTA has not yet been demonstrated in forest tree species, but it is likely to be true for outbred and highly heterozygous plants as reported by Van Eck et al (1994).

We have assumed that the QTL position is precisely known from the analysis conducted with the FS progeny. Alternatively, a more accurate location could be obtained by combining genotypic information at the flanking markers of both the FS and HS progenies. Methods developed by Luo and Kearsey (1989) or Knapp et al (1990) could be used to obtain such an estimate. While the precise location would be crucial for map-based cloning objectives, however, it should not be a major problem for marker-assisted selection. Indeed, the extreme markers bracketing an LOD 1.0 support interval could be used for ensuring successful selection for the favorable QTL allele.

We investigated the combined FS and HS strategy developed here, and illustrated the feasibility of the integration of RAPD markers in the maritime pine and eucalyptus tree-breeding programs (Plomion et al, 1996). Genetic gain and costs associated with the use of molecular markers in separate FS progenies were

evaluated and compared to other strategies that aim to obtain similar selection efficiency. MAS has proved to have great potential as a complementary tool in the breeding of elite populations of forest trees.

## Acknowledgments

We thank the anonymous referees and the editor for their helpful suggestions for improving the manuscript. We wish to thank A Rebaï and JM Elsen for their helpful discussions.

## REFERENCES

- Andersson L, Haley CS, Ellegren H, Knott SA, Johansson M, Andersson K, Andersson-Eklund L, Edfors-Lilja I, Fredholm M, Hansson I, Hakansson J, Lundström K (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* 263, 1771-1774
- Archibald AL, Haley CS, Brown JF, Couperwhite S et al (1995) The PiGMaP consortium linkage map of the pig (*Sus scrofa*). *Mamm Genome* 6, 157-175
- Avery PJ, Hill WG (1979) Distribution of linkage disequilibrium with selection and finite population size. *Genet Res Camb* 33, 29-48
- Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick BW, Ryan AM, Clayton D, Li L, Neibergs HL, Zhang N et al (1994) A genetic linkage map of the bovine genome. *Nat Genet* 6, 227-35
- Beavis WD, Grant D, Albertsen M, Fincher R (1991) Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. *Theor Appl Genet* 83, 141-145
- Beckman JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor Appl Genet* 67, 35-43
- Beckman JS, Soller M (1986) Restriction fragment length polymorphisms in genetic improvement of agricultural species. *Euphytica* 35, 111-124
- van der Beek S, van Arendonk JAM, Groen AF (1995) Power of two- and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. *Theor Appl Genet* 91, 1115-1124
- Binelli G, Bucci G (1994) A genetic linkage map of *Picea abies* Karts, based on RAPD markers, as a tool in population genetics. *Theor Appl Genet* 88, 283-288
- Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SL, Hawkins GA, Toldo SS, Fries R, Grosz MD, Yoo J et al (1994) A genetic linkage map for cattle. *Genetics* 136, 619-639
- Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. *Theor Appl Genet* 89, 167-178
- Broad TE, Hill DF (1994) Mapping the sheep genome: practice, progress and promise. *Br Vet J* 150, 237-252
- Cai Q, Guy CL, Moore GA (1994) Extension of the linkage map in *Citrus* using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. *Theor Appl Genet* 89, 606-614
- Carlson JE, Tulsieram LK, Glaubitz JC, Luk VWK, Kauffeld C, Rutledge R (1991) Segregation of random amplified DNA markers in F1 progeny of conifers. *Theor Appl Genet* 83, 194-200
- Conkle MT (1981) Isozyme variation and linkage in six conifer species. In: *Proc Symp Isozymes of North American Forest Trees and Forest Insects*, USDA For Serv Gen Tech Rep, PSW 48, 11-17

- Crawford AM, Montgomery GW, Pierson CA, Brown T, Dodds KG, Sunden SLF, Henry HM, Ede AJ, Swarbrick PA, Berryman T, Penty JM, Hill DF (1994) Sheep linkage mapping: nineteen linkage groups derived from the analysis of paternal half-sib families. *Genetics* 137, 573-579
- Devey ME, Fiddler TA, Liu BH, Knapp SJ, Neale BD (1994) An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. *Theor Appl Genet* 88, 273-278
- Echt CS, Kidwell KK, Knapp SJ, Osborn TC, McCoy TJ (1994) Linkage mapping in diploid alfalfa (*Medicago sativa*). *Genome* 37, 61-71
- van Eck HJ, Jacobs JME, Stam P, Ton J, Stiekema WJ, Jacobsen E (1994) Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. *Genetics* 137, 303-309
- Ellegren H, Chowdhary BP, Johansson M, Marklund L, Fredholm M, Gustavsson I, Andersson L (1994) A primary linkage map of the porcine genome reveals a low rate of genetic recombination. *Genetics* 137, 1089-1100
- Falconer DS (1989) *Introduction to Quantitative Genetics*, 3rd edition. Longman Scientific and Technical, Essex, UK
- Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X et al (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139, 907-920
- Graef GL, Fehr WR, Cianzo SR (1989) Relation of isozyme genotype to quantitative characters in soybean. *Crop Sci* 29, 683-688
- Grattapaglia D, Chaparro J, Wilcox P, McCord S, Werner D, Amerson H, McKeand, Bridgwater F, Whetten R, O'Malley D Sederoff R (1992) Mapping in woody plants with RAPD markers: Application to breeding and horticulture. In: *Proceedings of the Symposium: Applications of RAPD Technology to Plant Breeding*, Minneapolis, MN, Joint Plant Breeding Symposium: Crop Science, Horticultural Science, American Genetic Association, 37-40
- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *E urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Genetics* 137, 1121-1137
- Grattapaglia D, Bertolucci FL, Sederoff RR (1995) Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Theor Appl Genet* 90, 933-947
- Hastings A (1989) The interaction between selection and linkage in plant populations. In: *Plant Population Genetics, Breeding and Genetic Resources* (Brown, Clegg, Kahler, Weir, eds), Sinauer Associates Inc, Sunderland, MA, 163-180
- Hillel J, Schaap T, Haberfeld A, Jeffrey AJ, Plotzky Y, Cahaner A, Iavi U (1990) DNA fingerprints applied to gene introgression in breeding programs. *Genetics* 124, 783-789
- Kempthorne O (1957) *An Introduction to Genetic Statistics*. John Wiley, New York
- Knapp SJ, Bridges WC, Birkes D (1990) Mapping quantitative trait loci using molecular marker linkage maps. *Theor Appl Genet* 79, 583-592
- Knott SA, Haley CS (1992) Maximum-likelihood mapping of quantitative trait loci using full-sib families. *Genetics* 132, 1211-1222
- Kubisiak TL, Nelson CD, Nance WL, Stine M (1995) RAPD linkage mapping in a longleaf pine × slash pine F1 family. *Theor Appl Genet* 90, 1110-1127
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756
- Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185-199

- Leonards-Schippers C, Gieffers W, Schäfer-Pregl R, Ritter E, Knapp SJ, Salamini F, Gebhardt C (1994) Quantitative resistance to *Phytophthora infestans* in potato: a case study for QTL mapping in an allogamous plant species. *Genetics* 137, 67-77
- Liu B-H., Sederoff R, O'Malley D (1993) Linkage mapping using open-pollinated populations. In: *Proc 22nd Southern Forest Tree Improvement Conference*, Atlanta, GA, 14-17 June 1993, 489
- Luo ZW, Kearsey MJ (1989) Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. *Heredity* 63, 401-408
- Muona O (1982) Population structure of forest trees. *Silva Fennica* 16, 107-114
- Nelson CD, Nance WL, Doudrick RL (1993) A partial genetic linkage map of slash pine (*Pinus elliotti* Englem var *elliotti*) based on random amplified polymorphic DNA's. *Theor Appl Genet* 8, 145-151
- Nelson CD, Kubisiak TL, Stine M, Nance WL (1994) A genetic linkage map of longleaf pine based on random amplified polymorphic DNAs. *J Hered* 85, 433-439
- O'Malley DM, McKeand SE (1994) Marker assisted selection for breeding value in forest trees. *For Genet* 1, 231-242
- Plomion C, Bahrman N, Durel CÉ, O'Malley DM (1996) Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. *Heredity* 74, 661-668
- Plomion C, Durel CÉ, Verhaegen D (1996) Utilisation des marqueurs moléculaires dans les programmes d'amélioration génétique des arbres forestiers : exemple du pin maritime et de l'eucalyptus. *Ann Sci For* 53, 819-848
- Rebaï A, Goffinet B, Mangin B (1995) Comparing power of different methods for QTL detection. *Biometrics* 51, 87-99
- Rohrer GA, Leeson LA, Keele JW, Smith TP, Beattie CW (1994) A microsatellite map of the porcine genome. *Genetics* 136, 231-245
- Soller M (1978) The use of loci associated with quantitative traits in dairy cattle improvement. *Anim Prod* 27, 133-139
- Strauss SH, Lande R, Namkoong G (1992) Limitations of molecular-marker-aided selection in forest tree breeding. *Can J For Res* 22, 1050-1061
- Tanksley SD, Hewitt JD (1988) Use of molecular markers in breeding for soluble solids in tomato: a re-examination. *Theor Appl Genet* 75, 811-823
- Tulsieram LK, Glaubitz JC, Kiss G, Carlson JE (1992) Single tree genetic linkage analysis in conifers using haploid DNA from megagametophytes. *Biotechnology* 10, 686-690
- Weeden NF, Hemmat M, Lawson DM, Loghi M, Manganaris AG, Reish BI, Brown SK, Ye GN (1994) Development and application of molecular marker linkage maps in woody fruit crops. *Euphytica* 77, 71-75
- Weller JI, Kashi Y, Soller M (1990) Power of 'daughter' and 'granddaughter' designs for mapping of quantitative traits in dairy cattle using genetic markers. *J Dairy Sci* 73, 2525-2532
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18, 6531-6535
- Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the Tm-2 locus of tomato during backcross breeding. *Theor Appl Genet* 77, 353-359