DOI: 10.1534/genetics.104.028993

Quantitative Trait Loci (QTL) Detection in Multicross Inbred Designs: Recovering QTL Identical-by-Descent Status Information From Marker Data

Sébastien Crepieux,*[,1] Claude Lebreton,[†] Bertrand Servin[‡] and Gilles Charmet*

**UMR 1095 INRA-UBP, 63039 Clermont-Ferrand Cedex 2, France, [†]Limagrain Verneuil Holding, site d'ULICE, F-63204 Riom Cedex, France and [‡]INRA UMR de Génétique Végétale, INRA/UPS/INAPG, 91 190 Gif sur Yvette, France*

Manuscript received March 18, 2004
Accepted for publication August 10, 2004

ABSTRACT

Mapping quantitative trait loci in plants is usually conducted using a population derived from a cross between two inbred lines. The power of such QTL detection and the parameter estimates depend largely on the choice of the two parental lines. Thus, the QTL detected in such populations represent only a small part of the genetic architecture of the trait. In addition, the effects of only two alleles are characterized, which is of limited interest to the breeder, while common pedigree breeding material remains unexploited for QTL mapping. In this study, we extend QTL mapping methodology to a generalized framework, based on a two-step IBD variance component approach, applicable to any type of breeding population obtained from inbred parents. We then investigate with simulated data mimicking conventional breeding programs the influence of different estimates of the IBD values on the power of QTL detection. The proposed method would provide an alternative to the development of specifically designed recombinant populations, by utilizing the genetic variation actually managed by plant breeders. The use of these detected QTL in assisting breeding would thus be facilitated.

THE availability of molecular markers in the 1980s opened a new scope for quantitative genetics and breeding. It was anticipated that the manipulation of loci underlying quantitative traits (QTL) would be as easily feasible as with Mendelian factors. This, however, has generally not been the case, despite the large corpus of theoretical studies on marker-assisted selection (MAS; *e.g.*, Lande and Thompson 1990; Gimelfarb and Lande 1994, 1995; Hospital *et al.* 1997). The main reason is probably the cost of markers and the relatively low improvement in selection efficiency that leads MAS to be generally much more expensive than conventional breeding (Moreau *et al.* 2000). The other reason is that applied breeding programs and QTL research are often disconnected, *i.e.*, performed by different teams and using different plant material.

Generally, QTL analyses are carried out on a few progenies from crosses between a small number of distantly related lines, often including wild relatives. Such analyses mostly involve biparental progenies such as backcrosses (BC), doubled haploid lines (DH), F_2, or recombinant inbred lines (RILs). In the approaches based on this kind of plant material, the effect of an allele substitution at a candidate locus is tested. This is called the fixed-model approach (Xu and Atchley 1995) since it considers a fixed number of distinct alleles (most often two) at each putative QTL. Statistical methods for the QTL analysis of biparental populations underwent successive improvements through the advent of interval mapping (Lander and Botstein 1989) and its linearization (Haley and Knott 1992), composite interval mapping (Zeng 1993, 1994; Jansen 1993), and multiple-trait QTL mapping (Jiang and Zeng 1995; Korol *et al.* 1995).

In contrast, breeder's material is distinct from the biparental populations studied in many mapping experiments. Breeders generally handle many small-sized families derived from crosses between (often highly related) elite lines. The methods described above are poorly suited to such material. Moreover, there are many drawbacks for the breeder's use of the QTL found on biparental populations. First, when only two parents are considered, some markers and potential QTL are more likely to be monomorphic, even if parental lines are carefully selected for trait divergence. Since, by definition, QTL can be found only at polymorphic sites in the genome, the expected number of QTL detected with a biparental cross will be lower than that expected when analyzing several crosses at a time (assuming the total number of genotypes is not the limiting factor). The second drawback is that the QTL effect is estimated as a contrast between two alleles and in one genetic background only. In that context, the improvement of

[1]*Corresponding author:* Limagrain Verneuil Holding, site d'ULICE, av. G. Gershwin, BP173, F-63204 Riom Cedex, France.
E-mail: sebastien.crepieux@limagrain.com

a line by the introgression of a QTL allele into a completely new genetic background is rather unpredictable, because of possible epistatic interaction between QTL and genetic background. Finally, from an economic standpoint, the cost of creation of large single-cross progenies and specific trials for trait evaluation to perform QTL detection is quite high and often at the expense of other selection programs.

All these drawbacks reduce the breeder's interest in implementing such experimental designs when funding and work are constrained. Biparental crosses are usually preferred for more upstream studies, *e.g.*, genomics: the fine mapping of a QTL, which is a prerequisite for its positional cloning, is easier when fewer QTL are segregating. In contrast, the breeder's focus will be to characterize the effect of a wide range of alleles in his germplasm. Methods for simultaneous detection and manipulation of QTL in breeding programs would thus enhance the applicability of MAS.

Muranty (1996) suggested the use of progenies from several parents, to achieve a high probability of obtaining more than one allele at a putative QTL and also to have a more representative estimate of the variance accounted for by a QTL. Xu (1998) compared the QTL detection powers obtained with random-effect models and fixed effects and found similar values for individual family sizes as low as 25 individuals. However, in more unbalanced designs, the random-effect approach was presumed to be more suited as it can handle any arbitrary pedigree of individuals (Lynch and Walsh 1998; Xu 1998). Efficient methodologies for more fragmented populations in plants have been developed (for example, Xie *et al.* 1998; Yi and Xu 2001; Bink *et al.* 2002; Jansen *et al.* 2003), but their extension or implementation for any complex plant designs, implying a mixture of half-sib and full-sib families of different sizes, at any generation of selfing, is not straightforward. The identical-by-descent (IBD)-based variance component analysis is a powerful statistical method for QTL mapping in complex populations and can be used in pedigrees of arbitrary size and complexity (Almasy and Blangero 1998). These IBD-based variance component analyses are derived from the assumption that individuals of similar phenotype are more likely to share alleles that are identical by descent. The construction of IBD matrices for alleles at each position tested along the genome and the fitting of random-effect models (which assumes that QTL effects are normally distributed) offer an appropriate method to map QTL if the mapping population is large enough and if the progenies are connected in some way. In addition, these models do not need to assume a known, finite set of alleles at each putative QTL. Thus, they offer a less parameterized statistical environment in which to map QTL, because only the variances need to be estimated instead of every allele substitution effect. IBD-based variance component approaches mostly differ from one another in the computation of the IBD matrices (see George *et al.* 2000 for a review). The IBD status between two individuals can be precisely inferred if the relationship between the two individuals (*i.e.*, the pedigree) is known and if ancestors in the pedigree can be genotyped. Most of the existing methods to compute IBD probabilities [see, for instance, the software LOKI (Heath 1997), SOLAR (Almasy and Blangero 1998), and SIMWALK2 (Sobel *et al.* 2001)] were developed for human and animal genetics and are not directly applicable to the particularities of plants (inbreeding, self-pollination, controlled mating, selection . . .). Moreover, in such methods, if unknown relationships exist between parents of individuals of the mapping population, then parents are considered as founders. This statement yields systematically for nonsibs IBD likelihood of zero at any position on their genome, even if it is commonly assumed in plant breeding programs that most of the parents share common "unknown" ancestors (use of some "star" varieties; see Bernardo 1993, for example). Furthermore, in fragmented situations, *i.e.*, where there are many families of small sizes (especially when the genotyping takes place at a late stage in pedigree breeding, where we may easily end up with as few as one or two lines per cross), the IBD-likelihood matrix can be very sparse. Hence, much could be gained in exploring the actual between-family IBD likelihoods in cases where only little information is available.

In this article, we continued these developments and further assumed a nonzero IBD likelihood between nonsib lines. For that, we devised a method to take into account an estimate of the coefficients of coancestries between parents to build the IBD matrices and present a unified IBD-based variance component analysis framework, to map QTL in any kind of multicross designs involving self-pollinating species, at any generation.

To test the accuracy of the method, we considered the most general case of pedigree breeding programs, where many different two-parent crosses are performed, each yielding very small progenies in advanced generations of selfing. We then investigated the influence of different methods of IBD computation on the power and accuracy to detect QTL on different complex populations similar to those used in breeding.

METHODS

Plant breeding material—multicross inbred designs: In this article, we consider a mapping population composed of several subpopulations of small size. Each of these subpopulations is composed of as few as one offspring coming from a single cross between two inbred parents. For example, these subpopulations could be produced by several consecutive selfings (*e.g.*, RILs) or backcrossings. We use the terms "parent," "half-sib," and "full-sib" in a broadened sense. By parents, we mean the two inbred lines that are crossed with each other to start a new breeding cycle. By full-sib, we mean indi-

viduals derived from the same initial cross (*i.e.*, involving the same two parents), after any number of selfing and/or backcrossing generations. By half-sib, we mean individuals sharing one parent in common, after any number of selfing and/or backcrossing generations. Any individuals that do not share any parent in common are termed "unrelated." The definitions are more relevant to plants, since our phenotyped progenies may commonly be as far as six or seven generations from their parents. Nevertheless, in a general case, the genome of the individuals of the mapping population could be fixed (*i.e.*, lines), fully heterozygous (*i.e.*, F_1), or a mixing of fixed and heterozygous parts (*i.e.*, issued from successive backcrossing or selfing generations).

Mixed linear models: We assume that the quantitative trait value is a linear combination of fixed design effects, putative QTL (with additive or/and dominance effects), and additive polygenic effects. The polygenic effect is seen as the cumulative effect of all loci affecting the quantitative trait that are unlinked to the QTL. The model without dominance effect is

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{Zv} + \mathbf{e}, \quad (1)$$

where $\mathbf{y}$ is an $(m \times 1)$ vector of phenotypes, $\mathbf{X}$ is an $(m \times s)$ design matrix, $\boldsymbol{\beta}$ is a $(s \times 1)$ vector of fixed effects, $\mathbf{Z}$ is an $(m \times q)$ incidence matrix relating records to individuals, $\mathbf{u}$ is a $(q \times 1)$ vector of additive QTL effects, $\mathbf{v}$ is a $(q \times 1)$ vector of additive polygenic effects, and $\mathbf{e}$ is the residual vector. We assume the random effects $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{e}$ as uncorrelated and distributed as multivariate normal densities,

$$\mathbf{u} \sim (0, \mathbf{G}\sigma^2_{\mathbf{u}}), \quad \mathbf{v} \sim (0, \mathbf{A}\sigma^2_{\mathbf{v}}), \quad \mathbf{e} \sim (0, \mathbf{I}\sigma^2_{\mathbf{e}}),$$

with $\sigma^2_{\mathbf{u}}$, $\sigma^2_{\mathbf{v}}$, and $\sigma^2_{\mathbf{e}}$ being, respectively, the additive variance of the QTL, the polygenic variance, and the residual variance. $\mathbf{A}$ is the $(q \times q)$ additive genetic relationship matrix; $\mathbf{G}$ is the $(q \times q)$ (co)variance matrix for the QTL additive effects conditional on marker information; and $\mathbf{I}$ is the identity matrix.

The model without QTL segregating in the population is, with the same notations,

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zv} + \mathbf{e}. \quad (2)$$

Computation and implementation of G and A matrices: To solve the mixed-linear model, we need to know $\mathbf{A}$ and $\mathbf{G}$ matrices ($\mathbf{y}$, $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{I}$ are known) to estimate $\sigma^2_{\mathbf{u}}$, $\sigma^2_{\mathbf{v}}$, and $\sigma^2_{\mathbf{e}}$.

With the above definitions of the material, if we consider a pair of individuals from the mapping population, they may be (i) taken from the same subpopulation, in which case they are full-sibs, or (ii) taken from two different subpopulations. In this last case, if one of the parents is common to the two subpopulations, the two individuals will be half-sibs; if the parents of the two subpopulations are distinct, the two individuals are considered as unrelated.

First, we draw relevant calculations of the IBD values for each of these cases (Figure 1).

Computation of G matrices with parents considered as founders: *Exact IBD value between two individuals at a QTL:* Within each subpopulation, only two alleles are segregating at each locus, giving only three possible genotypes at the QTL, for example, Q_1Q_1, Q_1Q_2, and Q_2Q_2.

Suppose that one of the subpopulations is composed of two individuals (i and j) that are thus full-sibs. The IBD value between two full-sibs i and j at a QTL is measured as

$$\pi_{ij} = 2\theta_{ij} = \begin{cases} 2 & \text{for } Q_1Q_1 - Q_1Q_1 \text{ or } Q_2Q_2 - Q_2Q_2 \\ 1 & \text{for } Q_1Q_1 - Q_1Q_2,\ Q_2Q_2, -\ Q_1Q_2, \text{ or } Q_1Q_2 - Q_1Q_2 \\ 0 & \text{for } Q_1Q_1 - Q_2Q_2, \end{cases}$$

π_{ij} being the IBD value between individuals i and j, at a putative QTL (π_{ij} represent also the ijth elements of $\mathbf{G}$), and θ_{ij} being Malecot's (1948) coefficient of coancestry. If i and j are inbred, π_{ij} is interpreted as twice the coefficient of coancestry for the QTL (see Xie *et al.* 1998 for the interpretation of the inbred case).

In the same manner, the IBD values between two half-sibs i and j at a QTL are measured as

$$\pi_{ij} = 2\theta_{ij} = \begin{cases} 2 & \text{for } Q_2Q_2 - Q_2Q_2 \\ 1 & \text{for } Q_1Q_2 - Q_2Q_3 \\ 0 & \text{otherwise.} \end{cases}$$

Finally, if individuals i and j are non-sibs, and their parents are still supposed unrelated, they will share IBD probability of 0.

Inferring the IBD likelihood at a QTL from marker data: The IBD value is determined by the genotypes of two individuals at the QTL of interest. The actual QTL genotype of an individual, however, is in most cases not observable and must be inferred from flanking marker information (that we term I_M—this is represented in Figure 1 by A and A').

We denote the following probabilities, suited for all cases (full-sib and half-sib cases are particularities of the unrelated case), $p_{i2} = \Pr(Q_1Q_1|I_M)$, $p_{i1} = \Pr(Q_1Q_2|I_M)$, $p_{i0} = \Pr(Q_2Q_2|I_M)$ and $p_{j2} = \Pr(Q_3Q_3|I_M)$, $p_{j1} = \Pr(Q_3Q_4|I_M)$, and $p_{j0} = \Pr(Q_4Q_4|I_M)$. It should be noted that for half-sibs Q_4 is replaced by Q_2 and for full-sibs Q_3 is replaced by Q_1 and Q_4 by Q_2. We write $p_i = [p_{i2}\ p_{i1}\ p_{i0}]^T$ and $p_j = [p_{j2}\ p_{j1}\ p_{j0}]^T$.

Starting from Xie *et al.*'s (1998) notations addressing the case of full-sib individuals only, the conditional expectations of the IBD values are $\hat{\pi}_{ij} = E(\pi_{ij}|I_M) = p_i^T\mathbf{C}p_j$ for between individuals and $\hat{\pi}_{i,i} = E(\pi_{i,i}|I_M) = \mathbf{c}^T p_i$ for the individual with itself, where

$$\mathbf{C} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}.$$

We can easily extend the $\mathbf{C}$ matrix to generalize the

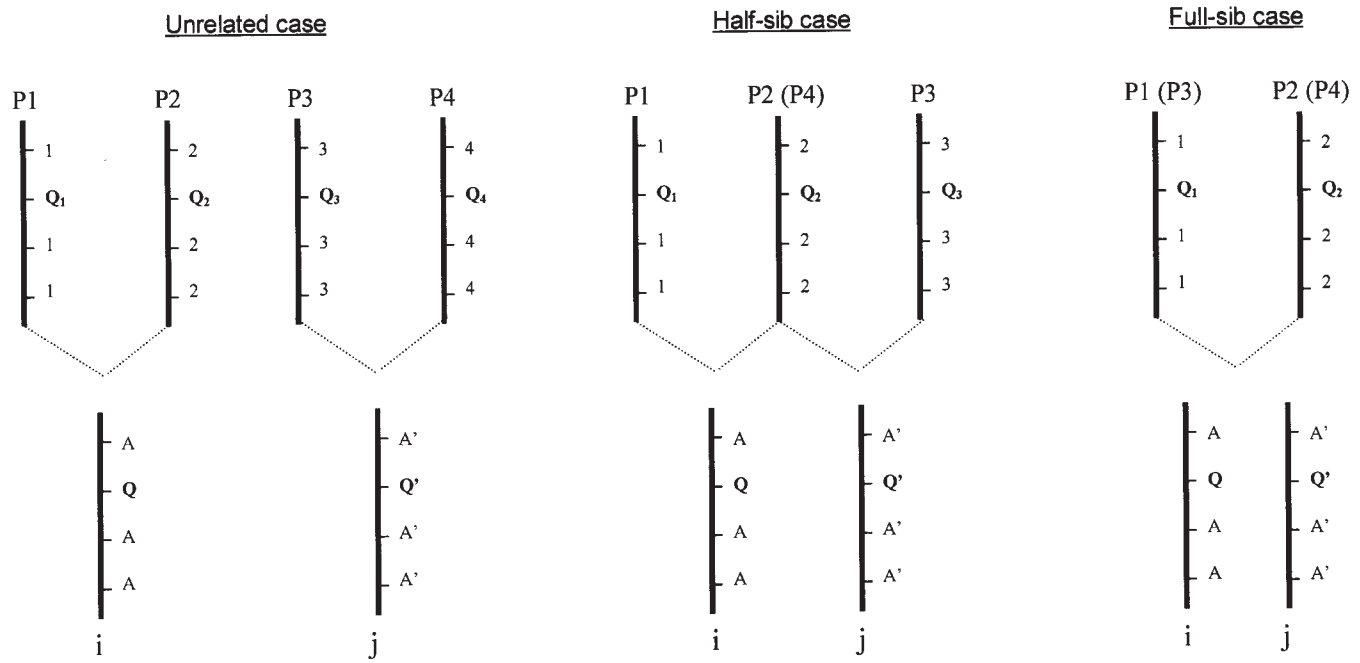

FIGURE 1.—If parents (P1, P2, P3, and P4) are considered as founders, only three types of relationships exist between individuals i and j of the mapping population. Notations Q_1, Q_2, Q_3, and Q_4 represent parents' QTL information while Q and Q' (unknown) represent progenies' i and j QTL information. In the same manner, 1, 2, 3, and 4 represent parents' marker alleles information while A and A' (supposedly known) represent progenies' i and j marker information. Q_1, Q_2, Q_3, and Q_4 are homozygous while Q and Q' can be heterozygous. The possible genotypes at the QTL for the three cases are as follows:

	Unrelated case	Half-sib case	Full-sib case
Q	Q_1Q_1, Q_1Q_2, and Q_2Q_2	Q_1Q_1, Q_1Q_2, and Q_2Q_2	Q_1Q_1, Q_1Q_2, and Q_2Q_2
Q'	Q_3Q_3, Q_3Q_4, and Q_4Q_4	Q_2Q_2, Q_2Q_3, and Q_3Q_3	Q_1Q_1, Q_1Q_2, and Q_2Q_2

full-sib and half-sib case by introducing the coancestries between parents P1-P3 and P2-P4, denoted by θ_{P1P3} and θ_{P2P4}. If parents are considered as founders, these coancestries can take only values 1 or 0. Thus, the new **C** matrix can then be rewritten as $\mathbf{C}_1$:

$$\mathbf{C}_1 = \begin{bmatrix} 2\theta_{P1P3} & \theta_{P1P3} & 0 \\ \theta_{P1P3} & \tfrac{1}{2}(\theta_{P1P3} + \theta_{P2P4}) & \theta_{P2P4} \\ 0 & \theta_{P2P4} & 2\theta_{P2P4} \end{bmatrix}.$$

Note that for the full-sib case P1 = P3 and P2 = P4, so that θ_{P1P3} and θ_{P2P4} are equal to one and the $\mathbf{C}_1$ matrix is similar to **C**. Similarly, the relevant **C** matrices for half-sib individuals can be obtained by replacing θ_{P1P3} by zero and θ_{P2P4} by one—or θ_{P1P3} by one and θ_{P2P4} by zero (and, for unrelated individuals, by replacing both θ_{P1P3} and θ_{P2P4} by zero).

This formula, using the $\mathbf{C}_1$ matrix (with the θ's being equal to 0 or 1 only) for computing the IBD values, is referred to as formula 1 in the rest of the article.

Computation of G matrices with parents not considered as founders: Using the above formula to compute $\hat{\pi}_{i,j}$'s, we assumed that parents of subpopulations were unrelated; *i.e.*, they did not share any common ancestors. Thus, to infer the IBD probabilities in the previous case, we did not need to have more genotypic information than that of the mapping population and of their parents. For the following, we still consider that the parents of the latest breeding cycle and the current $F(n)$-derived lines are the only genotyped material. However, we consider this time that the parents of the mapping population could come from previous generations of breeding. They are thus very likely to share common ancestors (due to the intensive use of some star varieties, for instance), even if those ancestors cannot be genotyped. Thus, for the full-sib case example, we could take into account the probability that the two parents share IBD QTL alleles. For the unrelated case, we could take into account the probability that P1-P3, P1-P4, P2-P3, or P2-P4 share IBD QTL so that $Q_1 \equiv Q_3$, $Q_1 \equiv Q_4$, $Q_2 \equiv Q_3$, and $Q_2 \equiv Q_4$. If we are able to estimate these probabilities, they could be used to improve the computation of $\hat{\pi}_{i,j}$'s. For the following, we supposed that estimates of these probabilities between all parents were available. We take the more general case, *i.e.*, the unrelated one, to draw a general formula that incorporates these estimates and that covers the three cases of relationships between individuals of the mapping population. We denote by θ_{P1P3}, θ_{P1P4}, θ_{P2P3}, and θ_{P2P4} the estimates of the coefficients of coancestries between the four parents.

First, we generalized above the **C** matrix to the known half-sib and full-sib individuals by introducing the coan-

cestries between parents P1-P3 and P2-P4, giving the $\mathbf{C_1}$ matrix.

Similarly, taking into account the coefficients between the parents P1 and P4 on one hand and P2 and P3 on the other hand, we can write the $\mathbf{C_2}$ matrix as

$$\mathbf{C_2} = \begin{bmatrix} 0 & \theta_{P2P3} & 2\theta_{P2P3} \\ \theta_{P1P4} & \tfrac{1}{2}(\theta_{P2P3} + \theta_{P1P4}) & \theta_{P2P3} \\ 2\theta_{P1P4} & \theta_{P1P4} & 0 \end{bmatrix}.$$

Finally, with these two matrices, we can draw a general formula for the conditional expectation of the IBD values between two individuals coming from four (distinct or not) inbred parents:

$$\hat{\pi}_{i,j} = E(\pi_{i,j}|I_M) = p_i^T \mathbf{C_1} p_j + p_i^T \mathbf{C_2} p_j = p_i^T (\mathbf{C_1} + \mathbf{C_2}) p_j;$$

i.e.,

$$\begin{aligned} \hat{\pi}_{i,j} = E(\pi_{i,j}|I_M) = {} & 2(\theta_{P1P3}[(p_{j2} + \tfrac{1}{2}p_{j1})(p_{i2} + \tfrac{1}{2}p_{i1})] \\ & + \theta_{P1P4}[(p_{j2} + \tfrac{1}{2}p_{j1})(p_{i0} + \tfrac{1}{2}p_{i1})] \\ & + \theta_{P2P3}[(p_{j0} + \tfrac{1}{2}p_{j1})(p_{i2} + \tfrac{1}{2}p_{i1})] \\ & + \theta_{P2P4}[(p_{j0} + \tfrac{1}{2}p_{j1})(p_{i0} + \tfrac{1}{2}p_{i1})]). \end{aligned}$$

The conditional expectation of the IBD for an individual with itself remains

$$\hat{\pi}_{i,i} = E(\pi_{i,i}|I_M) = 2p_{i2} + p_{i1} + 2p_{i0}.$$

In the rest of the article, this formula, using the $\mathbf{C_1}$ and $\mathbf{C_2}$ matrices to compute IBD values, is referred to as formula 2.

Please note that in the case of two full-sib individuals, the probability that the two parents P1 and P2 share initially IBD QTL is taken into account in formula 2 by replacing P3 by P1 and P4 by P2 (P1 and P2 are considered as the parents of the first full-sib, P3 and P4 as the parents of the second full-sib). Thus, both θ_{P1P3} and θ_{P2P4} will take values of one (accounting for the full-sib relationship with parents considered as founders—similar to the formula of Xie *et al.* 1998) while both θ_{P1P4} and θ_{P2P3} will be written as θ_{P1P2} (accounting for possible coancestry between parent P1 and P2).

Estimates of the coefficients of coancestries: With the above formula 2, it may be seen that accurate estimates of the coefficients of coancestries between parents of individuals i and j of the mapping population are needed for the computation of the **G** matrices (that are built at each scanned position). These coefficients need also to be estimated between all the individuals i and j of the mapping population, to account for polygenic variation through the relationship matrix **A**. There are two main ways to estimate these coefficients of coancestries. The first one is to compute Malecot's coefficients on the basis of the available declared pedigrees and come back to the pedigree of each variety as far as possible. For example, two parents of the mapping population with a grandparent in common will share an expected proportion of genome identical by descent of 2×0.125. Thus, this information would be used in formula 2 to improve the accuracy of the IBD estimate.

The second way to estimate these coefficients is to use the available molecular marker information. Nei and Li's (1979) formula can be used to calculate the genetic similarity index (GS): $GS = 2N_{ij}/(N_i + N_j)$, where N_{ij} is the number of alleles in common between genotypes i and j, and N_i and N_j are the total number of alleles observed for genotypes i and j, respectively.

Implementation of the IBD formula: We used the deterministic approach of the MDM program (Servin *et al.* 2002) to compute all the p_i and p_j probabilities, at any generation of selfing or backcrossing. IBD values were computed every 3 cM. Two flanking markers were used to infer the genotypes' probabilities. In the frequent case where the two parents shared the same marker alleles at one or two loci flanking the putative QTL position, the next closest markers to the interval were used. It can easily be demonstrated that the IBD values calculated at a putative QTL will be more precise if the flanking markers are highly polymorphic.

Solving of the mixed-linear models and test statistic under the null hypothesis: *Two-step IBD-based variance component method:* The method used to map QTL in a complex inbred pedigree is then similar to all interval-mapping-based variance component methods. It is composed of two steps (two-step IBD-based variance component method), as described in George *et al.* (2000). In step 1, we computed the **G** matrices according to the formula tested, for all the scanned positions. We then inverted and wrote them in ASREML (Gilmour *et al.* 1998) format for user-defined inverse (co)variance matrices. We also computed the appropriate additive relationship matrix **A**, inverted it, and wrote it in ASREML format. In step 2, ASREML provided restricted maximum-likelihood (REML) estimates of steps 1 and 2. To test for the presence of a QTL against no QTL at a particular chromosomal position, we used the log-likelihood-ratio test: $LR = -2\ln[L_0 \text{ (H0, no QTL present)} - L_1 \text{ (H1, QTL present)}]$, where L_1 and L_0 represent the likelihood values of steps 1 and 2 evaluated at the REML solutions, respectively.

Test statistic under the null hypothesis: The choice of a test statistic threshold is always challenging in this kind of situation. As mentioned by George *et al.* (2000) permutation testing is problematic for such IBD-based variance component analysis since it is unclear how to permute the data while retaining the association between polygenic variation and marker information. Many publications (Zeng 1994; Xu and Atchley 1995, for example) report that when a chromosomal interval is being scanned, the empirical distribution of LR follows a mixture of two chi-square distributions, with 1 and 2 d.f., respectively. Since this article deals with simulated data, it is possible to replicate data under the null hypothesis of no QTL segregating, construct the empirical distribution of LR, and derive an empirical threshold by

choosing the 95th percentile of the highest test statistic, generally over 500 or 1000 stochastic realizations. In this article, we calculated an empirical threshold for each set of parameters, and then we ran 1000 additional simulations with no QTL segregating on the scanned chromosome. We increased the polygenic variance such that the total genetic variance remained unchanged and determined the empirical threshold by choosing the 95th percentile from the list of 1000 runs. It should be noted that this threshold is not genomewise but is chromosomewise.

A SIMULATION STUDY: THE CASE OF A PEDIGREE BREEDING PROGRAM

We chose the case of pedigree breeding for the simulation study as it contained most of the difficulties generally encountered in inbred breeding programs: frequent lack of reliable pedigree information, beyond the parents (and thus unavailability of ancestor lines for genotyping); possible genotyping only of advanced generations of selfing, when the number of lines has decreased and the precision of trials increased, constraining the computation of IBD at the end of a breeding cycle, without any marker information between the initial cross and the resulting progenies (a breeding cycle comprises the initial crosses between many different parents to obtain the new improved lines after many generations of self pollination); the very high number of parents of the mapping population yielding very small full-sib families, and an uneven (L-shaped) distribution of half-sib family sizes; and the possible occurrence of mass selection for the choice of the parents at the start of a breeding cycle.

Simulation of the breeding program: An S-PLUS (2000) function was developed to reproduce the typical steps of pedigree-based plant breeding programs (see http://www.genetics.org/supplemental/ for a detailed description). Briefly, we started by creating founder lines at the beginning of breeding (beginning of 20th century, for instance). At this stage, the material was in complete linkage disequilibrium, with as many alleles as there were founder lines (for example, founder line 1 carried only allele "1" for all the markers and QTL . . .). In the first breeding cycle, we produced new germplasm by crossing the founder lines together. Then, during the following breeding cycles, we performed crosses in a pedigree-breeding fashion. First a large number of parents were used to obtain a reduced number of lines in advanced selfing generations (for example, 100 parents are crossed to obtain only 500 individuals at the end of a breeding cycle). Second, most of the current parents were chosen among the lines derived from the most recent breeding cycles while a small part was extracted from older breeding cycles (to represent nonelite germplasm). Third, crosses were unevenly distributed (elite germplasm was crossed more than nonelite, for example). Fourth, mass selection on the value of the quantitative trait was possible at each breeding cycle.

At every generation, a phenotype was simulated for each individual line on the basis of its main QTL and polygene alleles. We performed QTL detection on the last breeding cycle.

Note that at the beginning of our breeding programs, all the allele frequencies were equal, which was not the case after many generations due to genetic drift, nonpanmictic conditions, and selection. All the markers and QTL were in full linkage disequilibrium at G_0 but were not so after the breeding programs—the chromosomes having undergone many recombinations. Hence, as anticipated, a simple ANOVA was inefficient (results not shown).

Simulated populations: To illustrate the methodology, we focused only on two representative settings (two complex populations of different size), for which we varied a limited number of parameters. For both settings, we initially fixed the following parameters: 20 founder lines (that initially correspond to 20 different alleles at each marker and QTL, with 20 different allele effects at the QTL), 21 chromosomes of length 100 cM each with 11 markers spaced every 10 cM, a QTL segregating at position 45 (half-way between two markers) on chromosome 1, and a total genetic heritability (QTL and polygenes) of 0.5. We fixed the number of breeding cycles to 10 without selection and to 6 with selection (to retain genetic variance around the chromosome 1 QTL and around the polygenes). The number of polygenes varied from 40 for the cases without selection to 9 and 4 with selection, for QTL heritabilities 0.05 and 0.1. We chose these numbers of polygenes in the case of selection to set an equivalent heritability for each QTL and polygene to avoid the rapid fixation of chromosome 1 QTL.

Setting 1 is composed of 300 inbred lines derived from crosses between 50 parents chosen at random from the previous breeding cycles. Of 1225 different possible crosses [$(50 \times 49)/2$], 170 crosses per breeding cycle were simulated. Each cross gave, on average, 1.75 full-sibs and each parent was found, on average, in 12 progenies. We simulated two groups of mapping populations: group a was obtained without the influence of selection on the quantitative trait, and group b was obtained under the influence of selection on the quantitative trait for the choice of the parents at each generation.The heritability of the chromosome 1 QTL was fixed for each population at 0.1.

Setting 2 is composed of 500 individuals, derived from crosses between 100 parents. Of 4950 different possible crosses, 285 crosses per breeding cycle were simulated. Each cross gave, on average, 1.75 full-sibs and each parent was found, on average, in 10 progenies. We simulated different groups of populations, for different levels of QTL heritabilities and with and without the influence

TABLE 1

Main characteristics of the different mapping populations

Mapping populations	Occurrence of selection and simulated QTL and total genetic heritability	No. of breeding cycles	No. of polygenes	No. of marker alleles flanking the QTL	Effective no. of marker alleles
Setting 1 (50 parents, 300 progenies)	No, $h^2_{QTL} = 0.1$ $h^2_g = 0.486$ (0.083)	10	40	10	5.5
	Yes, $h^2_{QTL} = 0.1$ $h^2_g = 0.478$ (0.180)	6	4	6	2.8
Setting 2 (100 parents, 500 progenies)	No, $h^2_{QTL} = 0.05, 0.1, 0.2$ $h^2_g = 0.427, 0.435, 0.456$ ($\sim$0.050)	10	40	14	6.2
	Yes, $h^2_{QTL} = 0.05$ $h^2_g = 0.590$ (0.210)	6	9	13	4.5
	Yes, $h^2_{QTL} = 0.1$ $h^2_g = 0.550$ (0.210)	6	4	12	4

The fixed parameters are 20 founder lines (*i.e.*, initially 20 possible alleles at all the markers and QTL), 21 chromosomes of length 100 cM each with 11 markers spaced every 10 cM, and QTL segregating at position 45 on chromosome 1. The effective number of alleles is computed as $N_{eff} = 1/\Sigma_1^N f^2$, f being the allele frequencies and N the number of alleles. This effective number of alleles is averaged at the two markers flanking the QTL (located on position 45 cM).

of selection: group a was obtained without the influence of selection. We created the quantitative trait on the mapping population (10th generation of breeding) for QTL heritabilities 0.05, 0.1, and 0.2. Group b was obtained under the influence of selection on the value of the quantitative trait. We investigated two levels of QTL heritability: 0.05 (with nine polygenes of 0.05 each) and 0.1 (with four polygenes of 0.1 each).

Table 1 summarizes the main characteristics of the different mapping populations. It should be noted that there were initially 20 alleles for each marker and QTL but that this number was greatly reduced after 6–10 breeding cycles, due to genetic drift and/or selection pressure.

Methods compared: In this article, we investigated two different ways to infer the coefficients of coancestries. We thus termed formulas 2a and 2b as follows:

Formula 2a: Malecot's coefficients of coancestries are used to build **G** (through formula 2) and **A** matrices. For that, full pedigree is stored during simulations and used to compute parents' and progenies' coefficients of coancestries. The algorithm implemented is described in Lynch and Walsh (1998, p. 763).

Formula 2b: Marker-based estimates of the coefficients of coancestries on the whole genome are used to build **G** (through formula 2) and **A** matrices. They are computed using Nei and Li's (1979) formula.

The reference method in the simulation study is formula 1, which uses only half-sib and full-sib relationships, which are known with 100% certainty, to compute the IBD matrices and the relationship matrix. Thus, for formula 1, the **G** matrix will have terms different from 0 only when full-sib or half-sib relationships exist between two individuals of the mapping population, and the **A** matrix will take the expected proportion of genome shared by two individuals, *i.e.*, 2×0.5 if the two inbred individuals are full-sibs, 2×0.25 if the two inbred individuals are half-sibs, 0 otherwise.

In setting 1, we used alternatively formulas 1, 2a, and 2b, while we used only formulas 1 and 2b in setting 2, due to computation time required for obtaining Malecot's coefficients of coancestries for such important populations.

We tested every third centimorgan for the presence of a QTL. Under each condition, the detection was performed for 100 random replicates. Parameters estimates and their standard error are reported for all replicates.

RESULTS

The average likelihood-ratio test profiles (over 100 replicates) are presented in Figure 2 for both settings. There was a strong influence of the formula on the LR profile for both settings, either without the influence of selection (Figure 2, a and c) or with selection (Figure 2, b and d). As expected, there was also a strong influence of the magnitude of the QTL effect (*i.e.*, the heritability of the QTL) on the LR profile (Figure 2, c and d). Formula (2b)—which takes into account ancestor pedigree relationships as estimated by markers to infer the IBD values—outperformed in terms of detection power other formulas for both settings.

The ability of the three formulas to estimate the parameters of interest accurately can be judged from the results presented in Tables 2 and 3. The accuracy of the estimated QTL position increased with both the

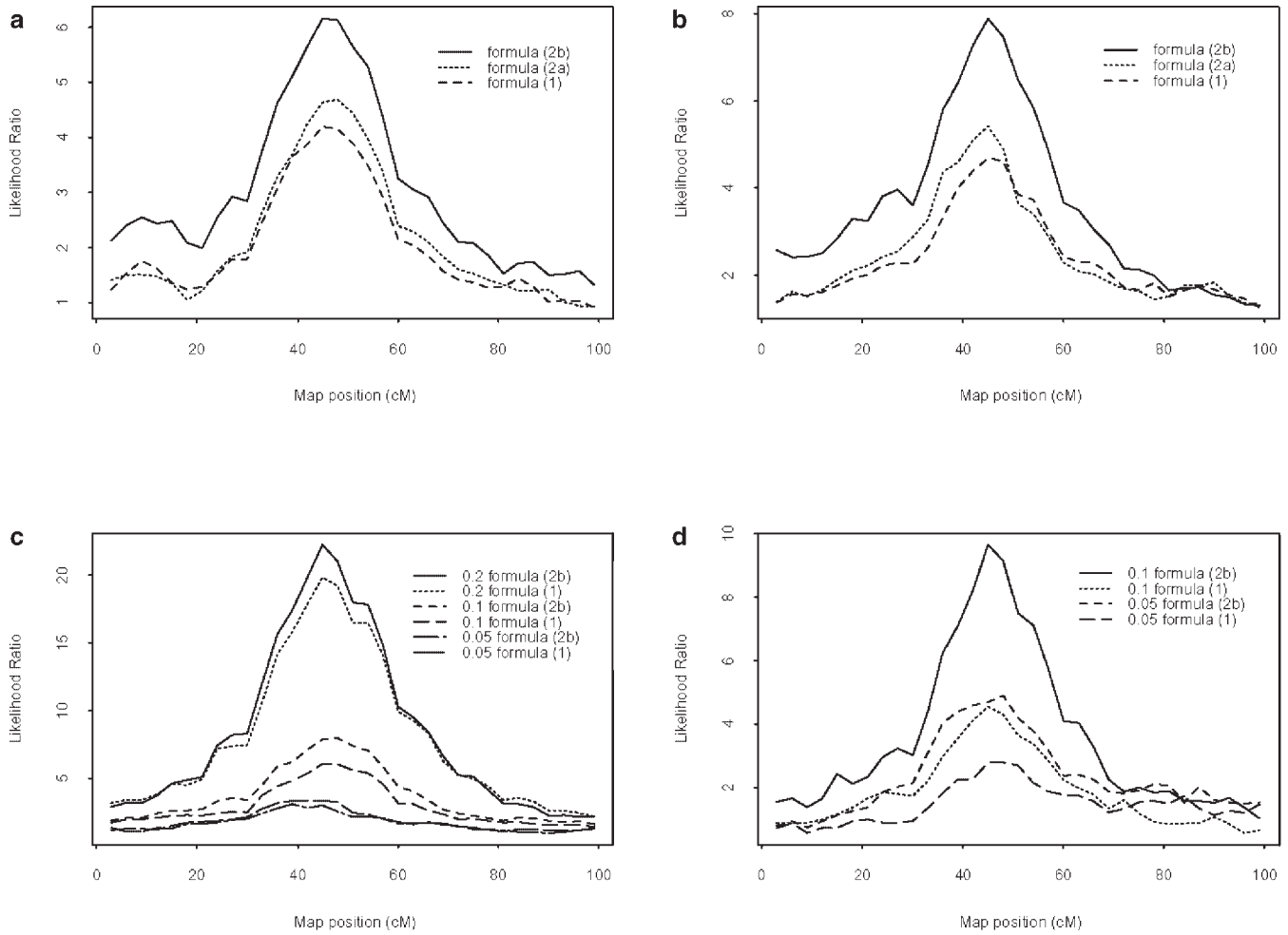

FIGURE 2.—Comparison of the LR profiles for (a) setting 1 under formulas 1, 2a, and 2b without selection; (b) setting 1 under formulas 1, 2a, and 2b with selection; (c) setting 2 under formulas 1 and 2b for three levels of QTL heritabilities (0.2, 0.1, and 0.05), without selection; and (d) setting 2 under formulas 1 and 2b for two levels of QTL heritabilities (0.1 and 0.05), with selection.

design of the population (higher QTL heritabilities, bigger mapping population) and the switch from formula 1 to formulas 2a and 2b. Selection also acted on the accuracy of the position estimates by reducing the confidence intervals under formula 2b. We also noted difficulties in estimating the QTL heritability accurately, as it was already shown in simulation studies by GRIGNOLA *et al.* (1996, 1997) and GEORGE *et al.* (2000). The

TABLE 2

Estimates of position, QTL heritability ($\hat{h}^2_{QTL}$), total genetic heritability ($\hat{h}^2_g$), and test statistic (LR) for setting 1

Setting 1	Tested formula	Position	$\hat{h}^2_{QTL}$	$\hat{h}^2_g$
Without selection	True values	45 cM	0.1	0.486 (0.083)
	Formula 1	46.65 (18.51)	0.117 (0.054)	0.454 (0.154)
	Formula 2a	47.14 (18.98)	0.126 (0.059)	0.468 (0.145)
	Formula 2b	45.18 (18.48)	0.138 (0.067)	0.479 (0.124)
With selection	True values	45 cM	0.1	0.478 (0.180)
	Formula 1	50.05 (19.35)	0.125 (0.059)	0.446 (0.178)
	Formula 2a	46.20 (19.14)	0.150 (0.068)	0.457 (0.155)
	Formula 2b	44.08 (17.19)	0.169 (0.079)	0.516 (0.189)

See Table 1 for the description of setting 1. Mean and standard deviations (in parentheses) are calculated among the 100 replicates.

TABLE 3

Estimates of position, QTL heritability ($\hat{h}^2_{QTL}$), total genetic heritability ($\hat{h}^2_g$), and test statistic (LR) for setting 2

Setting 2	Tested formula	True h^2_g	Position	$\hat{h}^2_{QTL}$	$\hat{h}^2_g$
			Without selection		
$h^2_{QTL} = 0.05$	Formula 1	0.427 (0.068)	49.44 (25.42)	0.067 (0.035)	0.427 (0.108)
	Formula 2b		46.45 (21.84)	0.085 (0.041)	0.432 (0.107)
$h^2_{QTL} = 0.1$	Formula 1	0.435 (0.051)	47.76 (20.88)	0.099 (0.042)	0.442 (0.081)
	Formula 2b		45.52 (17.72)	0.129 (0.056)	0.450 (0.083)
$h^2_{QTL} = 0.2$	Formula 1	0.456 (0.047)	44.91 (7.35)	0.192 (0.058)	0.451 (0.083)
	Formula 2b		45.73 (6.89)	0.228 (0.065)	0.467 (0.091)
			With selection		
$h^2_{QTL} = 0.05$	Formula 1	0.590 (0.210)	53.62 (23.71)	0.060 (0.037)	0.635 (0.171)
	Formula 2b		52.98 (21.09)	0.096 (0.055)	0.658 (0.178)
$h^2_{QTL} = 0.1$	Formula 1	0.550 (0.210)	48.10 (17.90)	0.121 (0.036)	0.657 (0.186)
	Formula 2b		46.76 (12.33)	0.155 (0.061)	0.638 (0.178)

See Table 1 for the description of setting 2. Mean and standard deviations (in parentheses) are calculated among the 100 replicates.

accuracy of the estimated QTL heritability was influenced by the initial effect of the QTL, by the switch from formula 1 to formula 2b, and by the occurrence of selection. For all designs, formula 2b led us to overestimate QTL heritabilities more than formula 1 did.

We report in Table 4 the average LR test statistics over all replicated simulations and the respective power estimates under the empirical chromosomewise threshold. The empirical threshold values were nearly equivalent for all the designs, which was not really surprising as the number of parameters being tested in the random model strategy did not vary. The values of the LR test and thus the power to detect QTL under the empirical threshold were influenced by the design of the population (higher QTL heritabilities, size of the mapping population, influence of selection) and by the switch from formula 1 to formula 2b. This switch to formula 2b gave an increase in the value of the test by a mean

TABLE 4

Observed 95th percentile likelihood ratios under the hypothesis of no QTL segregation, test statistic, and power to detect QTL

	Nonoccurrence of selection			Selection		
Formula	Threshold	Test statistic	Power (%)	Threshold	Test statistic	Power (%)
			Setting 1: 300 individuals, 50 parents			
Formula 1	4.04	5.66 (3.99)	60	4.12	5.44 (3.93)	45
Formula 2a	3.97	6.08 (4.57)	63	3.88	5.78 (4.59)	47
Formula 2b	4.06	8.16 (5.62)	78	3.91	8.88 (5.86)	85
			Setting 2: 500 individuals, 100 parents			
$h^2_{QTL} = 0.05$						
Formula 1	4.08	4.62 (3.52)	47	3.86	4.41 (3.94)	44
Formula 2b	3.96	5.23 (3.97)	58	3.44	7.35 (4.63)	70
$h^2_{QTL} = 0.1$						
Formula 1	4.08	7.75 (5.06)	71	3.66	6.35 (4.14)	67
Formula 2b	3.96	9.76 (6.71)	80	3.44	11.65 (6.22)	94
$h^2_{QTL} = 0.2$						
Formula 1	4.08	22.03 (10.9)	100	—	—	—
Formula 2b	3.96	23.69 (10.8)	100	—	—	—

See Table 1 for a description of settings 1 and 2. Threshold represents the empirical chromosomewise threshold calculated for 1000 replicates. Test statistic is the mean and standard deviation of the maximum of the LR test for the 100 replicates. Power is the percentage of replicates with maximum LR exceeding the empirical threshold. —, simulations are not performed under these conditions.

of 20%, yielding thus an increase in the detection power. The interest of formula 2b was further demonstrated with selection, for both settings: almost twice as many replicates were significant when IBD values were inferred by taking into account genetic similarities as estimated by markers as when using direct pedigrees (or Malecot's coefficients of coancestries for setting 1).

DISCUSSION

Many statistical methods already exist to map QTL in inbred plant material; however, most of these methods focus on a single biparental cross or on simple experimental populations such as diallel designs. Other methods have been developed to address more challenging population structures (Xie *et al.* 1998; Yi and Xu 2001; Bink *et al.* 2002, for example), but they do not appear to be easily extendable to highly fragmented and unbalanced populations, at any selfed or backcrossed generation, and they do not take into account the possibility for alleles to be IBD if ancestor pedigrees are not available. In this study, we extended the QTL mapping methodology proposed by Xie *et al.* (1998) to typical plant breeding populations made up of selfed (or backcrossed) individuals, which may have two parents in common, one parent in common, or parents more distantly related to each other or not related. Two sets of populations mimicking conventional breeding programs were simulated, in an effort to reproduce realistic conditions of marker and gene frequencies and linkage disequilibrium across the parental lines. The complex design of these populations (highly fragmented, with unbalanced contributions of the parents to the following generation and the influence of selection) was chosen to represent the more complex and more general scenario found in real plant breeding schemes, and thus results should be applicable to any simpler breeding design (for example, to diallel or factorial designs, which are particular cases of the complex simulated designs). We assessed, on these populations, different approaches to compute IBD values for QTL detection, while applying a two-step IBD-based variance component method.

In such multicross inbred designs, there is a strong within-family linkage disequilibrium that can be exploited by comparing the parents' genotypes with the current mapping population, which accumulated relatively few crossovers. Formula 1 is solely based on the utilization of this linkage disequilibrium, using only direct pedigrees (which lines are the parents of a given cross) to compute IBD values, considering that no relevant pedigree information was available from the parents of the current mapping population. Results obtained under formula 1 in terms of test statistic, power, and accuracy on the position estimates are close to those found in Xie *et al.* (1998) and Xu (1998) for populations of equivalent sizes, even if the structure of our simulated populations is a little different. In a second approach, we considered that estimates of the coefficients of coancestries were inferable between the parents of the mapping population, but that genotypic information from the parents' ancestors was not available. We integrated these coefficients of coancestries to the IBD computation, in formula 2. Thus this formula can be viewed, loosely speaking, as an attempt to merge, to some extent, several families together on the basis of the likelihood that the parents share the same alleles identical-by-descent at the putative locus. Then, in constructing the matrices of IBD values, the extent to which the **G** matrix was modified from formula 1 to formula 2 is quite large. The proportion of IBD values equal to zero in **G**, with formula 1—those values between non-sib lines—and replaced by nonzero values in formula 2, was equal to 87% for setting 1 and 91% for setting 2, with an average inferred IBD value of 0.11 between nonsibs. This leads to a substantial improvement of the accuracy of the position estimates and of the QTL detection power for all the designs, by extracting more information on IBD status between individuals. The power increase obtained by using formula 2b instead of formula 1 follows the same principle as that obtained by Xie *et al.* (1998) in his Table 4, when he switched from a 250×2 sampling strategy (250 families with two full-sib individuals each), for example, to a less fragmented 50×10. The power to detect QTL in IBD-based approaches increases with the proportion of nonnull PIBD in the **G** matrix. Thus, in the multicross design of Xie *et al.* (1998), nonzero diagonal boxes in the **G** matrix corresponding to the full-sib relationships make up an increasing proportion of the total **G** matrix when reducing the number of families (for example, $250 \times 2 \times 2 = 1000$ cells with full-sib relationships for a 250×2 sampling strategy instead of $50 \times 10 \times 10 = 5000$ cells for a 50×10 sampling strategy). This gives an increase in the level of information at each putative QTL and thus in the power of the test.

The superiority in terms of power of formula 2b compared to the other formulas is even higher in the situation of selection. One explanation is that, during selection, the same best alleles tend to be selected and this is so for every QTL in the genome, while the other alleles are discarded. The same phenomenon also takes place at the neutral markers because of linkage disequilibrium. This decrease in allele number increases the resemblance between individuals and reduces the effective population size. This also amounts to a decrease in the effective number of alleles and of parents. Hence, the assumption that alleles across the different parents are non-IBD, as implied by formula 1, gradually becomes even less justified as selection operates whereas formula 2b integrates the increasing proportion of the genome in common between the parents at the successive breeding cycles by taking into account their genetic similarities. Selection also generated a bias in the predicted

proportion of IBD alleles shared between parents when Malecot's coefficients were used. This bias induced by selection explained the inefficiency of formula 2a, which gave the same results as formula 1. Finally, under the influence of selection, the reduction in marker polymorphism across the parents (for setting 1, with selection, the effective number of alleles decreased from 5.3 down to 3.6 on average for all chromosome 1 markers) decreased the chance to have informative markers flanking the interval being scanned: thus, informative flanking markers had to be found further apart on average. This led, in turn, to lower accuracy of estimates of the putative allelic state of QTL. Under formula 2b, however, this reduction of the effective number of alleles had less influence on the chance to detect the QTL. Taking into account the increasing proportion of genome in common between the parents did more than compensate the decrease in the number of informative markers, in terms of QTL detection power.

We mention that the structure of breeding programs is not really appropriate for the computation of Malecot's coefficients of coancestries, first because the selection pressure during line development often generates biases in the predicted proportion of parental genomes shared by the current lines and second, because pedigrees noted by breeders or declared for variety registration before commercial release are often prone to errors. It has already been suggested by Bernardo (1993) that the use of molecular marker information to compute coefficients of coancestries between individuals in the case of plant breeding was more suitable than computing them by declared pedigrees. This property was also shown in this article for the use of genetic similarities instead of Malecot's coefficients of coancestries to improve the IBD computation. Sources of biases, either on marker information (presence of alike-in-state, *i.e.*, non-IBD alleles, uneven repartition of markers along the chromosomes) or on pedigrees (with a portion of wrong parents' pedigrees), were added to the settings. QTL analysis performed under these conditions showed that the use of marker information to compute genetic similarities always contributed more positively to the QTL detection power than the use of Malecot's coefficients (results not shown). This trend was not reversed, even in the case of an uneven distribution of polygenes (when only four or nine polygenes were spread on different chromosomes in the case of selection).

There is still some scope for a more accurate and probably less biased estimation of the coefficients of coancestries between parents and between individuals to estimate the parameters of the model more accurately and increase the QTL detection power. We could suggest, for example, subtracting from all genetic similarities an estimated proportion of alleles in common that supposedly unrelated lines have in common—by definition, these alleles in common would be identical by state only and not IBD. This method to infer the proportion of the IBD genome was suggested by Melchinger *et al.* (1991).

Another lead is to improve the efficiency of the model, for example, to account for multiple QTL. We would first analyze one chromosome at a time, introducing the appropriate IBD matrices into the linear mixed model (1). Once QTL detection is performed for all the chromosomes, we would extract the most significant QTL and introduce it as a covariate in a new linear mixed model (with two known random terms: the polygenic term and the most significant QTL). We would perform the analysis again, introducing the appropriate IBD matrices into this new model. If significant QTL still remained or appeared during the genome analysis, then the most significant one would be added to the model and the analysis carried out again until no more significant QTL appear. This procedure is described in Almasy and Blangero (1998) and is somewhat analogous to the composite interval mapping proposed by Zeng (1993) for biparental populations.

Alternatively, we could also improve the precision of the matrix **A** if its computation were based on the markers that are actually linked to some polygenes, *i.e.*, to some QTL, instead of using all the markers indiscriminately. This procedure could bring an advantage only if a few QTL explain the genetic variation as opposed to many with a small effect, all over the genome.

Our method did not take into account haplotype information on the carrier chromosome, as the goal in this study was to detect QTL at a low marker density. The method is typically a linkage method based concomitantly on the available information of the last breeding generation and on an estimate of the proportion of IBD alleles between parents, at any gene, based on marker information from the whole genome. But what would happen, for formula 2b, if genetic similarities between parents were computed on the scanned chromosome only? When a QTL experiment is launched on new germplasm, little is known about the genetic factors whose segregation is going to influence the trait most. Therefore, a genome-wide scan for QTL must be carried out, using a low-marker density first. Hence, using haplotype information as in Jansen *et al.* (2003) or Lund *et al.* (2003) would have been worse in this context—that of our study—since linkage disequilibrium between markers separated by 10 cM is too low to recognize conserved chromosome fragments from a putative common founder. Alternatively, using the restricted set of markers (to those of the scanned chromosome) to calculate our IBD values as in formula 2b, without attempting to identify conserved haplotypes, yields poorer detection power than using the complete marker set data (results not shown). This is due to the fact that, in situations of low linkage disequilibrium, adjacent markers with the densities mentioned above can be considered to segregate independently. Thus, restricting our marker set to those of the scanned chromosome amounts only to

decreasing our sample of probed loci used to infer the IBD value expectancy at any locus.

With the sort of experimental designs that we have simulated one could envisage using the two-stage IBD method with the improvements that we propose, to get a first estimate of the chromosome segment where the QTL lies. This would be done at a low marker density, highly polymorphic markers placed every 10–20 cM performing equally well (results not shown). Next, fine mapping of the QTL could be undertaken on this material with an increased marker density in the QTL's region. QTL-IBD probabilities between each pair of haplotypes would be calculated and linkage disequilibrium mapping could be performed, as described, for example, in Meuwissen and Goddard (2000). Other methods for fine mapping of a quantitative trait locus combining linkage and linkage disequilibrium mapping (Meuwissen *et al.* 2002; Lund *et al.* 2003) within the mixed model framework could also be used in such pedigree breeding material. This should increase the cost efficiency and the precision of QTL mapping in comparison to each method performed separately. If the study is not so much oriented toward fine mapping the QTL but more toward marker-assisted selection, the method of parental haplotype sharing proposed by Jansen *et al.* (2003) could allow identification of the haplotypes of minimum length that have the most promising effect. It would then directly provide markers to manipulate these haplotypes in breeding schemes, which is perhaps the main goal of QTL detection in such material.

The use of this methodology could increase the efficiency and cost effectiveness of quantitative trait loci mapping in applied contexts and could provide an alternative to the development of a specifically designed recombinant population, by exploiting the genetic variation used by plant breeders. It is in the typical breeding, non-purpose-built populations that the improvement we propose to the two-step IBD variance component method would provide the highest gain in QTL detection power. The methodology developed in this article is currently applied to the analysis of real wheat-breeding data.

The authors are grateful to L. Moreau and to the reviewers for helpful comments on the manuscript. They also thank F. Vear for her assistance with the English language. This research was supported by the Ministère de l'Economie, des Finances et de l'Industrie (Après Séquençage Génomique program no. 01 04 90 6058).

LITERATURE CITED

Almasy, L., and J. Blangero, 1998 Multipoint quantitative trait linkage analysis in general pedigrees. Am. J. Hum. Genet. **62:** 1198–1211.

Bernardo, R., 1993 Estimation of coefficient of coancestry using molecular markers in maize. Theor. Appl. Genet. **85:** 1055–1062.

Bink, M. C. A. M., P. Uimari, M. Sillanpää, L. Janss and R. Jansen, 2002 Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor. Appl. Genet. **104:** 751–762.

George, A. W., P. M. Visscher and C. S. Haley, 2000 Mapping quantitative trait in complex pedigrees: a two-step variance component approach. Genetics **156:** 2081–2092.

Gilmour, A. R., B. R. Cullis, S. J. Welham and R. Thompson, 1998 *ASREML. Program User Manual.* Orange Agricultural Institute, Orange, New South Wales, Australia.

Gimelfarb, A., and R. Lande, 1994 Simulation of marker-assisted selection in hybrid populations. Genet. Res. **63:** 39–47.

Gimelfarb, A., and R. Lande, 1995 Marker-assisted selection and marker-QTL associations in hybrid populations. Theor. Appl. Genet. **91:** 522–528.

Grignola, F. E., I. Hoeschele, Q. Zhang and G. Thaller, 1996 Mapping quantitative trait loci in outcross populations via residual maximum likelihood. II. A simulation study. Genet. Sel. Evol. **28:** 491–504.

Grignola, F. E., Q. Zhang and I. Hoeschele, 1997 Mapping linked quantitative trait loci via residual maximum likelihood. Genet. Sel. Evol. **29:** 529–544.

Haley, C. S., and S. Knott, 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity **69:** 315–324.

Heath, S. C., 1997 Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am. J. Hum. Genet. **61:** 748–760.

Hospital, F., L. Moreau, F. Lacoudre, A. Charcosset and A. Gallais, 1997 More on the efficiency of marker-assisted selection. Theor. Appl. Genet. **95:** 1181–1189.

Jansen, R. C., 1993 Interval mapping of multiple quantitative trait loci. Genetics **135:** 205–211.

Jansen, R. C., J.-L. Jannink and W. D. Beavis, 2003 Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. Crop Sci. **43:** 829–834.

Jiang, C., and Z-B. Zeng, 1995 Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics **140:** 1111–1127.

Korol, A., Y. Ronin and V. Kirzhner, 1995 Interval mapping of quantitative trait loci employing correlated trait complexes. Genetics **140:** 1137–1147.

Lande, R., and R. Thompson, 1990 Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics **124:** 743–756.

Lander, E., and D. Botstein, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121:** 185–199.

Lund, M. S., P. Sorensen, B. Guldbransten and D. A. Sorensen, 2003 Multitrait fine mapping of quantitative trait loci using combined linkage disequilibria and linkage analysis. Genetics **163:** 405–410.

Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.

Malécot, G., 1948 *Les Mathématiques de l'Hérédité.* Masson, Paris.

Melchinger, A. E., M. M. Messmer, M. Lee, W. L. Woodman and K. R. Lamkey, 1991 Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci. **31:** 669–678.

Meuwissen, T. H. E., and M. E. Goddard, 2000 Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. Genetics **155:** 421–430.

Meuwissen, T. H. E., A. Karlsen, S. Lien, I. Olsaker and M. E. Goddard, 2002 Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. Genetics **161:** 373–379.

Moreau, L., S. Lemarié, A. Charcosset and A. Gallais, 2000 Economic efficiency of one cycle of marker-assisted selection. Crop Sci. **40:** 329–337.

Muranty, H., 1996 Power of tests for quantitative trait loci detection using full-sib families in different schemes. Heredity **76:** 156–165.

Nei, M., and W. H. Li, 1979 Mathematical model for studying genetic variations in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5369–5373.

Servin, B., C. Dillmann, G. Decoux and F. Hospital, 2002 MDM a program to compute fully informative genotype frequencies in complex breeding schemes. J. Hered. **93** (3): 227–228.

Sobel, E., H. Sengul and D. E. Weeks, 2001 Multipoint estimation of identity-by-descent probabilities of arbitrary positions among marker loci on general pedigrees. Hum. Hered. **52** (3): 121–131.

S-PLUS, 2000 *S-PLUS Guide to Statistical and Mathematical Analyses.* MathSoft, Massachusetts Institute of Technology, Cambridge, MA.

Xie, C., D. D. G. Gessler and S. Xu, 1998 Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method. Genetics **149:** 1139–1146.

Xu, S., 1998 Mapping quantitative trait loci using multiple families of line crosses. Genetics **148:** 517–524.

Xu, S., and W. R. Atchley, 1995 A random model approach to interval mapping of quantitative trait loci. Genetics **141:** 1189–1197.

Yi, N., and S. Xu, 2001 Bayesian mapping of quantitative trait loci under complicated mating designs. Genetics **157:** 1759–1771.

Zeng, Z-B., 1993 Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA **90:** 10972–10976.

Zeng, Z-B., 1994 Precision mapping of quantitative trait loci. Genetics **136:** 1457–1468.

Communicating editor: C. Haley