Interval Mapping of Quantitative Trait Loci in Autotetraploid Species

C. A. Hackett,* J. E. Bradshaw† and J. W. McNicol*

*Biomathematics and Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland
and †Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland

ABSTRACT

This article presents a method for QTL interval mapping in autotetraploid species for a full-sib family derived by crossing two parents. For each offspring, the marker information on each chromosome is used to identify possible configurations of chromosomes inherited from the two parents and the locations of crossovers on these chromosomes. A branch and bound algorithm is used to identify configurations with the minimum number of crossovers. From these configurations, the conditional probability of each possible QTL genotype for a series of positions along the chromosome can be estimated. An iterative weighted regression is then used to relate the trait values to the QTL genotype probabilities. A simulation study is performed to assess this approach and to investigate the effects of the proportion of codominant to dominant markers, the heritability, and the population size. We conclude that the method successfully locates QTL and estimates their parameters accurately, and we discuss different modes of action of the QTL that may be modeled.

L INKAGE analysis and quantitative trait loci (QTL) mapping methods are now well established and widely used for diploid plant species, and there is an increasing interest in extending these methods to autopolyploid species, despite the complications of polysomic inheritance. Linkage maps have been calculated for autotetraploid potato (MEYER *et al.* 1998), autotetraploid alfalfa (BROUWER and OSBORN 1999; DIWAN *et al.* 2000), autohexaploid sweet potato (UKOSIT and THOMPSON 1997), and autooctaploid sugarcane (AL-JANABI *et al.* 1993; DA SILVA *et al.* 1993, 1995; RIPOL *et al.* 1999). Most of these studies used dominant markers or individual alleles from codominant markers that are present in one parent and segregate in a 1:1 ratio in the mapping population (simplex markers). Methods for calculating recombination frequencies between simplex markers were discussed by WU *et al.* (1992).

Unless a mapping population is very large, it is difficult to detect repulsion linkages between simplex markers in polyploids. YU and PAULS (1993) and HACKETT *et al.* (1998) examined the use of double-dose (duplex) markers as well as simplex markers to identify homologous chromosomes in autotetraploid populations, and MEYER *et al.* (1998) used this approach in their map of potato. DA SILVA (1993) and RIPOL *et al.* (1999) gave similar theoretical results for octaploid species and used this to obtain a map of sugarcane. Recently LUO *et al.* (2001) showed how to calculate recombination frequencies and LOD scores for all possible configurations of

codominant multiallelic markers in autotetraploid species and how this information may be used to construct a linkage map. They showed that codominant markers are, in general, more powerful for detecting linkage and give more precise estimates of the recombination fraction.

An important use of linkage maps is to locate major genes and QTL for important traits. Early studies of diploid species compared trait means for different phenotypes at a single marker using regression models, and some authors used the same approach in polyploid species. SILLS *et al.* (1995) used regression models to establish associations between four stalk traits in sugarcane and a set of simplex markers. A similar approach was used by MEYER *et al.* (1998) and BRADSHAW *et al.* (1998) to locate QTL in potato for quantitative resistance to late blight [*Phytophthora infestans* (Mont.) De Bary] and the white potato cyst nematode [*Globodera pallida* (Stone)], respectively. These authors used both simplex and duplex markers and used permutation tests (CHURCHILL and DOERGE 1994) to establish an appropriate significance level for testing multiple markers. In an autotetraploid cross, however, there could be up to eight different QTL alleles and the regression approach will give little insight into their individual effects and interactions. We also want to test for the presence of a QTL at locations between markers.

In this article we present an approach for QTL interval mapping in autotetraploid species in a full-sib family derived by crossing two parents. As for similar populations derived from outbreeding diploid parents, we use all the markers on a chromosome to estimate conditional probabilities for QTL genotypes. The trait values are then related to the QTL genotypes, using a mixture

Corresponding author: Christine Anne Hackett, Biomathematics and Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland. E-mail: christine@bioss.ac.uk

model. We present a simulation study to look at effects of the proportion of codominant to dominant markers, the heritability, and the population size on the ability of the model to locate QTL and to estimate their effects.

METHODS

The mapping population: The QTL mapping approach is developed for an F_1 population derived by crossing two parents, P_1 and P_2. The phenotypes of m molecular markers are assumed to be known for the parents and n offspring, and trait data are available for the parents and offspring. The parents can have up to eight distinct alleles at each marker or quantitative trait locus: These are represented by A–H or O for a "null" allele (CALLEN *et al.* 1993). In practice, most loci will have less than eight alleles, and the parental phenotypes will generally be compatible with more than one genotype. LUO *et al.* (2000) have shown how the probabilities of possible parental marker genotypes may be inferred from the parental and offspring phenotypes using Bayes' theorem.

Tetrasomic inheritance: The model for QTL mapping is developed by assuming random chromosomal segregation. The four homologous chromosomes are assumed to pair at random to give two bivalents, and crossing over is assumed to be restricted to within each bivalent. All bivalent pairings are assumed to be equally likely. The possibilities of nonrandom chromosomal pairing, or multivalent formation, are not considered here. We assume that there is no chromatid and no crossover interference. LUO *et al.* (2001) have shown how recombination frequencies and LOD scores between pairs of markers can be estimated for all possible phases, assuming random chromosomal segregation, and how the most likely phase may be identified. Using this information, markers can be partitioned into linkage groups, ordered, the distances between them calculated, and the phases of the markers in the parents can be deduced. We assume that this analysis has been carried out, so that the map of molecular markers is known.

Model for a quantitative trait: KEMPTHORNE (1957) discussed the partitioning of the genetical variance of polyploid individuals in a random mating population at equilibrium. He expressed the genotypic value Y_G of a tetraploid individual with genotype $A_i A_j A_k A_l$ as

$$Y_G = \mu + \alpha_i + \alpha_j + \alpha_k + \alpha_l + \beta_{ij} + \beta_{ik} + \beta_{il}$$
$$+ \beta_{jk} + \beta_{jl} + \beta_{kl} + \gamma_{ijk} + \gamma_{ijl} + \gamma_{ikl} + \gamma_{jkl} + \delta_{ijkl},$$
$$(1)$$

where μ is the population mean, the α's are the main effects of the alleles (analogous to additive effects in a diploid population), and β's, γ's, and δ's are the diallelic, triallelic, and tetraallelic interactions, respectively. The β's are analogous to dominance deviations in a diploid population, but there is no diploid analogy to

the triallelic and tetraallelic interactions. APPENDIX A compares the notation of model (1) with that used by various authors when considering two alleles at a tetraploid locus. For our full-sib family model, an individual will inherit alleles A_i and A_j, $1 \leq i \leq 3$, $i \leq j \leq 4$ from parent 1 and A_k and A_l, $5 \leq k \leq 7$, $k \leq l \leq 8$ from parent 2. There are therefore eight main effects, 28 diallelic interactions, 24 triallelic interactions, and 36 tetraallelic interactions in model (1), totaling more than the 36 genotypes available for model fitting. We rewrite model (1) for a full-sib family with indicator variables $X_i = 1/0$ corresponding to allele A_i present/absent for that individual. Model (1) becomes

$$Y_G = \mu + \sum_{i=1}^{8}\alpha_i X_i + \sum_{i=1}^{7}\sum_{j=i+1}^{8}\beta_{ij}X_i X_j + \sum_{i=1}^{3}\sum_{j=i+1}^{4}\sum_{k=5}^{8}\gamma_{ijk}X_i X_j X_k$$
$$+ \sum_{i=1}^{4}\sum_{k=5}^{7}\sum_{l=k+1}^{8}\gamma_{ikl}X_i X_k X_l + \sum_{i=1}^{3}\sum_{j=i+1}^{4}\sum_{k=5}^{7}\sum_{l=k+1}^{8}\delta_{ijkl}X_i X_j X_k X_l.$$
$$(2)$$

There is intrinsic aliasing of $\{X_i\}$ and higher products, so that some of the parameters are nonestimable. As each individual inherits precisely two alleles from each parent, we have the constraints

$$\sum_{i=1}^{4} X_i = 2, \quad \sum_{i=5}^{8} X_i = 2. \quad (3)$$

Substituting these into a model with main effects only gives

$$Y_G = \mu + 2\alpha_1 + 2\alpha_5 + \sum_{i=2}^{4}(\alpha_i - \alpha_1)X_i + \sum_{i=6}^{8}(\alpha_i - \alpha_5)X_i.$$
$$(4)$$

The estimable parameters are $(\mu + 2\alpha_1 + 2\alpha_5)$, $\alpha_2 - \alpha_1$, etc. To estimate the individual parameters, we impose the constraints $\alpha_1 = 0$, $\alpha_5 = 0$, sometimes referred to as cornerpoint constraints.

There are further constraints on the higher-order terms of Equation 2, for example,

$$X_2 + X_3 + X_4 - X_2 X_3 - X_2 X_4 - X_3 X_4 = 1, \quad (5)$$

and similarly for the other parent. In total, it is possible to fit six main effects, 13 biallelic interactions (2 for interactions between alleles from parent 1, 2 for parent 2, and 9 for interactions between alleles from different parents), 12 triallelic interactions, and 4 tetraallelic interactions. These total 35 effects, equal to the degrees of freedom among the 36 genotype means.

When fitting models in the simulation study, we concentrate on the main effects model (4), with cornerpoint constraints on the parameter estimates. However, there is no theoretical difficulty in including higher-order terms.

In practice, the offspring QTL genotypes are unknown, and the conditional probabilities of QTL genotypes must be estimated from the marker information. **A mixture model for QTL mapping:** Here we develop

a maximum-likelihood approach for fitting a single QTL model, considering one chromosome at a time. The analysis is an extension of that used by Jansen (1992). We assume that we have a population of n offspring and that for offspring i we observe trait value y_i and marker phenotype data o_i for the chromosome. Let $q_i \in Q_i$ be the set of possible QTL genotypes, and let $g_i \in G_i$ be the set of chromosome configurations that are compatible with phenotype o_i. By "chromosome configurations" we mean the marker genotypes and the parental chromosomes from which the marker alleles come, so that it is clear how the bivalents occurred to form that offspring and where recombinations occurred. We adopt an interval mapping approach, fitting a QTL at a set of locations along the chromosome and maximizing the likelihood for each location as a function of the QTL parameters $\theta = (\mu, \alpha_2, \alpha_3, \alpha_4, \alpha_6, \alpha_7, \alpha_8, \sigma^2)$, where μ and α_i are as in Equation 4 and σ^2 is the residual variance.

The likelihood of the trait and marker data is

$$L(\theta) = \prod_{i=1}^{n} f(y_i, o_i). \tag{6}$$

Now,

$$
\begin{aligned}
f(y_i, o_i) &= \sum_{g_i \in G_i} \sum_{q_i \in Q_i} f(y_i, o_i, g_i, q_i) \\
&= \sum_{g_i \in G_i} \sum_{q_i \in Q_i} f(y_i | o_i, g_i, q_i) P(o_i, g_i, q_i) \\
&= \sum_{g_i \in G_i} \sum_{q_i \in Q_i} f(y_i | q_i) P(q_i | g_i) P(g_i | o_i) P(o_i), \quad (7)
\end{aligned}
$$

assuming conditional independence of (i) the trait value and the marker data given the QTL genotype and (ii) the QTL genotype and the marker phenotype given the chromosome configuration.

We can maximize the log-likelihood by

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \theta} \ln L(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln P(o_i) + \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln f(y_i | o_i) \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln P(o_i) + \sum_{i=1}^{n} \frac{1}{f(y_i | o_i)} \frac{\partial}{\partial \theta} f(y_i | o_i). \quad (8)
\end{aligned}
$$

The first term on the right-hand side of (8) does not depend on the QTL parameters θ, and the second term may be written as

$$
\begin{aligned}
\sum_{i=1}^{n} \frac{1}{f(y_i | o_i)} \frac{\partial}{\partial \theta} f(y_i | o_i) &= \sum_{i=1}^{n} \frac{1}{f(y_i | o_i)} \frac{\partial}{\partial \theta} \sum_{g_i \in G_i} \sum_{q_i \in Q_i} f(y_i | q_i) P(q_i | g_i) P(g_i | o_i) \\
&= \sum_{i=1}^{n} \sum_{g_i \in G_i} \sum_{q_i \in Q_i} \frac{f(y_i | q_i) P(q_i | g_i) P(g_i | o_i)}{f(y_i | o_i)} \frac{\partial}{\partial \theta} \ln(f(y_i | q_i) P(q_i | g_i) P(g_i | o_i)) \\
&= \sum_{i=1}^{n} \sum_{g_i \in G_i} \sum_{q_i \in Q_i} P(q_i | g_i, y_i) P(g_i | o_i) \frac{\partial}{\partial \theta} \ln(f(y_i | q_i) P(q_i | g_i) P(g_i | o_i)) \\
&= \sum_{i=1}^{n} \sum_{g_i \in G_i} \sum_{q_i \in Q_i} P(q_i | g_i, y_i) P(g_i | o_i) \frac{\partial}{\partial \theta} \ln P(g_i | o_i) \\
&+ \sum_{i=1}^{n} \sum_{g_i \in G_i} \sum_{q_i \in Q_i} P(q_i | g_i, y_i) P(g_i | o_i) \frac{\partial}{\partial \theta} \ln P(q_i | g_i) \\
&+ \sum_{i=1}^{n} \sum_{g_i \in G_i} \sum_{q_i \in Q_i} P(q_i | g_i, y_i) P(g_i | o_i) \frac{\partial}{\partial \theta} \ln f(y_i | q_i). \quad (9)
\end{aligned}
$$

Only the third term of the final sum depends on θ and so contributes to the likelihood equation. The term represents a regression of the trait values on the QTL genotypes, weighted by the conditional probability of each QTL genotype. Jansen (1992, 1996) discussed how this may be solved by the expectation-maximization (EM) algorithm. The QTL genotype probabilities can be expressed as

$$P(q_i | y_i, g_i) P(g_i | o_i) = \frac{f(y_i | q_i) P(q_i | g_i) P(g_i | o_i)}{\sum_{q_j \in Q_i} f(y_i | q_j) P(q_j | g_i)}. \tag{10}$$

We can therefore use $P(q_i | g_i) P(g_i | o_i)$ as initial weights, maximize the likelihood conditional on these, update the QTL genotype probabilities using Equation 10, and repeat until the log-likelihood converges to a maximum. An alternative approach is to calculate the QTL genotype probabilities from the marker data only, regardless of the trait values. In this case we would use a single weighted regression with weights $P(q_i | g_i) P(g_i | o_i)$ rather than an iterative process to update the QTL genotype probabilities. We compare both approaches here.

Estimation of QTL genotype probabilities: In populations such as doubled haploids from inbred diploid lines, every marker is fully informative, and the conditional probability of a QTL genotype depends only on the genotypes of the markers flanking the possible QTL location. When markers are not fully informative, we need to consider the information from all markers on a chromosome to calculate the conditional probabilities of QTL genotypes. The QTL genotype probabilities $P(q_i | g_i) P(g_i | o_i)$ factor into two terms, and we consider them separately.

The conditional probability of the chromosome configuration, given the marker phenotypes, $P(g_i | o_i)$: One way to calculate these probabilities would be to use a hidden Markov model (HMM). These have been widely used for multipoint mapping in diploid species (*e.g.*, Lander and Green 1987; Jiang and Zeng 1997) and Xie and Xu (2000) tried to apply them to mapping in tetraploid species under random chromosomal segregation. There are theoretical problems with the latter's approach, however, due to the form used for the matrix of transition probabilities between neighboring markers, and as a consequence their HMM represents multivalent formations (Hackett 2001).

We preferred to use an alternative method via a "branch and bound" algorithm, where we search for chromosome configurations for each offspring that are compatible with the marker phenotype information, arise from a possible bivalent pairing in each parent, and have the minimum number of crossovers. appendix b describes the process of reconstruction for one individual from a cross between the two parents shown in Table 1 and Figure 1 illustrates the eight possible chromosome configurations for this individual that are compatible with the phenotypic information and have the minimum number of crossovers (six).

TABLE 1

Typical parental genotypes using codominant markers (simulations A, B, and E)

Locus	Position	Chromosomes:	Parent 1				Parent 2			
			1	2	3	4	5	6	7	8
L_1	0.0		C	B	O	D	E	A	D	C
L_2	10.7		C	D	D	A	C	D	A	O
L_3	14.2		E	E	D	D	C	O	D	A
L_4	18.9		D	C	D	O	E	O	C	E
L_5	30.8		C	D	C	O	A	E	D	D
L_6	41.4		D	A	D	C	D	B	C	E
L_7	50.2		A	O	C	B	E	O	B	B
L_8	63.5		B	E	E	A	A	O	A	O
L_9	74.5		D	E	E	O	B	C	B	B
L_{10}	84.7		B	E	B	E	E	O	C	O

If there are m marker loci, the recombination frequency between loci i and $i + 1$ is r_i, and there are x_i recombinations between them ($0 \leq x_i \leq 4$), then the probability of a configuration can be calculated as

$$P(g_i|o_i) = \prod_{i=1}^{m-1} r_i^{x_i}(1 - r_i)^{1-x_i}. \qquad (11)$$

This assumes that there is no interference and that recombinations occur independently between different pairs of loci. As all the configurations have the same number of crossovers, the probabilities of the different configurations

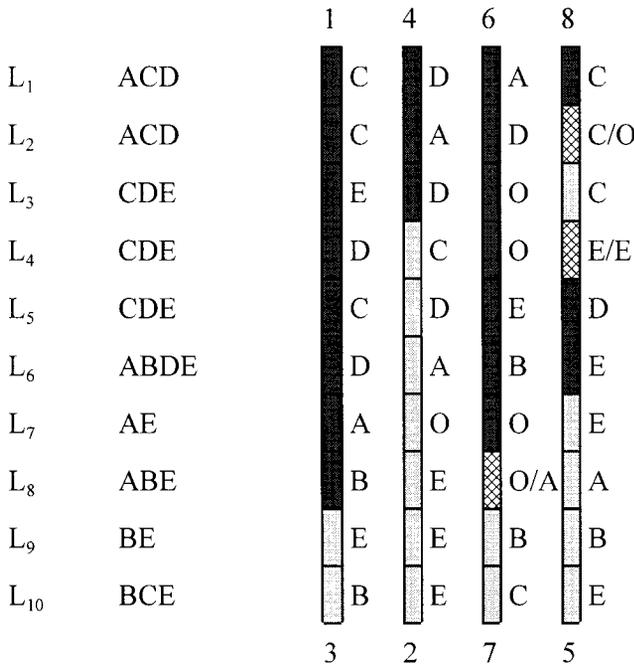|		1	4	6	8
L_1	ACD	C	D	A	C
L_2	ACD	C	A	D	C/O
L_3	CDE	E	D	O	C
L_4	CDE	D	C	O	E/E
L_5	CDE	C	D	E	D
L_6	ABDE	D	A	B	E
L_7	AE	A	O	O	E
L_8	ABE	B	E	O/A	A
L_9	BE	E	E	B	B
L_{10}	BCE	B	E	C	E
		3	2	7	5

FIGURE 1.—The graphical genotype of an individual from the parents shown in Table 1. Chromosomes 1–4 are from parent P_1, and chromosomes 5–8 are from parent P_2. The dark and light shadings distinguish the two chromosomes of each bivalent. Hatched areas show where the location of a recombination is uncertain. There are three regions of uncertainty, each with two possible alleles, giving $2^3 = 8$ configurations for this individual.

will be similar unless the marker spacing is very uneven. There is, of course, an approximation here as we are ignoring configurations that have more than the minimum number of crossovers compatible with the phenotypes. However, configurations with more crossovers will have smaller probabilities by Equation 11.

The probability of the QTL genotype, given the chromosome configuration $P(q_i|g_i)$: Once we have calculated a set of possible chromosome configurations for each individual, we can identify possible QTL genotypes and calculate their probability for putative QTL locations at a set of positions along the chromosome. We assume that there are no double crossovers between markers. The individual illustrated in Figure 1, for example, has inherited chromosomes 1, 2, 6, and 8 at marker loci L_5 and L_6 for all configurations. Therefore, for QTL locations between L_5 and L_6, we assume that the QTL genotype is $Q_1Q_2Q_6Q_8$, with probability 1. This individual has also inherited chromosomes 1, 2, 6, and 5 at locus L_7, with a crossover between chromosomes 5 and 8. For QTL locations between L_6 and L_7, there are two possible QTL genotypes, $Q_1Q_2Q_6Q_8$ and $Q_1Q_2Q_6Q_5$. The probability of the former genotype will decrease, and the probability of the latter will increase, as we consider locations at an increasing distance from L_6 and closer to L_7. To calculate these probabilities, we assume that crossovers follow a Poisson process with the probability of no crossovers in an interval of M morgans equal to e^{-M} and the probability of one crossover as Me^{-M}. If the positions of L_6 and L_7 are m_6 and m_7 and we want to calculate the probability associated with a QTL at position m_Q between them,

$P(Q_1Q_2Q_6Q_8|$flanking markers from chromosomes 1268, 1265)

$$= \frac{P(\text{no crossover between } m_6 \text{ and } m_Q; \text{ crossover between } m_Q \text{ and } m_7)}{P(\text{crossover between } m_6 \text{ and } m_7)}$$

$$= \frac{e^{-(m_Q-m_6)}(m_7 - m_Q)e^{-(m_7-m_Q)}}{(m_7 - m_6)e^{-(m_7-m_6)}}$$

$$= \frac{m_7 - m_Q}{m_7 - m_6}. \qquad (12)$$

For many positions the set of possible QTL genotypes

will depend upon the configuration; *e.g.*, for positions between loci L_8 and L_9 (configuration 3275), there will be two QTL genotypes ($Q_1Q_2Q_7Q_5$ and $Q_3Q_2Q_7Q_5$) for configurations where the configuration at L_8 is 1275 and four QTL genotypes ($Q_1Q_2Q_6Q_5$, $Q_1Q_2Q_7Q_5$, $Q_3Q_2Q_6Q_5$, and $Q_3Q_2Q_7Q_5$) for configurations where the configuration at L_8 is 1265. Crossovers between chromosomes 1 and 3 occur independently of crossovers between chromosomes 6 and 7, and so we have probabilities

$P(Q_1Q_2Q_6Q_5|\text{flanking markers from chromosomes 1265, 3275})$

$$= \frac{P(\text{no crossover between } m_8 \text{ and } m_Q; \text{ 2 crossovers between } m_Q \text{ and } m_9)}{P(\text{2 crossovers between } m_8 \text{ and } m_9)}$$

$$= \left(\frac{m_9 - m_Q}{m_9 - m_8}\right)^2. \tag{13}$$

Similarly,

$P(Q_1Q_2Q_7Q_5|\text{flanking markers from chromosomes 1265, 3275})$

$$= \frac{P(\text{1 crossover between } m_8 \text{ and } m_Q; \text{ 1 crossover between } m_Q \text{ and } m_9)}{P(\text{2 crossovers between } m_8 \text{ and } m_9)}$$

$$= \frac{(m_Q - m_8)(m_9 - m_Q)}{(m_9 - m_8)^2} \tag{14}$$

$$= P(Q_3Q_2Q_6Q_5|\text{flanking markers from chromosomes 1265, 3275})$$

and

$P(Q_3Q_2Q_7Q_5|\text{flanking markers from chromosomes 1265, 3275})$

$$= \frac{P(\text{2 crossovers between } m_8 \text{ and } m_Q; \text{ no crossover between } m_Q \text{ and } m_9)}{P(\text{2 crossovers between } m_8 \text{ and } m_9)}$$

$$= \left(\frac{m_Q - m_8}{m_9 - m_8}\right)^2. \tag{15}$$

From these formulas, we can calculate QTL genotype probabilities from Equation 10 and use them for a weighted regression to test for the significance of a QTL at position m_Q. A plot of the adjusted coefficient of determination R_a^2 against the position m_Q should show a maximum at the true QTL location.

Simulation study: A simulation study was carried out to investigate this approach for QTL mapping and to quantify the effects of the marker type, the trait heritability, and the population size. Two parents were simulated initially and these were crossed to give a population of 200 offspring by random chromosomal segregation. The first simulation consisted of one chromosome with 10 codominant markers, spaced at 10-cM intervals. There were five possible alleles (A–E) and a null allele (O) at each marker locus, with equal probability. The parental marker genotypes were simulated by sampling four alleles for each parent from the set of possible alleles, with replacement. One such parental configuration is shown in Table 1. A QTL was assumed to be situated halfway between markers L_2 and L_3 and to have eight different alleles. The QTL alleles were assumed to have additive effects, of sizes 0, 1, 1, 2, 0, -1, -1, and -2 for alleles Q_1–Q_8, respectively. The trait values for each offspring were calculated as an overall mean

of 10.0, plus the sum of the effects of their alleles, plus an environmental effect distributed as $N(0, \sigma^2)$. The value of σ^2 was chosen to give the desired trait heritability $h^2 = \sigma_G^2/(\sigma_G^2 + \sigma^2)$, where σ_G^2 is the genetic variance: $\sigma^2 = 4.0, 12.0$ correspond to heritabilities of 25 and 10%, respectively. The true QTL genotype of each individual was known, so that it was possible to compare parameter estimates from an unweighted regression on the true QTL genotype with the interval mapping approach of weighted regression on the possible QTL genotypes.

The above simulation has three random stages: simulation of the parents, simulation of the offspring given the parents, and simulation of the environmental error to add to the genotype values; and the study can be replicated at each stage. To see the effect of each level of replication, 10 pairs of parents were simulated, 20 sets of 200 offspring were simulated for each set of parents, and 20 sets of environmental error were simulated for each set of offspring, giving a total of 4000 sets of marker and trait data for analysis. Simulations A and B had heritabilities of 25 and 10%, respectively.

Most experimental data sets will be a mixture of codominant and dominant marker types, and in general the dominant markers are less informative. A further set of simulations (C) was generated to investigate this. The dominant markers were simulated as a mixture of simplex markers (AOOO × OOOO), duplex markers (AAOO × OOOO), and double-simplex markers (AOOO × AOOO), in the proportion found in potato by MEYER *et al.* (1998). Other configurations of dominant markers were excluded. Dominant and codominant markers were ordered so that there were two loci with dominant alleles present in each parent between the codominant loci. A typical parental configuration is shown in Table 2. For this configuration, loci L_1, L_5, L_9, and L_{14} are codominant. Between L_1 and L_5 there are three dominant markers (one double simplex, one duplex originating from parent 1, and one simplex originating from parent 2). Between L_5 and L_9 there are another three dominant markers (one double simplex and two duplex, one from each parent) and between L_9 and L_{14} there are four simplex markers, two from each parent. A QTL was positioned halfway between the first and second codominant markers for each configuration. A heritability of 10% was used for set C. A further set D was run to investigate the effect of reducing the population size from 200 to 100 offspring. Set D was otherwise the same as C.

We need a threshold for R_a^2, above which we declare a QTL present. For a single regression on a known QTL genotype in a population of size n, the threshold for declaring significance at a 5% level is the 95% point of an F distribution with 6 and $(n - 7)$ d.f. (Six QTL effects are fitted, rather than eight, due to the constraints discussed in Equation 4.) The F-statistic is related to R_a^2 by

$$R_a^2 = \frac{\nu_1(F - 1)}{\nu_1 F + \nu_2}, \tag{16}$$

TABLE 2

Typical parental genotypes using codominant and dominant markers (simulations C, D, and F)

Locus	Position	Chromosomes:	Parent 1				Parent 2			
			1	2	3	4	5	6	7	8
L_1	0.0		D	D	E	B	C	B	A	C
L_2	11.5		A	O	O	O	O	O	O	A
L_3	16.9		O	A	O	A	O	O	O	O
L_4	23.6		O	O	O	O	A	O	O	O
L_5	27.5		O	A	B	B	C	D	O	B
L_6	37.5		O	O	O	O	O	A	A	O
L_7	45.2		O	A	O	O	O	O	A	O
L_8	53.9		O	O	A	A	O	O	O	O
L_9	62.9		A	D	O	A	O	A	O	E
L_{10}	71.5		O	O	O	O	O	A	O	O
L_{11}	80.5		O	O	A	O	O	O	O	O
L_{12}	84.8		O	O	O	O	O	O	A	O
L_{13}	91.2		A	O	O	O	O	O	O	O
L_{14}	94.9		B	C	D	C	D	A	A	A

where ν_1 and ν_2 are the numerator and denominator degrees of freedom. For $n = 200$, the 95% point for the F-statistic is 2.146, corresponding to $R_a^2 = 3.3\%$, and for $n = 100$, the 95% point for the F-statistic is 2.198, corresponding to $R_a^2 = 6.8\%$. In QTL interval mapping, however, we consider the location with the maximum R_a^2 for a large number of linked positions and this is difficult to establish theoretically. Instead simulation sets E and F were run, with a similar pattern to simulations A and D, but the QTL effects were all set to zero. In this way, we can see how large a value of R_a^2 can be achieved by random variation and establish the distribution of R_a^2 under the null hypothesis of no QTL on the chromosome. If the observed peak of R_a^2 along a profile exceeds the 95% point of the distribution under the null hypothesis, we declare a QTL present with significance $p < 0.05$.

Computing: All routines were written in Fortran 90 (DIGITAL 1997). The branch and bound algorithm was adapted from a branch and bound algorithm for best subset selection by ROBERTS (1984).

RESULTS

Reconstruction of the chromosome configurations: The reconstruction of the chromosome configurations depends only on the marker data and is not affected by the heritability of the trait. The computer program to calculate the possible configurations had an upper limit of 5000 for each individual so that individuals with >5000 possible configurations were excluded from further analysis. No individuals were excluded, using the codominant markers for simulations A and B. The mean numbers of individuals excluded from simulations C and D were 8.8 (SE 0.93) out of 200 individuals and 4.8 (SE 0.96) out of 100 individuals, respectively.

For the individual illustrated in Figure 1, there are eight possible configurations. These agree, and are the same as the true genotypic configuration, for 37 of the 40 alleles. For the other 3 alleles one-half of the configurations have the simulated allele. The mean proportions of alleles that are correct for every chromosome configuration are summarized in Table 3 for simulations A, B, C, and D. The proportions of correct reconstructions were significantly higher for simulations A and B, using codominant markers only, but the proportions for simulations C and D were still high (0.81).

Interval mapping: For each trait, we obtained a profile of R_a^2 at a series of positions along the chromosome. Figure 2 shows such a profile for a trait from simulation A, with codominant markers and a heritability of 25%. We took the maximum of the profile to indicate the most likely location of the QTL. Table 4 summarizes the estimates of the QTL effects, R_a^2, and the residual mean square for the different simulation studies. For each study, the mean and standard errors over 4000 data sets (10 parental combinations × 20 offspring combinations × 20 traits) are presented. The row labeled wt4 is the sum of the probabilities of the true QTL genotype for each individual in the weighted regression

TABLE 3

The proportion of alleles reconstructed correctly in all chromosome configurations

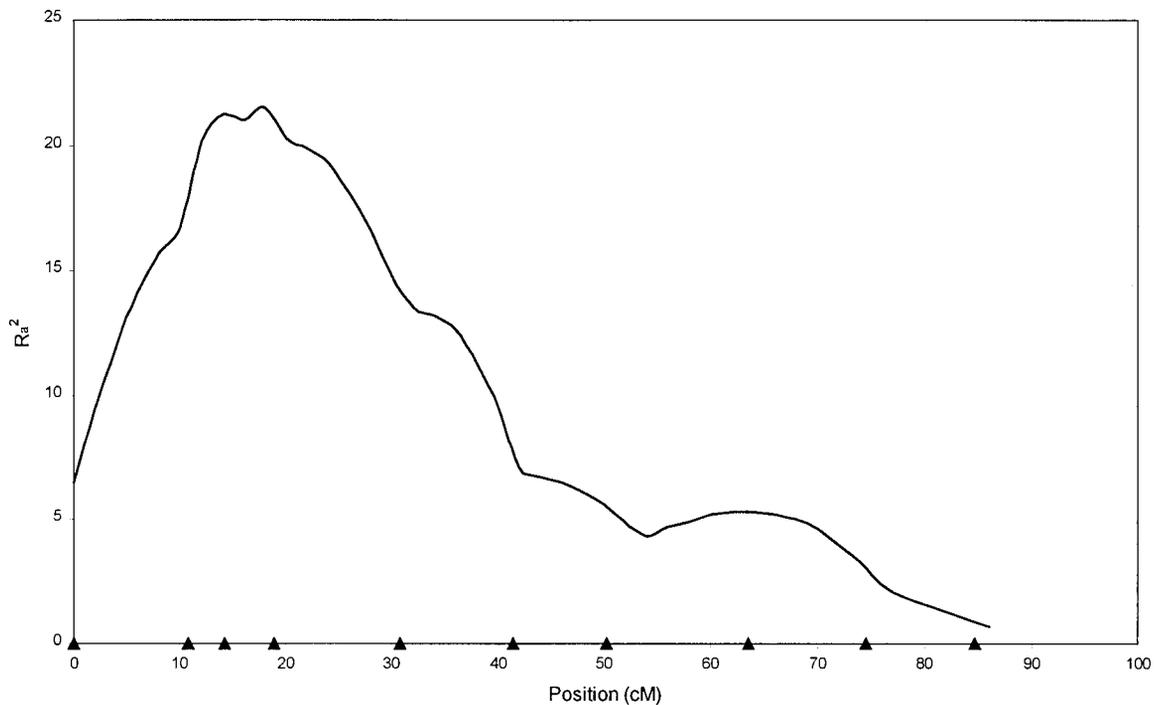| Simulation	Proportion correct	SE
A	0.88	0.0096
B	0.88	0.0093
C	0.81	0.0057
D	0.81	0.0059

FIGURE 2.—Profile of R_a^2 along the chromosome. The triangles show the positions of the marker loci.

step of the EM algorithm. If the true QTL genotype has been inferred for each individual with probability 1, then wt4 is equal to the number of individuals n. Similarly, wt3 is the sum of the probabilities of QTL genotypes that are correct for three-quarters of the alleles. The proportions of simulations indicating a QTL in the interval (5–25 cM) are shown.

Model fitting by iterative or noniterative weighted regression: As discussed above, the QTL model can be fitted at each position by a single weighted regression or by an iterative process, updating the QTL genotype probabilities using trait information as in Equation 10. The third and fourth columns of Table 4 compare the model parameters from these two methods of model fitting for simulation set A. The estimated value for the constant term was close to the true value in both cases, and the mean position, the proportion of QTL located to the region between 5 and 25 cM, and the total weight of correct and three-quarters correct QTL genotypes were very similar. For all the QTL allele effects, however, the iterative approach gave an estimate close to the true value, while the noniterative approach gave estimates whose absolute values were biased downward. The noniterative approach also underestimated the percentage variance accounted for and overestimated the residual variance. The same pattern was observed for all of the other simulation sets (results not shown). We conclude that the noniterative approach to model fitting is inadequate, and it is not considered further.

The threshold for declaring a QTL present: Simulation sets E and F (codominant markers and $n = 200$, and

dominant and codominant markers and $n = 100$, respectively, and with all QTL effects equal to zero) were used to investigate the distribution of R_a^2 when no QTL is segregating. For set E, the mean R_a^2 was 2.6 (SE 0.04), and the upper 95% point of the distribution was 6.5. We therefore took $R_a^2 = 6.5$ as the threshold above which a QTL was declared present for simulations A and B. Further simulations (results not shown) indicated that the same threshold was appropriate for simulation C, with a mixture of codominant and dominant markers. For set F, with 100 individuals, the mean R_a^2 was 5.1 (SE 0.18), and the upper 95% point of the distribution was 13.7. This is the appropriate threshold for simulation set D.

The effect of heritability: The fourth and fifth columns of Table 4 compare simulations with heritabilities of 25 and 10%, respectively. Means and standard errors were calculated over all 4000 data sets in each case. For a heritability of 25%, 3998/4000 data sets had R_a^2 greater than the threshold of 6.5. For a heritability of 10%, 3098/4000 data sets had $R_a^2 > 6.5$, indicating a significant QTL. The sixth column of Table 4 shows the means and standard deviations for these 3098 significant data sets. The standard error of the QTL location increased as the heritability decreased, although ~80% of the simulations were still found in the interval between 5 and 25 cM. The weight placed on the true QTL genotypes in the weighted regression decreased slightly for the lower heritability, and the weight on the three-quarters correct QTL genotype increased. The estimates of the QTL effects, R_a^2, and the residual mean square were

TABLE 4

Results of simulation studies

Parameter	True value	A, $h^2 = 25\%$, $n = 200$		B, $h^2 = 10\%$, $n = 200$		C, $h^2 = 10\%$, $n = 200$		D, $h^2 = 10\%$, $n = 100$	
		No iteration	With iteration	All	Significant	All	Significant	All	Significant
Constant	10	10.1 (0.04)	10.0 (0.03)	10.1 (0.05)	10.1 (0.06)	10.0 (0.09)	9.9 (0.10)	10.0 (0.10)	9.9 (0.14)
eff2	1	0.83 (0.018)	0.94 (0.015)	0.92 (0.017)	1.00 (0.017)	0.87 (0.041)	0.97 (0.036)	0.83 (0.041)	1.17 (0.059)
eff3	1	0.87 (0.020)	0.97 (0.020)	0.95 (0.024)	1.03 (0.025)	0.84 (0.059)	0.93 (0.058)	0.80 (0.055)	1.13 (0.082)
eff4	2	1.66 (0.040)	1.87 (0.028)	1.84 (0.034)	2.01 (0.036)	1.80 (0.035)	2.01 (0.036)	1.73 (0.036)	2.39 (0.045)
eff6	−1	−0.88 (0.022)	−0.96 (0.019)	−0.99 (0.015)	−1.07 (0.019)	−0.84 (0.042)	−0.94 (0.044)	−0.78 (0.045)	−1.08 (0.041)
eff7	−1	−0.88 (0.014)	−0.96 (0.012)	−0.96 (0.014)	−1.05 (0.015)	−0.84 (0.032)	−0.93 (0.031)	−0.81 (0.035)	−1.06 (0.049)
eff8	−2	−1.74 (0.031)	−1.91 (0.025)	−1.92 (0.031)	−2.08 (0.034)	−1.75 (0.067)	−1.95 (0.064)	−1.71 (0.073)	−2.33 (0.091)
Adjusted R^2	25 or 10	17.9 (0.57)	22.7 (0.44)	9.8 (0.14)	11.4 (0.10)	9.2 (0.25)	11.2 (0.15)	11.1 (0.31)	18.9 (0.14)
rms	4 or 12	4.4 (0.03)	4.1 (0.02)	12.1 (0.03)	11.9 (0.03)	12.1 (0.03)	12.0 (0.09)	11.9 (0.05)	11.0 (0.05)
wt4	—	135.2 (4.16)	136.6 (4.12)	125.3 (3.52)	127.6 (3.48)	100.8 (2.07)	102.8 (2.15)	45.7 (0.80)	49.2 (1.02)
wt3	—	54.7 (2.96)	53.7 (2.95)	59.2 (2.37)	58.4 (2.38)	69.0 (1.09)	68.6 (1.13)	35.2 (0.46)	34.6 (0.55)
Position	varies	13.0 (0.28)	13.2 (0.19)	15.9 (0.24)	15.0 (0.24)	18.5 (0.55)	17.2 (0.55)	25.8 (0.46)	21.1 (0.72)
Proportion in (5–25 cM)	—	0.95	0.95	0.78	0.82	0.66	0.70	0.47	0.57

The true values of the constant and the QTL effects are the same for all simulations. The true value of the adjusted R^2 is 25% for simulation A, 10% otherwise, and the corresponding values of the residual mean square (rms) are 4 and 12. The true QTL location is at 15 cM for simulations A and B and 17 cM for simulations C and D. Wt4 represents the sum of the weights given to the correct QTL genotype for each individual and wt3 represents the sum of the weights given to QTL genotypes that are correct for three-quarters of alleles.

still close to the true values. However, the mean QTL effects and R_a^2 for all 4000 data sets were consistently slightly lower than the true value, while they were consistently slightly higher for the 3098 significant data sets. The opposite trend was seen for the residual mean square. This is to be expected, as we are excluding the data sets with the lowest R_a^2 from the mean in the case of the significant sets.

The effect of marker type: The effect of the change from all codominant markers (simulation B) to a mixture of codominant and dominant markers (simulation C) can be seen by comparing columns five and seven (all 4000 data sets) and columns six and eight (significant data sets) of Table 4. In each case the heritability is 10%. For simulation C, 2848/4000 data sets had $R_a^2 > 6.5$, indicating a significant QTL. The proportion of QTL located in the interval (5–25 cM) fell to 0.66 (0.70 for the significant data sets) and the weight given to the correct QTL genotypes decreased. The absolute values of the QTL effects were biased downward for the mean over all 4000 data sets but less so for the significant data sets.

The effect of population size: The last four columns of Table 4 compare the effect of decreasing the population size from 200 to 100 (simulation D), for the situation of 10% heritability and a mixture of codominant and dominant markers. As discussed above, the threshold for declaring a significant QTL with a population of 100 is $R_a^2 > 13.7$, and 1233/4000 data sets had R_a^2 in this range. The absolute values of the QTL effects for a population of 100 (estimated over all 4000 data sets) were biased downward slightly more than for 200 individuals, and the standard errors were generally slightly larger. The weight given to the correct QTL genotypes decreased to below half of the population size. The mean of the QTL location changed markedly from 18.5 to 25.8 cM. However, an examination of the distribution of locations showed that the distribution was skewed. The median QTL position was 18.0 for all data sets and 16.0 for the significant data sets, closer to the true location. The estimates of the QTL effects for the significant data sets, and especially the mean R_a^2, were biased upward. This is to be expected, but is a source of potential bias in the analysis of experimental data sets.

Comparison with regression on true QTL genotype: In a simulation study such as this, the true QTL genotype is known and we can compare the parameter estimates from a regression of each trait on the true QTL genotype to that obtained by our interval mapping procedure. An examination of the estimates produced by regression on the true QTL genotypes shows that these estimates varied substantially. To illustrate this, Figure 3 shows the relationship between the estimate of R_a^2 from the interval mapping and that from regression on the true genotype for each set of simulations. The line $x = y$ is shown on each plot. There was a high correlation in each case, decreasing as the heritability, the informa-

tiveness of the markers, and the population size decreased.

DISCUSSION

In this article we proposed and tested a method for interval mapping of QTL in a full-sib population of an autotetraploid species. This method could also be extended for QTL mapping in plant species of higher ploidy. As with diploid species, the precision with which a QTL may be located is affected by the heritability of the trait, the size of the mapping population, and the informativeness of the markers. It is useful to have as high a proportion of codominant markers as possible, both for precision of QTL mapping and for linkage map construction (Luo *et al.* 2001).

The threshold at which a QTL was declared present was calculated by simulating data sets with QTL effects set to zero and examining the distribution of the adjusted coefficient of determination R_a^2. Using the 95% point of the distribution of R_a^2 gives a test of significance at a 5% level for the presence of a QTL. Using this threshold, the power to detect QTL varied considerably. For simulation set A, with codominant markers, a heritability of 25% and a population of 200 individuals, the power was >99%. This fell to 77% when the heritability was reduced to 10% (set B), to 71% when a mixture of dominant and codominant markers was used (set C), and to 31% when the population size was reduced to 100 (set D). For set D, the true value of R_a^2 is 10%. The mean value of R_a^2 from all the simulations in set D was 11.1% (SE 0.31), which is lower than the threshold for declaring significance (13.7%). The simulations for which a QTL is declared significant are those with values of R_a^2 in the upper tail of the distribution, which have mean 18.9% (SE 0.14), overestimating the true value. The effects of the QTL alleles are similarly overestimated. This problem is not confined to tetraploid analysis. Simulation studies by Beavis (1994) demonstrated the same problem in an F_2 population and showed how dramatically the variance accounted for by a QTL could be overestimated as the power to detect a QTL decreased. Utz *et al.* (2000) proposed the use of cross-validation to obtain unbiased estimates of the proportion of genotypic variance explained by a QTL. Power calculations and cross-validation could be usefully applied to analysis of experimental tetraploid populations to ensure that QTL detection and estimation are reliable.

We used a novel approach of reconstructing possible chromosome configurations from the observed marker phenotypes for each offspring. The branch and bound algorithm was used to identify configurations with the minimum number of crossovers consistent with the observed data. Configurations that did not come from bivalent pairings were rejected. This analysis was motivated by the need for QTL mapping studies in tetraploid
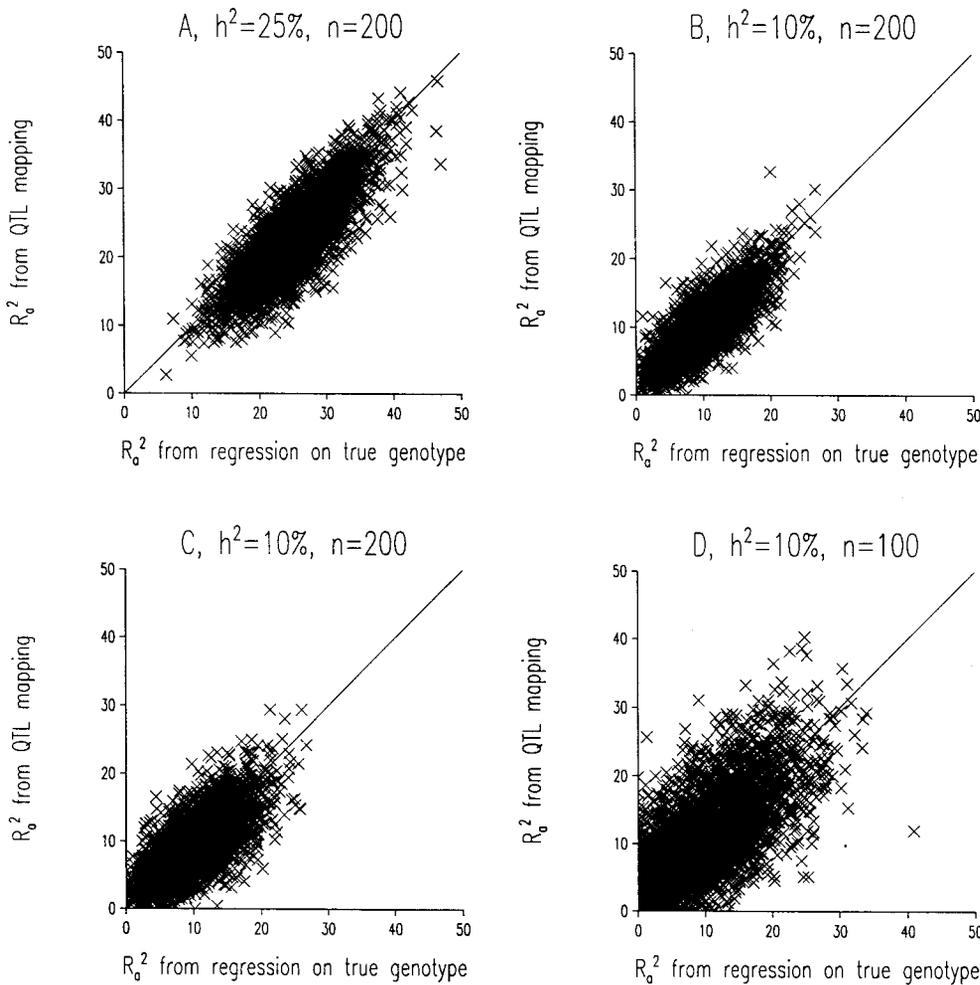
FIGURE 3.—Plots of the adjusted coefficient of determination R_a^2 from QTL mapping against R_a^2 from regression on the true QTL genotype, for each simulation.

potato. A recent ultrahigh-density diploid genetic linkage map of potato chromosome 1 found that chromatids had experienced 0, 1, or 2 recombination events during meiosis (E. ISIDORE, personal communication); that is, one or two chiasmata per chromosome pair had occurred during the meiosis. The same is likely to be true for chromosomes 2–12, given the lengths of the linkage groups (68–108 cM) found using molecular markers (VAN ECK *et al.* 1995). Hence it is not surprising that the limited cytological evidence available suggests that bivalents predominate in potato, although low frequencies of quadrivalents, trivalents, and univalents also occur (SWAMINATHAN and HOWARD 1953). As ring and chain quadrivalents can give an equal 2:2 distribution of homologues and hence balanced gametes, we must assume that departures from chromosomal segregation (multivalents and double reduction) will occur at a low frequency, which may or may not be higher for the two long chromosomes of potato (PIJNACKER and FERWERDA 1984). Hence when analyzing real data under the assumption of crossing over restricted to bivalents, some anomalous progeny may occur and need to be eliminated from the analysis and this may affect the power of QTL detection and bias the estimation of QTL

effects. However, our computer simulation, based on bivalent pairing, has shown that for some offspring multivalent configurations with fewer than the true number of recombinations could be constructed. It would be incorrect in these simulations to infer multivalent formation for such offspring, and the same problem may arise with real data. Clearly, more theoretical work and much more experimental data will be required to resolve these issues.

The accuracy with which the chromosome configurations were reconstructed depended on the type of markers used. In this study we considered codominant markers (for example, microsatellites) and dominant markers (for example, amplified fragment length polymorphisms). For the codominant markers, the dosages and configurations of alleles were obtained by random sampling with replacement from a maximum of five alleles and a null allele. The proportion of alleles reconstructed correctly was lower for a mixture of dominant and codominant markers than for codominant markers alone. The codominant markers with most alleles, and in particular those with most alleles in simplex configurations, gave offspring phenotypes that could occur in the fewest ways. Such markers are therefore the most useful for

chromosome reconstruction. Useful marker information could also be obtained by pyrosequencing single nucleotide polymorphisms to measure dosages of alleles for each offspring, which should be more informative in chromosome reconstruction than presence/absence data.

This simulation study, and in particular the chromosome reconstructions, assumed that the marker order was known without error. This may not be the situation for experimental data, and the reconstruction method could also be used to check and improve the locus ordering. The current strategy (LUO *et al.* 2001) is to calculate the recombination frequency and LOD score for all possible phases for each pair of markers on a linkage group and to use the recombination frequency and LOD score of the most likely phase for assembling the map of that group. The final step uses the module JMMAP of JoinMap (STAM and VAN OOIJEN 1995). This strategy was tested for a map of simplex and duplex dominant markers by HACKETT *et al.* (1998) and was found to give locus orders with high rank correlations with the true order for populations of 150 or more offspring. Once the offspring genotypes have been reconstructed on the basis of the order from JoinMap, the order could be investigated in two ways:

i. Drop each marker in the linkage group in turn and calculate the total number of crossovers for all the offspring. If the omission of any marker reduces the number of crossovers markedly compared to the order for the full group, then try other positions for this marker and reposition it where the total number of crossovers is lowest. Repeat for other markers if necessary.

ii. Examine the distribution of the crossovers for all the offspring and identify individuals with large numbers of crossovers. The marker data corresponding to the crossovers should be checked and corrected where necessary.

Theoretically, we could also use a computer-intensive search method to search directly for the locus order and parental phases that minimize the total number of crossovers in the offspring. However, there are a very large number of possible orders for a tetraploid cross [$m!/2$ orders for m loci, and up to $(4!)^2$ possible phases at each locus], so it is preferable to use pairwise information to reduce the search space as far as possible. There is a need for further research here.

This analysis was restricted to the case of additive effects of the QTL alleles. However, this model may be too simple. For example, many traits in potato display specific as well as general combining ability (BRADSHAW and MACKAY 1994). Furthermore, BARNES and HANSON (1967) postulated that downy mildew resistance in alfalfa is controlled by a gene that confers resistance only when in a triplex or quadruplex state. In theory, there is no problem in extending the weighted regression

model to fit separate means for each of the 36 QTL genotypes but a large population would be required to obtain good estimates of the means. One possible strategy is to use our additive model to locate regions of the genome associated with the trait. The mean trait value for each QTL genotype could then be calculated from the conditional probabilities of the genotypes for putative QTL locations in this region, to examine how they change. A range of models for the gene action (additive effects, simple models for dominance, etc.) could be tested to identify which are compatible with the means. If the additive model is rejected in favor of an alternative model, the interval mapping process should be repeated to see if the most likely QTL region changes. This strategy will be used to explore experimental data from autotetraploid potatoes at the Scottish Crop Research Institute. Computer routines may be obtained by contacting the first author.

LITERATURE CITED

AL-JANABI, S. M., R. J. HONEYCUTT, M. MCCLELLAND and B. W. S. SOBRAL, 1993 A genetic linkage map of *Saccharum spontaneum* L. 'SES 208'. Genetics **134:** 1249–1260.

BARNES, D. K., and C. H. HANSON, 1967 An illustrated summary of genetic traits in tetraploid and diploid alfalfa. U.S. Department of Agriculture Technical Bulletin 1370. U.S. Department of Agriculture, Washington, DC.

BEAVIS, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies, pp. 250–266 in *49th Annual Corn and Sorghum Industry Research Conference*. ASTA, Washington, DC.

BRADSHAW, J. E., 1994 Quantitative genetics theory for tetrasomic inheritance, pp. 71–99 in *Potato Genetics*, edited by J. E. BRADSHAW and G. R. MACKAY. CAB International, Wallingford, Oxon, UK.

BRADSHAW, J. E., and G. R. MACKAY, 1994 Breeding strategies for clonally propagated potatoes, pp. 467–497 in *Potato Genetics*, edited by J. E. BRADSHAW and G. R. MACKAY. CAB International, Wallingford, Oxon, UK.

BRADSHAW, J. E., C. A. HACKETT, R. C. MEYER, D. MILBOURNE, J. W. MCNICOL *et al.*, 1998 Identification of AFLP and SSR markers associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*) with a view to marker-assisted selection. Theor. Appl. Genet. **97:** 202–210.

BROUWER, D. J., and T. C. OSBORN, 1999 A molecular marker linkage map of tetraploid alfalfa (*Medicago sativa* L.). Theor. Appl. Genet. **99:** 1194–1200.

CALLEN, D. F., A. D. THOMPSON, H. A. Y. SHEN, H. A. PHILLIPS, R. I. RICHARDS *et al.*, 1993 Incidence and origin of "null" alleles in the $(AC)_n$ microsatellite markers. Am. J. Hum. Genet. **52:** 922–927.

CHURCHILL, G. A., and R. W. DOERGE, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138:** 963–971.

DA SILVA, J., 1993 A methodology for genome mapping of autopolyploids and its application to sugarcane (*Saccharum* spp.). PhD. Dissertation, Cornell University, Ithaca, NY.

DA SILVA, J. A. G., M. E. SORRELLS, W. L. BURNQUIST and S. D. TANKSLEY, 1993 RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome **36:** 782–791.

DA SILVA, J., R. J. HONEYCUTT, W. BURNQUIST, S. M. AL-JANABI, M. E. SORRELLS *et al.*, 1995 *Saccharum spontaneum* L. 'SES 208'

genetic linkage map combining RFLP- and PCR-based markers. Mol. Breed. **1:** 165–179.

DIGITAL, 1997 *DIGITAL Fortran Language Reference Manual.* Digital Equipment Corporation, Maynard, MA.

DIWAN, N., J. H. BOUTON, G. KOCHERT and P. B. CREGAN, 2000 Mapping of simple sequence repeat (SSR) DNA markers in diploid and tetraploid alfalfa. Theor. Appl. Genet. **101:** 165–172.

EASTON, H. S., 1976 Etude comparative d'effects génétique chez des plantes diploïdes et tetraploïdes isogéniques de *Festuca pratensis* Huds. Thèse de Doctorat d'Etat des Sciences Naturelles, Université de Paris-Sud, Paris, France.

HACKETT, C. A., 2001 A comment on Xie and Xu: 'mapping quantitative trait loci in tetraploid species.' Genet. Res. **78:** 187–189.

HACKETT, C. A., J. E. BRADSHAW, R. C. MEYER, J. W. MCNICOL, D. MILBOURNE et al., 1998 Linkage analysis in tetraploid species: a simulation study. Genet. Res. **71:** 143–154.

JANSEN, R. C., 1992 A general mixture model for mapping quantitative trait loci by using molecular markers. Theor. Appl. Genet. **85:** 252–260.

JANSEN, R. C., 1996 A general Monte Carlo method for mapping multiple quantitative trait loci. Genetics **142:** 305–311.

JIANG, C. J., and Z-B. ZENG, 1997 Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica **101:** 47–58.

KEMPTHORNE, O., 1957 *An Introduction to Genetic Statistics.* John Wiley & Sons, New York.

LANDER, E. S., and P. GREEN, 1987 Construction of multilocus genetic-linkage maps in humans. Proc. Natl. Acad. Sci. USA **84:** 2363–2367.

LUO, Z. W., C. A. HACKETT, J. E. BRADSHAW, J. W. MCNICOL and D. MILBOURNE, 2000 Predicting parental genotypes and gene segregation for tetrasomic inheritance. Theor. Appl. Genet. **100:** 1067–1073.

LUO, Z. W., C. A. HACKETT, J. E. BRADSHAW, J. W. MCNICOL and D. MILBOURNE, 2001 Construction of a genetic linkage map in tetraploid species using molecular markers. Genetics **157:** 1369–1385.

MEYER, R. C., D. MILBOURNE, C. A. HACKETT, J. E. BRADSHAW, J. W. MCNICOL et al., 1998 Linkage analysis in tetraploid potato and associations of markers with quantitative resistance to late blight (*Phytophthora infestans*). Mol. Gen. Genet. **259:** 150–160.

PIJNACKER, L. P., and M. A. FERWERDA, 1984 Giemsa C-banding of potato chromosomes. Can. J. Genet. Cytol. **26:** 415–419.

RIPOL, M. I., G. A. CHURCHILL, J. A. G. DA SILVA and M. SORRELLS, 1999 Statistical aspects of genetic mapping in autopolyploids. Gene **235:** 31–41.

ROBERTS, S. J., 1984 A branch and bound algorithm for determining the optimal feature subset of given size. Appl. Stat. **33:** 236–241.

SILLS, G. R., W. BRIDGES, S. M. AL-JANABI and B. W. S. SOBRAL, 1995 Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes and Jesw. ex Grassl). Mol. Breed. **1:** 355–363.

STAM, P., and J. W. VAN OOIJEN, 1995 *JoinMap Version 2.0: Software for the Calculation of Genetic Linkage Maps.* CPRO-DLO, Wageningen, The Netherlands.

SWAMINATHAN, M. S., and H. W. HOWARD, 1953 The cytology and genetics of the potato (*Solanum tuberosum*) and related species. Bibliogr. Genet. **16:** 1–192.

UKOSIT, K., and P. G. THOMPSON, 1997 Autopolyploidy versus allopolyploidy and low-density randomly amplified polymorphic DNA linkage maps of sweetpotato. J. Am. Soc. Hort. Sci. **122:** 822–828.

UTZ, H. F., A. E. MELCHINGER and C. C. SCHON, 2000 Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics **154:** 1839–1849.

VAN ECK, H. J., J. R. VAN DER VOORT, J. DRAAISTRA, P. VAN ZANDVOORT, E. VAN ENCKEVORT et al., 1995 The inheritance and chromosomal localization of AFLP markers in a noninbred potato offspring. Mol. Breed. **1:** 397–410.

WRIGHT, A. J., 1979 The use of differential coefficients in the development and interpretation of quantitative genetics models. Heredity **43:** 1–8.

WU, K. K., W. BURNQUIST, M. E. SORRELLS, T. L. TEW, P. H. MOORE et al., 1992 The detection and estimation of linkage in polyploids

using single-dose restriction fragments. Theor. Appl. Genet. **83:** 294–300.

XIE, C., and S. XU, 2000 Mapping quantitative trait loci in tetraploid populations. Genet. Res. **76:** 105–115.

YU, K. F., and K. P. PAULS, 1993 Segregation of random amplified polymorphic DNA markers and strategies for molecular mapping in tetraploid alfalfa. Genome **36:** 844–851.

Communicating editor: C. HALEY

APPENDIX A: RELATIONSHIP BETWEEN BIOMETRICAL MODELS

BRADSHAW (1994) compares three notations that have been used for the genetic values of tetraploid individuals. All these notations assume that two alleles, and hence five genotypes, are at a locus affecting the trait. Table A1 shows the notation used for each genotype by EASTON (1976) and WRIGHT (1979) and compares this to the expression derived from Equation 1 of this article. From the last column, the average of the two quadruplex genotypes is

$$m = \mu + 2\alpha_1 + 2\alpha_2 + 3\beta_{11} + 3\beta_{22} + 2\gamma_{111} + 2\gamma_{222}$$
$$+ \delta_{1111}/2 + \delta_{2222}/2.$$

We can equate the expression from EASTON (1976) to the difference between the genetic values in the last column and m to give the following correspondences:

$$d = \alpha_1 - \alpha_2 + 3(\beta_{11} - \beta_{22})/2 + \gamma_{111} - \gamma_{222} + (\delta_{1111} - \delta_{2222})/4$$

$$h = \beta_{12} - (\beta_{11} + \beta_{22})/2 + (\gamma_{112} + \gamma_{122} - \gamma_{111} - \gamma_{222})/2$$
$$+ \delta_{1122}/4 - (\delta_{1111} + \delta_{2222})/8$$

$$v = (3\gamma_{112} - 3\gamma_{122} - \gamma_{111} + \gamma_{222})/2 + (\delta_{1112} - \delta_{1222})/2 - (\delta_{1111} - \delta_{2222})/4$$

$$w = (\delta_{1112} + \delta_{1222})/2 - 3\delta_{1122}/4 - (\delta_{1111} + \delta_{2222})/8.$$

We see that d is a function of the additive effects α and higher-order terms, while h, v, and w depend on the diallelic, triallelic, and tetraallelic interactions and higher-order terms, respectively.

APPENDIX B: RECONSTRUCTION OF CHROMOSOME CONFIGURATIONS

Here we demonstrate the reconstruction of the possible chromosomes inherited by an offspring from the cross between parents P_1 and P_2, with genotypes given in Table 1. The phenotype of this offspring is shown in Table A2.

Consider locus L_1, with phenotype ACD. The A allele must have come from chromosome 6 (parent P_2). The individual does not have a B or E allele and therefore has not inherited chromosome 2 or 5 at this locus. The possibilities are (i) chromosomes 6 and 7 from P_2, together with either chromosomes 1 and 3 from P_1 (giving genotype ACDO); or (ii) chromosomes 1 and 4 from P_1 (giving genotype ACDD); or (iii) chromosomes 6 and 8 from P_2, together with either chromosomes 1 and 4 (giving genotype ACCD); or (iv) chromo-

TABLE A1

Comparison of biometrical models for the case of two alleles at a QTL

Genotype	EASTON (1976)	Current model
$A_1A_1A_1A_1$	$2d$	$\mu + 4\alpha_1 + 6\beta_{11} + 4\gamma_{111} + \delta_{1111}$
$A_1A_1A_1A_2$	$d + 3h + v + w$	$\mu + 3\alpha_1 + \alpha_2 + 3\beta_{11} + 3\beta_{12} + \gamma_{111} + 3\gamma_{112} + \delta_{1112}$
$A_1A_1A_2A_2$	$4h$	$\mu + 2\alpha_1 + 2\alpha_2 + \beta_{11} + 4\beta_{12} + \beta_{22} + 2\gamma_{112} + 2\gamma_{122} + \delta_{1122}$
$A_1A_2A_2A_2$	$-d + 3h - v + w$	$\mu + \alpha_1 + 3\alpha_2 + 3\beta_{12} + 3\beta_{22} + 3\gamma_{122} + \gamma_{222} + \delta_{1222}$
$A_2A_2A_2A_2$	$-2d$	$\mu + 4\alpha_2 + 6\beta_{22} + 4\gamma_{222} + \delta_{2222}$

The expressions given by EASTON (1976) are differences from the mean of the two quadruplex genotypes.

somes 3 and 4 (giving genotype ACDO). Table A2 shows the chromosome configurations giving rise to the phenotypes of this individual at each locus.

A branch and bound algorithm is then used to identify the chromosome configurations that give the minimum number of recombinations for the complete linkage group. An initial configuration is found by ordering the configurations according to the number of loci for which they are possible and selecting for each locus the most frequent compatible configuration. For this individual, the configurations 1257 and 2357 are jointly most frequent, each being possible for 5 of the 10 loci. Neither of these is possible for locus L_1, however. The initial order is

$$L_1 \quad 1\,4\,6\,7$$
$$L_2 \quad 1\,2\,5\,7$$
$$L_3 \quad 1\,2\,5\,7$$
$$L_4 \quad 1\,2\,5\,7$$
$$L_5 \quad 1\,4\,6\,7$$
$$L_6 \quad 3\,2\,6\,8$$
$$L_7 \quad 1\,2\,5\,6$$
$$L_8 \quad 1\,2\,5\,7$$
$$L_9 \quad 3\,2\,5\,7$$
$$L_{10} \quad 1\,2\,5\,7$$

with 12 recombinations. The algorithm searches for configurations with the minimum number of crossovers. It is not necessary to test every combination in Table A2: If a combination for, say, L_1–L_5 has more recombina-

tions than the current minimum, then this is rejected without considering L_6–L_{10}.

The minimum number of recombinations for this individual is six, and there are 20 configurations with this minimum. Up to now, the question of whether the configuration may be produced by bivalent pairing has been ignored, but now this is checked for each configuration.

One possible configuration with six recombinations is

$$L_1 \quad 1\,4\,6\,7$$
$$L_2 \quad 1\,4\,6\,7$$
$$L_3 \quad 1\,4\,6\,5$$
$$L_4 \quad 1\,2\,6\,5$$
$$L_5 \quad 1\,2\,6\,8$$
$$L_6 \quad 1\,2\,6\,8$$
$$L_7 \quad 1\,2\,6\,5$$
$$L_8 \quad 1\,2\,7\,5$$
$$L_9 \quad 3\,2\,7\,5$$
$$L_{10} \quad 3\,2\,7\,5.$$

If we consider the first four loci, these suggest that the chromosomes are paired as $1 + 3$ and $2 + 4$ from P_1 and $5 + 7$ and $6 + 8$ from P_2. However, L_5 and L_6 have chromosomes 6 and 8 from P_2, and L_8–L_{10} have chromosomes 5 and 7 together. This configuration is rejected as incompatible with bivalent pairing. However, the configuration

$$L_1 \quad 1\,4\,6\,8$$
$$L_2 \quad 1\,4\,6\,8$$
$$L_3 \quad 1\,4\,6\,5$$

TABLE A2

Possible chromosome configurations

Locus	Phenotype	Possible chromosome configurations
L_1	ACD	1367, 1467, 1468, 3468
L_2	ACD	1257, 1267, 1278, 1357, 1367, 1378, 1456, 1467, 1468, 2357, 2456, 2457, 2458, 3456, 3457, 3458
L_3	CDE	1257, 1356, 1357, 1456, 1457, 2356, 2357, 2456, 2457
L_4	CDE	1256, 1257, 1258, 1268, 1278, 1357, 1378, 1457, 1478, 2356, 2357, 2358, 2368, 2378, 3457, 3478
L_5	CDE	1267, 1268, 1367, 1368, 1467, 1468, 2367, 2368, 3467, 3468
L_6	ABDE	1268, 2368
L_7	AE	1256
L_8	ABE	1256, 1257, 1258, 1267, 1278, 1356, 1357, 1358, 1367, 1378
L_9	BE	2357, 2358, 2378, 2457, 2458, 2478, 3457, 3458, 3478
L_{10}	BCE	1257, 1267, 1278, 1357, 1457, 1467, 1478, 2357, 2367, 2378, 3457, 3467, 3478

L_4 1 2 6 5
L_5 1 2 6 8
L_6 1 2 6 8
L_7 1 2 6 5
L_8 1 2 7 5
L_9 3 2 7 5
L_{10} 3 2 7 5

is compatible throughout with the chromosomes pairing as 1 + 3, 2 + 4, 6 + 7, and 5 + 8. Of the 20 configurations with six recombinations for this individual, 8 were compatible with bivalent pairings. They can be summarized as

L_1 1 4 6 8
L_2 1 4 6 5/8
L_3 1 4 6 5
L_4 1 2 6 5/8
L_5 1 2 6 8
L_6 1 2 6 8
L_7 1 2 6 5
L_8 1 2 6/7 5
L_9 3 2 7 5
L_{10} 3 2 7 5.

This is represented as a graphical genotype in Figure

1. The 8 configurations coincide for 37 of the 40 alleles, but there is uncertainty about the other 3 alleles. For example, there is definitely a recombination between chromsomes 6 and 7 between L_7 and L_9, but it is uncertain on which side of L_8 it occurred. One of the 8 configurations is the same as the simulated genotype for this individual.

If locus L_3 was excluded from the analysis, there would be no evidence to establish the first two crossovers between chromosomes 5 and 8. Loci L_1, L_3, and L_5 have unique alleles on chromosome 5, and this individual carries the unique allele from L_3 but not from L_1 or L_5. The true recombinations were between L_1 and L_2 and between L_4 and L_5, but the phenotypes observed for L_2 and L_4 are both compatible with inheriting chromosomes 6 and 8 from parent P_2. Without L_3, the minimum recombination configuration would have four recombinations, fewer than that simulated.

Occasionally the minimum recombination configurations have fewer than the simulated number of recombinations, but are all incompatible with bivalent pairing. In this case configurations with minimum + 1, minimum + 2, etc., recombinations are considered until compatible configurations are found.