Genet. Res., Camb. (2000), **76**, *pp.* 105–115. *With 2 figures.* Printed in the United Kingdom © 2000 Cambridge University Press

Mapping quantitative trait loci in tetraploid populations

CHONGQING XIE† AND SHIZHONG XU*
Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA

(*Received 21 June 1999 and in revised form 18 August 1999*)

Summary

Knowledge of quantitative trait locus (QTL) mapping in polyploids is almost void, albeit many exquisite strategies of QTL mapping have been proposed and extensive investigations have been carried out in diploid animals and plants. In this paper we develop a simple algorithm which uses an iteratively reweighted least square method to map QTLs in tetraploid populations. The method uses information from all markers in a linkage group to infer the probability distribution of QTL genotype under the assumption of random chromosome segregation. Unlike QTL mapping in diploid species, here we estimate and test the compound 'gametic effect', which consists of the composite 'genic effect' of alleles and higher-order gene interactions. The validity and efficiency of the proposed method are investigated through simulation studies. Results show that the method can successfully locate QTLs and separates different sources (e.g. additive and dominance) of variance components contributed by the QTLs.

1. Introduction

Polyploids are very common in plants, especially in angiosperms. Many new species have been produced in connection with tetraploids and higher polyploids. The frequent occurrence of polyploidization in nature and the widespread distribution of polyploids itself suggest that polyploids play an important role in evolution of the vegetative kingdom. Perhaps 50% of all angiosperms and 44–95% of ferns and fern allies are of polyploid origin (Soltis & Soltis, 1995). Among these, some are of economic importance, such as cultivated potato, sugarcane and alfalfa. In addition, polyploid forms of plants can adapt to more extreme conditions than their diploid relatives and are thus of great practical value. Evidence from molecular data has recently revealed that the amount of genetic diversity present in polyploids is remarkably higher than that in diploids (Soltis & Soltis, 1995; Song *et al.*,). Thus, it is useful to develop linkage maps directly at the polyploid level. This will in turn provide insight into the studies of the parentage formation and genome constitution of polyploid species.

Construction of linkage maps in polyploids is more challenging than that in diploids due to the complex nature of multiple alleles in polyploids. Therefore, many studies on constructing linkage maps have focused on diploid relatives of polyploids, such as in potato and alfalfa. However, there are several reasons why mapping at high ploidy levels is necessary. First, polyploidization and subsequent evolution of polyploid genomes is an extremely dynamic process (Soltis & Soltis, 1995). Song *et al.* (1995) detected extensive changes in the nuclear genome of synthetic polyploids in each of the first five generations. Hence, the linkage maps constructed in diploid relatives are expected to differ from those for polyploids. Secondly, not all polyploids have a diploid relative available in nature and genetic analyses of these polyploids must be carried out in the polyploid form. Thirdly, many

* Corresponding author. Tel: +1 (909) 787 5898. Fax: +1 (909) 787 4437. e-mail: xu@genetics.ucr.edu
† Current address: Dr Chongqing Xie, Pioneer Hi-Bred International Inc., 7300 N.W. 62nd Avenue, Johnston, IA 50131, USA.

cultivated crops are polyploid species and the management of breeding programmes in these species is actually carried out at the polyploid level, not at the diploid level. Finally, directly mapping quantitative trait loci (QTLs) at the polyploid level allows the detection of higher-order interaction between alleles.

Recently, significant efforts have been made to develop linkage maps for some polyploid species, such as alfalfa, cultivated potato and sugarcane (Yu & Pauls, 1993; Al Janabi *et al*., 1993; Da Silva *et al*., 1993; Grivet *et al*., 1996). The mapping strategies used in these studies are based on the idea of single-dose restriction fragments as proposed by Ritter *et al*. (1990) in diploid parents and extended by Wu *et al*. (1992) to tetraploid species. Single-dose restriction fragments occur only in individuals with genotype *Mmmm*, where *M* and *m* stand for the dominant and recessive alleles, respectively. A parent with such a single dose of the dominant allele is called the simplex parent. A simplex parent can produce two kinds of gametes (*Mm* and *mm*) with a 1:1 ratio. Crossing between two simplex parents will generate progeny with a band presence to absence ratio of 3:1, equivalent to the well-known Mendelian ratio observed in the progeny of F2 derived from crossing of two inbred diploid grandparents. Therefore, mapping QTLs in tetraploids can follow existing procedures practised in diploid organisms. Yu & Pauls (1993) and Da Silva *et al*. (1993) extended this method to linkage analysis in crosses of duplex × duplex or simplex × duplex, where a duplex parent is defined as a parent with genotype *MMmm*. Hackett *et al*. (1998) found that accuracy of linkage estimates depends on the type of markers involved. The simplex–simplex coupling pairs are most reliable, whereas the simplex–simplex repulsion pairs and duplex–duplex pairs in any configuration but coupling are least reliable. The single-dose restriction fragment technique is particularly appropriate for polymerase chain reaction (PCR)-based dominant markers. However, because multiplex provide no information on segregation, these types of markers mask most of the polymorphism occurring between parents. In view of this, Milbourne *et al*. (1997) and Meyer *et al*. (1998) suggested combining high polymorphic markers, such as simple sequence repeats (SSRs) or microsatellites, with dominant markers for the construction of linkage maps in tetraploid potato.

Although co-dominant markers are expensive to genotype, they provide more information about the allelic inheritance of a gene than dominant markers. While many exquisite theories of QTL mapping have been developed and extensive investigations have been done in diploid organisms, methods of QTL mapping using co-dominant markers have been lacking in polyploids. In this study, we take a tetraploid full-sib family as an example to show that QTL mapping in polyploids is as convenient as that in diploids. Under random chromosomal segregation, we use a multipoint mapping method to infer the distribution QTL genotype and then integrate it with Kempthorne's (1955) theory for the decomposition of genetic variance. Instead of determining the 'genic effect' of allelic inheritance, we treat each gamete (the combination of two alleles within the same locus) as a segregation unit and then determine its 'gametic effect'. In other words, the genotypic value of an individual is determined entirely by the gametes that make up the genotype of the individual. Under this model, we develop an iteratively reweighted least square method of QTL mapping in an outbred full-sib family.

2. Statistical methods

(i) *Genetic model*

The mapping population consists of a single family of full-sibs derived from the cross of two outbred parents. An existing linkage map is assumed to be available for many polymorphic co-dominant markers. The marker linkage phases in the parents are assumed to be known or can be deduced from grandparents or through their progeny. Genotypic information of these markers is then used to infer the distribution of the genotype of a putative QTL linked with these markers.

Because parents are assumed to be outbred, each parent carries four different alleles at the QTL in question. These four alleles are not identical-by-descent and thus each one can be uniquely assigned a label. Let $Q_1^m Q_2^m Q_3^m Q_4^m$ and $Q_1^f Q_2^f Q_3^f Q_4^f$ be the genotypes of the male and female parents, respectively. Each parent can produce six possible gametes ($C_4^2 = 6$). The configurations of the six gametes produced by the male parent are

$$\{Q_1^m Q_2^m \;\; Q_1^m Q_3^m \;\; Q_1^m Q_4^m \;\; Q_2^m Q_3^m \;\; Q_2^m Q_4^m \;\; Q_3^m Q_4^m\}$$

and those produced by the female parents are

$$\{Q_1^f Q_2^f \;\; Q_1^f Q_3^f \;\; Q_1^f Q_4^f \;\; Q_2^f Q_3^f \;\; Q_2^f Q_4^f \;\; Q_3^f Q_4^f\}.$$

With random union between the male and female gametes, there are 36 possible genotypes in the progeny. Denote G_{ik} for $i, k = 1, \ldots, 6$ as the genotypic value of an individual with a genotype composed of the ith gamete from the male parent and the kth gamete from the female parent. For example, G_{35} is the genotypic value of the genotype composed of the 3rd gamete from the male parent and the 5th gamete from the female parent, i.e. $Q_1^m Q_4^m Q_2^f Q_4^f$. Let y_j be the

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

phenotypic value of the jth progeny, for $j = 1, \ldots, n$. The usual linear model for y_j appears:

$$y_j = \mu + \sum_{i=1}^{6}\sum_{k=1}^{6} X_{ik(j)} G_{ik} + \epsilon_j, \tag{1}$$

where μ is the mean of the population ϵ_j is the environmental error distributed as $N(0, \sigma_\epsilon^2)$, and $X_{ik(j)}$ is an indicator variable defined as

$$X_{ik(j)} = \begin{cases} 1 & \text{if } j \text{ is of genotype } G_{ik} \\ 0 & \text{if } j \text{ is not of genotype } G_{ik}. \end{cases}$$

The model is not of full rank because

$$\sum_{i=1}^{6}\sum_{k=1}^{6} X_{ik(j)} = 1$$

for all j. Therefore, the population mean μ is usually suppressed from the model, in which case the genotypic values are expressed as deviations from the mean. The genotypic value, however, is a composite term, consisting of a gametic effect from each parent and an interaction effect between the two gametes. Specifically, G_{ik} is expressed as

$$G_{ik} = \alpha_i^m + \alpha_k^f + \delta_{ik} \quad \text{for } i, k = 1, \ldots, 6, \tag{2}$$

where α_i^m and α_k^f are the effects of gamete i from the male parent and gamete k from the female parent, and δ_{ik} is the interaction effect between the two gametes. Because each gamete is made up of two alleles, the gametic effects, as α_i^m and α_k^f, are also composite terms consisting of two allelic effects and the interaction (dominance) effect between the two alleles. The interaction effect δ_{ik} also consists of higher-order gene interactions, such as trigenic and quadrigenic effects.

The genotype of QTL is not observable, and thus $X_{ik(j)}$ is missing. However, the distribution of $X_{ik(j)}$ can be inferred from linked markers. Let $p_{ik(j)} = \Pr(X_{ik(j)} = 1 \mid I_{M(j)})$ be the probability that $X_{ik(j)} = 1$ conditional on marker information $I_{M(j)}$. We have

$$E(X_{ik(j)} \mid I_{M(j)}) = p_{ik(j)},$$

$$\mathrm{Var}(X_{ik(j)} \mid I_{M(j)}) = p_{ik(j)}(1 - p_{ik(j)})$$

and

$$\mathrm{Cov}(X_{ik(j)}, X_{ik(j)} \mid I_{M(j)}) = p_{ik(j)} p_{ik(j)}.$$

method of computing $p_{ik(j)}$ will be described later.

Model (1) can be approximated by substituting $E(X_{ik(j)} \mid I_{M(j)})$ in place of $X_{ik(j)}$:

$$y_j = \sum_{i=1}^{6}\sum_{k=1}^{6} E(X_{ik(j)} \mid I_{M(j)}) G_{ik} + e_j. \tag{3}$$

Note that the residual e_j is not the same as ϵ_j in (1); instead, it is now distributed as a mixture of normal distributions with a heterogeneous variance. Let us define $\mathbf{X}_j$ as a 36×1 vector consisting of all the 36 $X_{ik(j)}$ values which are arranged in the appropriate order, and β is a 36×1 vector corresponding to the 36 G_{ik} values. Further, denote $\mathbf{U}_j = E(\mathbf{X}_j \mid I_{M(j)})$ as a vector of conditional expectations and $\Sigma_j = \mathrm{Var}(\mathbf{X}_j \mid I_{M(j)})$ as a 36×36 conditional variance–covariance matrix. We have an alternative expression for (3):

$$y_j = \mathbf{U}_j \beta + e_j.$$

Comparing this equation with (1), we can see that

$$e_j = (\mathbf{X}_j - \mathbf{U}_j)\beta + \epsilon_j. \tag{4}$$

Therefore, e_j has an expectation of zero and a variance of

$$\mathrm{Var}(e_j) = \beta^{\mathrm{T}} \Sigma_j \beta + \sigma_\epsilon^2,$$

where the first term $\beta^{\mathrm{T}} \Sigma_j \beta$ reflects the inflation of the residual error variance due to uncertainty of the QTL genotype (Xu, 1995, 1998*a*, *b*). An estimate of β can be obtained from the heterogeneous residual variance model using the iteratively reweighted least squares method described by Xu (1998*a*, *b*).

(ii) *Parameter estimation*

Under the above formulation, we first estimate the 36 genotypic values β, which have been expressed as deviations from the mean, and then convert them into the required genetic effects – a total of 48 genetic parameters. Therefore, the model is overparameterized and, as a result, we are forced to make the following constraints to the parameters:

$$\sum_i \alpha_i^m = \sum_k \alpha_k^f = \sum_i \delta_{ik} = \sum_k \delta_{ik} = 0.$$

When tracing the gametic effects of parents, the linear model and the above constraints are identical to a two-way ANOVA. After the constraints, there are 5 df left for the six gametic effects for each parent and 25 df for the interaction between the two parents. Hence, there are only 35 estimable genetic effects which are obtained using the linear contrasts of the 36 genotypic values. The 35 linear contrasts are

$$\begin{cases} \alpha_i^m = \bar{G}_{i.} - \bar{G}_{..} & \text{for } i = 1, \ldots, 5 \\ \alpha_k^f = \bar{G}_{.k} - \bar{G}_{..} & \text{for } k = 1, \ldots, 5 \\ \delta_{ik} = G_{ik} - \bar{G}_{i.} - \bar{G}_{.k} + \bar{G}_{..} & \text{for } i, k = 1, \ldots, 5, \end{cases} \tag{5}$$

where

$$\bar{G}_{i.} = \frac{1}{6}\sum_{j=1}^{6} G_{ij}, \quad \bar{G}_{.k} = \frac{1}{6}\sum_{j=1}^{6} G_{jk}$$

and

$$\bar{G}_{...} = \frac{1}{36}\sum_{ik} G_{ij} = 0.$$

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

Note that we have used the same symbols for both the linear contrasts and the original allelic effects, although they have different meanings. Treating the linear contrasts as genetic parameters and denoting the vector of these genetic parameters by

$$\theta = \{\alpha_1^m \dots \alpha_5^m, \alpha_1^f \dots \alpha_5^f, \delta_{11} \dots \delta_{55}\}$$

we can express the genetic parameters by

$$\theta = \mathbf{H}\beta = [\mathbf{H}_1//\mathbf{H}_2//\mathbf{H}_3]\beta,$$

where $\mathbf{H}$ is a 35×36 matrix containing the coefficients of the contrasts, $\mathbf{H}_1$ is a submatrix (the first five rows) of $\mathbf{H}$, $\mathbf{H}_2$ is a submatrix (rows 6–10) of $\mathbf{H}$, $\mathbf{H}_3$ is a submatrix (rows 11–35) of $\mathbf{H}$, and the symbol '//' represents vertical matrix concatenation borrowed from the SAS/IML language (SAS Institute, 1990). The pattern of the arrangement of elements in matrix $\mathbf{H}$ can be found by examining (5).

The variance of the total gametic effects, analogous to the 'additive effects' in diploids, is defined as follows:

$$\sigma_A^2 = \sigma_m^2 + \sigma_f^2 = \frac{1}{5 \times 36^2}\beta^{\mathrm{T}}(\mathbf{H}_1^{\mathrm{T}}\mathbf{H}_1 + \mathbf{H}_2^{\mathrm{T}}\mathbf{H}_2)\beta \tag{6}$$

where σ_m^2 and σ_f^2 are the paternal and maternal components of the gametic variance. The variance of interaction, analogous to the 'dominance variance' in diploid organisms, is given by

$$\sigma_D^2 = \frac{1}{25 \times 36^2}\beta^{\mathrm{T}}\mathbf{H}_3^{\mathrm{T}}\mathbf{H}_3\beta. \tag{7}$$

Alternatively, σ_D^2 can be obtained by

$$\sigma_D^2 = \frac{1}{36}\sum_{i,j} G_{ij}^2 - \sigma_A^2.$$

Asymptotically unbiased estimates of the above variance components are (Seber, 1977)

$$\hat{\sigma}_A^2 = \frac{1}{5 \times 36^2} \times \{\hat{\beta}^{\mathrm{T}}(\mathbf{H}_1^{\mathrm{T}}\mathbf{H}_1 + \mathbf{H}_2^{\mathrm{T}}\mathbf{H}_2)\hat{\beta} - \mathrm{Tr}[\mathbf{H}_1^{\mathrm{T}}\mathbf{H}_1 + \mathbf{H}_2^{\mathrm{T}}\mathbf{H}_2)\mathbf{V}_{\hat{\beta}}]\}, \tag{8}$$

where $\mathbf{V}_{\hat{\beta}} = \mathrm{Var}(\hat{\beta})$, the variance–covariance matrix of the estimated genotypic values. Similarly, an asymptotically unbiased estimate of σ_D^2 is

$$\hat{\sigma}_D^2 = \frac{1}{25 \times 36^2}\{\hat{\beta}^{\mathrm{T}}\mathbf{H}_3^{\mathrm{T}}\mathbf{H}_3\hat{\beta} - \mathrm{Tr}(\mathbf{H}_3^{\mathrm{T}}\mathbf{H}_3\mathbf{V}_{\hat{\beta}})\}. \tag{9}$$

(iii) *Tests of hypotheses*

The overall hypothesis to be tested is

$$H_0: \mathbf{H}\beta = 0.$$

We use an F-test statistic,

$$F = \hat{\beta}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}[\mathbf{H}(\mathbf{U}^{\mathrm{T}}\mathbf{V}_{\hat{\beta}}^{-1}\mathbf{U})^{-1}\mathbf{H}^{\mathrm{T}}]\mathbf{H}\hat{\beta}, \tag{10}$$

where $\mathbf{U} = [\mathbf{U}_1//\mathbf{U}_2// \dots //\mathbf{U}_n]$ and $\mathbf{U}_j$ has been defined in an earlier section.

The overall test statistic can be partitioned into three subtests, each testing a particular variance component. The test statistic for the kth component is

$$F_k = \hat{\beta}^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}[\mathbf{H}_k(\mathbf{U}^{\mathrm{T}}\mathbf{V}_{\hat{\beta}}^{-1}\mathbf{U})^{-1}\mathbf{H}_k^{\mathrm{T}}]\mathbf{H}_k\hat{\beta}, \tag{11}$$

where F_1 tests for the paternal component, F_2 for the maternal component and F_3 for the interaction.

(iv) *Inferring QTL genotype from linked markers*

We now go back to the conditional expectations and variance–covariance matrix of the QTL genotype indicator $\mathbf{X}_j$ given marker information. At a marker locus, a randomly selected parent does not necessarily segregate for four distinguished alleles. Two or more alleles may be identical-by-state, resulting in a partially informative or non-informative marker. Therefore, a multipoint method that simultaneously uses all markers in the same linkage group is used to infer the distribution of $\mathbf{X}_j$. To simplify the derivation, we first present the interval mapping procedure that uses only two flanking markers. We can then easily extend the interval mapping to multipoint mapping. Define $M_1^m M_2^m M_3^m M_4^m$ and $M_1^f M_2^f M_3^f M_4^f$ as the marker genotypes of the male and female parents, respectively, at the left-hand side of the QTL, and $N_1^m N_2^m N_3^m N_4^m$ and $N_1^f N_2^f N_3^f N_4^f$ as those at the right-hand side. Under random chromosomal segregation, each parent will produce six possible gametes at each marker, resulting in a total of 36 possible genotypes in the progeny. Define $M \in \{M_1, M_2, \dots, M_{36}\}$ as the set of 36 possible genotypes where $M_1 = M_1^m M_2^m M_1^f M_2^f$ etc. are properly ordered in the set, and also define $N \in \{N_1, N_2, \dots, N_{36}\}$ and $Q \in \{Q_1, Q_2, \dots, Q_{36}\}$ in the same manner. Interval mapping requires computing $\Pr(Q = Q_i \mid M = M_j, N = N_k)$ for $i, j, k = 1, \dots, 36$. This is achieved by using the Bayes theorem:

$$\Pr(Q = Q_i \mid M = M_j, N = N_k) = \frac{\Pr(Q = Q_i)\Pr(M = M_j \mid Q = Q_i)\Pr(N = N_k \mid Q = Q_i)}{\sum_{i=1}^{36}\Pr(Q = Q_i)\Pr(M = M_j \mid Q = Q_i)\Pr(N = N_k \mid Q = Q_i)}, \tag{12}$$

where $\Pr(Q = Q_i) = 1/36$ for $i = 1, \dots, 36$ is the prior probability, and $\Pr(M = M_j \mid Q = Q_i)$ or $\Pr(N = N_k \mid Q = Q_i)$ is the transition probability from a QTL genotype to a marker genotype. The transition matrix between M and Q is given by

$$\mathbf{T}_{MQ} = \mathbf{T} \otimes \mathbf{T},$$

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

where $\otimes$ denotes the Kronecker product of two matrices and

$$\mathbf{T} = \begin{bmatrix} (a+b)/d & (b+c)/d & (b+c)/d & (b+c)/d & (b+c)/d & 2c/d \\ (b+c)/d & (a+b)/d & (b+c)/d & (b+c)/d & 2c/d & (b+c)/d \\ (b+c)/d & (b+c)/d & (a+b)/d & 2c/d & (b+c)/d & (b+c)/d \\ (b+c)/d & (b+c)/d & 2c/d & (a+b)/d & (b+c)/d & (b+c)/d \\ (b+c)/d & 2c/d & (b+c)/d & (b+c)/d & (a+b)/d & (b+c)/d \\ 2c/d & (b+c)/d & (b+c)/d & (b+c)/d & (b+c)/d & (a+b)/d \end{bmatrix}.$$

Table 1. *Observed 95th and 99th empirical threshold values obtained from 1000 replicated simulations under the hypothesis of no QTL segregation. The length of the linkage map is 100 cM*

Linkage map interval	Allele	$\alpha = 0{\cdot}05$				$\alpha = 0{\cdot}01$			
		T	T_m	T_f	$T_{m\times f}$	T	T_m	T_f	$T_{m\times f}$
10 cM	4	62·70	13·72	14·15	48·38	72·00	18·51	18·62	54·77
	6	63·21	14·10	14·49	49·82	72·43	18·18	19·33	56·18
	8	63·19	13·49	14·46	49·12	73·86	18·74	18·75	58·40
5 cM[a]	6	61·59	13·41	13·55	47·43	68·07	19·14	18·66	53·16
20 cM	6	60·28	12·52	12·98	47·17	67·23	16·83	17·13	52·42

[a] When the marker interval is 5 cM, the length of the map is only 50 cM so that the total number of markers remains the same as in other settings.

Table 2. *Mean test statistics for overall* (T) *paternal* (T_m) *maternal* (T_f) *and interaction* ($T_{m\times f}$) *effects from 100 replicated simulations, with standard errors given in parentheses*

h^2		T	T_m	T_f	$T_{m\times f}$
No. of alleles per marker					
0·10	4	71·31 (13·50)	18·88 (7·74)	17·63 (9·18)	32·09 (9·54)
	6	73·53 (15·54)	19·21 (8·94)	18·63 (7·84)	31·84 (9·33)
	8	72·01 (13·48)	18·17 (8·61)	18·56 (8·36)	32·22 (8·75)
0·20	4	102·35 (21·30)	33·77 (11·26)	34·14 (11·16)	26·83 (7·59)
	6	104·57 (19·95)	35·92 (11·94)	33·97 (12·52)	27·80 (8·77)
	8	105·88 (21·32)	34·61 (10·75)	35·30 (14·59)	28·00 (6·94)
0·40	4	207·98 (34·75)	80·56 (19·39)	80·19 (17·76)	26·25 (6·98)
	6	210·08 (37·82)	80·06 (20·22)	81·30 (22·73)	27·29 (8·44)
	8	207·18 (31·36)	81·84 (20·59)	80·74 (19·62)	25·75 (6·70)
Length of marker interval (cM)[a]					
0·10	5	73·08 (14·29)	19·63 (8·57)	20·96 (8·65)	29·29 (7·42)
	20	68·94 (14·33)	17·64 (8·19)	16·72 (7·53)	30·94 (8·83)
0·20	5	109·95 (22·30)	36·80 (12·82)	37·42 (11·22)	27·07 (7·16)
	20	98·31 (20·66)	30·45 (10·97)	33·62 (11·33)	26·53 (7·17)
0·40	5	227·16 (38·22)	88·34 (18·50)	91·80 (23·28)	26·12 (8·03)
	20	181·00 (32·48)	68·38 (20·63)	66·88 (18·04)	26·04 (7·82)
Interaction effect[b]					
0·10	$\sigma^2_\delta = 0{\cdot}056$	69·44 (12·40)	12·16 (6·46)	11·74 (6·11)	43·87 (9·31)
0·20	$\sigma^2_\delta = 0{\cdot}125$	103·63 (22·56)	18·80 (7·89)	20·26 (99·12)	60·46 (15·54)
0·40	$\sigma^2_\delta = 0{\cdot}334$	191·35 (33·28)	42·08 (12·87)	41·28 (14·80)	99·10 (21·87)
Sample size (n)					
0·20	200	82·14 (18·35)	21·83 10·87)	23·08 (12·28)	31·00 (9·99)
	300	104·57 (19·95)	35·92 (11·94)	33·97 (12·52)	27·80 (8·77)
	500	145·30 (27·20)	56·80 (17·32)	53·49 (15·83)	26·89 (8·62)

[a] The length of the chromosome segment is 50 cM when the marker interval is set at 5 cM.
[b] Interaction variance is set at $\sigma^2_\delta = 2\sigma^2_m = 2\sigma^2_f = 0{\cdot}056$, 0·125 *and* 0·334 *for* $h^2 = 0{\cdot}10$, 0·20 and 0·40, respectively.

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

Table 3. *Statistical power* (%) *of QTL detection at type I error rates of* $\alpha = 0{\cdot}05$ *and* $\alpha = 0{\cdot}01$ *for testing the overall* (T), *paternal* (T_m), *maternal* (T_f) *and interaction* ($T_{m\times f}$) *effects* (*obtained from 100 replicated simulations*)

		T		T_m		T_f		$T_{m\times f}$	
h^2		$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$	$\alpha = 0{\cdot}05$	$\alpha = 0{\cdot}01$
No. of marker alleles									
0·10	4	72·0	42·0	70·0	45·0	57·0	42·0	5·0	1·0
	6	73·0	47·0	67·0	47·0	66·0	41·0	5·0	1·0
	8	77·0	41·0	71·0	44·0	69·0	44·0	6·0	1·0
0·20	4	98·0	95·0	97·0	92·0	99·0	94·0	0·0	0·0
	6	100·0	97·0	99·0	96·0	99·0	91·0	2·0	1·0
	8	100·0	97·0	100·0	97·0	97·0	89·0	1·0	0·0
0·40	4	100·0	100·0	100·0	100·0	100·0	100·0	0·0	0·0
	6	100·0	100·0	100·0	100·0	100·0	99·0	2·0	0·0
	8	100·0	100·0	100·0	100·0	100·0	100·0	1·0	0·0
Length of marker interval (cM)[a]									
0·10	5	78·0	57·0	75·0	48·0	73·0	57·0	2·0	0·0
	20	69·0	52·0	68·0	52·0	64·0	48·0	3·0	0·0
0·20	5	100·0	100·0	100·0	97·0	100·0	96·0	0·0	0·0
	20	98·0	95·0	97·0	94·0	98·0	94·0	0·0	0·0
0·40	5	100·0	100·0	100·0	100·0	100·0	100·0	1·0	0·0
	20	100·0	100·0	100·0	100·0	100·0	100·0	0·0	0·0
Dominance effect[b]									
0·10	$\sigma^2_\delta = 0{\cdot}050$	71·0	42·0	32·0	19·0	32·0	12·0	24·0	14·0
0·20	$\sigma^2_\delta = 0{\cdot}125$	100·0	99·0	73·0	57·0	65·0	40·0	65·0	43·0
0·40	$\sigma^2_\delta = 0{\cdot}334$	100·0	100·0	100·0	99·0	99·0	97·0	100·0	99·0
Sample size (n)									
0·20	200	81·0	71·0	74·0	59·0	72·0	55·0	5·0	0·0
	300	100·0	97·0	99·0	96·0	99·0	91·0	2·0	1·0
	500	100·0	100·0	100·0	100·0	100·0	100·0	2·0	2·0

[a,b] See the annotation of Table 2.

The value of each element is a function of the recombination fraction (r) between the QTL and the marker in question. Under the Haldane mapping function, $a = (1-r)^2$, $b = r(1-r)/3$, $c = r^2/9$ and $d = 1-2r/3+4r^2/9$. The numerator in (12) can be expressed in matrix notation:

$$\Pr(M = M_j \mid Q = Q_i)\Pr(N = N_k \mid Q = Q_i) = \mathbf{1}^{\mathrm{T}}\mathbf{D}_M\mathbf{T}_{MQ}\mathbf{D}_{(i)}\mathbf{T}_{QN}\mathbf{D}_N\mathbf{1}, \quad (13)$$

where $\mathbf{1}$ is a 36×1 vector with unity elements, $\mathbf{D}_M$ is a diagonal matrix with the jjth diagonal element equal to 1 and all other elements equal to 0, $\mathbf{D}_N$ is similarly defined but with the kkth diagonal element equal to 1 and all others equal to 0, and $\mathbf{D}_{(i)}$ is a diagonal matrix with the iith diagonal element equal to 1 and all others equal to 0. Note that $\mathbf{D}_M$ represents the data from marker M. Because the jth genotype of marker M is observed, the jjth element of $\mathbf{D}_M$ is filled by 1. The kkth diagonal element of $\mathbf{D}_N$ is filled by 1 because the kth genotype of marker N has been observed. We choose $\mathbf{D}_{(i)}$ because it is the probability of the ith genotype of the QTL that is of interest.

One advantage of the matrix notation comes from its ease of handling missing and partially informative markers. If a marker is not fully informative, more than one possible genotype among the 36 are compatible with the data. In this case, we can easily take all the compatible genotypes into consideration by replacing $\mathbf{D}_M$ by a $\mathbf{D}$ matrix with all diagonal elements corresponding to the positions of the compatible genotypes filled by one and all other elements filled by 0. If both parents are homozygotes at a marker locus, or the genotype is missing at a particular marker for an individual, all genotypes are compatible, and thus $\mathbf{D}_M$ is simply an identity matrix, i.e. all diagonal elements are one. With the matrix notation, dominant markers present no problem.

The most important advantage of the above treatment is the ability to perform multipoint mapping. If markers are not fully informative, non-flanking markers also provide information for the genotype distribution of QTL. Multipoint mapping using all markers simultaneously will significantly increase the power of QTL detection. Assuming that there are m markers in the linkage group and QTL is located between markers k and $k+1$, the general formula for the multipoint method is

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

Table 4. *Mean estimates of QTL parameters* (*QTL position*, *heritability*, *paternal*, *maternal and interaction variances*, *as well as the residual variance*) *obtained from 100 replicated simulations when interaction effects is absent*. *The standard errors are given in the parentheses*

h^2	Marker allele	QTL position (cM)	$\hat{h}^2$	$\hat{\sigma}^2_m$	$\hat{\sigma}^2_f$	$\hat{\sigma}^2_\delta$	$\hat{\sigma}^2_\epsilon$
Parameter:		25·00	0·100	0·056	0·056	0·000	1·000
0·10	4	26·54 (16·35)	0·082 (0·045)	0·046 (0·035)	0·039 (0·036)	0·034 (0·056)	0·935 (0·090)
	6	28·84 (12·00)	0·102 (0·038)	0·054 (0·033)	0·055 (0·028)	0·033 (0·058)	0·958 (0·096)
	8	25·12 (11·55)	0·098 (0·049)	0·055 (0·041)	0·053 (0·038)	0·035 (0·053)	0·974 (0·084)
Parameter:		25·00	0·200	0·125	0·125	0·000	1·000
0·20	4	24·84 (6·57)	0·202 (0·056)	0·124 (0·057)	0·128 (0·051)	0·011 (0·048)	0·980 (0·102)
	6	25·60 (5·61)	0·196 (0·053)	0·126 (0·052)	0·118 (0·051)	0·013 (0·045)	0·980 (0·088)
	8	25·44 (5·97)	0·194 (0·059)	0·124 (0·054)	0·117 (0·059)	0·015 (0·041)	0·985 (0·088)
Parameter:		25·00	0·400	0·334	0·334	0·000	1·000
0·40	4	24·83 (5·00)	0·374 (0·055)	0·296 (0·086)	0·303 (0·089)	0·008 (0·047)	0·990 (0·077)
	6	24·40 (4·29)	0·392 (0·062)	0·328 (0·090)	0·324 (0·095)	0·009 (0·184)	0·993 (0·092)
	8	25·52 (3·51)	0·393 (0·059)	0·331 (0·094)	0·325 (0·096)	0·004 ((0·040)	0·996 (0·091)

$$\Pr(Q = Q_i \mid I_M) = \frac{\Pr(Q = Q_i)\Pr(I_M \mid Q = Q_i)}{\sum_{i=1}^{36} \Pr(Q = Q_i)\Pr(I_M \mid Q = Q_i)}, \quad (14)$$

where I_M is a generic symbol for marker information (all markers) and

$$\Pr(I_M \mid Q = Q_i) = \mathbf{1}^{\mathrm{T}}\mathbf{D}_1\mathbf{T}_{12}\mathbf{D}_2 \ldots \mathbf{D}_k\mathbf{D}_{kQ}\mathbf{D}_{(i)}\mathbf{T}_{Q(k+1)} \times \mathbf{D}_{k+1} \ldots \mathbf{D}_{m-1}\mathbf{T}_{(m-1)m}\mathbf{D}_m\mathbf{1}, \quad (15)$$

where $\mathbf{D}_k$ is the data matrix for marker k. Partially important markers are more common than fully informative markers in outbred parents. This is particularly true in tetraploid mapping presented here. Therefore, multipoint mapping is essential in tetraploids. Equation (15) is essentially derived using the hidden Markov model (HMM). Further references for multipoint mapping using HMM are Lander & Green (1987), Kruglyak & Lander (1995), Jiang & Zeng (1997), Xu & Gessler (1998) and Xie & Xu (1999).

3. Simulation studies

To explore the properties of QTL mapping in tetraploids, we conducted a series of simulation experiments. Each gamete consists of two alleles, and the two alleles act together as a unit. The six possible gametes in each parent are considered as six 'alleles' each of which is assigned a value. In the simulation experiments, the six 'allelic effects' of each parent are fixed (not randomly sampled), and their values are determined so that their variance equals a preassigned genetic variance. We then applied Kempthorne's (1954) method for multiple alleles in diploids (cf. Li, 1957) to the tetraploid genetic analysis.

In most cases, we simulated one chromosome of length 100 mM with 11 markers evenly spaced along the chromosome (equivalent to 10 cM marker intervals). One QTL was simulated at position 25 cM in a full-sib family of size 300. To determine the marker linkage phases in the parents, we first randomly chose four grandparents from a reference hypothetical population that is in Hardy–Weinberg equilibrium, and then generated two parents for mating. This is similar to a four-way cross, but with four heterozygous lines, i.e. grandparents. Marker alleles at each locus in each of the four grandparents were sampled at random from the base population with an equal frequency. The variance of the environmental effect was set to $\sigma^2_\epsilon = 1{\cdot}0$.

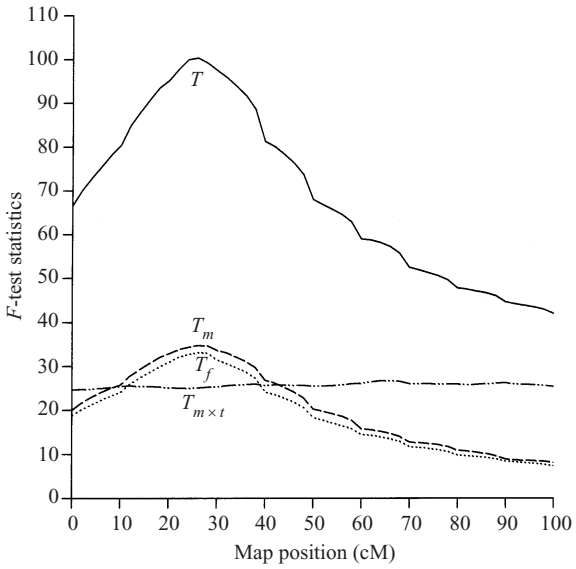

Fig. 1. Comparison of the *F*-test statistics of QTL mapping in a full-sib family of size 300. Eleven markers each having six alleles are evenly spaced along a 100 cM long chromosome. A single additive QTL is at position 25 cM and explains 20% of the total variation. T is the overall test for the presence of QTL; T_m is the test for QTL segregation in the male parent; T_f is the test for QTL segregation in the female parent; and $T_{m\times f}$ is the test for the interaction.

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

Table 5. *Mean estimates of QTL parameters obtained from 100 replicated simulations when interaction effect is present. The standard errors are given in the parentheses*

h^2	Marker allele	QTL position (cM)	$\hat{h}^{2a}$	$\hat{h}^2_G$	$\hat{\sigma}^2_m$	$\hat{\sigma}^2_f$	$\hat{\sigma}^2_\delta$	$\hat{\sigma}^2_e$
Parameter:		25·00	0·050	0·100	0·028	0·028	0·056	1·000
0·05	6	28·70 (15·21)	0·048 (0·029)	0·102 (0·042)	0·023 (0·025)	0·027 (0·021)	0·058 (0·047)	0·963 (0·080)
Parameter:		25·00	0·100	0·200	0·063	0·063	0·125	1·000
0·10	4	25·54 (6·43)	0·106 (0·40)	0·208 (0·070)	0·066 (0·037)	0·066 (0·044)	0·129 (0·076)	0·986 (0·095)
	6	24·74 (5·40)	0·096 (0·041)	0·211 (0·076)	0·055 (0·035)	0·064 (0·079)	0·145 (0·079)	0·974 (0·109)
	8	24·88 (6·13)	0·100 (0·042)	0·205 (0·066)	0·061 (0·043)	0·067 (0·040)	0·132 (0·077)	0·995 (0·095)
Parameter:		25·00	0·200	0·400	0·167	0·167	0·334	1·000
0·20	6	25·40 (3·55)	0·195 (0·049)	0·387 (0·085)	0·169 (0·071)	0·160 (0·068)	0·329 (0·140)	1·007 (0·096)

[a] $\hat{h}^2_G$ is broad-sense heritability.

To evaluate the performance of QTL mapping, we varied the following factors successively: (1) the size of the QTL, measured by QTL additive and dominance variances, (2) the amount of marker information, indicated by the number of marker alleles and the marker density, (3) sample size of the family, and (4) one versus two QTLs. In the two-QTL analysis we generated the phenotypic data with one QTL segregating in the male parent and the other in the female parent. When sampling the marker alleles to make a marker genotype, we sampled the alleles from three different hypothetical base populations: (1) population one consists of four alleles each equal frequency (1/4), (2) population two consists of six alleles with equal frequency (1/6), and (3) population three consists of eight alleles with equal frequency (1/8). Note that even under the most informative situation of eight alleles, within each repeat of the simulations, the two parents at any markers are hardly fully informative (with probability $8!(\frac{1}{8})^8 = 0{\cdot}0024192$ for the two parents to carry eight different alleles). Therefore, multipoint method plays an important role in the simulation studies. Under each condition the simulation was repeated 100 times. The standard error of an estimate is calculated from the standard deviation of the estimates among 100 replicates.

Statistical power is determined by counting the number of runs out of 100 replicates which have a test statistic greater than an empirical threshold value. To estimate the threshold values, we ran an additional 1000 simulations under the null model (with no QTL segregation). The empirical threshold values under each condition were then obtained by determining the 95th and the 99th percentiles of the highest test statistics from the list of 1000 runs under the null model and are presented in Table 1. These threshold values are slightly larger than those of the χ^2 distribution for the overall test $T = F$ (with 35 df), for test of the paternal segregation $T_m = F_1$ (with 5 df), for test of the maternal segregation $T_f = F_2$ (with 5 df) and for test of the interaction $T_{m\times f} = F_3$ (with 25 df).

(i) *QTL detection*

The test statistics and the powers of QTL detection over 100 replicated simulations are summarized in Tables 2 and 3, respectively. As expected, the test statistic and power are increased with the increase in the size of QTL, marker information content and the family size.

Fig. 1 gives the plots of the mean test statistics against the map position over 100 replicates for a QTL explaining 20% trait variation. The true position of the simulated QTL is at position 25 cM. The profiles of the test statistics behave exactly as we expected. The overall test (T) for the presence of a

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

Table 6. *Mean estimates of QTL parameters obtained from 100 replicated simulations* (*comparing different marker intervals*). *The standard errors are given in the parentheses*

h^2	Marker allele	QTL position (cM)	$\hat{h}^2$	$\hat{\sigma}^2_m$	$\hat{\sigma}^2_f$	$\hat{\sigma}^2_\delta$	$\hat{\sigma}^2_\epsilon$
Parameter:		25·00	0·100	0·056	0·056	0·000	1·000
0·10	5 cM	25·99 (8·97)	0·102 (0·044)	0·054 (0·040)	0·059 (0·036)	0·014 (0·050)	0·978 (0·085)
	20 cM	29·84 (19·75)	0·091 (0·059)	0·049 (0·038)	0·044 (0·044)	0·016 (0·086)	0·921 (0·120)
Parameter:		25·00	0·200	0·125	0·125	0·000	1·000
0·20	5 cM	25·26 (3·79)	0·199 (0·053)	0·122 (0·051)	0·129 (0·049)	0·004 (0·034)	0·992 (0·094)
	20 cM	26·04 (7·47)	0·212 (0·061)	0·128 (0·070)	0·139 (0·053)	0·019 (0·055)	0·937 (0·116)
Parameter:		25·00	0·400	0·334	0·334	0·000	1·000
0·40	5 cM	24·66 (1·83)	0·394 (0·056)	0·329 (0·089)	0·331 (0·092)	0·003 (0·042)	1·000 (0·093)
	20 cM	25·52 (5·84)	0·406 (0·077)	0·331 (0·113)	0·323 (0·110)	0·005 (0·107)	0·940 (0·128)

Table 7. *Mean estimates of QTL parameters obtained from 100 replicated simulations* (*comparing different sample sizes*). *The standard errors are given in the parentheses*

Family size	QTL position (cM)	$\hat{h}^2$	$\hat{\sigma}^2_m$	$\hat{\sigma}^2_f$	$\hat{\sigma}^2_\delta$	$\hat{\sigma}^2_\epsilon$
Parameter:	25·00	0·200	0·125	0·125	0·000	1·000
200	25·74 (11·36)	0·195 (0·092)	0·114 (0·078)	0·118 (0·089)	0·050 (0·106)	0·920 (0·114)
300	25·60 (5·61)	0·196 (0·053)	0·126 (0·052)	0·118 (0·051)	0·013 (0·045)	0·980 (0·088)
500	25·74 (4·66)	0·188 (0·042)	0·122 (0·044)	0·112 (0·044)	0·006 (0·028)	0·998 (0·070)

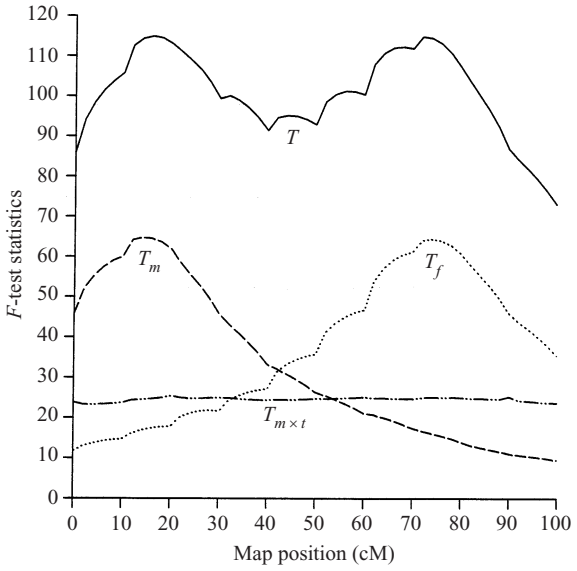

Fig. 2. Profiles of the *F*-test statistics for QTL mapping in a full-sib family of size 300. Eleven markers each having six alleles are evenly spaced along a 100 cM chromosome. Two QTLs are simulated, one is at position 15 cM and segregates in the male parent only, and the other is at 75 cM and segregates in the female parent only. See the legend of Fig. 1 for T, T_m, T_f and $T_{m\times f}$.

QTL has the highest test statistic, whereas the curves for T_m and T_f have similar heights due to $\sigma^2_m = \sigma^2_f = 0{\cdot}125$, and the shape of $T_{m\times f}$ is flat because $\sigma^2_\delta = 0{\cdot}0$. It can be seen that $T_{m\times f}$ gives a rather high signal although the interaction effect is zero. However, the $T_{m\times f}$ test is not significant because it has a high critical value due to the large number of degrees of freedom (df = 25).

(ii) *QTL parameter estimation*

Under $\sigma^2_\delta = 0{\cdot}0$, the QTL additive variance was examined at three levels: $\sigma^2_A = 0{\cdot}111$, 0·250 and 0·667, corresponding to a QTL heritability of $h^2 = \sigma^2_A/(\sigma^2_A + \sigma^2_E) = 0{\cdot}10$, 0·20 and 0·40, respectively. The average values of estimated QTL parameters and standard errors of the estimates are given in Table 4. The size of the QTL notably affects the precision of the estimated QTL position. In the case of low QTL heritability ($h^2 = 0{\cdot}10$), the estimated QTL position is biased towards the centre. When the interaction effect exists, the QTL position and various variance components are also successfully estimated, implying a fair partitioning of the additive and interaction variances (Table 5).

The number of alleles at each marker locus has a small effect on the estimates of various variance components and the heritability. However, it has a relatively large impact on the precision of the estimated position of the QTL. A large number of alleles indicates high information content and thus can reduce the standard error of the estimated position (see Table 4). The length of the marker interval also

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

reflects marker information content. We simulated three levels of marker interval length: 5, 10 and 20 cM per interval. The map lengths simulated are 100 cM for the 10 cM and 20 cM intervals, and 50 cM for the 5 cM interval case. Each marker locus has 6 alleles in the base population. As expected, high marker density can reduce the standard errors of the estimated QTL position and genetic variance components (Table 6).

The mean estimates of the QTL position and variance components under three levels of family sizes (200, 300 and 500) are given in Table 7. As expected, sample size has a profound effect on the performance of the method.

(iii) *Analysis of two QTLs*

Here we simulated a 100 cM long chromosome segment with 11 evenly spaced markers each having six alleles. The first QTL was put at position 15 cM and segregates only in the male parent, and the second QTL was put at 75 cM and segregates only in the female parent. The two QTLs jointly explain 40% of the total phenotypic variance. We used the single-QTL model to analyse the data. The mean test statistics (obtained from 100 replicated simulations) are plotted against the map position (Fig. 2). The overall test (T) for the presence of QTLs shows two peaks and has a signal twice as high as either T_m or T_f. The T_m test statistic indicates a QTL segregating in the male parent whereas the T_f test statistic indicates a QTL segregating in the female parent. The $T_{m\times f}$ curve is flat because the interaction effect has been set to zero. We did not examine the multiple QTL model or the composite mapping approaches (Jansen, 1993; Zeng, 1994) which are designed to search for multiple QTLs.

4. Discussion

The tetraploid mapping procedure is developed using an outbred full-sib family in which the two parents can carry up to eight different alleles at each marker and QTL. In practice, however, it is rarely true that the eight alleles are all different at any marker. Essentially, all markers are partially informative. the multipoint method presented in this study has provided an automatic mechanism for handling partially informative markers. The genetic model, however, still assumes eight different alleles at the QTL in question. This presents no problem because if the number of alleles at the QTL is less than eight, we will have less than 36 distinguishable genotypes in the progeny and some of the linear contrasts (genetic effects) will have a zero expectation. The test statistics for those contrasts are expected to be non-significant. This is equivalent to the situation of eight different alleles but some of the allelic differences being infinitesimal. There is no logical problem in statistics to estimate and test an effect with zero expectation. It can decrease the statistical power, however, if one knows exactly the number of distinguishable genotypes but still pretends there are 36 distinguishable genotypes. If the history of the base population in which the parents are sampled is known, the number of alleles may be known and this information should be taken into account. For example, if the progeny are derived from selfing a non-inbred parent, the maximum number of alleles in the family is immediately known (four alleles), and the maximum possible number of genotypes in the progeny will be $6(6+1)/2 = 21$ instead of 36. If the progeny are derived from selfing a hybrid of two inbred lines, then the number of alleles is two and the maximum number of genotypes in the progeny is six – far fewer than 36.

One advantage of polyploid mapping over diploid mapping comes from the increased chance of sampling a non-homozygous parent. If a gene is segregating in a base population but the sampled parents are homozygous, then the genetic variance at this locus cannot be detected, no matter how large the family size is. This has been explained as the drift error (Xu, 1996). In tetraploid mapping, this drift error can be substantially reduced compared with diploid mapping. For instance, the probability of sampling a heterozygous diploid (two alleles per locus) parent from a base population with three ($K = 3$) equally frequent alleles is $(K^{2-1}-1)/K^{2-1} = 2/3$). However, the probability of sampling a non-homozygous tetraploid (four alleles per locus) parent from the same base population is $(K^{4-1}-1)/K^{4-1} = 26/27$.

Although the model presented in this study is a single-QTL model, it can be extended to handle multiple QTLs with no theoretical difficulty. The model can even be modified to estimate and test potential epistatic effects. One practical problem of multiple QTL mapping is the inconvenience of jointly searching for the number of QTLs and their locations. The problem has not been completely solved even with simple line crossing experiments in diploid species. Bayes' method of QTL mapping has been investigated in diploid organisms (Satagopan *et al.*, 1996); Sillanpaa & Arjas, 1998) and its application to tetraploid mapping certainly represents the direction of future research.

Existing methods of QTL mapping in tetraploids rely primarily on dominant markers (Hackett *et al.*, 1998). The efficiency of these methods depends on the marker genotypes selected for the parents. Selecting the simplex–simplex pair of parents proves to be efficient and any other types of pairs are all inefficient. If one knows, *a priori*, that a QTL is sitting near a marker with both parents being simplex, then the analysis seems to be meaningful; otherwise, chromosomal scanting is required. The probability of selecting

https://doi.org/10.1017/S0016672399004395 Published online by Cambridge University Press

two parents with sufficient number of markers of the required simplex–simplex configuration can be extremely small. Therefore, co-dominant markers are necessary for tetraploid mapping. Genotyping a large number of co-dominant markers is still expensive, but this economic limitation will soon disappear. This study has provided the statistical tools with which we are now ready to analyse real data as they become available.

We wish to thank Dr Rhonda Meyer for helpful discussion on the marker system in polyploids and Dr Damian Gessler for comments on the manuscripts. We also thank two anonymous reviewers whose comments and criticisms have greatly improved the presentation of the manuscript. This research was supported by the National Institutes of Health Grant GM55321-03 and the USDA National Research Initiative Competitive Grants Program 97-35205-5075 to S.X.

References

Al Janabi, S. M., Honeycutt, R. J., McClelland, M. & Sorral, B. W. S. (1993). A genetic linkage map of *Saccharum spontaneum* L. 'SES 2008'. *Genetics* **134**, 1249–1260.

Da Silva, J. A. G., Sorrells, M. E., Burnquist, W. L. & Tanksley, S. D. (1993). RFLP linkage map and genome analysis of Saccharum spontaneum. *Genome* **36**, 782–792.

Grivet, L., D'Hont, A., Roques, D., Feldmann, P., Lanaud, C. & Glaszmann, J. C. (1996). RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* **142**, 987–1000.

Hackett, C. A., Bradshaw, J. E., Meyer, R. C., McNicol, J. W., Milbourne, D. & Waugh, R. (1998). Linkage analysis in tetraploid species: a simulation study *Genetical Research* **71**, 143–154.

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

Jiang, C. & Zeng, Z.-B. (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines *Genetica* **101**, 47–58.

Kempthorne, O. (1954). The correlation between relatives in a random mating population. *Proceedings of the Royal Society of London*, *Series B* **143**, 103–113.

Kempthorne, O. (1955). The correlation between relatives in a simple autotetraploid population. *Genetics* **40**, 168–174.

Kruglyak, L. & Lander, E. S. (1995). Complete multipoint sib-pair analysis of qualitative and quantitative traits. *American Journal of Human Genetics* **57**, 439–454.

Lander, E. S. & Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Sciences of the USA* **84**, 2363–2367.

Li, C. C. (1957). The genetic variance of autotetraploids with two alleles. *Genetics* **42**, 583–592.

Meyer, R. C., Milbourne, D., Hackett, C. A., Bradshaw, J. E., McNicol, J. W. & Waugh, R. (19998). Linkage analysis in tetraploid potato and associations of markers with quantitative resistance to late blight (*Phytophthora infestans*). *Molecular and General Genetics* **259**, 150–160.

Milbourne, D., Meyer, R. C., Bradshaw, J. E., Baird, E., Bonar, N., Provan, J., Powell, W. & Waugh, R. (1997). Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. *Molecular Breeding* **3**, 127–136.

Ritter, E., Gebhardt, C. & Salamini, F. (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* **125**, 645–654.

SAS Institute (1990). *SAS/IML Software*. Cary, NC: SAS Institute.

Satagopan, R. J., Yandell, B. S., Newton, M. A. & Osborn, T. C. (1996). A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. *Genetics* **144**, 805–816.

Seber, G. A. F. (1977). *Linear Regression Analysis*. New York: Wiley.

Sillanpaa, M. J. & Arjas, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* **148**, 1373–1388.

Soltis, D. E. & Soltis, P. S. (1995). The dynamic nature of polyploid genomes. *Proceedings of the National Academy of Sciences of the USA* **92**, 8089–8091.

Song, K., Lu, P., Tang, K. & Osborn, T. C. (1995). Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences of the USA* **92**, 7719–7723.

Wu, K. K., Burnquist, W., Sorrells, M. E., Tew, T. L., Moore, P. H. & Tanksley, S. D. (1992). The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics* **83**, 294–300.

Xie, C. & Xu, S. (1999). Mapping quantitative trait loci with dominant markers in fourway crosses. *Theoretical and Applied Genetics* **98**, 1014–1021.

Xu, S. (1995). A comment on the simple regression method for interval mapping. *Genetics* **141**, 1657–1659.

Xu, S. (1996). Mapping quantitative trait loci using four-way crosses. *Genetical Research* **68**, 175–181.

Xu, S. (1998*a*). Further investigation on the regression method of mapping quantitative trait loci. *Heredity* **80**, 364–373.

Xu, S. (1998*b*). Iteratively reweighted least squares mapping of quantitative trait loci. *Behavior Genetics* **28**, 341–355.

Xu, S. & Gessler, D. D. G. (1998). Multipoint genetic trait loci with dominant markers in four-way crosses. *Theoretical and Applied Genetics* **83**, 1014–1021.

Yu, K. F. & Pauls, K. P. (1993). Segregation of random amplified polymorphic DNA markers and strategies for molecular mapping in tetraploid alfalfa. *Genome* **36**, 844–851.

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