Approximate Thresholds of Interval Mapping Tests for QTL Detection

Ahmed Rebaï, Bruno Goffinet and Brigitte Mangin

*INRA, Centre de Toulouse, Station de Biometrie et d'Intelligence Artificielle, Auzeville B.P. 27,
31326 Castanet-Tolosan, France*

ABSTRACT

A general method is proposed for calculating approximate thresholds of interval mapping tests for quantitative trait loci (QTL) detection. Simulation results show that this method, when applied to backcross and F_2 populations, gives good approximations and is useful for any situation. Programs which calculate these thresholds for backcross, recombinant inbreds and F_2 for any given level and any chromosome with any given distribution of codominant markers were written in Fortran 77 and are available under request. The approach presented here could be used to obtain, after suitable calculations, thresholds for most segregating populations used in QTL mapping experiments.

THE use of genetic markers has opened new horizons for the identification and the location of Mendelian factors involved in the expression of quantitative traits, so-called QTL (quantitative trait loci), in different segregating populations (backcross, F_2, . . .). Recently, many efficient methods considering pairs of neighboring linked markers have been developed (CARBONELL *et al.* 1992; HALEY and KNOTT 1992; JANSEN 1992; KNAPP *et al.* 1990; LANDER and BOTSTEIN 1989; MARTINEZ and CURNOW 1992; MORENO-GONZALEZ 1992) to infer the position and the effect of a putative QTL lying between such markers. The interval mapping method described by LANDER and BOTSTEIN (1989) (LB) uses maximum likelihood estimation and provides a likelihood ratio test for the presence of a QTL between the markers considered. The linearized method of KNAPP *et al.* (1990) and other related approaches (HALEY and KNOTT 1992) have been shown to be asymptotically equivalent to interval mapping (REBAÏ *et al.* 1994) (RGM).

The interval mapping test performed along the genome could be defined as the supremum of a stochastic process and the distribution of the test statistic is not known in general. This has prevented many works for determining the exact threshold corresponding to a given significance level. LB have stated that, for a backcross (BC) population of large size and in the case where the density of markers in the genome is infinite, the test varies according to the square of an Orenstein-Uhlenbeck process (LEADBETTER *et al.* 1983). They also gave a formula which is appropriate for threshold calculations under the conditions above and proposed to use extensive numerical simulations to determine the thresholds for intermediate marker densities. Threshold calculations for interval mapping tests in other progeny types, where more than one parameter are involved, have essentially been done using simulations (VAN OOIJEN 1992). RGM applied the DAVIES (1977) approach

to get quite good approximations to the threshold of the interval mapping test in a BC population for a single interval of codominant markers.

In this report we generalize the approach previously described (RGM) for many intervals and other progenies for which there are two or more tested parameters. Simulations were also carried out to check the validity of the approximations used. We will first consider the case where the QTL is characterized by only one estimable parameter (its principle effect) which applies to BC and equivalent populations and then discuss the case where the QTL has additivity and dominance parameters, with application to F_2 (two parameters) and diallel-derived populations.

THRESHOLDS IN THE SINGLE PARAMETRE CASE

The Davies approximation: Consider a chromosome of length P in recombination units having $m + 1$ mapped markers from a BC population (equivalent to double haploid lines or F_1 testcross) and so m intervals of length $p_i, (i = 1 \ldots m)$ each. In this case the test for QTL detection involves only one parameter which is the principal effect of the putative QTL. The model used to estimate a (half the difference between genotypic values of QTL genotypes) is the same as that in RGM but any other parametrisation is equivalent. The thresholds calculated in this and following sections are thus suitable for flanking marker methods based on regression or likelihood analysis if only codominant markers are used. At each position x of the chromosome one can compute the likelihood ratio test (LRT(x)) for the hypothesis H_0: $a = 0$ as in RGM or the LOD$(x) = $ LRT$(x)/(2 \ln 10)$. We have shown (RGM) that LRT(x) was asymptotically equivalent to a test statistic $T^2(x)$ where

$$T(x) = \frac{\hat{a}}{\sqrt{\widehat{\text{Var}}(\hat{a})}} \tag{1}$$

A. Rebaï, B. Goffinet and B. Mangin

TABLE 1

Thresholds at 5% level for the interval mapping test calculated for a chromosome of 100 cM and different numbers of markers in BC populations

Threshold	No. of markers									
	1001	101	51	41	26	21	11	6	3	2
Simulations[a]	9.05[b]	8.95	8.70	8.56	8.45	8.28	7.60	6.90	6.15	
DAVIES approximation	12.0	9.74	9.09	8.88	8.43	8.20	7.58	6.92	6.07	5.22
FEINGOLD et al.	9.06	8.53	8.26	8.16	7.91	7.79	7.31	6.78	6.12	5.78

[a] From 10,000 replications with population of size 200.
[b] Calculated using proposition 2 of LB.

has asymptotically a standard Gaussian distribution. $T^2(x)$ follows a χ^2 with one d.f. under H_0. To find an appropriate threshold $C = c^2$ corresponding to a global level α for the test $T_I = \text{Sup}(T^2(x), 0 \leq x \leq p)$ where p is the length of a single given interval, we have used DAVIES' (1977) bound

$$\frac{\alpha}{2} = \Pr(\sup_{0 \leq x \leq p} T(x) > c)$$

$$\approx \Phi(-c) + \frac{1}{2\pi} \exp\left(-\frac{1}{2}c^2\right) \int_0^p \sqrt{-\rho_{11}(x, p)} \, dx \tag{2}$$

where $\rho_{11}(x, p)$ is the autocorrelation function defined by

$$\rho_{11}(x, p) = \left[\frac{\partial^2 \text{Cov}(T(x), T(y))}{\partial y^2}\right]_{y = x}$$

and $\Phi(x)$ is the cumulative normal distribution function. In the case of BC populations and ignoring double recombination we found (RGM):

$$\int_0^p \sqrt{-\rho_{11}(x, p)} \, dx = 2 \, \text{Arctan}\left(\sqrt{\frac{p}{1 - p}}\right).$$

The problem in generalizing this approach to a chromosome is that the derivative of $T(x)$ is not continuous on $[0, P]$, the discontinuities occurring on the marker positions. DAVIES (1987) observed that approximation (2) is still appropriate when the derivative has a finite number of jumps. In regard to this, approximation (2) could be extended to the chromosome by

$$\frac{\alpha}{2} = \Pr(\sup_{0 \leq x \leq p} T(x) > c)$$

$$\approx \Phi(-c) + \frac{1}{2\pi} \exp\left(-\frac{1}{2}c^2\right) \sum_{i=1}^m \int_0^{p_i} \sqrt{-\rho_{11}(x, p_i)} \, dx \tag{3}$$

The integrals are evaluated for each interval and c calculated by successive approximations for a given value of α. Note that $\sum_i p_i$ is not equal to P unless the frequency of double recombination is neglected. The approximation could also be used with distances from any mapping function but we have taken the recombination fraction as the distance (double recombination ignored). Calculations based on distances from other mapping func-

tions or taking into account double recombination give similar values. In the case where m is high (number of jumps in the derivative tends to infinity) conditions for the DAVIES' approximation are not well satisfied and other approximations are available.

Other approximations: LB have proved that T_I is a stationary Gaussian process with a covariance function satisfying, for any mapping function,

$$\text{Cov}(T(x), T(y)) = R(t) = 1 - 2t + o(t^2) \quad \text{as} \quad t \to 0$$

$$\text{where} \quad t = |x - y|.$$

Based on this result LB showed that the threshold for the LOD test for an infinitely dense-map could be approximated by their proposition 2: the threshold C for a genome of H chromosomes and genetic length G (in Morgans) is obtained by solving the equation $\alpha \approx (H + 2GC)\chi^2(C)$ where $\chi^2(.)$ denotes the inverse cumulative distribution function of the χ^2 distribution with 1 d.f. FEINGOLD et al. (1993) proposed another approximation for quite dense maps which is also appropriate in our context. In our case it is:

$$\alpha \approx 1 - \Phi(c) + 2Lc\phi(c)\nu[2c\sqrt{\Delta}] \tag{4}$$

where L is the genetic length of the chromosome expressed in Morgans ($L = -\frac{1}{2}\ln(1 - 2P)$ if we use the Haldane (1919) mapping function), $\phi(x)$ is the probability density function for the standard normal distribution, Δ the average marker spacing in cM and $\nu(\mu)$ is a special function used by SIEGMUND (1985, p. 82) which can be approximated by $\nu(\mu) = \exp(-0.583\mu) + o(\mu^2)$.

Results: Results for $\alpha = 5\%$ from calculations based on both approximations (3) and (4) with those from simulations (10,000 replications) using the likelihood-based interval mapping method are given in Table 1. Values of $C = c^2$, the threshold for the test T_I are presented for $L = 100$ cM and various numbers of equidistant markers. We see that (3) gives the best approximation except for very high marker densities where (4) is better. This is because approximation (4) performs well when $c\sqrt{\Delta}$ converges to a finite limit so that the argument of ν is not extremely large. The DAVIES' approximation seems to be appropriate for threshold calculations for most experimental marker maps (26 to 3 markers per chromosome) and could be applied even

TABLE 2

Marginal and conditional distributions of QTL genotypes given the genotypes of the flanking codominant markers in F_2 populations with no double crossing over

Marker genotypes	Frequency[a]	Q_1Q_1 [b]	Q_1Q_2	Q_2Q_2
$A_1A_1B_1B_1$	$(1-p)^2/4$	1	0	0
$A_1A_1B_1B_2$	$p(1-p)/2$	$1-r$	r	0
$A_1A_1B_2B_2$	$p^2/4$	$(1-r)^2$	$2r(1-r)$	r^2
$A_1A_2B_1B_1$	$p(1-p)/2$	r	$1-r$	0
$A_1A_2B_1B_2$	$\frac{1}{2}[p^2+(1-p)^2]$	$r(1-r)/q$	$1-2r(1-r)/q$	$r(1-r)/q$
$A_1A_2B_2B_2$	$p(1-p)/2$	0	$1-r$	r
$A_2A_2B_1B_1$	$p^2/4$	r^2	$2r(1-r)$	$(1-r)^2$
$A_2A_2B_1B_2$	$p(1-p)/2$	0	r	$1-r$
$A_2A_2B_2B_2$	$(1-p)^2/4$	0	0	1

[a] p is the recombination fraction between the markers A and B.
[b] x is the recombination rate between the putative QTL and the left marker A, $r = x/p$ and $q = 1 + (1 - 1/p)^2$.

when the markers are not equally spaced. Calculations at the 1% level (results not shown) yield the same conclusions. Note that our simulations gave values which are different from those of DARVASI *et al.* (1993) (multiply their values by $2 ln 10 \approx 4.6$ to get them on a χ^2 scale). In fact, for 6 markers they proposed 6.58 and our simulations gave 6.9 with [6.73, 7.07] as a 95% confidence interval. This is probably due to: first, the difference in size of the population used (200 here against 1000 in DARVASI *et al.*) and second to different likelihood maximization algorithms. DARVASI *et al.* used a Newton-Raphson algorithm where the derivatives of the likelihood function were calculated analytically whereas we used the procedure E04AJF from the NAG library (1990) based on a quasi-Newton algorithm which proceeds numerically and has different convergence criteria.

THE MULTIPARAMETER CASE

General theory: Consider a population of size n and the corresponding linear regression model $Y = X\beta + e$ where $Y = (Y_1 \ldots Y_n)'$ the observation vector, X the (n, s) incidence matrix of known coefficients depending on the position (x) of test on the marker interval considered, β the column vector of the s unknown parameters and e the vector of residuals with means 0 and covariance matrix $\sigma^2 I$. Let $X = [X_0 \mid X_1]$ and $\beta = [\beta_0 \mid \beta_1]'$ where β_1 is the k vector of the QTL parameters (*e.g.*, additivity and dominance) and β_0 are the vector of other parameters (*e.g.*, general mean). X_0 is the submatrix (having elements 0 or 1) relative to β_0 and X_1 the submatrix corresponding to β_1 which elements depend on x and are calculated using Table 2. Suppose we are interested in testing H_0: $\beta_1 = 0$ against $\beta_1 \neq 0$ then the appropriate test statistic is (GRAYBILL 1976):

$$T(x) = \frac{Y'(XX^- - X_1 X_1^-)Y}{\sigma^2}$$

where $XX^- = X(X'X)^{-1}X'$. When the number of observations is large (say $n \geq 200$) σ^2 could be replaced by $\hat{\sigma}^2$, its linear model estimator, without any modification in the distribution of $T(x)$ which follows a χ^2_k under H_0.

Let us denote $Q = XX^- - X_1 X_1^-$ and $\lambda_1, \ldots, \lambda_k$ the non-zero eigenvalues of $Q(\partial Q/\partial x)^2 Q$ [see DAVIES (1987) for the underlying theory]. The significance level for $T(x)$ performed along the entire chromosome could be approximately calculated by:

$$\alpha = Pr(Sup(T(x)_{0 \leq x \leq P}) > C)$$

$$\approx Pr(\chi^2_k > C) \tag{5}$$

$$+ \frac{C^{((k-1)/2)}e^{-(C/2)}}{\sqrt{\pi 2^k}\Gamma((k+1)/2)} \sum_{i=1}^{m} \int_0^{p_i} E(\| \eta(x, p_i) \|) \, dx$$

where $\eta_j (j = 1, \ldots, k)$ are independent centered normal random variables with variances given by the λ_j and $\| \eta \| = (\eta'\eta)^{1/2}$. DAVIES (1987) used the approximation:

$$E(\| \eta(x, p_i) \|) \tag{6}$$

$$= (2\pi)^{-1/2} \int_0^\infty \left[1 - \prod_{j=1}^{k} (1 + \lambda_j(x)t)^{-1/2} \right] t^{-3/2} \, dt.$$

In the case where $k = 2$ the approximation above becomes:

$$E(\| \eta(x, p_i) \|) = \left(\frac{2\lambda_1}{\pi}\right)^{1/2} \xi\left(1 - \frac{\lambda_2}{\lambda_1}\right) \tag{7}$$

where $\lambda_1 \geq \lambda_2$ and ξ denotes a complete elliptic integral of the second kind (ABRAMOWITZ and STEGUN 1972, formula 17.3.3):

$$\xi(x) = \int_0^1 \left(\frac{1 - xt^2}{1 - t^2}\right)^{1/2} dt = \int_0^{\pi/2} (1 - x \sin^2\theta)^{1/2} \, d\theta. \tag{8}$$

Expressions (6) and (7) are due to HARVEY (1965). $\xi(x)$ could not be expressed analytically as a function of x but some approximations are available. One of these is (ABRAMOWITZ and STEGUN 1972, formula 17.3.35):

$$\xi(x) = [1 + a_1 z + a_2 z^2] + [b_1 z + b_2 z^2]\ln(1/z) + \epsilon(x), \epsilon(x)$$

$$< 4.10^{-5} \tag{9}$$

where $z = 1 - x$, $a_1 = 0.46301$, $a_2 = 0.10778$, $b_1 =$

TABLE 3

Thresholds at 5% level for the interval mapping test calculated for a chromosome of 100 cM and
different numbers of markers in F_2 population

Threshold	No. of markers							
	101	51	41	21	11	6	3	2
Simulations[a]		13.37	13.23	11.42	10.97	10.14	9.02	8.21
Davies approximation	14.19	13.29	12.89	11.70	10.65	9.73	8.63	8.42

[a] From 5,000 replications with population size of 200.

0.24527 and $b_2 = 0.04124$.

Alternatively, Davies (1987) showed that

$$E(\| \eta(x, p_i) \|) = E\left(\left| \frac{\partial T(x)^{1/2}}{\partial x} \right| \right) \pi^{1/2} \frac{\Gamma((k + 1)/2)}{\Gamma((k/2)}$$

and proposed to use the following approximation for the significance level

$$\alpha \approx \Pr(\chi_k^2 > C) \tag{10}$$

$$+ C^{((k-1)/2)} 2^{-(k/2)} \frac{e^{-(C/2)}}{\Gamma(k/2)} \sum_{i=1}^{m} | T(p_i)^{1/2} - T(0)^{1/2} |.$$

This formula can be used for approximate simulation-based calculation of the significance value for any experimental design. It is particularly useful when analytical expressions of the eigenvalues could not be obtained.

Application to F_2 populations: The model for F_2 populations is described in CARBONELL et al. (1992). Expectations of the genotypic classes for a pair of codominants markers where double recombination is neglected are given in the APPENDIX (A1) (based on Table 2). To simplify the evaluation of the eigenvalues we started by computing

$$A = X_1' X_1, \qquad B = \left(\frac{\partial X_1'}{\partial x}\right)\left(\frac{\partial X_1}{\partial x}\right), \qquad F = \left(\frac{\partial X_1'}{\partial x}\right) X_1.$$

These three matrices were found to be diagonal and therefore we just need to compute eigenvalues of

$$D = A^{-1}B - (A^{-1})^2 F^2 \tag{11}$$

as proposed by Davies (1987). Expressions for λ_1, λ_2, found after some formal algebra calculations using MAPLE (Char et al. 1988), are given in the APPENDIX (A1). The ratio $\lambda = \lambda_2/\lambda_1$ is always less than 1 and has its maximum value at $x = 0$ or p and its minimum for $x = p/2$. Approximation (5) applies here as:

$$\alpha \approx \Pr(\chi_2^2 > C) + \sqrt{\frac{C}{2\pi^3}} e^{-(C/2)} \sum_{i=1}^{m} \int_0^1 f(r, p_i) \, dr \tag{12}$$

where $f(r, p_i)$ is approximated by

$$f(r, p_i) = \sqrt{\lambda_1}([1 + a_1\lambda + a_2\lambda^2] + [b_1\lambda + b_2\lambda^2]\ln(1/\lambda))$$

where a_j and b_j are the constants defined in (9). Calcu-

lations from (12) were done using the eigenvalues found when double recombinations with no interference is assumed, which are too complicated to be included in this study. We have written a Fortran 77 program which computes, for any given level α and any set of linked markers on a chromosome, the threshold for QTL detection test in F_2 populations when additivity and dominance are considered. Simulations were done using the likelihood approach of interval mapping described by VAN OOIJEN (1992). The results, given in Table 3, show that the Davies' bound is a good approximation to the threshold, especially for marker densities which are customary in QTL mapping experiments (a marker very 5 to 20 cM). Thresholds were also calculated for the 1% level and results (not shown) confirm the good quality of the Davies-based approximations. VAN OOIJEN (1992) gave a value of 11.74 for the threshold at 5% level in F_2 population for a chromosome of 120 cM with markers each 5 cM based on 16,000 simulations. Our approximation gives 12.08 which is rather satisfactory.

TOWARD A GENERAL THRESHOLD APPROACH

The Davies' bound given by formula (5) could be computed for any experimental population and would give good approximations for the thresholds of tests when the population size is not too small ($n \geq 200$). However formal calculations needed could be difficult to carry out, particularly when three or more parameter are involved as in the case for diallel schemes. If the eigenvalues or the integral (6) could not be directly expressed as a function of x one could use either numerical approximations or expression (10) to carry out easy simulations.

We have studied the situation of F_3 progenies derived from a diallel cross between four inbreds as described in REBAÏ and GOFFINET (1993). The test $T = Y'(XX^- - X_0X_0^-)Y/\sigma^2$ which follows a χ_5^2, was used. Thresholds were calculated for a chromosome of 100 cM and different numbers of codominant markers each having four alleles using simulations with and without approximation (10). To use (10) one just needs to simulate the population under interest and to calculate for each replication the test statistics on marker positions (which is easier and faster than scanning all the chromosome), to compute the sum in (10) and then to calculate C by approximation (10) for the desired level α. The thresh-

TABLE 4

Thresholds for the interval mapping test calculated for a chromosome of 100 cM with 11, 6 and 3 markers in F_3 populations derived from a diallel cross

| | No. of markers at levels: | | | | | |
| | 11 | | 6 | | 3 | |
Threshold	5%	1%	5%	1%	5%	1%
Simulations	15.90	19.81	15.09	18.62	13.96	16.29
DAVIES simulations[b]	14.08	18.22	13.51	17.66	12.47	15.14

See text for further details.
[a] Based on approximation (10).

old is then the mean, over replications, of C values. Results for the diallel over 1,000 simulations with $\sigma^2 = 1$ and a total population size of 600 are shown in Table 4. The deviation between simulated and approximated values is larger than that previously calculated for BC or F_2, probably in relation with the small number of replications and individuals (six connected F_3 populations each with 100 individuals). Nevertheless (10) gives a quite satisfactory approximation and would be very useful when investigating the thresholds for tests in complex crossing designs.

Consider a species having H chromosomes. To ensure an overall significance level α for the test of QTL over the genome one could take a level α_i per chromosome. Traditionally α_i are taken all equal: $\alpha_i = 1 - (1 - \alpha)^{1/H}$. However α_i, the per-chromosome level, could be chosen by a manner which takes into account the relative lengths of all chromosomes. We hope to raise some new issues on this in future work. The level α_i, once chosen, is used to compute the corresponding threshold for the QTL detection test by the approach described in this paper.

Fortran 77 programs, using subroutines from the NAG library (1990), which calculate the thresholds for any given level and any distribution of codominant markers on a chromosome for BC, recombinant inbreds and F_2 populations have been written, compiled and successfully tested. Note that for recombinant inbreds the same approximation (3) is used, but there are minor transformations to obtain the integral [see APPENDIX (A2)].

For illustration consider the case of a genome with 12 chromosomes each of length 100 cM (similar to the tomato) and all having markers every 20 cM. To ensure a global significance level of 5% we use a nominal level of 0.5% for each chromosome and found thresholds of 11.64 (2.5 in LOD unit) and 14.65 (3.18) for BC and F_2 populations, respectively. These values are greater than those classically used for interval mapping (2.4 and 3, respectively). Moreover our method can provide a specific threshold for each chromosome relating to the actual distribution of markers on it. We are happy to make our programs available to interested readers.

We think that the use of the DAVIES' bounds is a good tool to calculate approximate thresholds for interval mapping tests in a wide variety of experimental designs.

LITERATURE CITED

ABRAMOWITZ, M., and I. A. STEGUN, 1972 Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. Dover, New York.

CARBONELL, E. A., T. M. GERIG, E. BALANSARD and M. J. ASINS, 1992 Interval mapping in the analysis of nonadditive quantitative trait loci. Biometrics 48: 305–315.

CHAR, B. W., K. O. GEDDES, M. B. GONNET, M. B. MONAGAN and S. M. WATT, 1988 MAPLE Reference Manual, Ed. 5. Symbolic Computation Group, Waterloo, Ontario, Canada.

DARVASI, B., A. WEIREB, V. MINKE, J. I. WELLER and M. SOLLER, 1993 Detecting Marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134: 943–951.

DAVIES, R. B., 1977 Hypothesis testing when nuisance parameter is present only under the alternative. Biometrika 64: 247–254.

DAVIES, R. B., 1987 Hypothesis testing when nuisance parameter is present only under the alternative. Biometrika 74: 33–43.

FEINGOLD, E., P. O. BROWN, and D. SIEGMUND, 1993 Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. Am. J. Hum. Genet. 53: 234–251.

HALDANE, J. B. S., 1919 The combination of linkage values and the calculation of distances between the loci of linked factors. J. Genet. 8: 299–309.

HALDANE, J. B. S., and C. H. WADDINGTON, 1931 Inbreeding and linkage. Genetics 16: 357–374.

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

HARVEY, J. R., 1965 Fractional moments of a quadratic form in noncentral normal random variables. Ph.D. Thesis, North Carolina State University, Raleigh, N.C.

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

KNAPP, S. J., W. C. BRIDGES and D. BIRKES, 1990 Mapping quantitative trait loci using molecular marker linkage maps. Theor. Appl. Genet. 79: 583–592.

LANDER, E. S., and D. BOTSTEIN, 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185–199.

LEADBETTER, M. R., G. LINDGREN and H. ROOTZEN, 1983 Extremes and Related Properties of Random Sequences and Processes. Springer, New York.

MARTINEZ, O., and R. N. CURNOW, 1992 Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. 85: 480–488.

MORENO-GONZALEZ, J., 1992 Genetic models to estimate additive and non-additive effects of marker-associated QTL using multiple regression techniques. Theor. Appl. Genet. 85: 435–444.

NUMERICAL ALGORITHMS GROUP, 1990 The NAG Fortran Library Manual–Mark 14. NAG Ltd., Oxford.

REBAÏ, A., and B. GOFFINET, 1993 Power of tests for QTL detection using replicated progenies derived from a diallel cross. Theor. Appl. Genet. **86:** 1014–1022.

REBAÏ, A., B. GOFFINET and B. MANGIN, 1994 Comparing power of different methods for QTL detection. Biometrics (*in press*).

SIEGMUND, D., 1985 *Sequential analysis: tests and confidence intervals.* Springer, New York.

VAN OOIJEN, J. W., 1992 Accuracy of mapping quantitative trait loci in autogamous species. Theor. Appl. Genet. **84:** 803–811.

Communicating editor: B. S. WEIR

APPENDIX

A1: The linear model for F_2 populations is $Y = X\beta + e$, where $\beta = (\mu, a, d)'$ and $\beta_1 = (a, d)'$. The reduced form of X is (Table 2):

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1-r & 1-2r & r & 0 & -r & 2r-1 & r-1 & -1 \\ 0 & r & 2r(1-r) & 1-r & s & 1-r & 2r(1-r) & r & 0 \end{pmatrix}$$

where $r = x/p$, $s = (1 - 2r + 2r^2)/(1 - (1 - 1/p)^2)$. In this case the vector of residuals e has variance matrix $\sigma^2 V$, where V^{-1} is a 9×9 diagonal matrix whose elements are the expected individual numbers in the nine marker classes (see Table 2). We then have, taking $\sigma^2 = 1$,

$$Q = V^{-1}(X(X'V^{-1}X)^{-1}X' - X_1(X_1'V^{-1}X_1)^{-1}X_1')V^{-1}.$$

Unfortunately calculations based on this expression of Q are too complicated. We then used an approximation which consists of working with a centered model $Z = X_1\beta + e$ with $Z = Y - X_0\hat{\beta}_0$, where $\hat{\beta}_0$ is the least square estimator of β_0 from the full model. So one can use

$$Q = V^{-1}(X_1(X_1'V^{-1}X_1)^{-1}X_1')V^{-1}.$$

We then verified that $A = X_1'V^{-1}X_1$, $B = (\partial X_1'/\partial x)V^{-1}(\partial X_1/\partial x)$ and $F = (\partial X_1'/\partial x)V^{-1}X_1$ are all diagonal and used (11) to derive eigenvalues. We found

$$\lambda_1 = \frac{-4(p^2 + (1 - p)^2)(a_0 + a_1 r + a_2 r^2 + a_3(2r^3 - r^4))}{(b_0 + b_1 r + b_2 r^2 + b_3(2r^3 - r^4))^2}$$

$$a_0 = -1 + 2p - p^2 - 3p^3 + 2p^4$$

$$a_1 = 4p(2 - 6p + 9p^2 - 4p^3)$$

$$a_2 = -4p(3 - 9p + 14p^2 - 7p^3)$$

$$a_3 = 4p(1 - 3p + 5p^2 - 3p^3)$$

$$b_0 = 2 - 6p + 8p^2 - 3p^3$$

$$b_1 = -4(1 - 3p + 4p^2 - p^3)$$

$$b_2 = 4(1 - 2p + 2p^+ 2p^3)$$

$$b_3 = -4p(1 - 2p + 3p^2)$$

$$\lambda_2 = \frac{4p(1 - p)}{(1 - 4pr + 4pr^2)^2}.$$

A2: For recombinant inbred lines produced by successive selfings, the recombination between two markers is $R = 2r/(1 + 2r)$ where r is the actual recombination rate (HALDANE and WADDINGTON 1931). Substituting x and p by their respective transforms $2t/(1 + 2t)$ in the covariance function of the BC (REBAÏ *et al.* 1994) leads to

$$\text{Cov}(T(x), T(y)) = \frac{p(1 - 2x - 2y + 4xy) + 8xy}{\sqrt{(p(1 - 2x)^2 + 8x^2)(p(1 - 2y)^2 + 8y^2)}}$$

and

$$\rho_{11}(x, p) = -\frac{8p}{(p(1 - 2x)^2 + 8x^2)^2}.$$

Analytical calculation of the integral gives

$$\int_0^p \sqrt{-\rho_{11}(x, p)} \, dx = \text{Arctan}\left(\sqrt{\frac{p}{2}}\right) + \text{Arctan}\left(\sqrt{\frac{p}{2}}(2p + 3)\right).$$