

Comparison of several confidence intervals for QTL location

B. MANGIN* & B. GOFFINET

Institut National de la Recherche Agronomique, Station de Biométrie et d'Intelligence Artificielle, BP 27, 31326, Castanet-Tolosan Cedex, France

The confidence interval for the map location of quantitative trait loci (QTL) is a very important quantity for geneticists. The one LOD support interval has been proved to be a biased confidence interval. Moreover the distribution of the LOD score has been shown to depend on the value of the QTL effect, which is why the LOD score cannot be used to build an unbiased confidence interval when the value of the QTL effect is unknown. A new confidence interval based on a maximum likelihood ratio test and using statistics whose asymptotic distribution does not depend on the QTL effect, has been proposed and proved to lead to an asymptotically similar confidence interval. The major difficulty of this method is the computation of the correct threshold for the maximum likelihood ratio test. An approximation for the threshold is proposed in this paper. When the value of the QTL effect is known, an unbiased confidence region could be built using the LOD score. A simulation study is carried out to compare the average length of this region, which is unobtainable for an unknown value of the QTL effect, to the average length of the asymptotically similar confidence interval. It shows that the required property of similarity does not increase the confidence interval length significantly for QTL having a small effect, and leads to an increase of about 5cM length for a 1 Morgan chromosome when the value of the QTL is great. An empirical symmetrical confidence interval could be constructed with the empirical distribution function of the maximum likelihood estimation for the QTL position. The simulation study shows that, when a QTL is detected, the average length of the asymptotically similar confidence interval could be half the length of the empirical symmetrical one. This great difference can be explained by the fact that the asymptotically similar confidence interval is dependent of the interval mapping test of detection whereas the empirical symmetrical one is independent of the actual data.

Keywords: confidence interval, local asymptotic framework, QTL location, similar test.

Introduction

The advent of maps of molecular markers enables geneticists to detect and map loci affecting quantitative traits (QTL). Since the work of Lander & Botstein (1989) a widely applied QTL mapping method is interval mapping, which leads to maximum likelihood estimation of the QTL map location. Lander & Botstein (1989) proposed to use as a confidence interval of the QTL position on the chromosome a one LOD support interval. Van Ooijen (1992) studied this type of confidence interval for backcross and F_2 populations. He found that

a two LOD support interval is necessary to ensure a 95 per cent confidence interval in his simulation study.

Mangin *et al.* (1994) proved that, from a theoretical point of view, a confidence interval could not be built out of the LOD score. The reason is that the LOD score is not similar to the value of the QTL effect, that is, the distribution of the LOD score depends on the value of the QTL effect, and therefore information on this value is necessary to build an unbiased confidence interval. We propose to build a new confidence interval using statistics for which the asymptotic distribution does not depend on the value of the QTL effect. With these statistics, at each position along the chromosome, a maximum

*Correspondence. E-mail: mangin@toulouse.inra.fr.

likelihood ratio test is computed to answer the question of the presence of a putative QTL. The set of positions not 'rejected' is an asymptotically similar confidence interval.

One of the major difficulties for a practical use of this method is the computation of the correct threshold for the maximum likelihood ratio test to ensure a given type I error. The first aim of this paper is to propose an approximate analytical threshold that could be computed for any map.

The length of the asymptotically similar confidence interval is of interest. For breeding purposes, as well as for the molecular biologist, a large confidence interval is unusable. In order to test if the required property of similarity does not lead to a too large interval, we build, when the value of the QTL effect is known, an unbiased LOD support interval. With a simulation study, we compare the length of the asymptotically similar confidence interval to the length of this unobtainable (when the value is unknown) LOD support interval. Another simulation study is carried out in order to compare the length of the Darvasi *et al.* (1993) empirical symmetrical confidence interval to the length of the asymptotically similar confidence interval. It shows that our confidence interval leads to more optimistic results, for the correct determination of the QTL location, than those published by Darvasi *et al.* (1993).

Model and asymptotically similar confidence interval

In a backcross population of size n , we consider a QTL present at the position d on a chromosome of length L . The trait value has a normal distribution with means μ_A and μ_B for the two QTL genotypes present in the backcross population and the same variance σ^2 for both genotypes. We will use $a = \mu_A - \mu_B$, which is the value of the QTL effect, and $\mu = (\mu_A + \mu_B)/2$, the global mean. For each individual $k = 1, \dots, n$, the phenotypic value of the trait y_k and a set of marker genotypes $M_{j,k}$ for $j = 1, \dots, J$ taking values A or B depending on the allele marker, are scored. In the following, we assume no interference in recombination events and therefore use Haldane's mapping function.

A similar confidence interval for a QTL position, built out of a test statistic T and a significance level α , is the set of values d_0 such that $T(d_0) \leq c_\alpha(d_0)$, where $c_\alpha(d_0)$ verifies $P(T(d_0) \leq c_\alpha(d_0)) = 1 - \alpha$ whatever the values of the nuisance parameters (μ , a , σ^2).

Note that the property of similarity for a confidence interval is a strong property that implies unbiasedness. The nuisance parameters μ and σ^2 could be estimated consistently, so there is no difficulty in ensuring similar (or invariant) procedures for them. The main difficulty is to obtain similar statistics for the a parameter, because a cannot be estimated consistently for QTL having a small effect. The idea of Cox & Hinkley (1974) for obtaining similar statistics is, in our model, to use only that part of the data which is conditional on sufficient statistics for the a parameter when the QTL is supposed to be located at d_0 . The second idea is to work in a local asymptotic framework for a , that is $a\sqrt{n}$ tends to a finite constant when n goes to infinity. Following these two ideas, Mangin *et al.* (1994) found asymptotically and locally similar statistics that are

$$Z_j(d_0) = \frac{1}{\sqrt{\hat{\sigma}^2}} \left(\frac{S_j}{1 - 2r_{j,d_0}} - \frac{S_{j+1}}{1 - 2r_{j+1,d_0}} \right),$$

where r_{j,d_0} denotes the recombination rate between the marker j and a QTL located at d_0 , $\hat{\sigma}^2$ denotes the classical estimator of the variance and

$$S_j = \frac{\sqrt{n}}{2} \left(\frac{\sum_k y_k \mathbf{1}_{[M_{j,k}=A]}}{\sum_k \mathbf{1}_{[M_{j,k}=A]}} - \frac{\sum_k y_k \mathbf{1}_{[M_{j,k}=B]}}{\sum_k \mathbf{1}_{[M_{j,k}=B]}} \right),$$

where $\mathbf{1}_{[M_{j,k}=.]}$ is the indicator of the events $[M_{j,k} = .]$. Note that the S_j statistics are the least square statistics for parameter a at each marker when they are divided by $\sqrt{\hat{\sigma}^2}$.

A score test of detection $T_Z(d_0)$, asymptotically equivalent to the maximum likelihood ratio test, is then computed out of the $Z_j(d_0)$ statistics, leading to

$$T_Z(d_0) = \sup_{0 \leq d \leq L} \frac{W^2(d, d_0)}{\text{Var}_\infty(W(d, d_0))},$$

where $W(d, d_0)$ is a linear function of $Z_j(d_0)$ and $\text{Var}_\infty(\cdot)$ denotes the asymptotic variance. Expressions of $W(d, d_0)$ and $\text{Var}_\infty(W(d, d_0))$ are given in the Appendix.

The asymptotically similar confidence interval is then defined by the set of points not 'rejected'.

Mangin *et al.* (1994) proved that under the hypothesis 'The QTL is located at position d_0 ', $Z(d_0)$ is asymptotically normal with mean 0 and variance V (given in Appendix). So $T_Z(d_0)$ is asymptotically the supremum of a χ_1^2 process.

Threshold approximation

To obtain a $1 - \alpha$ confidence interval it is necessary to compute the threshold function $c_x(d_0)$. This function depends on the length of the chromosome and the marker map. Mangin *et al.* (1994) gave values of threshold function for various numbers of equally spaced markers, obtained by simulations. However, an upper bound function, denoted $c_x''(d_0)$, could be found using Davies' (1977) approximation. This leads to the computation of $c_x''(d_0)$ as the solution of the following equation:

$$\alpha = \Phi(-c_x''(d_0)^2) + \frac{1}{2\pi} \exp\left(-\frac{c_x''(d_0)^2}{2}\right) \sum_j \int_0^{L_j} \sqrt{-\rho_j(t, d_0)} dt,$$

where L_j is the length of the j th interval, Φ denotes the standard normal cumulative distribution function and

$$\rho_j(t, d_0) = \frac{\partial^2 \text{Cor}_\alpha(W(t, d_0), W(t', d_0))}{\partial t'^2} \Big|_{t'=t}$$

for t, t' in the j th interval, where $\text{Cor}_\alpha(\cdot)$ denotes the asymptotic correlation. Using MAPLE (Char *et al.*, 1988), we found for the integral in each interval an analytical expression, given in the Appendix for the cases where d_0 is not on a marker.

Note that Rebaï *et al.* (1994) used the same approximation to compute an upper bound threshold for the interval mapping test of detection and showed by simulations that the upper bound threshold is very close to the correct threshold for most experimental marker maps (maps with markers every 50 cM to 5 cM).

The $c_x(d_0)$ threshold function has discontinuity points at each marker. This can be explained by the fact that when testing a putative QTL on a marker, the genotypes of individuals observed are known exactly, so such positions are more informative than the other positions of the linkage group. In practice, considering a discontinuous threshold function could lead to a confidence interval with holes at markers inside the interval. This type of interval could seem very strange for a user, so we prefer to build a continuous threshold function considering at marker points the right and left limits of the function.

Figure 1 presents the threshold functions $c_x''(d_0)$ for 3, 6 and 11 markers equally spaced on a 1 Morgan chromosome for a 10 per cent probability type I error. The plotted points are the empirical thresholds obtained by simulations over 50 000 repli-

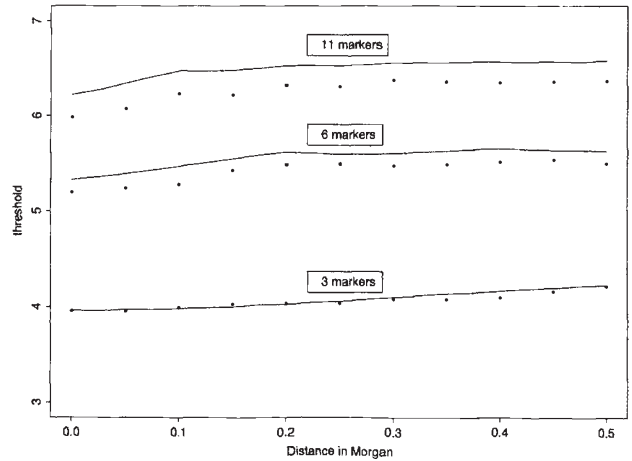


Fig. 1 Upper threshold function (lines) and empirical threshold (points) for 3, 6 and 11 markers equally spaced in a 1 Morgan chromosome.

cations. For equally spaced marker maps, the threshold function is symmetrical about the middle of the chromosome, so half of the threshold function is drawn. The difference between $c_x''(d_0)$ and the empirical threshold increases with the number of markers. The use of $c_{0.1}''(d_0)$ gives approximately a 90.5 per cent confidence interval for the 6 marker map and a 91 per cent confidence interval for the 10 marker map. This change in the coverage probability could be considered as negligible compare to the computation facilities provided by the use of $c_x''(d_0)$. Instead of time consuming simulations an analytical function could be programmed that provides a threshold function for any map. However, for dense maps, other approximate thresholds should be studied.

Comparison with other confidence intervals

An important point raised in our previous paper (Mangin *et al.*, 1994) is the possibility of a loss in power when using the asymptotically similar test for constructing a confidence interval. Such a loss could imply an increase of the length of the confidence interval if it is compared to an unbiased confidence interval built out of statistics that are not imposed to be similar. Indeed, the asymptotically similar confidence interval is built with statistics that use only the part of the data that does not depend on the nuisance parameters so objectively it loses part of the information. On the other hand, the LOD score is not similar but uses the whole data information. When information is available on the value of the QTL effect, it could be used to find a correct thresh-

old for the LOD score in such a way that an unbiased confidence interval is built out of the LOD score and this threshold. Obviously, the most precise piece of information is the actual value of the QTL effect, and the objective of this section is to compare the length of the asymptotically similar confidence interval to the length of the LOD support interval, in the idealized case where the value of the QTL effect is known.

Darvasi *et al.* (1993) proposed an empirical symmetrical confidence interval obtained with the empirical distribution of the maximum likelihood estimation for the QTL position. Their simulation study gave a very pessimistic impression on the possibility of correctly locating a QTL for experimental populations. We perform a part of their simulation study to compare the asymptotically similar confidence interval to the empirical symmetrical one.

Comparison with the support interval built out of the LOD score and using information on the value of the QTL effect

There are many difficulties in studying analytically the length of this support interval and the length of the asymptotically similar confidence interval, so simulations were carried out to obtain insights into the response.

The LOD support interval uses the statistics

$$T(d_0) = \log_{10} \frac{\sup_{\mu, \sigma^2, a, d} L_{\mathcal{M}}(Y; \mu, \sigma^2, a, d)}{\sup_{\mu, \sigma^2, a} L_{\mathcal{M}}(Y; \mu, \sigma^2, a, d = d_0)},$$

where $L_{\mathcal{M}}(Y; \mu, \sigma^2, a, d)$ denotes the likelihood of the observations given the information markers \mathcal{M} .

When the value of the QTL effect, a , is known, a threshold noted c_{α}^a could be found, that verifies

$$P(T(d_0) > c_{\alpha}^a) = \alpha.$$

The unbiased confidence interval built out of $T(d_0)$ and this threshold is the set of points d_0 not 'rejected'.

There are two major difficulties for the comparison of the lengths of the two intervals. The first is that we can obtain disconnected sets of points as a confidence interval and the second is that we must perform the comparison for exactly the same coverage probability. To avoid disconnected sets of points, we compare the sets

$$I_{\alpha} = [\min_d (T(d) < c_{\alpha}^a), \max_d (T(d) < c_{\alpha}^a)]$$

and

$$\tilde{I}_{Z, \alpha} = [\min_d (T_Z(d) < c_{\alpha}^u(d)), \max_d (T_Z(d) < c_{\alpha}^u(d))]$$

that give connected confidence intervals for both methods. In each simulation, we try many values of thresholds c_{α}^a and $c_{\alpha}^u(\cdot)$. We then perform the comparison for the thresholds that give approximately the same empirical coverage probabilities for both the intervals.

Table 1 presents for different values of parameter a the empirical average length of the two intervals and the empirical power for the LOD score test of detection, obtained with 1000 replications. Empirical powers are given for a 5 per cent (1 per cent) probability type I error per linkage group. Thresholds for the LOD score test of detection, that ensure a given type I error, are calculated using the proposition of Rebaï *et al.* (1994). For each replication, data for 200 backcross progeny were simulated with six markers equally spaced on a 1 Morgan chromosome and a QTL located in the middle of the chromosome.

Except for $a = 0.3$ the average lengths of the two intervals are significantly different. The asymptotically similar confidence interval is, on average, about 5 cM bigger for QTL detected with a 90 per cent power. This increase in the length is the price to be paid to be sure of getting an asymptotically unbiased confidence interval when the value of the QTL is unknown.

Comparison with Darvasi's results

We carried out part of the Darvasi *et al.* (1993) simulation study. This has 500 and 1000 backcross progenies with $\mu = 0$, $\sigma^2 = 1$ and $a = 0.25$ for 10, 20 and 50 cM map density and a QTL located in the middle interval at a relative position of 0, 1/4 and 1/2. Our goal is to compare the average length given by the asymptotically similar confidence interval to the length of the 95 per cent empirical symmetrical confidence interval given by Darvasi *et al.* (1993). In their paper, they proposed two other types of confidence interval, and it is clear that they are biased because they too are symmetrical and generally give a lower length.

We do not try to get exactly a 95 per cent confidence interval but prefer to use a naive procedure that is to fix $\alpha = 0.05$ and compute $\tilde{I}_{Z, 0.05}$ with the upper bound function $c_{0.05}^u$. We know that this procedure is conservative for two reasons: the threshold is an upper bound and a connected confidence interval

Table 1 Empirical average length (in cM) of I_z and $\bar{I}_{z,z}$, and empirical power of LOD score test of detection over 1000 replications. Data for 200 backcross progeny were simulated with a 100 cM chromosome with markers each 20 cM. The QTL was located in the middle of the chromosome, $\mu = 0$, $\sigma^2 = 1$ and empirical coverage probability was 90 per cent

a	I_z	$\bar{I}_{z,z}$	Power	
			Type I error	
	Average length	Average length	5%	1%
0.3	70.1 ± 1.9	72.1 ± 1.9	36%	17%
0.4	58.3 ± 1.8	63.8 ± 1.8	59%	32%
0.5	43.1 ± 1.5	50.1 ± 1.5	80%	59%
0.6	34.7 ± 1.2	40.7 ± 1.3	93%	81%
1	16.9 ± 0.4	21.7 ± 0.4	100%	100%

Table 2 Empirical average length (in cM) and coverage probability (in %) for $\bar{I}_{z,0.05}$ for 500 backcross progeny with $a = 0.25$ over 1000 replications and given the fact that a QTL is detected with the LOD score test of detection for a 5% (1%) probability type I error. Empirical power of the LOD score test of detection for both the levels is also given

	Length of marker intervals in cM								
	10			20			50		
	0	1/4	1/2	0	1/4	1/2	0	1/4	1/2 ^a
	50	47.5	45	40	45	50	50	37.5	25 ^b
<i>All replications</i>									
Average length	72	74	75	67	75	75	69	80	83
Coverage	97	96	95	97	97	96	98	96	96
<i>Only replications with QTL detected (Type I error 5%)</i>									
Power	61	59	58	66	57	57	69	52	41
Average length	56	58	58	52	59	59	59	67	64
Coverage	96	95	92	95	95	94	98	94	91
<i>Only replications with QTL detected (Type I error 1%)</i>									
Power	37	36	35	44	34	33	48	26	21
Average length	43	46	47	43	47	47	53	58	56
Coverage	95	94	91	94	93	91	98	92	87

^aRelative location of the QTL in the marker interval.
^bDistance of the QTL from one end of the chromosome.

is considered. So we compute for each simulation the empirical coverage probability.

We use the simulations to obtain information about a natural procedure involving building the confidence interval only if a QTL is detected using the LOD score test of detection for a 5 per cent (1 per cent) probability type I error per linkage group. The results for 500 backcross progeny obtained with 1000 replications are presented in Table 2.

In this case, Darvasi *et al.* (1993) found an empirically symmetrical confidence interval of length equal to 90 cM for an idealized dense map and varying from 61 to 95 cM for the map density studied. It is clear that this confidence interval is too large. One reason is of course the required property of symmetry. But the main reason why this confidence interval gives a very pessimistic impression for the possibility of accurately locating a QTL is that it is

independent of the interval mapping test of detection. That is not the case for $\bar{I}_{Z,0.05}$. Its average length given that a QTL is detected (for a 1 per cent probability type I error) is about half (for 10 cM and 20 cM map density) that of the empirical symmetrical one, whereas the coverage probability remains acceptable. This important decrease of the average

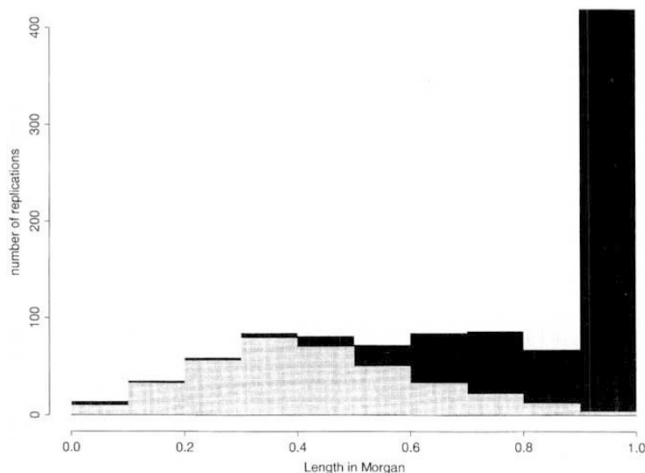


Fig. 2 Histogram for the distribution of $\bar{I}_{Z,0.05}$ length (black + grey boxes) and histogram for the distribution of $\bar{I}_{Z,0.05}$ length given the fact that a QTL is detected for a 1 per cent probability type I error (grey boxes).

length could be explained by the fact that when the QTL is not detected $\bar{I}_{Z,0.05}$ is often equal to the entire chromosome. Figure 2 illustrates this point, presenting for a QTL located in the middle of the chromosome, a map density of 10 cM and 500 backcross progeny, the histogram of $\bar{I}_{Z,0.05}$ length distribution over 1000 replications (grey + black boxes) and the histogram of $\bar{I}_{Z,0.05}$ length distribution for the replicates for which a QTL is detected (grey boxes).

Table 3 presents the results for 1000 backcross progeny obtained with 1000 replications. In that case, the two confidence intervals seem to be more similar. The reason is that the power of the interval mapping test of detection is about 90 per cent (80 per cent) for a 5 per cent (1 per cent) probability type I error. However, the average length of the asymptotically similar confidence interval for detected QTL with a 1 per cent probability type I error, is about 15 cM smaller than the average length of the empirical symmetrical confidence interval.

Discussion

Software for the computation of the asymptotically similar confidence interval for a backcross population is not yet available but could be programmed

Table 3 Empirical average length (in cM) and coverage probability (in %) for $\bar{I}_{Z,0.05}$ for 1000 backcross progeny with $a = 0.25$ over 1000 replications and given the fact that a QTL is detected with the LOD score test of detection for a 5% (1%) probability type I error. Empirical power of the LOD score test of detection for both the levels is also given

	Length of marker intervals in cM								
	10			20			50		
	0	1/4	1/2	0	1/4	1/2	0	1/4	1/2 ^a
	50	47.5	45	40	45	50	50	37.5	25 ^b
<i>All replications</i>									
Average length	44	48	50	44	52	55	52	67	69
Coverage	96	96	96	96	95	95	98	97	97
<i>Only replications with QTL detected (Type I error 5%)</i>									
Power	92	90	89	91	88	87	95	96	72
Average length	40	43	44	38	46	50	50	61	59
Coverage	96	95	95	95	94	95	98	96	95
<i>Only replications with QTL detected (Type I error 1%)</i>									
Power	80	76	75	81	73	68	84	64	50
Average length	34	37	38	34	41	43	47	56	54
Coverage	95	94	94	95	94	94	98	94	94

^aRelative location of the QTL in the marker interval.

^bDistance of the QTL from one end of the chromosome.

without great difficulty. At each position along the chromosome, the computations needed for constructing this interval are algebraic computations and one maximization of a one-parameter function. The time required for computation is not great compared to the interval mapping test of detection.

The methodology that leads to the asymptotically similar confidence interval could be extended to other breeding populations. For F_2 populations, asymptotically sufficient statistics using the mean differences between homozygous and heterozygous individuals at each marker could be found. Statistics, whose asymptotic distributions do not depend on the additivity and dominance parameters, are linear combinations of these sufficient statistics. Work is needed to derive their algebraic expression and the upper bound threshold, using Davies's (1987) approximation for the supremum of the χ^2_2 process. For more complicated breeding populations, such as the diallele cross described in Rebaï & Goffinet (1993), or for dominant markers, the methodology could be generalized using the linear regression model for QTL detection with flanking markers (Knapp *et al.*, 1990; Haley & Knott, 1992). Note that even in backcross or F_2 populations with codominant markers, if there is a lot of missing marker data, the construction of the asymptotically similar confidence interval must use the generalized methodology via the linear regression model. Linear regression with flanking markers has been proved to be asymptotically and locally equivalent to an interval mapping method based on the maximum likelihood ratio test (Rebaï *et al.*, 1995).

For QTLs having great effect, i.e. those detected with power equal to one, the $T(d_0)$ statistic is asymptotically similar, so that the LOD support interval is asymptotically unbiased (Mangin *et al.*, 1994) and compares well to the asymptotically similar confidence interval. The question is how could we know if the detected QTL is a QTL with a great effect? Common thinking is that insights of the response could be found by looking at the LOD curve. A sharp LOD curve with a great maximum is probably linked to a large QTL. Simulation studies could give knowledge like, for example, a maximum LOD score reaching 10 with fewer than 500 individuals (15 with fewer than 1000 individuals) is the sign of a great value for the QTL effect, because these quantities are never observed with QTLs having small effect, but theoretical work is needed in this field. This question will become more and more important if the use of multiple QTL methods as described by Jansen (1993) and Zeng (1993) becomes common. Indeed, use of markers as cofactors tends to reduce

the environmental variance as power of detection increases. We can say, using a figure of speech, that a QTL with a small effect in a single QTL analysis could become a QTL with a great effect in a multiple QTL analysis.

This paper has studied the confidence interval for the QTL location map. An unbiased confidence interval for the value of the QTL effect is also of interest and requires new methodology to take into account QTL having small effect.

References

- CHAR, B. W., GEDDES, K. O., GONNET, G. H., MONAGAN, M. B. AND WATT, S. M. 1988. *MAPLE Reference Manual*. Symbolic Computation Group, Waterloo, Ontario, Canada.
- COX, D. R. AND HINKLEY, D. V. 1974. *Theoretical Statistics*. Chapman and Hall, London.
- DARVASI, A., WEINREB, A., MINKE, V., WELLER, J. I. AND SOLLER, M. 1993. Detecting marker QTL gene effect and map location using a saturated genetic map. *Genetics*, **134**, 943–951.
- DAVIES, R. B. 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**, 247–254.
- DAVIES, R. B. 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **74**, 33–43.
- HALEY, C. S. AND KNOTT, S. A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315–324.
- JANSEN, R. C. 1993. Interval mapping of multiple quantitative trait loci. *Genetics*, **135**, 205–211.
- KNAPP, S. J., BRIDGES, W. C. AND BIRKES, D. 1990. Mapping quantitative trait loci using molecular linkage maps. *Theor. Appl. Genet.*, **79**, 583–592.
- LANDER, E. S. AND BOTSTEIN, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185–199.
- MANGIN, B., GOFFINET, B. AND REBAÏ, A. 1994. Constructing confidence intervals for QTL location. *Genetics*, **138**, 1301–1308.
- REBAÏ, A. AND GOFFINET, B. 1993. Power of tests for QTL detection using replicated progenies derived from diallel cross. *Theor. Appl. Genet.*, **86**, 1014–1022.
- REBAÏ, A., GOFFINET, B. AND MANGIN, B. 1994. Approximate thresholds of interval mapping tests for QTL detection. *Genetics*, **138**, 235–240.
- REBAÏ, A., GOFFINET, B. AND MANGIN, B. 1995. Comparing power of different methods for QTL detection. *Biometrics*, **51**, 87–99.
- VAN OOIJEN, J. W. 1992. Accuracy of mapping quantitative trait loci in autogamous species. *Theor. Appl. Genet.*, **84**, 803–811.
- ZENG, Z. B. 1993. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proc. Natl. Acad. Sci. U.S.A.*, **136**, 1457–1468.

Appendix

[A1] The expression of $W(d, d_0)$ is

$$W(d, d_0) = \begin{cases} v_{i_{d_0}} Z_{i_{d_0}}(d_0) + v_{i_d} Z_{i_d}(d_0) + \sum_{j=i_d+1}^{i_{d_0}-1} Z_j(d_0) & \text{for } i_d < i_{d_0} \\ Z_{i_{d_0}}(d_0) & \text{for } i_d = i_{d_0} \\ \tilde{v}_{i_{d_0}} Z_{i_{d_0}}(d_0) + \tilde{v}_{i_d} Z_{i_d}(d_0) + \sum_{j=i_{d_0}+1}^{i_d-1} Z_j(d_0) & \text{for } i_d > i_{d_0} \end{cases}$$

with

$$v_{i_{d_0}} = \frac{x_{i_{d_0}}^2 + 1, d_0 (1 - x_{i_{d_0}}^2, d_0)}{x_{i_{d_0}}^2, d_0 (1 - x_{i_{d_0}}^2 + 1, d_0) + x_{i_{d_0}}^2 + 1, d_0 (1 - x_{i_{d_0}}^2, d_0)}$$

$$v_{i_d} = \frac{x_{i_d, d}^2 (1 - x_{i_d+1, d}^2)}{1 - x_{i_d, d+1}^2}$$

$$\tilde{v}_{i_{d_0}} = 1 - v_{i_{d_0}}$$

$$\tilde{v}_{i_d} = \frac{x_{i_d+1, d}^2 (1 - x_{i_d, d}^2)}{1 - x_{i_d, d+1}^2},$$

where $x_{a,b} = 1 - 2r_{a,b}$ and i_d (i_{d_0}) denotes either the interval of the point d (d_0) or the left marker of this interval.

Because V is a diagonal matrix with diagonal element equal to

$$V_{jj} = \frac{1}{x_{j, d_0}^2} - 2 \frac{x_{j+1}}{x_{j, d_0} x_{j+1, d_0}} + \frac{1}{x_{j+1, d_0}^2}$$

the algebraic expression of $\text{Var}_\infty (W(d, d_0))$ is easily found.

[A2] For d_0 not on a marker

$$\int_0^{L_j} \sqrt{-\rho_j(t, d_0)} dt = \begin{cases} 0 & \text{for } j = i_{d_0} \\ \arctan \left(\sqrt{\frac{A_j}{\text{Var}_\infty(Z_j(d_0))}} \right) & \text{else} \end{cases}$$

where

$$A_j = \begin{cases} v_{i_{d_0}}^2 \text{Var}_\infty(Z_{i_{d_0}}(d_0)) + \sum_{k=j+1}^{i_{d_0}-1} \text{Var}_\infty(Z_k(d_0)) & \text{for } j < i_{d_0} \\ \tilde{v}_{i_{d_0}}^2 \text{Var}_\infty(Z_{i_{d_0}}(d_0)) + \sum_{k=i_{d_0}+1}^{j-1} \text{Var}_\infty(Z_k(d_0)) & \text{for } j > i_{d_0}. \end{cases}$$

Proof: the case $j = i_{d_0}$ is trivial so we leave it.

First it can be proved that instead of working in terms of distance t , any one-to-one twice-derivable func-

tion of t could be used for the computation of the integral because

$$\frac{\partial \text{Cor}_\infty (W(t, d_0), W(t', d_0))}{\partial t'} \Big|_{t'=t} = 0.$$

Using Haldane’s map function and

$$\begin{cases} t \rightarrow u = \frac{e^{-4t} - x_{j,j+1}^2}{1 - x_{j,j+1}^2} & \text{for } j < i_{d_0} \\ t \rightarrow u = \frac{e^{-4(L_j-t)} - x_{j,j+1}^2}{1 - x_{j,j+1}^2} & \text{for } j > i_{d_0}. \end{cases}$$

leads to

$$\int_0^{L_j} \sqrt{-\rho_j(t, d_0)} dt = \int_0^1 \sqrt{-\frac{\partial^2 C_j(u, u')}{\partial u'^2} \Big|_{u'=u}} du,$$

where

$$C_j(u, u') = \frac{A_j + \text{Var}_\infty (Z_j(d_0)) uu'}{\sqrt{A_j + \text{Var}_\infty (Z_j(d_0)) u^2} \sqrt{A_j + \text{Var}_\infty (Z_j(d_0)) u'^2}}.$$

If $A_j \neq 0$, differentiating twice $C_j(u, u')$ and integrating gives the result.

A_j could be equal to 0, only when $v_{i_{d_0}} = 0$ and $j = i_{d_0} - 1$ or $\tilde{v}_{i_{d_0}} = 0$ and $j = i_{d_0} + 1$. These two cases correspond to d_0 on a marker. This remark finishes the proof.