Genet. Res., Camb. (2004), **83**, *pp.* 41–47. *With* 1 *figure.*
DOI: 10.1017/S0016672303006554 Printed in the United Kingdom

A comparison between methods for linkage disequilibrium fine mapping of quantitative trait loci

JIHAD M. ABDALLAH[1]*, BRIGITTE MANGIN[2], BRUNO GOFFINET[2], CHRISTINE CIERCO-AYROLLES[2] AND MIGUEL PÉREZ-ENCISO[1]

[1] *Station d'amélioration génétique des animaux, Institut national de la recherche agronomique, Auzeville BP 27, 31326 Castanet-Tolosan Cedex, France*

[2] *Unité de biométrie et intelligence artificielle, Institut national de la recherche agronomique, Auzeville BP 27, 31326 Castanet-Tolosan Cedex, France*

(*Received 7 July 2003 and in revised form 1 September 2003*)

Summary

We present a maximum likelihood method for mapping quantitative trait loci that uses linkage disequilibrium information from single and multiple markers. We made paired comparisons between analyses using a single marker, two markers and six markers. We also compared the method to single marker regression analysis under several scenarios using simulated data. In general, our method outperformed regression (smaller mean square error and confidence intervals of location estimate) for quantitative trait loci with dominance effects. In addition, the method provides estimates of the frequency and additive and dominance effects of the quantitative trait locus.

1. Introduction

Linkage disequilibrium (LD), or non-random allelic association between loci, has become an important fine mapping tool after the initial success in mapping Mendelian disease genes (e.g. Hästbacka *et al.*, 1992; Kerem *et al.*, 1993; Snell *et al.*, 1989). Initially applied to binary traits, there is a growing interest among breeders and geneticists in methods that use LD for mapping quantitative trait loci (QTLs) (Farnir *et al.*, 2002; Meuwissen and Goddard, 2000; Slatkin, 1999). Traditionally, linkage analysis was used in mapping QTLs, which relies on following the segregation of the phenotype and marker alleles in structured pedigrees. The main difficulty with quantitative traits is the weak phenotype–genotype relationship. As for complex discrete traits, quantitative traits are influenced by environmental factors and, usually, by multiple genes. In addition, the resolution of linkage analysis approach is limited by the low number of recombinations in the pedigree data (usually two to three generations are available; Meuwissen and Goddard, 2000) and the accuracy for the QTL position is generally within several cM. The main advantage of LD mapping over linkage analysis is that it makes use of historical recombinations, resulting in higher resolution of gene location.

In this paper, we generalize an existing method used in mapping disease genes (Terwilliger, 1995) to allow for mapping of QTLs. In his work, Terwilliger (1995) proposed a powerful likelihood test that is not restricted by the number of marker alleles or the number of markers considered jointly. The test is based on maximizing the likelihood over an association parameter λ, defined as the proportion of increase of marker allele i in disease chromosomes, relative to its population frequency. Our extension of the method provides estimates of genetic effects and frequency of the QTL under additive and dominance models. We considered three analyses in our method, single marker, two marker and six marker analyses, and made paired comparisons among them and between these analyses and single marker regression.

2. Materials and methods

(i) *Single marker analysis*

Let us assume that a QTL is segregating in a population with two alleles, Q and q, with frequencies P_Q and $1-P_Q$, where allele Q resulted from a mutation t

* Corresponding author. Tel: +33 5 61 28 51 82. Fax: +33 5 61 28 53 53. e-mail: abdallah@germinal.toulouse.inra.fr

Table 1. *Conditional probabilities of QTL alleles on marker alleles, with allele i associated with the* Q *allele*

	Marker allele	
QTL allele	i	$j\ (j \neq i)$
Q	$P_Q + [\lambda(1-p_i)P_Q/p_i]$	$P_Q - \lambda P_Q$
q	$(1-P_Q) - [\lambda(1-p_i)P_Q/p_i]$	$(1-P_Q) + \lambda P_Q$
Total	1	1

generations ago, and that this allele has a positive effect on a quantitative trait. Consider also a linked polymorphic marker M with m alleles. Suppose that, initially, marker allele i was completely associated with the Q allele, the likelihood is

$$L_i = \prod_{n=1}^{N} \left(\sum_{k=1}^{3} P_{in}(g_k) \phi\,(y_n; \mu_k, \sigma^2) \right),$$

where: N is the number of observations and subscript k takes values 1, 2 and 3 for QTL genotypes QQ, Qq, and qq, respectively; $P_{in}(g_k)$ is the probability of the nth individual having QTL genotype g_k conditional on marker allele i being associated with the mutant QTL; $\phi(.)$ is the probability density function of a normal distribution; y_n is the phenotypic record; σ^2 is a common within-genotype variance; and μ_k is the genotypic mean for the kth QTL genotype. Using Falconer parameterization of genotypic means (Falconer and Mackay, 1996), $\mu_1 = \mu + a$, $\mu_2 = \mu + d$ and $\mu_3 = \mu - a$, with a and d defined as the additive and dominance deviations from μ, the mean of the homozygote QTL genotypes. Here, we assume a panmictic population in Hardy–Weinberg equilibrium. The $P_{in}(g_k)$ probabilities are calculated as follows

$$P_{in}(g_1) = P_i(Q|M_n^1)\, P_i(Q|M_n^2),$$

$$P_{in}(g_2) = P_i(q|M_n^1)\, P_i(Q|M_n^2) + P_i(Q|M_n^1)\, P_i(q|M_n^2),$$

$$P_{in}(g_3) = P_i(q|M_n^1)\, P_i(q|M_n^2),$$

where M_n^1 and M_n^2 are the alleles of marker M on the male and female haplotypes. Notice here that the phase is not required to be known. The probabilities of QTL alleles conditional on marker alleles (Table 1) were derived following the same parameterization as in Terwilliger (1995). For example, if the marker allele on M_n^1 is i and the allele on M_n^2 is j $(j \neq i)$ for an individual then the probability $P_{in}(g_1) = [P_Q + \lambda(1-p_i)\, P_Q/p_i]\ (P_Q - \lambda P_Q)$, where p_i is the frequency of allele i in the population, and λ is the proportion of excess of allele i on Q bearing chromosomes. Specifically, λ is defined such that $P(i|\mathrm{Q}) = p_i + \lambda(1-p_i)$.

Finally, as the particular allele i is unknown, Terwilliger proposed to integrate it out as

$$L = \sum_{i=1}^{m} p_i L_i,$$

where the sum is over all marker alleles. The likelihood is maximized over the parameters λ, P_Q, μ, a, d and σ^2. The likelihood ratio test statistic (LRT) is calculated as

$$LRT = -2 \ln \left(\frac{\max L(H_0)}{\max L(H_1)} \right),$$

where $L(H_0)$ is the likelihood evaluated under the null hypothesis of no linkage disequilibrium $(\lambda = 0)$ and $L(H_1)$ is the likelihood evaluated under the alternative hypothesis $(\lambda > 0)$. The analysis is repeated for all markers in the hypothesized region of the QTL and the location of the marker with maximum LRT over all markers is taken as an estimate of the location of the QTL.

(ii) *Multiple marker analysis*

Terwilliger (1995) modelled the parameter λ as a function of θ (the recombination fraction), α (the proportion of Q alleles originally in association with allele i) and t (number of generations since the mutation occurred) as $\lambda = \alpha(1-\theta)^t$. For the multiple marker case, Terwilliger proposed that the likelihood can be computed by multiplying together the likelihoods for each marker and using the previous relationship. The recombination fraction, θ, between any map position and each marker locus can be determined using an appropriate mapping function (e.g. Haldane's function). The parameters t and α can be fixed or estimated as nuisance parameters. Terwilliger (1995) indicated that it is both powerful and conservative to treat them in the latter manner. The combined likelihood is maximized over α and t in addition to P_Q, μ, a, d and σ^2 for every postulated QTL position. The null hypothesis is $\lambda = 0$ (i.e. $\alpha = 0$ or $t = \infty$) and the LRT is computed as before. The maximum LRT over all map positions is the most likely estimate of the QTL location. This way of forming the likelihood is approximate, because it does not take into account the correlations (i.e. LD) among marker loci.

(iii) *Regression analysis*

The phenotypic trait value y_n of individual n is regressed on the number of copies x_n of allele i of marker M according to the regression model

$$y_n = x_0 + \sum_i b_i x_{ni} + e_n,$$

where x_0 is the population mean of the quantitative trait, b_i is the regression coefficient on allele i of marker M and e_n is the residual error of the nth individual. The F statistic to test significant association of marker M with QTL is obtained by testing the model above against the model $y_n = x_0 + e_n$; that is, we test the overall association of marker alleles on the trait. The corresponding P values (the probability of an F value as large as or larger than the observed F statistic given the null hypothesis of no association (Weisberg, 1985)) are obtained using the appropriate degrees of freedom. The location of the marker that shows the lowest P value is taken as the estimate of the QTL location.

(iv) *Simulations*

The simulation strategy was the same as in Abdallah *et al.* (2003) and full details are given there. Briefly, a set of equally spaced markers (0·25 cM, 1·0 cM or 2·0 cM) were simulated on a chromosomal region of 10 cM in a founder population of 200 individuals. In subsequent generations, offspring haplotypes were sampled by the gene dropping method. Recombinations were modelled using Haldane's mapping function (Haldane, 1919). After 20 generations of random mating, a QTL mutation with a positive effect on the trait was introduced in one haplotype of a single random individual at position 3·6 cM. This results in complete initial LD between the QTL locus and other loci in the region. Data on QTL and marker loci were recovered from generation 120 (100 generations after the mutation was introduced).

We used either biallelic (single-nucleotide polymorphism (SNP)) or multiallelic (microsatellite (MST)) markers. We assumed five alleles per MST marker. Initially, all markers had equal allele frequencies. The alleles of MST markers were allowed to mutate at a rate of 10^{-4} per generation using a stepwise mutation model (i.e., an allele increased or decreased its count by one). Mutation was assumed to be negligible for SNP markers. Replicates were discarded when fixation occurred for the QTL or any of the markers. We also discarded replicates when the frequency of the Q allele was less than 0·05 because rare QTL alleles account for a small proportion of the variance and are not of interest in mapping studies.

The phenotype, y_n, of the quantitative trait for an individual was simulated as $y_n = g_n + e_n$, where g_n is the genetic value of the QTL genotype (a, d or $-a$) of the nth individual and e_n is an environmental value drawn from a normal distribution with mean 0 and variance of 1·0. We considered values of $a = 1{\cdot}0$, $d = 0$ (no dominance) and $d = 1{\cdot}0$ (complete dominance).

The location of the QTL was estimated using single marker regression analysis, single marker Terwilliger analysis (T1) and multiple marker Terwilliger analysis. In multiple marker analysis, we tested positions every 0·2 cM on the chromosomal region using the closest two markers to the position (T2) and closest six markers (T6). All analyses were performed on the same replicate. Bias in the QTL location was calculated as the average of the signed difference between the location estimate and the true location. Lower and upper limits of 90% confidence intervals (CI) were determined empirically by the 5th and 95th percentiles. The estimates of the QTL location were compared among the four analysis methods using the mean square error (MSE) (i.e. $\sum_{i=1}^{r} (\hat{s}_i - s)^2 / r$, where r is the number of replicates, $\hat{s}_i$ is the estimated QTL location and s is the true QTL location). Similarly, for analysis using Terwilliger method, MSE was calculated for estimates of P_Q, a, d and t. The MSE contains information about both the bias and the variance of the location estimate. Differences between methods in MSE of QTL location estimates were tested using a paired t test (tests of normality of distributions by Shapiro–Wilk test showed no significant deviations from normality).

The time required for the maximization process to converge is proportional to the number of parameters and the number of markers in the analysis. In order to complete our simulations in a realistic time, we tried two algorithms from the NAG library (Numerical Algorithms Group, 1990). The first (Routine E04CCF) uses the Simplex method and the second (Routine E04JYF) is a quasi-Newton algorithm based on estimating the gradient and curvature of the function. Routine E04CCF provided larger LRT values more often than did routine E04JYF for the same replicates and so was adopted in our simulations. We expect that the use of a more powerful algorithm (like simulated annealing algorithms) might result in more accurate parameter estimates but would require much longer time for convergence.

3. Results and Discussion

Bias in QTL location estimates is in Table 2. All mapping analyses showed significant ($P < 0{\cdot}05$) bias in location estimates towards the 'right-hand' side of the true QTL location (except for T1 with marker spacing of 0·25 and $d = 1{\cdot}0$, where bias was not significant). The bias was smaller when the QTL had a dominant effect than when it had a strict additive action. The bias increased as marker spacing increased. Part of the bias in the location is due to the presence of the true QTL on the left-hand side and not in the middle of the chromosomal region, which explains why the mean location estimate fell to the right of the true location. However, the contribution of intrinsic bias (the bias caused by limiting the length of the region between 0 cM and 10 cM) to the total bias was found to be negligible. The expected intrinsic bias was

Table 2. *Bias in QTL location estimates obtained using regression and Terwilliger-based QTL mapping methods*

Marker type	Marker spacing	d*	Analysis method‡ R	T1	T2	T6
SNP	0·25 cM	0·0	0·21	0·36	0·34	0·35
		1·0	0·15	0·08[ns]	0·13	0·17
	1 cM	0·0	0·68	0·48	0·60	0·49
		1·0	0·28	0·23	0·32	0·30
	2 cM	0·0	0·77	0·81	0·80	1·14
		1·0	0·42	0·21	0·43	0·38
MST	0·25 cM	0·0	0·10[ns]	0·17	0·12	0·20
		1·0	0·16	0·07[ns]	0·13	0·14
	1 cM	0·0	0·37	0·41	0·39	0·33
		1·0	0·24	0·15	0·17	0·19
	2 cM	0·0	0·50	0·50	0·45	0·45
		1·0	0·44	0·34	0·19	0·29

* Dominance effect of QTL.
‡ R, regression analysis; T1, Terwilliger-based single marker analysis; T2, Terwilliger-based analysis with two markers; T6, Terwilliger-based analysis with six markers.
[ns] Bias is not significantly different from 0 ($P \geqslant 0{\cdot}05$); otherwise, bias is significant ($P < 0{\cdot}05$).

calculated assuming a normal distribution with mean 3·6 and variance approximated by the variance of the location estimate.

Estimates of P_Q and additive effects of QTLs (data not shown) were also biased ($P < 0{\cdot}05$). Estimates of P_Q were biased upwards and additive effects were biased downwards, but bias was generally smaller when the QTL allele Q is dominant ($d = 1{\cdot}0$). Estimates of dominance effects were not biased when the QTL was dominant but were biased upwards when the QTL was not dominant ($d = 0$).

MSE values of QTL location estimates are in Table 3. The MSE includes the variance of the location estimate plus the variance caused by the bias of the estimate. Although the bias in location estimates was significant, its contribution to the MSE was small (1–14 %). As with bias, the MSE increased as marker spacing increased. The MSE was lower when the QTL had additive and dominance effects compared with additive effects only. Clearly, MST markers had lower MSE than SNP markers for all mapping analyses.

Paired differences in MSE of QTL location estimates between mapping analyses are in Table 4. With few exceptions, the regression method had higher MSE than Terwilliger-based methods (T1, T2 and T6) but differences in MSE were significant ($P < 0{\cdot}05$) only when the QTL was dominant. T2 generally had lower MSE than T1, with one exception (marker spacing of 0·25 cM using MSTs and $d = 1{\cdot}0$). No consistent trend was found for differences between T1 and T6. T2 had significantly lower MSE than T6 when marker spacing was 2 cM.

Table 3. *MSEs of QTL location estimates obtained using regression and Terwilliger-based QTL mapping methods*

Marker type	Marker spacing	d*	Analysis method‡ R	T1	T2	T6
SNP	0·25 cM	0·0	3·04	3·64	3·28	3·12
		1·0	2·26	1·85	1·76	1·67
	1 cM	0·0	6·15	5·81	5·71	5·64
		1·0	4·44	3·99	3·53	3·85
	2 cM	0·0	7·93	7·93	7·75	8·88
		1·0	6·49	5·30	4·57	6·15
MST	0·25 cM	0·0	1·60	1·91	1·67	1·59
		1·0	1·26	0·82	1·09	1·03
	1 cM	0·0	3·83	4·12	3·58	4·18
		1·0	3·33	2·61	2·08	2·13
	2 cM	0·0	5·39	5·34	4·96	5·31
		1·0	4·32	3·48	3·11	3·72

* Dominance effect of QTL.
‡ R, regression analysis; T1, Terwilliger-based single marker analysis; T2, Terwilliger-based analysis with two markers; T6, Terwilliger-based analysis with six markers.

MSEs of P_Q, a and d were generally large. For a 2-cM marker spacing, these ranged from 0·006 to 0·07 for P_Q, from 0·30 to 1·24 for a, and from 0·51 to 2·00 for d. The MSE decreased in simulations with marker spacing of 0·25 cM and ranged from 0·002 to 0·06, 0·22 to 0·72 and 0·29 to 1·07 for P_Q, a, and d, respectively. Among Terwilliger analyses, T1 had a lower MSE for a than T2 and T6 but a higher MSE for P_Q.

Estimates of t, the number of generations since the mutation occurred, were generally biased downwards (data not shown). The bias ranged from $-62{\cdot}5$ to $+18{\cdot}4$ generations for T2 and $-71{\cdot}1$ to $-9{\cdot}0$ generations for T6. More importantly, the estimates of t had very large variances and therefore large confidence intervals. Estimates by T6 were often less variable than estimates by T2. Standard deviations ranged from 50·5 to 581·6 for T2 and from 28·0 to 312·0 for T6. Estimates using MST markers had smaller standard deviations than SNP markers. For both marker types, estimates had smaller variances when the QTL had additive and dominance effects than when it had additive effects only. No clear trends in bias or variances of estimates were found for marker distances. The behaviour of estimates of t might be due to the decay model used (i.e. modelling λ in relation to α and t). Terwilliger (1995) stated that, if map distances are overestimated, the estimate of t would be small and that, if they are underestimated, the estimate of t is large. In our results, estimates of QTL location were biased upwards and that might explain the downward bias in t. We should note here that, in our simulations, estimates of t had

Table 4. *Differences in MSEs of QTL location estimates between regression and Terwilliger-based QTL mapping methods*

Marker type	Marker spacing	d[1]	R − T1[2]	R − T2	R − T6	T1 − T2	T1 − T6	T2 − T6
SNP	0·25 cM	0·0	−0·59**[3]	−0·24	−0·08	0·36	0·52	0·16
		1·0	0·42*	0·50**	0·60**	0·09	0·18	0·10
	1 cM	0·0	0·34	0·44	0·51	0·10	0·17	0·07
		1·0	0·45*	0·91***	0·59	0·46*	0·13	−0·32
	2 cM	0·0	0·00	0·18	−0·95	0·17	−0·96*	−1·13***
		1·0	1·20***	1·92***	0·34	0·73**	−0·86*	−1·58***
MST	0·25 cM	0·0	−0·31	−0·08	0·01	0·24	0·32	0·08
		1·0	0·44**	0·17	0·23	−0·27***	−0·22*	0·06
	1 cM	0·0	0·30	0·25	−0·35	0·55	−0·06	−0·60**
		1·0	0·66***	1·15***	1·37***	0·50***	0·71***	0·21
	2 cM	0·0	0·05	0·43	0·09	0·38	0·04	−0·34
		1·0	0·78**	1·08***	0·33	0·30	−0·45	−0·76**

[1] Dominance effect of QTL.
[2] R, regression analysis; T1, Terwilliger-based single marker analysis; T2, Terwilliger-based analysis with two markers; T6, Terwilliger-based analysis with six markers.
[3] Difference is significantly different from 0. *** $P<0{\cdot}01$; ** $P<0{\cdot}05$; * $P<0{\cdot}10$.

Table 5. *Empirical 90% confidence limits of QTL location estimates obtained using regression and Terwilliger-based QTL mapping methods*

Marker type	Marker spacing	d*	Analysis method‡ R	T1	T2	T6
SNP	0·25 cM	0·0	1·2–7·5	1·5–8·3	1·4–8·0	1·4–7·6
		1·0	1·5–6·5	1·5–6·0	1·6–6·2	1·8–6·4
	1 cM	0·0	1·0–9·0	0·0–9·0	0·6–9·0	0·4–9·0
		1·0	0·0–8·0	1·0–8·0	1·2–7·8	0·8–8·0
	2 cM	0·0	0·0–10·0	0·0–10·0	0·0–9·8	0·4–9·8
		1·0	0·0–10·0	0·0–8·0	0·2–8·2	0·0–9·4
MST	0·25 cM	0·0	1·8–6·0	2·0–6·0	1·8–6·0	2·2–6·4
		1·0	2·0–5·5	2·3–5·0	2·2–5·6	2·4–5·4
	1 cM	0·0	1·0–8·0	1·0–9·0	1·2–8·0	0·6–8·0
		1·0	1·0–7·0	1·0–7·0	1·2–6·2	1·0–6·2
	2 cM	0·0	0·0–8·0	0·0–9·1	0·7–8·7	0·6–9·3
		1·0	0·0–8·0	2·0–8·0	1·2–7·8	0·6–8·0

* Dominance effect of QTL.
‡ R, regression analysis; T1, Terwilliger-based single marker analysis; T2, Terwilliger-based analysis with two markers; T6, Terwilliger-based analysis with six markers.

very small effects on the values of maximum LRT (i.e., the likelihood conveys very little information about t).

Table 5 shows the empirical 90% CI for QTL location estimates. A larger CI indicates more uncertainty about the QTL's location. Here, the CIs obtained for marker spacing of 1·0 cM and 2·0 cM were large (5–10 cM). Such a resolution is less than desired for fine mapping. However, the use of a denser map (marker spacing of 0·25 cM) resulted in smaller CI (2·7–6·8 cM). Smaller spacing than 0·25 cM might be required to improve further the resolution of the QTL location. However, dense maps (less than 0·25 cM) are available in human genetic studies but not in animals, in which marker density is usually more than 1·0 cM. The results in Table 5 reflected the trends in MSE (Table 3). Terwilliger-based methods had smaller CIs than regression analysis when the QTL was dominant. This was more evident for marker spacing of 2 cM. Use of biallelic markers (SNPs) provided larger CIs than multiallelic markers (MSTs). However, SNPs are more abundant in genetic maps than MSTs and it is difficult to have MST maps as dense as for SNPs.

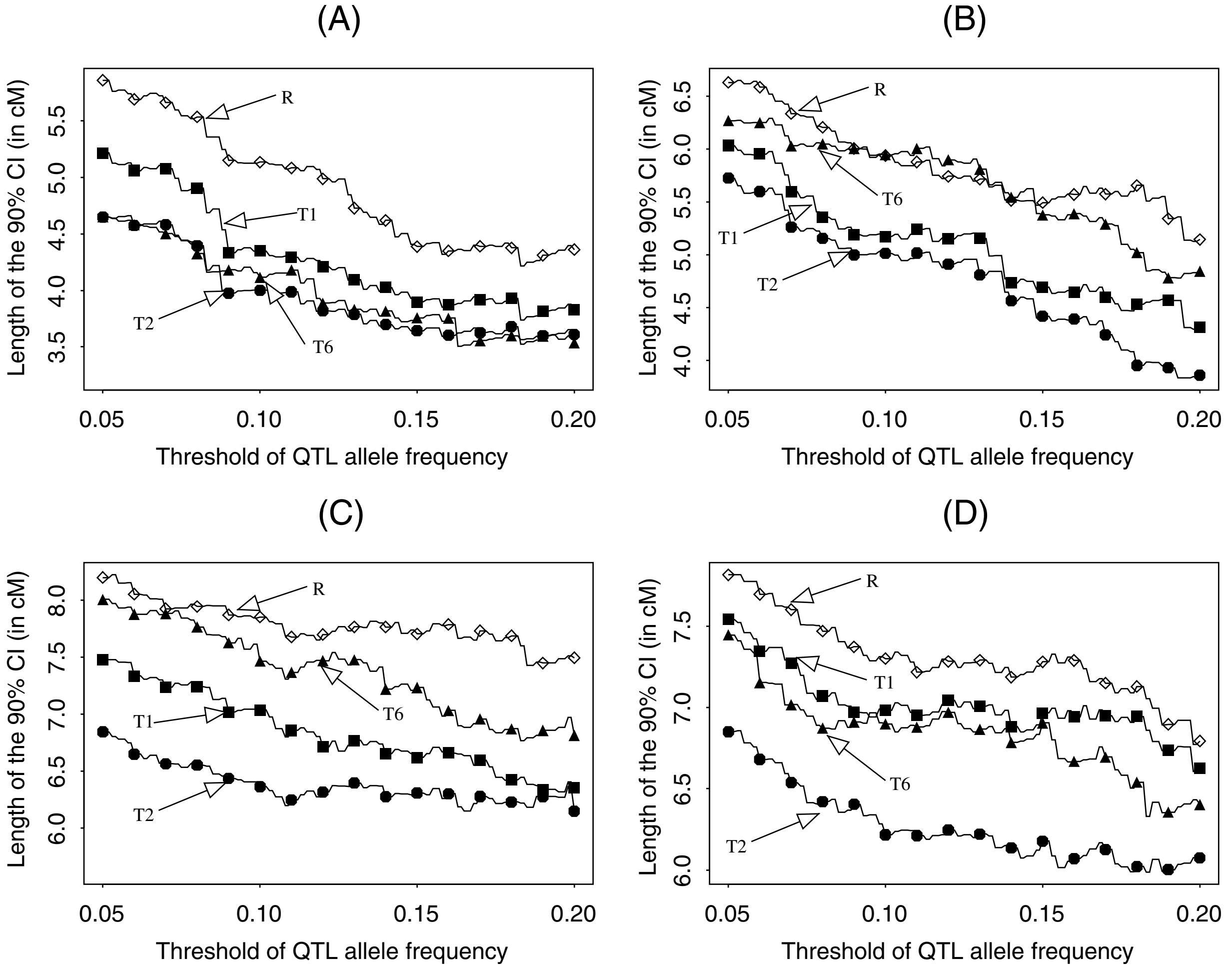

Fig. 1. Length of approximate normal 90 % confidence intervals as a function of threshold of QTL frequency (frequency of Q allele); for example, a threshold of 0·05 means all replicates with QTL frequency equal or greater than 0·05. In these simulations, QTL frequency ranged between 0·05 and 0·50. (A) Marker spacing of 1 cM using MSTs. (B) Marker spacing of 2 cM using MSTs. (C) Marker spacing of 2 cM using SNPs. (D) Marker spacing of 2 cM using SNPs, but marker alleles had minimum frequency of 0·05. In all four simulations, $a = 1{\cdot}0$ and $d = 1{\cdot}0$. R = regression analysis; T1 = Terwilliger-based single marker analysis; T2 = Terwilliger-based analysis with two markers; T6 = Terwilliger-based analysis with six markers.

MSEs and length of CI of all parameter estimates decreased as the frequency of Q allele, P_Q, increased. This is due to the increase of the variance explained by the QTL. Fig. 1 illustrates this by showing the length of 90 % CI of the estimated location as a function of threshold of P_Q. Here, we used approximate normal CIs that were not different from empirical ones but the trends were clearer. Fig. 1A, B shows simulations with MST markers with, respectively, 1·0 cM and 2·0 cM spacing. Fig. 1C, D shows simulations with SNP markers with 2 cM spacing, but the minimum frequency of marker alleles was set to 0·05 cM for simulations in Fig. 1D. In all the cases, the T2 test gave the smallest CI, whereas regression analysis gave the largest ones almost every time. The CI obtained by T1 test were almost halfway between those of T2 and regression. The trend of T6 test was less stable. In Fig. 1A, the CI of T6 looked like those obtained with T2, whereas T6 behaved more like regression when marker spacing increased (Fig. 1B). Limiting the minimum frequency of alleles did not really change the length of CI except for T6, which was the only method with a clear decrease in CI length from Fig. 1C to Fig. 1D.

We have seen that Terwilliger-based QTL analyses outperformed regression analysis (lower MSE and smaller CI) in mapping QTLs with additive and dominance effects, although the precision was comparable for a strict additive action. The advantage of Terwilliger analyses over regression analysis is that they allow an explicit modelling of the additive and dominance effects of the QTL. The disadvantage of likelihood-based methods (as for Terwilliger analysis) is that the maximization process might fail to converge to the global maximum.

Of the Terwilliger analyses, T2 performed the best overall. Generally, one would expect the use of more markers to result in more precise estimates of the location as more information is used. It seems that, in the case of T6, this is hindered by the difficulty of the maximization process. Terwilliger (1996) noted that the admissible proportion of the total parameter space becomes smaller and smaller with the increase in number of markers. The result is that, sometimes, the global maximum is not achieved, particularly with small data sets or with the presence of rare marker alleles. To test the effect of rare marker alleles, we carried out simulations such that the minimum frequency of any marker allele was at least 0·05. Setting a minimum on allele frequencies resulted in a decrease of 0·18 to 0·70 in MSE of the QTL location for T6. The length of CI for T6 became smaller than that for T1 and the advantage of T2 over T6 was attenuated (see Fig. 1D).

Another factor is that, in T6, the assumptions of marker independence are more flagrantly violated. Accounting for correlations between markers is complicated, especially when one considers more than two markers jointly (a simple measure of LD among several markers needs to be implemented). In fact, this is the subject of future work to improve the precision of the method. In our analysis here, we did not perform any test of significance and neither did we make any assumptions about the distribution of the statistics used. We simply estimated the position of the QTL by taking the maximum value of the statistic over all test positions. Regardless of the distribution of the statistic, the threshold values for test of significance can be determined from the data using permutation (e.g. Churchill and Deorge, 1994).

One factor that affects the efficiency of all LD mapping methods is the high variability of LD. In a previous study and using the same simulation strategy, we found that it is common to have the strongest association (or maximum disequilibrium) with the more distant markers from the QTL (Abdallah *et al.*, 2003). The variability of LD depends on allele frequency and high variability is found for extreme frequencies.

4. Conclusions

We presented a maximum-likelihood method for LD mapping of QTLs that is a generalization of the method of Terwilliger (1995) used in mapping disease genes. The method uses information from single and multiple markers with no restriction on the number of markers used. However, the use of two markers resulted in smaller MSE and CI of QTL location estimate than either single- or six-marker strategies. The advantage of the method over regression analysis seemed to be in modelling additive and dominance effects of the QTL. The method provides estimates of frequency and additive and dominance effects of the QTL, but these estimates had large variability.

A FORTRAN F-90 program (QTLTER) to carry out the described analyses is available and can be requested from the corresponding author. Funding for this work was provided by project 20 (2001–2002) of the Bureau des resources génétiques and a project within Action en bioinformatique of the Ministère de la Recherche (France). We thank the editor and two anonymous referees for their suggestions, which helped to improve the manuscript.

References

Abdallah, J. M., Goffinet, B., Cierco-Ayrolles, C. & Pérez-Enciso, M. (2003). Linkage disequilibrium fine mapping of quantitative trait loci: a simulation study. *Genetics Selection Evolution* **35**, 513–532.

Churchill, G. A. & Deorge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th edn. Essex, UK: Longman.

Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Moisio, S., Simon, P., Wagenaar, D., Vilkki, J. & Georges, M. (2002). Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics* **161**, 275–287.

Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distance between loci of linked factors. *Journal of Genetics* **8**, 299–309.

Hästbacka, J., de la Chappelle, A., Kaitila, I., Sistonen, P., Weaver, A. & Lander, E. (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics* **2**, 204–211.

Kerem, B. S., Romens, J. M., Buchanan, J. A., Markiewics, D., Cox, T. K., Lehesjoki, A., Koskiniemi, J., Norio, R., Tirrito, S., Sistonen, P., Lander, E. S. & de la Chapelle, A. (1993). Localization of the *EPM1* gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Human Molecular Genetics* **2**, 1229–1234.

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

Numerical Algorithms Group (1990). *The NAG Fortran Library Manual-Mark 19*. Oxford, UK: Numerical Algorithms Group.

Slatkin, M. (1999). Disequilibrium mapping of a quantitative-trait locus in an expanding population. *American Journal of Human Genetics* **64**, 1765–1773.

Snell, R., Lazarou, L., Youngman, S., Quarrell, O., Wasmuth, J., Shaw, D. & Harper, P. (1989). Linkage disequilibrium in Huntington's disease: an improved localisation for the gene. *Journal of Medical Genetics* **26**, 673–675.

Terwilliger, J. D. (1995). A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *American Journal of Human Genetics* **56**, 777–787.

Terwilliger, J. D. (1996). Reply to Sham *et al. American Journal of Human Genetics* **58**, 1095–1096.

Weisberg, S. (1985). *Applied linear regression*, 2nd edn. New York: John Wiley.