Mapping and analysis of quantitative trait loci in *Lycopersicon* (tomato) with the aid of genetic markers using approximate maximum likelihood methods

J. I. Weller*

Department of Genetics, the Hebrew University of Jerusalem, Jerusalem 91904, Israel.

1691 F-2 progeny of a cross between *Lycopersicon esculentum* and *L pimpinellifolium* grown under field conditions were scored for 18 quantitative traits of economic interest and 10 segregating genetic markers. Each parental strain was homozygous for one allele of each marker. Four of the markers were electrophoretic, and six were morphological. Three pairs of the genetic markers were linked. An algorithm described previously based on maximum likelihood technique was used to estimate the parameters of loci affecting the quantitative traits linked to the genetic markers and the recombination distance between quantitative trait loci and marker loci. The parameters of quantitative trait loci linked to two genetic markers were also estimated by solving for gene effect and recombination frequency from the independent equations derived from each marker. In general there was close correspondence between estimates obtained from the two methods. Except for cases where highest likelihood was obtained at complete linkage, results of the approximate maximum likelihood technique were within the parameter space, *i.e.* recombination frequencies between zero and 0·5 and positive variance estimates. Unreasonable results were obtained when the assumptions of the method were violated. These results support those presented previously based on simulated results and a positive control indicating that, for samples of this size, accurate estimates are derived by the maximum likelihood technique. A genetic map of quantitative trait loci is presented.

INTRODUCTION

Most traits of economic importance in higher organisms can be classified as quantitative in nature, *i.e.*, the observed phenotype is the joint result of the roughly additive effect of a large number of genetic and environmental factors. Theoretical (Elston and Steward, 1973; Hill, 1975; Jayakar, 1970; Soller *et al.*, 1976; Soller *et al.*, 1979), and applied studies (Tanksley *et al.*, 1982; Vallejos and Tanksley, 1983; Zhuchenko, *et al.*, 1979*b*; Weller *et al.*, 1986) have demonstrated the utility of marker-based methods for the indentification of quantitative trait loci (QTL).

Weller (1986) presented an algorithm based on an approximate maximum likelihood (ML) technique to estimate the linkage distance between a QTL and a linked genetic marker, and the biometric parameters of the QTL. Accurate estimates

were derived with this method for simulated data for QTL with effect greater than 1·0 phenotypic standard deviation, and for a known QTL affecting plant height linked to an electrophoretic marker. In this study results are presented on the application of this method to twelve quantitative traits measured in a population of 1691 F-2 progeny of a cross between *Lycopersicon esculentum* and *L. pimpinellifolium*. These results are also compared to estimates of QTL effect and location obtained by an alternate method applicable when two genetic markers were both linked to the same QTL.

MATERIAL AND METHODS

Description of the experiment. Experimental details were presented by Weller *et al.* (1986). Strain LA13 of *L. esculentum* (provided by Dr C. M. Rick) was crossed with strain CIAS27 of *L. pimpinellifolium* (provided by Dr R. Frankel), with the latter as male parent. Eighteen quantitative

traits were measured on the field grown plants. The parental strains differed in six morphological and four electrophoretic markers. The morphological markers were all recessive, but four of them, anthocyanineless (*a*), potato leaf (*c*), dwarf (*d*) and lutescent (*l*) were discernable in month-old seedlings. Heterozygous and homozygous dominants were distinguished by progeny test. It was not feasible to progeny test for the remaining two morphological markers, yellow fruit (*r*) and pigmentless epidermis (*y*), and therefore they were not included in the ML analysis. Of the electrophoretic markers three were peroxidase loci (*per*-1, *per*-3, and *per*-b), and one was an esterase locus (*est*).

Description of the maximum likelihood analysis. The approximate ML algorithm used to determine the recombination proportion between the QTL and the genetic marker (r_L), and the biometric parameters of the QTL was described by Weller (1986). The algorithm was applied only in those cases where a significant effect on a quantitative trait in the F-2 population was associated with one of the eight genetic markers where all three genotypes were determined. The ML method is contigent upon the following assumptions: (a) The distribution of the trait values within each QTL genotype is normal. (b) The effect on the quantitative trait associated with the genetic marker is due to only one QTL. (c) The effect of the QTL on the quantitative trait can be both additive and multiplicative, but not of a higher order. As noted by Zhuchenko *et al.* (1979*a*), multiplicative QTL effects will result in differing within genotype variances for the quantitative trait.

Six of the quantitative traits measured: number of nodes to first flower cluster, plant size, ratio of fruit to vegetative growth, flowering earliness, fruit earliness, and number of locules per fruit, were unsuitable for analysis by ML because their distributions were not continuous. All the remaining traits except plant height had significant skewness or kurtosis. Since the ML analysis method assumes an underlying normal distribution, the following transformations were applied:

$$y = (ax + b)^c \qquad (1)$$

$$y = d(\log(x + e)) + f \qquad (2)$$

where *y* is the transformed trait value, *x* is the untransformed trait value, and *a*, *b*, *c*, *d*, *e* and *f* are constants. Equation (1) was used for moderately skewed distributions ($g1$ values below 1·0)

and internode length, and equation (2) was used for more highly skewed distributions. The constants *b*, *c* and *e* were chosen to minimise skewness in the distribution of transformed values. The constants *a*, *d* and *f* did not affect the nature of the distribution and were included so that the range of values for *y* would be convenient for analysis:

To test the validity of the second assumption of the ML technique, the expected number of linked detectable QTL per genetic marker for each trait was estimated from the frequency of genetic markers without significant effects. Assuming a Poisson distribution of effects, the expected number of linked detectable QTL per genetic marker is equal to $-\ln p$, where *p* is the proportion of genetic markers without significant effects. The validity of the third assumption was not tested prior to the analyses.

When approximate ML was obtained at a non-zero recombination value the alternative hypothesis: complete linkage, was tested by the likelihood ratio between approximate ML and the likelihood at complete linkage. The hypothesis of complete linkage was rejected if the likelihood ratio was greater than 5·0, which corresponds in a normal distribution to the difference in likelihood between the mean and a point 1·8 standard deviations distant.

When approximate ML was obtained at complete linkage there was no obvious alternative hypothesis, and therefore an approximate "confidence interval" was estimated. The parameter space having a likelihood $> \frac{1}{5}$ the likelihood at complete linkage was searched for the combination giving the highest value for recombination proportion. This recombination proportion, denoted r_C, was taken as the boundary value of the confidence interval.

As shown previously on simulated data (Weller, 1986), the ability of the method to distinguish between complete and partial linkage is positively correlated with the magnitude of the effect of the QTL. This result was tested on the marker-trait combinations with approximate ML at complete linkage, by computing the correlation between r_C and the effect of the QTL, estimated as $|\bar{X}_{M11} - \bar{X}_{M22}|$, where $\bar{X}_{M11}$ and $\bar{X}_{M22}$ are means for the quantitative trait in F-2 standard deviation units of the individuals homozygous for alleles 1 and 2 of the marker.

The regression of $|\bar{X}M11 - \bar{X}M22|$ on r_C was used to estimate the power of the method as a function of QTL effect, *i.e.*, the ability of the method to detect incomplete linkage if present.

This regression was denoted as follows:

$$r_C = g + h(|\bar{X}_{M11} - \bar{X}_{M22}|) \qquad (3)$$

where g is the y-intercept and h is the regression. Assuming symmetry, the recombination proportion with a marginal power of 0·95 to reject the hypothesis of complete linkage for a given marker-linked effect can be computed from the left-hand side of equation (3).

The method was also tested on a sample of marker-trait combinations that violated the assumptions on which the method was based.
Other statistical analyses. Two additional sources of information were used to corroborate the results of the maximum likelihood algorithm. Zhuchenko *et al.* (1979*a*) have shown that partial linkage between a QTL and a genetic marker should result in a difference in skewness between the F-2 groups homozygous for the opposite alleles of the genetic marker. The homozygote with the higher mean should have the lower skewness value. Therefore skewness, as estimated by $g1$ was measured for each quantitative trait within each marker locus genotype. To test significance, the difference between $g1$ values for the two marker homozygous groups was divided by the approximate standard error of the difference:

$$S.E. = (6/n_1 + 6/n_2)^{1/2} \qquad (4)$$

where n_1 and n_2 are the sample sizes of the two homozygous groups, respectively (Weller *et al.*, 1986).

In addition, the ten genetic markers included three linked pairs. An estimate of linkage distance, denoted r_B, was derived, if the difference between the double homozygous means was greater than the difference between the single homozygous means for both loci. This method was used for all eighteen quantitative traits, including the six traits with discrete distributions. It was assumed that: (a) only one QTL was responsible for the effect, (b) this QTL is located in the chromosomal segment between the markers, and (c) recombination intereference was not found in this chromosomal segment, *i.e.*, $r = r_1 + r_2$, where r is the recombination frequency between the two markers and r_1 and r_2 are the recombination proportions between each of the genetic markers and the QTL. The following equations can then be formulated:

$$\bar{X}_{M11N11} = (1 - r_1 r_2)\mu_{A11} + 2r_1 r_2(1 - r_1 r_2)\mu_{A12}$$
$$+ (r_1 r_2)^2 \mu_{A22} \qquad (5)$$

$$\bar{X}_{M22N22} = (1 - r_1 r_2)\mu_{A22} + 2r_1 r_2(1 - r_1 r_2)\mu_{A12}$$
$$+ (r_1 r_2)^2 \mu_{A11} \qquad (6)$$

where $\bar{X}_{M11N11}$ and $\bar{X}_{M22N22}$ are the means of the F-2 individuals homozygous for the linked markers M and N; μ_{A11}, μ_{A12}, and μ_{A22} are the means of the three QTL genotypes, respectively; and the other terms are as defined above. The product $r_1 r_2$ will be maximum when $r_1 = r_2$. Even if the recombination proportion between the genetic markers is 0·4, $r_1 r_2$ will be no greater than 0·04, and $(r_1 r_2)^2$ will be no greater than 0·0016. Assuming that there were no events of double recombination within this chromosomal segment, the following approximate equality can then be derived:

$$\bar{X}_{M11N11} - \bar{X}_{M22N22} = \mu_{A11} - \mu_{A22} \qquad (7)$$

Then after Soller *et al.* (1976),

$$\bar{X}_{M11} - \bar{X}_{M22} = (1 - 2r_1)(\mu_{A11} - \mu_{A22}) \qquad (8)$$

$$\bar{X}_{N11} - \bar{X}_{N22} = (1 - 2r_2)(\mu_{A11} - \mu_{A22}) \qquad (9)$$

where $\bar{X}_{M11}$ and $\bar{X}_{M22}$ are the means of the individuals homozygous for alleles 1 and 2 of marker M, respectively; and $\bar{X}_{N11}$ and $\bar{X}_{N22}$ are the means of the individuals homozygous for alleles 1 and 2 of marker N, respectively. Substituting from equation (7), and solving for r_1 and r_2:

$$r_1 = 1/2 - (\bar{X}_{M11} - \bar{X}_{M22})/2(\bar{X}_{M11N11} - \bar{X}_{M22N22}) \qquad (10)$$

$$r_2 = 1/2 - (\bar{X}_{N11} - \bar{X}_{N22})/2(\bar{X}_{M11N11} - \bar{X}_{M22N22}) \qquad (11)$$

If $\bar{X}_{M11}$ was significantly different from $\bar{X}_{M22}$, but $\bar{X}_{N11}$ was not significantly different from $\bar{X}_{N22}$, then r_1 was computed from equation (10) and r_2 was computed as $r - r_1$. If both markers had significant effects on the trait, then both r_1, and r_2 were computed from equations (10) and (11). The final estimate for r_1 was the mean of this estimate and $r - r_2$, with r_2 estimated from equation (11).

RESULTS

The transformations used and the biometric parameters of the transformed values are listed in table 1. Although the transformations were constructed to minimize skewness, kurtosis was also reduced in the transformed distributions. After transformation, kurtosis was above 0·5 for only three traits: internode length, the number of flowers per cluster and fruit index. The ML algorithm is therefore inappropriate for these traits.

The marker loci having significant effects on the quantitative traits are given in Weller *et al.*, 1986. The expected number of detectable QTL per

Table 1 Descriptive statistics of the quantitative traits in transformed units

Trait	Transformation*	Mean	S.D.	Range	Skewness	Kurtosis
1. Internode length	$(x.10^6)^{0.2}$	18·57	1·03	11·5–22·7	0·005	3·407
2. Flowers per cluster	$2 \log (x - 3·5)$	30·68	1·88	20·4–38·9	0·003	1·606
3. Stem width	$(65·8x - 526)^{0.55}$	30·31	9·32	0·0–60·0	−0·006	0·036
4. Plant height†	—	180·10	59·60	20·0–300·0	−0·061	0·447
5. Leaf weight	$20 \log (x + 0·1) + 20$	21·73	5·07	4·0–37·0	0·001	0·035
6. Fruit weight	$20 \log (x - 0·88) + 4$	15·96	4·36	1·1–32·4	0·001	0·079
7. Fruit index	$(100x)^{0.7}$	23·12	0·87	18·1–27·9	0·004	2·123
8. Weight of 20 seeds	$(10x - 100)^{0.52}$	56·26	7·43	31·3–80·3	0·001	0·336
9. Total soluble solids	$x^{0.56}$	2·65	0·28	1·7–3·6	0·003	0·001
10. pH	$(928x - 3249)^{0.6}$	57·45	7·17	32·1–83·0	−0·027	0·285
11. Vitamin C	$30 \log (x + 15) - 30$	19·23	4·30	7·0–35·0	0·006	0·151
12. Sugar concentration	$30 \log (x + 0·6) + 3$	16·98	6·23	2·0–32·0	0·000	0·147

* x is the trait value in untransformed units.
† A transformation was not applied for plant height as the distribution of this trait was nearly normal.

genetic marker, computed from the frequency of markers without significant effects is given in table 2. For fruit weight, this quantity could not be estimated since all 10 markers had significant effects on this trait, but must be >2·3, which is the value obtained if 9 of the 10 markers were to have significant effects. Thus for fruit weight and also for total soluble solids (TSS), for which 8 of the markers had significant effects, the assumption that there is only one QTL in the vicinity of the genetic marker with a significant effect, is probably incorrect.

Results for marker-trait combinations with approximate ML obtained at recombination values greater than zero are presented in table 3. The sample size, the skewness difference between the homozygous genotypes in standard error units, the r_B value for cases where pairs of linked markers

Table 2 Expected number of detectable QTL per genetic marker*

Traits	Assumed number of effects per marker
1. Internode length (cm)	0·4
2. Number of flowers per cluster	0·4
3. Stem width (cm)	0·5
4. Plant height (cm)	0·7
5. Leaf weight (gm)	0·7
6. Fruit weight (gm)	—
7. Fruit index (%)	0·7
8. Weight of 20 seeds (mg)	0·5
9. Total soluble solids (%)	1·2
10. pH	0·4
11. Vitamin C (mg/100 ml)	0·4
12. Reducable sugar (g/100 ml)	0·5

* This estimate was computed from the frequency of markers without significant effects, as described in the text.

were available, the observed means and variances of the marker locus genotypes for the quantitative traits, the means and variances of the genotypes of the QTL estimated at approximate ML, the r_L value, the log likelihoods at complete linkage and at ML, and the ratio of likelihoods between ML and complete linkage are listed.

The first five combinations in table 3 are for cases in which pairs of linked marker loci allow calculation of r_B. The effect of *est* on plant height has been discussed in detail previously (Weller, 1986). The ML results are in close correspondence to the hypothesis that the QTL for this effect is the recessive locus *d*, located a distance of 38 map units from the locus *est*. r_B values were not computed for the effect of *per-b* on plant height, and the effect of *d* on seed weight, because in both cases, the difference between the double homozygotes of the linked pair of loci was no greater than the effect associated with the marker that had a significant effect. This indicates that the QTL is not located between the two markers, and hence the r_B estimate is not appropriate. Of the three cases where positive r_B values were computed there was a significant discrepancy only for the effect of *per*-3 on leaf weight. In this case the difference between the two estimates was 14 map units, which is probably well within the confidence interval of the estimates, as the sample size for this combination was only 654 plants. The likelihood ratios for the effects of *c* and *per*-3 on leaf weight, and the effect of *c* on *p*H were <5·0. Thus the hypothesis of complete linkage cannot be rejected for these combinations, despite that for two of these loci, r_L was above 0·30.

Significant skewness differences were found only for the effects of *est* and *l* on plant height. The skewness difference for *per-b* on plant height

Table 3 Maximum likelihood results for marker-trait combinations which correspond to assumptions of the model, and with ML at partial linkage

Trait[a]	Marker	Skewness difference[b]	Estimated recombination proportion[c]	Mean[d] pp	pe	ee	Variances pp	pe	ee	log likelihood[e]	likelihood ratio
Stem width (731)	per-3	1·2	0·26 / 0·26	2·9 / 2·7	3·0 / 3·1	3·2 / 3·3	0·8 / 0·5	1·0 / 1·2	0·7 / 0·4	-427·4 / -426·4	9·55
Plant height (1486)	est	2·9**	0·39 / 0·38	187·5 / 199·9	178·2 / 199·0	167·7 / 117·2	2829·3 / 1920·7	3320·0 / 2300·0	3411·4 / 1191·6	-3753·0 / -3747·5	$2·57 . 10^5$
Plant height (1389)	per-b	1·8	0·00 / 0·43	178·7 / 230·5	175·4 / 172·0	161·2 / 105·5	3192·8 / 1754·1	3196·0 / 1200·0	3128·8 / 808·9	-3269·3 / -3263·7	$3·47 . 10^5$
Leaf weight (654)	per-3	0·3	0·20 / 0·34	20·7 / 18·7	21·8 / 20·9	23·3 / 27·0	26·5 / 27·7	25·6 / 14·0	27·0 / 14·7	-865·9 / -865·3	4·47
Seed weight (678)	d	1·4	0·00 / 0·34	54·6 / 53·9	55·5 / 54·4	59·8 / 63·8	53·1 / 65·0	51·1 / 24·0	50·6 / 28·0	-997·6 / -996·4	14·45
Plant height (1681)	l	2·2*	— / 0·39	194·0 / 241·6	181·4 / 183·0	165·4 / 111·6	3042·9 / 1302·2	3171·2 / 1000·0	3145·6 / 979·7	-3973·6 / -3965·0	$3·23 . 10^8$
Plant height (1637)	per-1	0·4	— / 0·31	176·2 / 150·1	183·9 / 214·0	171·4 / 137·4	3006·2 / 1912·0	3177·5 / 1900·0	3368·2 / 2101·7	-3856·5 / -3853·3	$1·86 . 10^3$
Leaf weight (1526)	c	0·6	— / 0·36	20·9 / 19·7	21·5 / 20·5	22·7 / 26·0	25·4 / 27·7	24·6 / 17·0	26·2 / 16·7	-2009·7 / -2009·4	2·04
Seed weight (678)	a	-0·8	— / 0·34	57·9 / 62·9	56·1 / 55·0	53·7 / 50·0	54·9 / 38·2	52·8 / 24·0	52·8 / 48·5	-1002·9 / -1001·9	9·33
pH (1654)	c	-0·7	— / 0·10	56·8 / 56·9	56·5 / 56·0	59·1 / 59·8	54·1 / 54·9	50·5 / 50·0	46·7 / 43·1	-769·6 / -769·5	1·23
Vitamin C (1624)	c	0·9	— / 0·40	19·7 / 20·4	19·4 / 20·4	18·7 / 15·6	18·7 / 27·8	18·6 / 8·0	19·1 / 10·5	-2026·8 / -2024·2	398
Reducible sugar (1646)	l	1·5	— / 0·17	18·4 / 18·4	18·0 / 19·1	13·8 / 11·4	34·2 / 34·9	33·9 / 28·0	38·3 / 24·4	-2285·2 / -2283·2	112

[a] The number of plants included in each analysis are listed under the trait name in parentheses.

[b] Skewness effects are measured as the difference in the g1 values divided by the standard error of the difference. *, significant at $p < 0.05$; **, significant at $p < 0.01$.

[c] The r_B value is listed in the first row, and the r_L value is listed in the second row of each trait-marker combination.

[d] pp, pimpinellifolium allele homozygotes; pe, heterozygotes; and ee, esculentum allele homozygotes. The first row of means and variances for each marker-trait combination are the observed statistics for the marker genotypes, and the second row are the ML estimates for the linked QTL.

[e] The first row of each marker trait combination is the likelihood obtained assuming complete linkage between the QTL and the genetic marker, and the second column is the likelihood at ML.

was 1·8. Although this value is not significantly different from zero, it is in the direction postulated for partial linkage, in correspondence with the ML results. Among the 12 combinations listed, these three had the highest likelihood ratios. Only two combinations listed in this stable had skewness differences in the opposite direction postulated for partial linkage, and in both of these cases, the differences were not significantly different from zero.

The likelihood ratio for the effect of l on plant height was much higher than for any of the other combinations listed in this table. Furthermore, using the values generated from the ML algorithm, a difference of 2·5 phenotypic standard deviations was obtained between the homozygote means of the QTL—greater than the effect of dwarf on plant height. It may be that the model is inappropriate in this case, possibly due to a nonlinear effect of this locus, or of a linked QTL on plant height.

Marker-trait combinations which gave highest likelihood at complete linkage are presented in table 4. The same statistics listed in table 3 are presented where appropriate. Since ML was obtained at complete linkage for these combinations, the mean and variance estimates for the genetic marker are also the estimates for the QTL. Therefore in addition to these values, the r_C values, the means and variances that would be postulated for the QTL assuming this recombination value, the log likelihoods for both alternatives and the ratio of the likelihood at complete linkage to the latter likelihhod, are also listed. Since the change in likelihood was non-linear and dependent on the value of three variables, it was not possible to obtain the exact recombination value with a likelihood equal to $\frac{1}{5}$ of the maximum likelihood. Even though some of the likelihood ratios listed are much greater than 5·0, in all cases decreasing the recombination proportion by 0·01 resulted in a likelihood $>\frac{1}{5}$ of the ML likelihood. As expected for complete linkage between the QTL and the genetic marker, none of the skewness differences listed in table 4 are significantly different from zero, and three out of eight are in the opposite direction postulated by partial linkage.

r_C values for the combinations presented in table 4 ranged from 0·01, for the effect of d on plant height to 0·32, for the effect of per-1 on reducible sugar concentration, with a mean of 0·24. The mean value for $|\bar{X}_{M11} - \bar{X}_{M22}|$ was 0·49 F-2 standard deviations. The correlation between r_C and $|\bar{X}_{M11} - \bar{X}_{M22}|$ was −0·86. This result, which is significant at $p < 0.01$ with 6 degrees of freedom, supports previous simulation results that the ability

of the method to distinguish between complete and partial linkage is mainly a function of QTL effect (Weller, 1986). The regression of marker effect on r_C was −0·23, with a y-intercept of 0·35. Thus for a gene with an effect of equal to the phenotypic standard deviation, and ML at complete linkage, the expectation for r_C is 0·12. Assuming symmetry, the confidence interval in an experiment of this size ($n = 1600$) for recombination between a marker and a QTL of this magnitude with ML at partial linkage can be estimated as ±12 cM.

r_L was estimated for five additional marker-trait combinations in which the assumptions of the method were violated: the effects of d and l on internode length, the effect of per-1 of fruit index, the effect of d on fruit weight, and the effect of a on TSS. Internode length and fruit index had kurtotic distributions after transformation, and for TSS and fruit weight the assumed number of effects per marker was greater than one. Approximate ML was obtained with a recombination proportion greater than 0·3 for all three combinations involving traits with kurtotic distributions. Furthermore, likelihood ratios were all above 10^{11}, as compared to a maximum value of $3 \cdot 10^8$ in table 3. For the effects of d on fruit weight and a on TSS, approximate ML was found at complete linkage. The effects of a on TSS and d on fruit weight were 0·4 and 1·0 F−2 standard deviation units, respectively. Nevertheless, r_C values were 0·35 and 0·14, respectively. This may be due to the presence of several QTL, which together yield a "diffuse" effect. Despite the fact that ML was obtained at complete linkage, the skewness difference for the effect of a on TSS was significant and in the direction indicating partial linkage.

In addition to the 20 marker-trait combinations listed in tables 3 and 4, an additional 28 putative QTL were mapped by linkage to pairs of linked genetic markers using the r_B estimate, as described above. The locations of all 48 mapped QTL are shown in fig. 1. Most of the QTL mapped are located on the first three chromosomes. This is due to the fact that the linked pairs of markers were located on these chromosomes, so that both mapping techniques could be utilised for these chromosomes. Clusters of three or more adjacent effects are located at positions 30 and 61 on Chromosome I; 66 and 70 at Chromosome II; and 29 on Chromosome III. Three of these clusters are located adjacent to the morphological markers: y, d, and r, and are likely to be pleiotropic effects of these markers. It is possible that some of the other clusters may be the result of a single QTL with effects on several traits. No more than one putative QTL mapped

Table 4 Maximum likelihood results for marker-trait combinations which correspond to the assumptions of the model, and with ML at complete linkage

Trait[a]	Marker	Skewness difference[b]	Estimated recombination proportion[c]	Mean[d]			Variances			log likelihood	likelihood ratio
				pp	pe	ee	pp	pe	ee		
Stem width (1686)	d	0.7	0.00	31.4	31.3	2.70	93.1	87.3	64.7	-968.9	
			0.12	31.3	32.0	2.56	96.5	81.0	50.6	-969.8	9.12
Plant height (1681)	d	0.5	0.00	202.4	199.2	109.1	2030.4	1984.3	995.8	-3744.3	
			0.01	202.4	199.0	107.8	1920.7	2300.0	1191.6	-3746.7	240·
Plant height (731)	per-3	-0.1	0.18	169.9	183.2	185.0	3051.6	3443.0	3666.7	-1734.9	
			0.28	155.8	190.0	186.1	2237.9	3000.0	4214.9	-1735.8	8.32
Reducible sugar (1646)	d	-0.6	0.05	17.7	17.3	15.4	37.4	39.8	35.3	-2314.5	
			0.24	17.9	18.0	13.6	38.1	35.0	26.9	-2315.4	7.41
Stem width (1686)	l	0.8	—	31.2	30.5	29.1	90.9	84.5	85.6	-983.2	
			0.40	37.1	28.5	26.2	62.5	78.0	69.0	-984.0	5.89
Leaf weight (1526)	l	0.3	—	22.5	22.2	20.0	27.1	24.6	22.7	-2002.7	
			0.24	22.4	23.0	17.8	31.8	18.0	13.6	-2004.1	26.40
Reducible sugar (1646)	a	0.2	—	17.7	17.3	15.4	38.7	38.9	35.7	-2314.7	
			0.28	17.9	18.4	12.8	44.3	29.0	24.8	-2315.6	7.76
Reducible sugar (1594)	per-1	-0.5	0.00	17.1	17.3	15.9	42.9	37.5	36.8	-2274.6	
			0.32	16.5	18.8	13.0	52.9	29.0	26.5	-2275.9	22.38

[a] The number of plants included in each analysis are listed under the trait name in parentheses.
[b] Skewness effects are measured as the difference in the g1 values divided by the standard error of the difference.
[c] The r_B value is listed in the first row of each marker-trait combination, and the r_C value is listed in the second row.
[d] pp, pimpinellifolium allele homozygotes; pe, heterozygotes; and ee, esculentum allele homozygotes. The first row of means and variances for each marker-trait combination are the observed statistics for the marker genotypes, and the second row are the estimates derived assuming the recombination proportion in the previous column.
[e] The first row of each marker trait combination is the likelihood obtained assuming complete linkage between the QTL and the genetic marker, and the second column is the likelihood corresponding to the recombination proportion listed in the same row.
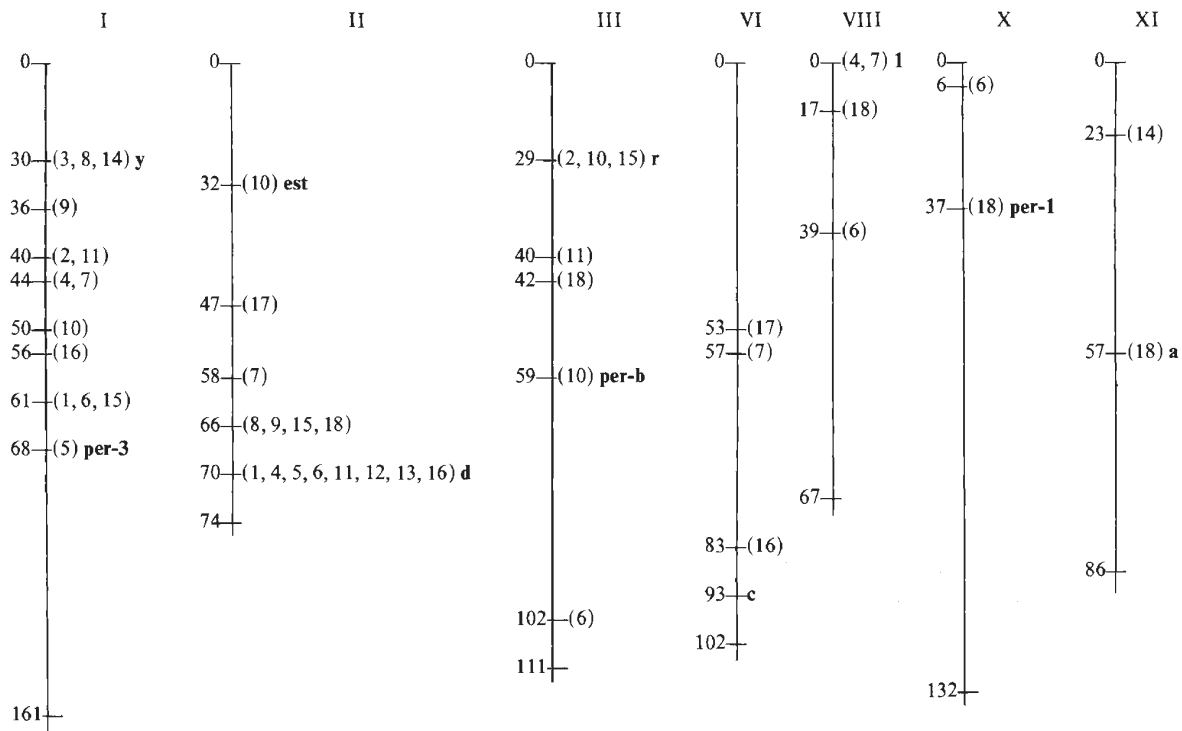
Figure 1 Location of the putative QTL on the tomato genetic map. Only the seven chromosomes with genetic markers segregating in the F-2 population are shown. The map locations are listed to the left of each chromosome, in cM, and trait numbers of the mapped QTL are listed to right in parenthesis. The genetic markers are listed after the QTL in bold type. The quantitative traits by number are:
1. Internode length 2. Number of nodes to the first flower cluster 3. Number of flowers per cluster 4. Stem width 5. Plant size 6. Plant height 7. Leaf weight 8. Ratio of fruit to vegetative growth 9. Flowering earliness 10. Fruit earliness 11. Fruit weight 12. Fruit index 13. Number of locules per fruit 14. Weight of 20 seeds 15. Total soluble solids 16. pH of the juice 17. Vitamine C concentration 18. Reducible sugar concentration.

next to any of the electrophoretic markers. The two effects linked to *l* are also likely to be pleiotropic effects as both of the traits in question, stem width and leaf weight, are clearly affected by this gene.

DISCUSSION

This study represents the first attempt at an extensive mapping of QTL in an economic organism. Previous more limited studies using a variant of the r_B method have been applied by Tanksley *et al.* (1982) and Vallejos and Tanksley (1983). Of the 48 putative QTL mapped in this study, only 22 showed complete linkage with the genetic markers. Thus if the map locations are correct, the remaining 26 effects are not pleiotropic effects of the genetic markers. Although four of the morphological markers apparently had more than one pleiotropic effect on the quantitative traits

measured, this was not found for any of the electrophoretic markers. This supports the contention of Tanksley, *et al.* (1982) that electrophoretic markers may be more suitable for this type of analysis because they are less likely to have pleiotropic effects on the traits of interest.

In general there was good correspondence between the results for r_L, r_B and the skewness differences between the homozygote distributions. r_B estimates are likely to be more reliable, than the r_L estimates, but can be calculated only when two linked genetic markers bracket the QTL.

The approximate ML results were generally reasonable when the first two assumptions of the method were not violated. Exceptions were the effect of *per-b* and *l* on plant height. In both cases a linked locus located about 40 cM from the marker was found to be much more likely than the null hypothesis of complete linkage. The postulated effects were greater than the effect of *d* on plant height, and the likelihood ratios were much higher

than for the other combinations. It is possible that in these cases, the third assumption of the ML technique was violated, *i.e.* the effect of the QTL may be non-linear. This is not surprising since interactions have been found between loci affecting quantitative traits (Tanksley *et al.*, 1982; Weller *et al.*, 1986).

Recombination proportion at approximate MI was below 0·2 for only two of the combinations listed. This is unexpected in view of results on simulated data (Weller, 1986) which in most cases demonstrated a bias in the direction of underestimation of recombination. As noted by Weller, *et al.* (1986), since 180 marker-trait combinations were tested, a few "spurious" significant effects can be expected by chance. Prior knowledge is not available as to the expectation for r_L when the main effect associated with the marker is an artifact.

In simulation studies of QTL with effects equal to 1·0 phenotypic standard deviation unit and complete linkage to the marker, mean estimated recombination proportion was 0·036 and the standard deviation of estimates was 0·033 (Weller, 1986). An approximate upper confidence limit; the mean plus two standard deviations, would yield a recombination proportion of 0·102. This is comparable to the value of 0·12 estimated in this study from field data. Thus the results presented here support the conclusion of Weller (1986) that for a QTL of 1·0 standard deviation units the method is able to differentiate with good power between complete linkage between the QTL and the marker and a recombination proportion of 0·2, but not for loci of smaller magnitude.

The results obtained for the combinations that violated the assumptions of the method were distinctly different from those obtained for combinations which did not violate the assumptions of the method. However, it is not clear why there was a significant skewness effect associated with the effect of the locus *a* on TSS.

At present the main obstacle in the application of this technique to the location of genes affecting traits of economic importance, is the lack of suitable genetic markers differentiating the inbred lines used to form the F-2 populations of interest. It is possible that in the near future, this problem may be overcome by the use of restriction fragment length polymorphisms which are likely to be quite common throughout the genome (Beckmann and

Soller, 1983). Once techniques become available for the mass screening of individuals for these markers, both the r_L and r_B mapping techniques described in this study should be useful in determing the approximate map locations of marker-linked QTL.

REFERENCES

BECKMANN, J. AND SOLLER, M. 1983. Restriction fragment length polymorphism in genetic improvement: methodologies, mapping and costs. *Theor. Appl. Genet.*, *67*, 35–43.

ELSTON, R. C. AND STEWARD, J. 1973. The analysis of quantitative traits for simple genetic models from parental, F1 and backcross data. *Genetics*, *73*, 695–711.

HILL, A. P. 1975. Quantitative linkage: a statistical procedure for its detection and estimation. *Ann. Human Genet.*, *38*, 439–450.

JAYAKAR, S. D. 1970. On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. *Biometrics*, *26*, 451–464.

SOLLER, M., BRODY, T. AND GENIZI, A. 1976. On the power of experimental marker-linked quantitative effects in crosses between inbred lines. *Theor. Appl. Genet.*, *47*, 35–39.

SOLLER, M., BRODY, T. AND GENIZI, A. 1979. The expected distribution of marker-linked quantitative effects in crosses between inbred lines. *Heredity*, *43*, 179–190.

TANKSLEY, S. D., MEDINA-FILHO, H. AND RICK, C. M. 1982. Use of naturally occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity*, *49*, 11–25.

VALLEJOS, C. E. AND TANKSLEY, S. D. 1983. Segregation of isozymes markers and cold tolerance in an interspecific backcross of tomato. *Theor. Appl. Genet.*, *66*, 241–247.

WELLER, J. I., SOLLER, M. AND BRODY, T. 1986. Linkage analysis of quantitative traits in an interspecific cross of tomato (*L. esculentum* × *L. pimpinellifolium*) by means of genetic markers. *Genetics*, in press.

WELLER, J. I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics*, *42*, 627–641.

ZHUCHENKO, A. A., KOROL, A. B. AND ANDRYUSCHCHENKO, V. K. 1979a. Linkage between loci of quantitative characters and marker loci. *Genetika*, *14*, 771–778.

ZHUCHENKO, A.A., SAMOVOL, A. P., KOROL, A. B. AND ANDRYUSHCHENKO, V. K. 1979b. Linkage between loci of quantitative characters and marker loci. II. Influence of three tomato chromosomes on variability of five quantitative characters in backcross progenies. *Genetika*, *15*, 672–683.