J. Genet. Vol. 72, Nos. 2 & 3, December 1993, pp. 73–83. © Printed in India.

COMMENTARY

RFLP approach to breeding for quantitative traits in plants—a critique

V. ARUNACHALAM and SHANTI CHANDRASHEKARAN
Division of Genetics, Indian Agricultural Research Institute, New Delhi 110012, India

MS received 24 October 1992; revised 26 April, 25 November 1993

Abstract. A number of recent papers suggest that use of RFLPs as markers offers a clear advantage in breeding for improvement in quantitative traits (QTs). The concepts underlying establishment of linkage between RFLP markers and QT loci stem from three papers on tomato published during 1987–1991. Essentially, continuously varying QT phenotypes are assigned to RFLP genotype classes, which can be considered to be determined by a single, diallelic gene with codominant alleles. Linkage is inferred through statistical analysis. Similarly, interaction between markers and QT is also tested by an analysis of variance. Here, the statistical methods employed in these three papers to detect linkage are critically evaluated, especially because subsequent investigations take the concepts developed in these papers as proven. In this paper, we examine the three fundamental papers *de novo*. We scrutinize the methods employed and the inferences drawn to bring to light what we believe are conceptual drawbacks.

Keywords. RFLP; molecular markers; quantitative traits; linkage; analysis of variance; plant breeding.

1. Introduction

Among molecular markers, DNA restriction fragment length polymorphisms (RFLPs) occupy a prominent place in providing profitable clues to breeding for quantitative traits (QTs). Successful experiments, mostly with tomato, have led to the conviction that molecular markers can be used to detect, map and monitor genes controlling quantitative variation (Tanksley and Hewitt 1988). RFLPs are inherited in a manner similar to Mendelian genes and can, to that extent, be called 'genes' (there are differences, however; one among them is the fact that RFLP 'genes' do not have a phenotypic expression).

The following properties of RFLP markers have been emphasized:

(i) Alleles of a given RFLP are codominant. This makes it easy to construct saturated linkage maps and locate, in principle, RFLP markers closely linked to genes controlling QTs.
(ii) They are environment-independent; hence decisions based on use of such markers are uninfluenced by phenotypic variation.
(iii) Genome saturation by markers can, in principle, dissect a QT 'into discrete genetic factors (quantitative trait loci, QTL)'.
(iv) When QTL show high epistatic interaction, direct selection for them is usually difficult and inefficient. But indirect selection for QTL using linked and non-epistatic RFLP 'genes' is possible.

Pioneering work on the use of RFLPs in QT improvement was done in tomato. Three fundamental papers set out the basis for the utility of RFLPs in predictable

QT improvement. They are by Osborn *et al.* (1987), Tanksley and Hewitt (1988) and Paterson *et al.* (1991), and we shall refer to them as OSB, TAN and PAT respectively. We examine them to evaluate the possible utility of RFLPs in practical plant breeding. Since our viewpoints can be presented in the light of those papers, we make no attempt to survey the fast-growing RFLP–QT literature.

2. Polygenes and QTs

The idea that 'genes concerned with quantitative variables' can be handled by classical genetics and that phenotypic variation can be cast into discrete groups attributable to some marker genotypes was put forth by Thoday (1961). The procedures outlined start with the assumption that 'a chromosome gives a higher value of a metric character than that given by a homologous chromosome marked with the recessive major genes'. Using this assumption, progeny evaluation in test crosses and F_2 are described to locate polygenes governing the QT. The methods are the same as those presented in elegant detail by Mather (1951), to which Thoday's 1961 paper also refers. There are, however, major difficulties, which stem from certain assumptions: (i) QTs (under polygenic control) behave like a major gene with codominant alleles, allowing grouping of QT phenotypes in three marker-based genotypic classes. (ii) Major-gene control implies absence of epistatic interactions in QT phenotypic expression—a situation that is far from reality. (iii) A genotypic distribution assumed to be stationary can have a dynamic phenotypic distribution owing to environmental contributions. The resulting new QT phenotypes would change the frequencies of the three marker-based genotypic classes and result in a dynamic genotypic distribution.

3. RFLP–QT linkage in tomato

3.1 *Paper by Osborn et al. (1987)—OSB*

OSB examines linkage between two RFLP markers and genes controlling soluble-solid content (SS) in tomato. A wild tomato species, *Lycopersicon chmielewskii* (LA1028), with SS ~ 10%, was crossed with cultivated tomato, *L. esculentum* (VF36), with SS ~ 5%, and through continuous backcross–sib mating (BC_5S_5) a line, LA1563, generally resembling the phenotype of VF36 but with 7 to 8% SS, was derived. These three lines, along with a breeding line H2038, were screened for RFLPs with four restriction enzymes and 60 cDNA clones used as probes. Two clones, ptHL75 and ptHL104, identified one RFLP pattern in LA1563 and LA1028, and an alternative one in H2038 and VF36. The RFLP allele in LA1563 was designated *c* and the one in H2038, *e*. The loci marked by the two clones were named A_1 and A_2. The F_2 segregating population from the cross H2038 × LA1563 was screened for RFLP alleles and the SS content of individual F_2 plants measured to identify possible linkage between SS and A_1 and A_2 'genes'.

A total of 165 F_2 plants were scored for A_1 and 147 for A_2. One hundred and fortytwo F_2 plants were scored for both A_1 and A_2. A fit to 1:2:1 was observed for both A_1 and A_2, independently suggesting that the RFLP alleles *c* and *e* in each were codominant, confirming the basic assumption. Frequencies of the nine genotypes with respect to the two RFLP 'genes' A_1 and A_2 were obtained and linkage

between A_1 and A_2 tested as follows: (i) by testing the frequencies of the marginal totals of A_1 and A_2 genotypes for conformity with the ratio 1:2:1 by a chi-square test, each with two degrees of freedom (d.f.); (ii) by obtaining a residual chi-square (4 d.f.) by subtracting from the total chi-square (8 d.f.) the chi-square due to the A_1 marginal totals conforming to 1:2:1 (2 d.f.) and similarly that due to the A_2 marginal totals (2 d.f.); (iii) by testing the residual chi-square (4 d.f.) and interpreting it as a test for linkage. The non-significant residual chi-square was taken to suggest that the RFLP 'genes' A_1 and A_2 were unlinked. This result was in agreement with the concept of non-epistasis associated with RFLP markers.

In a study in which 233 unrelated Canadian individuals were sampled for health parameters, Zerba *et al.* (1991) found significant disequilibrium (both quadri-allelic and tri-allelic) for two marker pairs despite absence of significant deviations from Hardy–Weinberg equilibrium for either marker. Inadequate sample size, as in many RFLP studies, and data exclusion due to methodological constraints of inability to successfully type all members of the sample for every RFLP (as in OSB) were thought to be possible major reasons. In particular, when non-evolutionary sampling factors were involved, such as those described above, the phenotypic variation associated with a marker would be estimated with bias and 'will remain a problem in the search for genotype–phenotype' relationships.

3.1a *The small F_2 sample*: The observed and expected frequencies (on the null hypothesis H_0: A_1 and A_2 are unlinked) of F_2 plants defined by the marker genes A_1 and A_2, from OSB, are given in table 1. Regardless of the status of linkage between A_1 and A_2, it is known from elements of two-gene theory that, in the case considered in OSB, the frequencies of genotypes in which $A_1^e A_1^e$ is common should conform to the expected values (see table 1); in other words, the frequencies of genotypes $A_1^e A_1^e A_2^e A_2^e$, $A_1^e A_1^e A_2^e A_2^c$ and $A_1^e A_1^e A_2^c A_2^c$ should not differ significantly from the expected values 8·88, 17·75 and 8·88 by the chi-square test. The same should be true of genotypes in which $A_1^c A_1^c$, $A_2^e A_2^e$ or $A_2^c A_2^c$ is common. Tests show that there are

Table 1. Analysis of observed (O) and expected (E) frequencies of genotypes with reference to marker genes A_1 and A_2 among 142 F_2 plants in data from Osborn *et al* (1987).

	$A_1^e A_1^e$		$A_1^e A_1^c$		$A_1^c A_1^c$		MT_R			
	O	E	O	E	O	E	O	E	χ^2_R	p
$A_2^e A_2^e$	15	8·88	19	17·75	9	8·88	43	35·5	4·31	0·12
$A_2^e A_2^c$	15	17·75	30	35·50	23	17·75	68	71·0	2·83	0·25
$A_2^c A_2^c$	13	8·88	14	17·75	4	8·88	31	35·5	5·39	0·07
MT_C	43	35·5	63	71·0	36	35·5	142			
χ^2_C		6·56		1·73		4·24				
p		0·04		0·44		0·13				
								$\chi^2_{MT_R}$	2·28	0·33
								$\chi^2_{MT_C}$	2·49	0·43

Null hypothesis H_0: class frequencies rowwise or columnwise conform to the expected values on two unlinked genes; MT, marginal total; $\chi^2_{(\,)}$, chi-square corresponding to H_0; R, row; C, column; χ^2_{MT}, chi-square for marginal totals conforming to 1:2:1; p, probability of exceeding $\chi^2_{(\,)}$.

difficulties in accepting the hypothesis for genotypes $A_1^e A_1^e$ __ ($p = 0{\cdot}04$) (table 1). There is thus reason to doubt the unbiasedness of the 142 F_2 sample. The fact that the frequencies of genotypes $A_1^e A_1^c$ __ and $A_2^e A_2^c$ __ conform to the expected frequencies on the hypothesis of two unlinked genes does not help to clear the doubt. It would have been ideal had the possibilities of detecting the frequencies of coupling and repulsion double heterozygotes with reference to the two marker loci been explored to obtain a direct measure of linkage.

3.1b *Method used to detect linkage*: Data in OSB were analysed for linkage by testing the 4-d.f. residual chi-square by the method of Mather (1951). Mather's method applies only to those situations where four phenotypic classes (and not nine genotypic classes) can be identified because of complete dominance at the two loci. Two fundamental differences between RFLPs and conventional markers are worthy of mention: (i) RFLP-marked individuals can only be genotyped and cannot be phenotyped. (ii) RFLP alleles are codominant; with no phenotypic expression, the question of 'dominance' does not arise.

Classification into RFLP phenotype classes in situations where the alleles do not have a phenotypic expression is not permissible. However, for the sake of illustration, we temporarily ignore this problem and compute the four 'phenotypic class' frequencies by adding the observed frequencies of the corresponding genotypes in each phenotypic class. The chi-square tests for linkage (see Mather (1951) for details) then result in the following: chi-square value for A_1 phenotypes (A_1^e:A_1^c) departing from 3:1 (1 d.f.) = 0·009, with $p = 0{\cdot}87$ in the sense given in table 1; chi-square value for A_2 phenotypes (A_2^e:A_2^c) departing from 3:1 (1 d.f.) = 0·761, with $p = 0{\cdot}31$; and chi-square (linkage) (1 d.f.) = 3·008, with $p = 0{\cdot}09$. We note, contrary to the conclusions in OSB, that there was a reasonable chance that A_1 and A_2 were linked.

3.1c *Method to detect linkage between an RFLP marker and SS*: The F_2 sample plants were first grouped into three RFLP marker-based genotypic classes with reference to A_1 or A_2. An analysis of variance (ANOVA) of SS content in 165 plants with respect to A_1 genotype has been given in OSB. The two d.f. between the three genotypes $A_1^e A_1^e$, $A_1^e A_1^c$ and $A_1^c A_1^c$ have been partitioned into a difference between $A_1^e A_1^e$ and $A_1^c A_1^c$ (1 d.f.) and the contrast $A_1^e A_1^c$ vs $1/2(A_1^e A_1^e + A_1^c A_1^c)$ (1 d.f.). They were tested as F-ratios using intragenotypic-class variation (162 d.f.).

The interpretations are constrained by the following points:

(i) Grouping of the continuous variation in the QT phenotypes into three marker-based genotypic classes due to the single diallelic gene A_1 raises points similar to those in section 3.1b.

(ii) The variation in SS from 165 F_2 plants (164 d.f.) was partitioned as that due to the three marker-based genotypic classes (2 d.f.) and that due to the remainder (164–2 = 162 d.f.). The latter was used as error to test the significance of variation among the three genotypic classes by an F-test. In principle there would be genetic variation in F_2 and, therefore, the sum of squares with 162 d.f. could not be an error sum of squares in the strict sense. The validity of the F-test and consequent inferences may not be justified.

(iii) The single-d.f. comparison, heterozygote vs average of homozygotes, tests only for dominance; by this test applied to the SS content of the groups defined by A_1

alone, A_2 alone, or A_1 and A_2 jointly, dominance was found to be absent. Thus, in addition to the property of codominance of RFLP alleles at markers A_1 and A_2, it so happened that the SS content of F_2 plants genotyped by A_1 or A_2 did not show dominance either, and the magnitude of interlocus interactions (epistasis) determined as contrasts from analysis of variance using data for genotypic classes combined on loci A_1 and A_2 was non-significant. This would imply that genetic effects on SS attributable to particular alleles at A_1 and A_2 are additive; in other words, the average SS content of, for example, the group with genotype $A_1^c A_1^c A_2^c A_2^e$ would be given as the sum of the average SS contents of the genotypic groups $A_1^c A_1^c$ __ and __ $A_2^c A_2^e$. We recollect that the F_2 sample on which the results are based was small in size and deficient in required basic characteristics (see section 3.1a).

(iv) A number of inferences in OSB suffer consequently, and two problems are indicated below.

The divergence in SS content between the high-SS parent LA1563 (SS ~ 6·7%) and the low-SS parent H2038 (SS ~ 5·4%) was low from plant-breeding standards, in contrast to the divergence in SS content between LA1028 (~ 10%) and VF36 (~ 5%), with which earlier breeding work was initiated. The F_1 mean (5·68) was less than mid-parent (6·07), and the F_2 mean (5·96) was marginally higher than the F_1 mean and slightly lower than the mid-parent value. None of the mean SS values of genotypes based individually on A_1 or A_2 loci exceeded that of the better parent, LA1563. The inference that the RFLP alleles in LA1563 were linked to genes significantly increasing SS is superficial, and needs further evaluation.

3.2 *Paper by Tanksley and Hewitt (1988)—TAN*

TAN examines the results given in OSB in greater detail using one isozyme and four RFLP markers. However, most of the difficulties with OSB, explained above, remain.

Further, continuous variation in SS or fruit pH was assigned a priori to three genotypic classes with respect to the isozyme marker Aco2 and an RFLP marker, CD56, individually. If the total number of SS phenotypes and their mean and range of variation in each class had been indicated, it would have been possible to check how many 'low'-SS phenotypes were classified into the *c/c* class and so on. However, the significant variation among the *e/e*, *e/c* and *c/c* classes was taken as evidence for the contribution of the concerned marker to QT differences. The way *F*-ratio was computed and the simple *t*-test of two sample means to infer QT increase limit the validity of the inferences. Further, it was tacitly assumed that the variation among the individual F_2 plants that contributed to the mean phenotypic value of genotypic class *e/e* (remembering that F_2 genotypes were grouped into it on the presence of the marker) was due to chance and hence constituted a part of the 'error'. This is an assumption difficult to defend. The contribution of a marker to an increase or decrease in QT was judged on the basis of *F* and *t*-tests that used different errors with different d.f. (in contrast to normal statistical tests of inference). Such inferences, further interpreted as indicative of correlation of marker segments with QTs, cannot be strong.

In a similar manner, a result of non-significance for the 2-way interaction Aco2 × CD56 for SS and pH was interpreted as absence of epistatic interaction between the markers for the QTs. The interaction term is usually calculated from this table. Here

Table 2. Frequencies of marginal classes of genotypes based on markers Aco2 and CD56, averaged over three populations, in F_2 data from tables 3 and 5 of Tanksley and Hewitt (1988).

		Aco2			
		e/e	*e/c*	*c/c*	Total
CD56	*e/e*				126
	e/c				215
	c/c				129
Total		195	302	147	470 644

the total number of F_2 plants by pooling those based on Aco2 (470) and CD56 (644) were unequal. Thus the error d.f. used to test Aco2 and CD56 main effects and the Aco2 × CD56 interaction were different. This renders the partitioning of total variation non-orthogonal, diluting interpretations based on it. We have computed the marginal class frequencies corresponding to Aco2 and CD56 from table 5 of TAN; these are shown in table 2 here. The total of nine individual cells based jointly on Aco2 and CD56 should be equal to 470, suggesting that some F_2 plants could not be scored for both markers. It is also possible that plants for which association between SS (or pH) and Aco2 or CD56 was inferred individually were a different set from those in the 2-way—Aco2, CD56—table shown in table 2 here. This adds another dimension of difficulty in collating results on the main effects of the two markers and on epistasis.

3.3 *Comparison of the results of OSB and TAN*

One of the inferences was that 'the two segments detected by Osborn *et al.* (1987) are likely a subset of the segments reported here' (TAN). Examination of data in TAN led to the conclusion that the statement 'higher-SS tomatoes could be bred by selection of the linked RFLP marker' made in OSB could be premature. Further, use of the markers Aco2 and CD56 led also to increased pH, which is an undesirable trait. This dissuades one from using these markers in breeding programmes. Thus genetic background and phenotypic expression could escape being accounted for by use of RFLP markers. The conclusions in TAN countering the earlier observation of Ellis (1986), that RFLP markers could only be used to follow the segregation of reasonably closely linked genes where segregating alleles conferred very different phenotypes, remain unfounded.

3.4 *Interval mapping*

OSB and TAN were among the first few reports to explore the potential for field application of RFLP-based QT selection. Subsequent work on diagnostic methods

to identify cases of RFLP–QT linkage was expected to steer clear of the weaknesses of these first papers. Interval mapping (Lander and Botstein 1986a,b, 1989) is such a method, devised mainly to map complex genetic traits in humans (particularly loci concerned with diseases). The genetic constitution of the parents with respect to the markers, namely whether in *cis* or *trans* 'phase', is usually unknown. It is hence not possible to count recombinants to estimate linkage. The recombination fraction r is therefore indirectly estimated by calculating an odds ratio. This is the ratio of the probability that the observed data would arise under a given hypothesis (say 7·5% linkage, or $r = 0{\cdot}075$) to the probability that they would arise under an alternative hypothesis (say no linkage, $r = 0{\cdot}5$). When the odds are large, usually 1000:1, in favour of linkage, the linkage hypothesis is accepted. The odds ratio is estimated for several hypothetical values of r. The most likely value (or the maximum likelihood estimate) is the one at which the odds ratio is the highest. The $\log_{10}$ of the odds ratio, called the LOD score, is used in general in place of the odds ratio.

This concept was improved and modified to achieve interval mapping of QTs. Once a saturated RFLP map is available, it becomes possible to estimate the odds that a QT is linked to two linked RFLPs simultaneously against the odds that it is not. When LOD scores are high enough to prove linkage (depending upon appropriate r values), the QT can be mapped to the physical interval between the two RFLPs. Problems of identifying optimum sample size and a norm for LOD score for deciding the location of QTL in the interval have also been addressed by Lander and Botstein (1986a,b, 1989).

However, basic problems that limit their applicability to practical plant breeding appear to remain, a few of which are presented below.

(i) LOD scores are calculated under assumptions for genetics of QT that are not often valid in practice. Very simply, the model used to obtain the expected phenotypic value, viz. $\phi_i = a + bg_i + e_m$ (Lander and Botstein 1989), is $P = G + E$, which does not hold for QTs of agricultural importance. The assumptions made, that QT loci are additive and independent, are also difficult to sustain.

(ii) Inferences in OSB and TAN suffer from the small and unrepresentative sample size used in F_2. The minimum sample size calculated from the expected value of the LOD score is open to similar questions. It is also a moot point whether size alone is crucial or the sample constitution too is important. For instance, if we agree that, from an F_2 population in the one-gene-two-allele case, a four-plant sample is the minimum required to discover three dominant and one recessive genotypes, we would endlessly argue whether any one sample picked at random from an F_2 population of, say, 500 plants would be the right sample. On the other hand, the argument that every sample is random and would have the desired 3:1 constitution of dominant: recessive genotypes also is difficult to sustain. The problem stems essentially from the application of 'deterministic' or 'infinite population size' properties (of F_2) to finite population sizes (of F_2), and worsens further when small samples are drawn from them.

(iii) To increase the power of QTL mapping, a method of selective genotyping, in which extreme individuals 'whose genotypes could be most clearly inferred from their phenotypes' (Lander and Botstein 1989) are used, is recommended. The preference, while being helpful in mapping QTL in situations of low progeny size, may result in a sample that is not representative of the F_2 variation. This reduces its practical utility in actual QT improvement.

Theory using mathematical/statistical approximations provides support to the procedure of interval mapping (see the appendix in Lander and Botstein 1989), but difficulties in translating these results to QT improvement in the field remain, as discussed below.

3.5 *Paper by Paterson et al.* (1991)—*PAT*

In this exhaustive paper, the authors incorporate information from the earlier OSB and TAN. The material used was from a cross between a cultivated line of *L. esculentum*, UC204B (denoted E), and a wild accession of *L. cheesmanii*, LA483 (denoted CM). Three hundred and fifty individual plants and their plant-to-progeny F_3 families (each containing 10 plants) were evaluated for RFLP markers and the QTs SS, fruit mass and fruit pH in three environments (F_2 in Davis, California; F_3 in Gilroy, California, and Rehovot, Israel). The studies could therefore examine RFLP genotyping in relation to phenotyping and phenotypic expression of QTs in three environments.

Use of a larger set of RFLP markers than in TAN led to identification of 11, seven and nine QTL influencing mass per fruit, SS and fruit pH respectively. It must, however, be remembered that the wild and the cultivated parents used in this study were different from those used in OSB and TAN. This disallows (i) confirmation of the results of OSB and TAN across environments and (ii) comparison of the methods of assigning values for additive and dominance effects by constructing QTL likelihood maps and examining the proportion of phenotypic variation accounted for by them. Interval mapping methodology and its applications are the major distinguishing features of PAT.

Paterson *et al.* used LOD scores to test adequacy of a fit to three single-gene models of each QTL examined. In this context, the observations of Risch (1992), that violation of the Mendelian assumption in LOD score analysis may have serious consequences, merit attention; in particular, the rate of 'false positives' (Paterson *et al.* 1988, and PAT) could increase in non-Mendelian cases. An examination of LOD score statistics for non-Mendelian diseases revealed that while a high LOD score is unlikely in the absence of linkage, it is equally unlikely in the presence of linkage when a locus has only a minor effect and 'it is the ratio of these two likelihoods that determines the probability that the result may be a false positive'. More important and relevant to QT linkage is the observation that, 'in the absence of any prior evidence for the major effect of a single locus (and particularly when there is evidence against such a major effect), a positive linkage should be viewed as a hypothesis-generating result (that is, existence of a major locus) rather than as a hypothesis-testing result. It is a hypothesis that can be confirmed or refuted by subsequent studies. Hence, for such cases, replication is indispensable.' These arguments suggest the need for additional evidence from practical breeding.

3.6 *Mapping of genomic segments contributing to SS in interspecific crosses of Lycopersicon*

To maximize the chances of detecting genomic segments with a high probability of linkage to segments determining SS (and avoid analysis of a large number of spurious

genomic segments), the authors of OSB and TAN tested only those RFLPs that were common to *L. chmielewskii* LA1028 and LA1563. OSB reported detection of two *chmielewskii* segments in LA1563 and TAN reported three. Since OSB did not report chromosomal location of the two segments, it is not possible to determine if they correspond to or map near the three segments of TAN. Even assuming that they do, this method is biased towards detection of genomic segments, potentially linked to SS genes, only from *L. chmielewskii*. The method does not allow detection of *L. esculentum* segments potentially linked to SS genes in this species. This drawback is partly removed in PAT, where mapping was done on the basis of data for the F_2 population of a cross between *L. esculentum* and *L. cheesmanii* and SS genes were found to map to four regions on three chromosomes (figures 4 and 5 in PAT).

RFLP probes are expected to tag and help in cloning loci for QTs across genetically broad-based material. Judged from this viewpoint, an obvious question arises: Would SS genes in the genus *Lycopersicon* map to the same or similar orthologous genomic regions regardless of the genetic background or the environment? Let us examine the SS loci located from RFLP experiments and conventional ones on four tomato crosses. In all the four crosses *L. esculentum* was one of the parents. *L. chmielewskii* LA1028 was common in three of the four crosses, and *L. esculentum* VF36 in two of the four (table 3). Locations of SS genes, expressed as likelihood intervals or approximate positions in centimorgan (cM), could be traced to seven regions on five chromosomes. Of the seven regions, one region on chromosome 6 was common to crosses b, c and d, and one on chromosome 10 was common to crosses a and d (table 3). Therefore, out of a total of 28 possible positions for seven SS loci only five were common to at least two. PAT data show that the positions of SS genes varied even with environment (figure 4 in PAT).

This demonstrates that even for crosses between related species (such as the ones in a to d, table 3) QT polygenes will have to be located every time a new species is employed as a parent in a breeding programme and in each environment. This proposition is monetarily, time-wise and logically impractical even in elementary backcross and F_2-breeding schemes such as the ones used in OSB, TAN and PAT. Breeding schemes to pyramid QTL use much more complex mating processes and the

Table 3. Chromosomal likelihood intervals (in cM) for SS genes in tomato in four papers.

Chromosome number	TAN[a]	PAT[b]	Paterson *et al.* 1988[c]	Rick 1974[d]
2	—	35·3–73·1	—	—
3	—	40·4–71·6	—	—
6	—	0·0–28·6	—	—
	—	37·7–69·7	57·5–67·1	≈61*
7	24–46	—	—	—
	—	—	53·5–68·4	—
10	0–5	—	—	≈0*

a: *L. chmielewskii* LA1028 × *L. esculentum* VF36; BC_5S_5
b: *L. esculentum* UC204B × *L. chmielewskii* LA483; F_2
c: *L. esculentum* UC82B × *L. chmielewskii* LA1028; BC_1
d: *L. chmielewskii* LA1028 × *L. esculentum* VF36, VF45; BC_5S_5
*Approximate location in cM; –, genes for SS not located.

time taken for the potential of RFLP-based selection to be demonstrated in the field may be long.

4. Requirements of RFLP-assisted plant breeding methods for QT improvement

It may be worth repeating the basic requirements of breeding methods to assess the possible role of RFLP-assisted selection in QT improvement.

1. To discover RFLP markers and to identify linkage to QTL, specific genetic material, such as inbred (homozygous) parents, their backcross or F_2 generations, and, most often, lines fixed in higher pedigree generations (BC_5S_5 in tomato in OSB, TAN and PAT for example), are necessary.
2. QTL identified in a crop plant at a given location (space environment) should have near-stationary map positions across seasons (time environment) irrespective of changing genetic background.

If these requirements are not satisfied, step 1 must be repeated every time RFLP marking of QTL is attempted.

3. For any crop plant, improvement of yield warrants simultaneous improvement in a number of QTs. This is a major reason why plant breeding methodologies are built around phenotypic values, variation and selection, and take time. Compared to RFLP-based techniques, field search and evaluation strategies are technically simple and several times less expensive in terms of cost, time, and requirement of trained personnel and infrastructure.
4. Making inferences from a large population in segregating generations requires a quantitative assay since plants containing the desirable gene combinations in the field occur at low frequency. RFLP-based assays cannot be done economically on a scale comparable to field assays of phenotypes. In this context, calculations of minimum sample size required for interval mapping also cannot help (see point (ii) under section 3.4).
5. Breeding methods to pyramid QT genes in a desired agronomic background are far more complex than the backcross or F_2 selection in which RFLP-based selection for QT has so far been attempted. Extending the methodology to such complex mating systems and selection will throw up fresh difficulties, such as multiple-cross parents, distorted segregation not fitting known theoretical patterns, the need to handle large samples, and so on.

The experiments reported in OSB, TAN and PAT are noteworthy as pioneering attempts towards RFLP-assisted QT improvement. However, careful consideration of critical views (for example Ellis 1986; Blake *et al.* 1991) and our foregoing analysis suggest that a large number of attempts using RFLP-based genotyping need to be done using a choice of parents and selection of recombinants in the field in the segregating generations before the potential of molecular marker-assisted QT improvement can be realized.

Acknowledgements

We sincerely thank Dr Deepak Pental (Tata Energy Research Institute, New Delhi) and Dr K. Vinod Prabhu (Division of Genetics, IARI, New Delhi) for discussions and constructive criticism.

References

Blake T., Lybeck N. and Hayes P. 1991 Good, bad and untested ideas in RFLP and QTL analysis. *Plant Breeding Abstr.* 61: 1–7

Ellis T. H. N. 1986 Restriction fragment length polymorphism markers in relation to quantitative characters. *Theor. Appl. Genet.* 72: 1–2

Lander E. S. and Botstein D. 1986a Mapping complex genetic traits in humans: New methods using a complex RFLP linkage map. *Cold Spring Harbor Symp. Quant. Biol.* 11: 49–62

Lander E. S. and Botstein D. 1986b Strategies for studying heterogeneous genetic traits in humans by using a linkage map of RFLPs. *Proc. Natl. Acad. Sci. USA* 83: 7353–7357

Lander E. S. and Botstein D. 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199

Mather K. 1951 *Measurement of linkage in heredity* (London: Methuen)

Osborn T. C., Alexander D. C. and Fobes J. F. 1987 Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. *Theor. Appl. Genet.* 73: 350–356

Paterson A. H., Lander E. S., Hewitt J. D., Paterson S., Lincoln S. E. and Tanksley S. D. 1988 Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721–726

Paterson A. H., Damon S., Hewitt J. D., Zamir D., Rabinowitch H. D., Lincoln S. E., Lander E. S. and Tanksley S. D. 1991 Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. *Genetics* 127: 181–197

Rick C. M. 1974 High soluble solids content in large fruited tomato lines derived from a wild green fruited species. *Hilgardia* 42: 493–510

Risch N. 1992 Genetic linkage: Interpreting LOD scores. *Science* 255: 803–804

Tanksley S. D. and Hewitt J. 1988 Use of molecular markers in breeding for soluble solids content in tomato—a re-examination. *Theor. Appl. Genet.* 75: 811–823

Thoday J. M. 1961 Location of polygenes. *Nature* 191: 368–370

Zerba K. E., Kessling A. M., Davignon J. and Sing C. F. 1991 Genetic structure and the search for genotype–phenotype relationships: An example from disequilibrium in the *ApoB* gene region. *Genetics* 129: 525–533