Theor Appl Genet (1989) 78: 613–618

Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci

J. Jensen

Agricultural Research Department, Risø National Laboratory, DK-4000 Roskilde, Denmark

Received May 20, 1989; Accepted July 31, 1989
Communicated by P. M. A. Tigerstedt

Summary. A new method is described to obtain maximum likelihood estimates of recombination frequencies between quantitative trait loci (QTL) and marker gene loci; it is based on Fisher's method of scoring and numerical differentiation. The method is applied to data from chromosome-doubled monoploid lines of barley originating from the F_1 generation of a cross between two well-adapted barley varieties. The lines segregated for marker gene loci *ddt* (DDT resistance) and *s* (short rachilla hairs) on chromosome 7. The quantitative trait of single-kernel weight was found statistically significantly associated with locus *s*, but not with locus *ddt*. The association is ascribed to a QTL designated *Kw1*. It could not be ascribed to pleiotropism at locus *s* since the recombination frequency between *s* and *Kw1* (0.26 ± 0.09) differed significantly from zero. The recombination frequencies between *Kw1* and *ddt* and between *ddt* and *s* were 0.42 ± 0.07 and 0.31 ± 0.03, respectively, suggesting the locus order *ddt, s, Kw1*. The segregation ratio for alleles in locus *Kw1* was estimated to be 43 : 57, which is not significantly different from a 1 : 1 ratio. Means and standard deviations of single-kernel weight for lines with either of the two *Kw1* alleles were estimated; the *Kw1* locus accounted for 25% of the variance of the single kernel weight.

Key words: Barley – Chromosome-doubled monoploids – Gene effects – Mixed normal distributions – Single kernel weight

Introduction

The observation of a statistically significant association of a quantitative character with the segregation of a marker gene locus may be the first step in a genetical mapping of a locus controlling such a character. The association indicates either that part of the variation of the quantitative character is due to a pleiotropic effect of the marker gene or that the marker gene is linked with one or more segregating loci controlling part of the quantitative character, an effective factor (Mather and Jinks 1971). When the effective factor is a single locus, although it may be complex, Geldermann (1975) described it as a quantitative trait locus (QTL). The strength of the association of a marker gene with a QTL depends on the degree of linkage and of the size of the effects of the QTL alleles.

Genetic techniques for locating a QTL were reviewed in the book by Thompson and Thoday (1979). Different statistical methods to estimate recombination frequencies between a marker gene and a QTL based on variance, on moments of higher order, and on the method of maximum likelihood have been treated theoretically (Choo 1983; Ginzburg 1983; Haseman and Elston 1972; Hill 1975; Jayakar 1970; Lander and Botstein 1989; Weller 1986; Zhuchenko et al. 1978). However, recombination estimates involving a QTL based on experimental data have been reported only for tomato (Paterson et al. 1988; Weller 1986, 1987; Zhuchenko et al. 1978, 1979).

The present paper describes a maximum likelihood method to estimate recombination frequencies between marker genes and a QTL, as well as the effects and segregation ratio of the linked QTL. The method is applied to a QTL controlling single kernel weight in chromosome-doubled monoploid barley material.

Materials and methods

A total of 198 chromosome-doubled monoploid barley lines produced by the bulbosum method (Jensen 1976) from the F_1

generation of three crosses, Risø 'Mutant 5678' × 'Foma', Risø 'Mutant 6018' × 'Foma', and 'Mutant SR7' × 'Carlsberg II', were studied. 'Mutant 5678' and 'Mutant 6018' were induced in the variety 'Carlsberg II' and carry the barley powdery mildew resistance mutant genes *ml-o5* and *ml-o6*, respectively; 'Mutant SR7' was induced in the variety 'Foma' and has the resistance mutant allele *ml-o10* (cf. Jørgensen 1976). The *ml-o* locus is located on chromosome 4. The *ml-o* genes have the following pleiotropic effects, depending on gene background: necrotic leaf spotting, lowered grain yield, and reduced single seed weight (Schwarzbach 1976).

The chromosome-doubled monoploids are supposed to represent a random sample of the genomes from the female gametes of the F_1. Apart from undetected mutations possibly induced simultaneously with the *ml-o* genes, the chromosome-doubled lines segregated only for the difference in the *ml-o* locus, in addition to differences between the genomes of the homozygotic varieties 'Foma' and 'Carlsberg II'.

'Carlsberg II' has long rachilla hairs and is resistant to the insecticide DDT. The characters are controlled by marker genes *S* and *ddt*, respectively. 'Foma' has genes *s* and *Ddt* and, furthermore, has a lower single-kernel weight than 'Carlsberg II'. Loci *s* and *ddt* are both located on the linkage map of barley chromosome 7 at positions 0 and 61.8 centimorgans, respectively (Jensen 1981).

From each cross 36 resistant and 30 susceptible lines were randomly selected. The lines as well as the parents were grown in three replicates in field plots, each of size 5.3 m^2.

The single-kernel weight was determined by weighing 400–500 kernels taken at random from each plot. The length of rachilla hairs was determined by visual inspection. The reaction to DDT was found by applying DDT to seedlings grown in a greenhouse (Jensen 1979).

The single-kernel weight of each plot was adjusted for systematic effects caused by the position of the plot in the field, as well as for the effect of the alleles in the *ml-o* powdery mildew resistance locus, and for the effect of the three crosses (ascribed to accidental mutations).

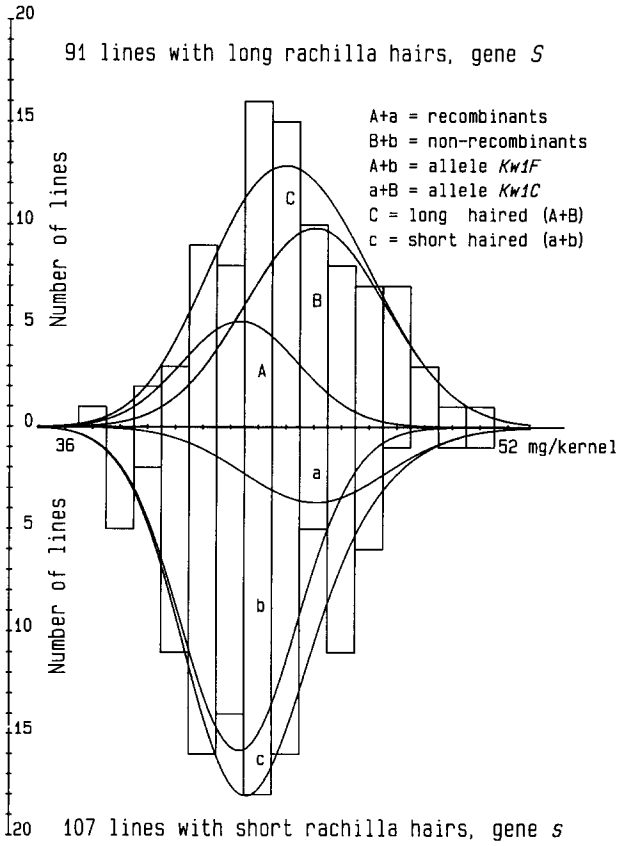

Fig. 1. The distribution of single-kernel weight of lines having long and short rachilla hairs is shown in the histograms, above and below the horizontal single-kernel weight axis, respectively. The normal distributions *A*, *a*, *B*, *b*, and the mixed normal distributions *C* and *c* are calculated from the estimated parameters

Results

Detection of associations

Single kernel weight was found statistically significantly associated with the character length of rachilla hairs determined by alleles of locus *s* ($\chi^2_{6df} = 16$, testing the independence of segregation of locus *s* and a grouped distribution of single kernel weight). The histograms in Fig. 1 show the association. The mean single-kernel weights of lines with allele *s* and *S* were 42.88 ± 0.25 mg and 44.18 ± 0.28 mg, respectively.

The association indicates either that the alleles in locus *s* have a pleiotropic effect on single kernel weight, or that locus *s* is linked with a quantitative trait locus (QTL) named *Kw1* (kernel weight). The alleles of *Kw1*, assumed to occur in 'Carlsberg II' and 'Foma', are designated *C* and *F*, respectively. Single-kernel weight has previously been reported associated with locus *s* (Bal et al. 1959); it has been suggested that it is controlled by a locus designated *Kw* (Robertson et al. 1965) like the single-kernel weight loci on chromosomes 1 and 2 (Smith 1951; Robertson et al. 1955). The segregation for locus *ddt* was not statistically significantly associated with single-kernel weight.

Estimation of recombination frequency and gene effect

The recombination frequencies between the assumed locus for single-kernel weight, *Kw1*, and loci *s* and *ddt* were estimated as follows: the expected frequency of the eight different genotypes in terms of the recombination frequencies p_1, p_2, and p_3 between *Kw1* and *ddt*, *ddt* and *s*, and *Kw1* and *s*, respectively, are shown in Table 1. Furthermore, a parameter *h*, which expresses the proportion of lines with the *Kw1C* allele, is included in order that the segregation ratios for locus *Kw1* can be tested. It is assumed that the single-kernel weights of the two groups of lines having alleles *Kw1C* and *Kw1F* are normally distributed with mean values M_C and M_F and standard deviations S_C and S_F, respectively. The likelihood

function L can then be written as:

$$L=\prod_{i=1}^{4}\prod_{j=1}^{A_i}\frac{1}{\sqrt{2\pi}}\left[\frac{h_{g_i}}{S_c}\exp\left\{-\frac{1}{2}\left(\frac{x_{i,j}-M_c}{S_c}\right)^2\right\}+\frac{(1-h)\,g_{i+4}}{S_F}\exp\left\{-\frac{1}{2}\left(\frac{x_{i,j}-M_F}{S_F}\right)^2\right\}\right]$$

where A_i is the number of lines of the $i=1$, 2, 3, and 4 marker gene genotypes *ddt,S; ddt,s; Ddt,S,* and *Ddt,s*, respectively ($\Sigma A_i=198$), and $X_{i,j}$ is the observed single-kernel weight of line no. j within lines having marker gene genotype no. i.

The likelihood function is maximized with respect to the eight parameters by Fisher's method of scoring for parameters using the observed amount of information (Bailey 1961). A small general program for a microcomputer was written in BASIC to perform the necessary calculations, including the single and double partial differentiation of the function by simple numerical methods. The program interactively allows for maximizing a likelihood function for a variable subset of parameters by keeping the remaining parameters constant. The function was maximized for each of the different parameters a number of times in order to obtain convergence when using poor starting values in the iteration procedure. In that way reasonable starting values were obtained for the final iteration of all parameters at once.

Due to rounding errors, the number of significant digits in the estimated parameters and their covariances or standard deviations may only be, respectively, about half and a quarter of those with which the computer works. In the present study the computer worked with 16 digits.

The parameters estimated are shown in Table 2. The proportion (h) of lines having the *Kw1C* allele is estimated as 43% ±8%. This value does not deviate significantly from the expected 50% of a single-gene segregation ratio. The recombination frequency of 26% ±9% between *s* and the hypothesized locus *Kw1* is significantly different from zero. It is therefore unlikely that pleiotropic effects of the alleles in locus *s* are responsible for the associations. However, the possibility exists that alleles in locus

Table 1. The expected frequencies of gametes of chromosome-doubled monoploid genotypes

Gamete type	Expected frequency
Kw1C, ddt, S	$g_1=h(2-p_1-p_2-p_3)/2$
Kw1C, ddt, s	$g_2=h(-p_1+p_2+p_3)/2$
Kw1C, Ddt, S	$g_3=h(p_1+p_2-p_3)/2$
Kw1C, Ddt, s	$g_4=h(p_1-p_2+p_3)/2$
Kw1F, ddt, S	$g_5=(1-h)(p_1-p_2+p_3)/2$
Kw1F, ddt, s	$g_6=(1-h)(p_1+p_2-p_3)/2$
Kw1F, Ddt, S	$g_7=(1-h)(-p_1+p_2+p_3)/2$
Kw1F, Ddt, s	$g_8=(1-h)(2-p_1-p_2-p_3)/2$

Table 2. Parameters estimated by the likelihood function. Extra decimals are provided in order to enable the reader to check later calculations

Recombination frequency	*Kw1* ~ *ddt*	p_1	0.423740 ±0.071767	
	ddt ~ *s*	p_2	0.313131 ±0.032959	
	Kw1 ~ *s*	p_3	0.262310 ±0.092356	
Proportion of lines with allele *C* of locus *Kw1*		h	0.424976 ±0.080932	
	Mean		Standard deviation	
Single-kernel weight (mg)	M_C	45.033118 ±0.646916	S_C	2.578596 ±0.375863
	M_F	42.328541 ±0.360398	S_F	2.095417 ±0.224603

Covariance/correlation Matrix[a]

	p_1	p_2	p_3	h	M_C	M_F	S_C	S_F
p_1	+0.005150	+0.000528	+0.003532	−0.000123	+0.010196	−0.006958	−0.003531	−0.006027
p_2	+0.22	+0.001086	−0.000032	0.000000	0.000000	0.000000	0.000000	0.000000
p_3	+0.53	−0.01	+0.008530	−0.002821	+0.039515	−0.015937	−0.018657	−0.009520
h	−0.02	0.00	−0.38	+0.006550	−0.030451	−0.002497	+0.009307	−0.000534
M_C	+0.22	0.00	+0.66	−0.58	+0.418500	−0.107671	−0.162682	−0.054911
M_F	−0.27	0.00	−0.48	−0.09	−0.46	+0.129887	+0.076427	+0.043094
S_C	−0.13	0.00	−0.54	+0.31	−0.67	+0.56	+0.141273	+0.025660
S_F	−0.37	0.00	−0.46	−0.03	−0.38	+0.53	+0.30	+0.050446

[a] Covariance, variance and correlation coefficient, respectively, above, on and below the diagonal

s together with one or more linked genes on chromosome 7 have effects on single-kernel weight.

It is also seen from Table 2 that the recombination frequency between *Kw1* and *ddt* is the largest one; however, statistically it is no larger than the recombination frequency between *ddt* and *s*. The most likely locus order, *ddt-s-Kw1*, therefore needs further confirmation.

The recombination percentage of 31.3% ± 3.3% between *ddt* and *s* could have been calculated in a much simpler way from the direct observations; however, the more complex method used also gives its covariances (Table 2) with the other parameters. The covariances, or correlation coefficients, are low to the two other recombination frequencies and, as expected, zero to the remaining parameters (h, M_C, M_F, S_C, and S_F). Based on the estimated recombination frequencies, the coefficient of coincidence is calculated as 0.92. This is of an order that could be expected in barley when dealing with three single genes. The mean single-kernel weight of lines with genes *Kw1C* and *Kw1F* differs by 2.7 mg. This is three times the difference between the single-kernel weights of the two parents, 'Carlsberg II' and 'Foma', which were 45.0 ± 0.24 mg and 44.1 ± 0.18 mg, respectively, when adjusted for external effects (plot location, etc.). This suggests segregation for additional, independently inherited genes.

Histograms showing the distribution of single-kernel weight of the lines with short and long rachilla hairs are given in Fig. 1. The figure further shows both the normal density distributions based on the estimated parameters as well as the mixed density distributions. The latter distributions look rather normal and have only one mode, which should also be expected when the difference between the means relative to the standard deviations is as small as here (Behboodian 1970).

Table 3 gives various expectations calculated on the basis of the estimates given in Table 2. The expectations are checked against the corresponding values obtained from an analysis of variance (exp versus obs). The expected total mean value of single-kernel weight and total variance is exactly the same as found by the analysis of variance. However, it is possible to divide the variance into: (a) between alleles (*Kw1C* and *Kw1F*) and (b) within alleles. The first value accounts for 25% of total variance. For each of the four marker gene genotypes, the observed and expected values for (1) number of lines, (2) means of single-kernel weights, and (3) their standard deviations did not deviate much, indicating a good fit of the model and a good precision of estimates.

Table 3. Comparison of directly calculated statistics (obs) with expected values (exp) calculated on the basis of the *Kw1* model, with parameters as estimated parameters from the likelihood function

Total aggregate statistics values			
Mean (mg)	$M_T = M_C h + M_F(1-h)$	obs = exp = 43.48	
Variance between genotypes (*C* and *F*)	$V_b = (M_C - M_T)^2 h + (M_F - M_T)^2 (1-h)$	exp = 1.7875	
Variance within genotype	$V_w = S_C^2 h + S_F^2 (1-h)$	exp = 5.3505	
Sum of variances	$V_T = V_b + V_w$	obs = exp = 7.1380	

Within marker gene genotypes					
i	No. of lines			Mean value of single kernel weight (mg)	
	Genotype	obs	exp[a]	obs[b]	exp[c]
1	*ddt, S*	61	63.3	44.20 ± 0.29	44.13 ± 0.34
2	*ddt, s*	32	33.4	42.71 ± 0.45	42.85 ± 0.42
3	*Ddt, S*	30	28.6	44.14 ± 0.59	44.22 ± 0.51
4	*Ddt, s*	75	72.7	42.95 ± 0.29	42.91 ± 0.29

[a] The expected number of lines is calculated as 198 $(g_i + g_{i+4})$

[b] These means and their standard deviations are from a grouped analysis of variance

[c] The expected means and standard deviations are calculated similarly to the total aggregate statistics, but h and $1-h$ are replaced by $g_i/(g_i+g_{i+4})$ and $g_{i+4}/(g_i+g_{i+4})$, respectively, and the sum of variances for each genotype is divided by the expected number of lines (see footnote a) and the square root is taken to obtain the standard deviation

Discussion and conclusion

The recombination percentage of 26 between *s* and *Kw1* corresponds to a map distance of 29 centimorgans (cM). This suggests that locus *Kw1* is located close to locus *r* (rough awn) which is located on the chromosome 7 map at 30 cM from locus *s* (Jensen 1981). The *Kw1* locus may be the same locus as the one which Bal et al. (1959) associated with locus *s*.

The estimated recombination percent between *s* and *Kw1* has a rather high standard deviation, 9.24. However, if the genotypes at locus *Kw1* could have been classified exactly, 198 lines and a recombination percent of 26.2 give a standard deviation of only 3.13. In other words, the exactness of the estimates corresponds to what could have been obtained with only 23 lines, if single-kernel weight was a qualitative character. This illustrates that the actual number of lines used, 198, with the present experimental design is near to the lower limit of the necessary number. The variance of the single kernel weight of the lines and the coefficient of variation were only 0.35% and 1.4%, respectively. The variance measured on a spaced single plant basis is about 25 times as high (unpublished), indicating the advantage of using chromosome-doubled monoploid lines in replicated field plots.

The fact that the estimated segregation ratio 43 : 57 for locus *Kw1* did not deviate significantly from the expected 1 : 1 ratio indicates that no serious bias in transmission or viability is associated with *Kw1*. It indicates further that the alleles in locus *Kw1* do not interact with alleles in other segregating loci. However, it does not tell us that the lines segregate for only one locus associated with *s* that affects single kernel weight. It is not possible to find out if locus *Kw1* consists of a block of linked loci. But a large block is expected to give a low coincidence coefficient, and a coincidence coefficient higher than expected at the Kosambi level was obtained ($0.9 > 2 \times 0.432$). Even for a qualitative character, it may be difficult to determine if it is controlled by a single locus or by a block of linked loci.

Kernels with long rachilla hairs (*S*) probably weigh somewhat more than kernels with short hairs (*s*) due to the weight of the hairs. However, hair weight accounts for probably less than 1% of the estimated allele effect. The lines probably segregate for many other loci with similar small allele effects.

It is reasonable to assume that loci exist that contain alleles with all grades of effects, many with nearly no effect and a decreasing number with increasing effects. Furthermore, it is reasonable to assume that alleles segregating for a quantitative character are randomly distributed in the two genomes in a diploid organism. Table 3 shows that the alleles in locus *Kw1* account for 25% of the variance between lines. This does not leave much room for the lines to segregate for loci with major effects on single-kernel weight that are not located on chromosome 7. However, such loci are much more likely to occur than two such linked loci on chromosome 7.

We have assumed that for each *Kw1* allele the single-kernel weight was normally distributed. However, if a secondary locus placed on a chromosome different from chromosome 7 segregated for alleles with major effects on single kernel weight, the requirements of normal distributions would not be fulfilled. If the effects of unlinked QTLs are smaller than those of the linked QTL, the deviation from normality will probably have little effect on the estimates; this has to be studied further.

The parents of the crosses studied were not very closely related, but morphologically they are very similar and agronomically they are well adapted. If one uses completely unrelated and very dissimilar parents, segregation at many loci is expected. With an increasing number of segregating loci the probability of interacting loci and of linked loci with reduced viability also increases. Such problems occur in studies of conventional marker gene loci, and it may also be a serious problem in the study of QTL.

A verification of the efficiency of the method of estimation and of the results obtained was made by applying the estimation method separately to each of the three replicates. The estimates for each replicate were all in good agreement with the results given above based on the mean single-kernel weight for each line. The method of estimating recombination frequency to a QTL has also been tested on simulated data. In addition, the computer program was tested on the data of ash content of peat from a bog given by Hall (1952). Estimates identical to those obtained by Hasselblad (1966), who also used a maximum likelihood method on the data of Hall, were obtained.

The iteration procedure to solve the maximum likelihood equations requires considerably more computer power when one of the loci is a QTL; nevertheless, the computation can still be made without too much effort on a microcomputer programmed in an interpretive language like BASIC. This is in contrast to the view of Weller (1986), who thought that it was too cumbersome and time-consuming to obtain iterative solutions of the maximum likelihood equations; however, he was working with F_2 populations. Instead, Weller proposed an approximate maximum likelihood method which he nevertheless still thought required a significant investment in computing time of a mainframe computer.

Acknowledgements. I thank my colleagues J. H. Jørgensen, H. P. Jensen, and B. Kjær for supplying me with their scored data. Their data and their interest in the association between qualitative and quantitative characters inspired me to investigate the possibility of identifying QTLs and estimating their recombination with marker gene loci. The critical reading of the manuscript and valuable suggestions of H. Doll, J. Hilden, and I. Linde-Laursen are very much appreciated.

References

Bailey NT (1961) Introduction to the mathematical theory of genetical linkage. Clarendon Press, Oxford

Bal BS, Suneson CA, Ramage RT (1959) Genetic shift during 30 generations of natural selection in barley. Agron J 51: 555–557

Behboodian J (1970) On the modes of a mixture of two normal distributions. Technometrics 12:131–139

Choo TM (1983) Doubled haploids for locating polygenes. Can J Genet Cytol 25:425–429

Geldermann H (1975) Investigations on inheritance of quantitative characters in animals by gene markers. 1. Methods. Theor Appl Genet 46:319–330

Ginzburg EKH (1983) Possible localization of genes controlling a quantitative character in self-pollinators. Genetika 19: 577–583 (in Russian)

Hall A (1952) Statistical theory with engineering applications. Wiley, New York, pp 152–158

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Hasselblad V (1966) Estimation of parameters for a mixture of normal distributions. Technometrics 8:431–444

Hill AP (1975) Quantitative linkage:a statistical procedure for its detection and estimation. Ann Hum Genet 38:439–449

Jayakar SD (1970) On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. Biometrics 26:451–464

Jensen CJ (1976) Barley monoploids and doubled monoploids: Techniques and experience. In: Gaul H (ed) Barley genetics III. Thiemig, Munich, pp 316–345

Jensen J (1979) Locating of a high-lysine gene and the DDT-resistance gene on barley chromosome 7. Euphytica 28:47–56

Jensen J (1981) Construction of a barley chromosome 7 linkage map. In: Asher MJC, Ellis RP, Hayter AM, Whitehouse RNH (eds) Barley genetics IV. Edinburgh University Press, Edinburgh, pp 927–939

Jørgensen JH (1976) Identification of powdery mildew resistant barley mutants and their allelic relationship. In: Gaul H (ed) Barley genetics III. Thiemig, Munich, pp 446–455

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Mather K, Jinks JL (1971) Biometrical genetics. 2nd edn. Chapman and Hall, London

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Robertson DW, Wiebe GA, Shands RG (1955) A summary of linkage studies in barley: Supplement II, 1947–1953. Agron J 47:418–425

Robertson DW, Wiebe GA, Shands RG, Hagberg A (1965) A summary of linkage studies in cultivated barley, *Hordeum* species: Supplement III, 1954–1963. Crop Sci 5:33–43

Schwarzbach E (1976) The pleiotropic effects of the *ml-o* gene and their implications in breeding. In: Gaul H (ed) Barley genetics III. Thiemig, Munich, pp 440–445

Smith L (1951) Genetics and cytology of barley. Bot Rev 17:133–202

Thompson JN Jr, Thoday JM (1979) Quantitative genetic variation. Academic Press, New York

Weller JI (1986) Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42:627–640

Weller JI (1987) Mapping and analysis of quantitative trait loci in *Lycopersicon* (tomato) with the aid of genetic markers using approximate maximum likelihood methods. Heredity 59:413–421

Zhuchenko AA, Korol AB, Andryushchenko VK (1978) Linkage between the loci for quantitative characters and marker loci. I. Model. Genetika 14:771–778 (in Russian)

Zhuchenko AA, Samoval AP, Korol AB, Andryushchenko VK (1979) 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 (in Russian)