Genet. Res., Camb. (1993), **62**, *pp.* 195–203 *With 2 text-figures* Copyright © 1993 Cambridge University Press

Detection of quantitative trait loci from frequency changes of marker alleles under selection

PETER D. KEIGHTLEY[1]* AND GRAHAME BULFIELD[2]
[1] *Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, Scotland*
[2] *AFRC Roslin Institute (Edinburgh), Roslin, Midlothian, EH25 9PS, Scotland*

(*Received 17 March 1993 and in revised form 18 August 1993*)

Summary

A method was developed to estimate effects of quantitative trait loci (QTL) by maximum likelihood using information from changes of gene frequency at marker loci under selection, assuming an additive model of complete linkage between markers and QTL. The method was applied to data from 16 molecular and coat colour marker loci in mouse lines derived from the F_2 of two inbred strains which were divergently selected on 6-week weight for 21 generations. In 4 regions of the genome, marker allele frequencies were more extreme than could be explained by sampling, implying selection at nearby QTL. An effect of about 0·5 standard deviations was located on chromosome 11, and accounted for nearly 10% of the genetic variance in the base population. QTL with effects as small as 0·2 phenotypic standard deviations could be detected. For typing of a given number of individuals, the power of detection of QTL is very high compared to, for example, analysis of an F_2 population. The joint effects of linkage and selection were investigated by Monte Carlo simulation. Marker gene frequencies change little as a consequence of selection at a QTL unless the marker and QTL are less than about 20 cM apart.

1. Introduction

The problem of detection and nature of individual loci affecting polygenic traits (quantitative trait loci, QTL: strictly regions of the genome influencing quantitative traits) led to the development of methods to measure the effects of chromosome segments delineated by visible genetic markers (Sax, 1923; Thoday, 1961). Recent discoveries that the genome contains an effectively limitless supply of easily scorable neutral molecular genetic markers have led to a resurgence of interest in QTL mapping. The standard experimental strategy is to type individuals for molecular markers distributed throughout the genome in an F_2 or backcross of two genetically different strains in order to determine the degree of association between regions of the genome and the quantitative trait value. This technique, or variants of it, is seeing wide application in many plant species (Paterson *et al.* 1988; Stuber *et al.* 1992), and increasingly in animals (Jacob *et al.* 1991; Todd *et al.* 1991; Georges, 1993). In most situations large numbers of individuals require to be typed to allow a reasonable power of detection (Soller & Beckmann, 1990). An increase in this power can be achieved by selectively typing extreme individuals for the quantitative trait (Lebowitz, Soller & Beckmann, 1987). Much greater phenotypic differences between groups containing different marker alleles can be generated by artificial selection on the quantitative trait over many generations. Garnett & Falconer (1975) attempted to locate some of the genes responsible for the selection response in Falconer's (1973) 'Q' mouse lines, which were divergently selected on 6-week body weight, by comparing electrophoretic and coat colour marker allele frequencies in high and low lines. One marker (*Hbb*) out of 9 scored showed a statistically significant association with body size, as one allele was fixed in replicates selected for large size and was at a low frequency in replicates selected for small size.

The selection experiment described here (the 'X' lines) was initiated specifically for the location of QTL affecting body size in the mouse. There is a long history of selection experiments on this trait in the mouse (Falconer, 1989). In the majority of cases base populations have been synthesized from crosses between several inbred and sometimes outbred strains, and realized heritabilities as high as 40% are not atypical. However, the base population for the present experiment was the F_2 of two well-characterized

* Corresponding author.

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press

inbred strains, C57BL/6J and DBA/2J. Linkages between marker loci and QTL are detected as changes in marker allele frequencies as a consequence of selection at QTL, and their detection is made easier because the initial gene frequency was 0·5 and there was strong linkage disequilibrium in the base population. Although the parental strains were almost indistinguishable in body size, the selection lines differed by between 4 and 5 phenotypic standard deviations (10 g) at generation 21, implying that segregation and recombination have revealed considerable quantities of 'hidden' quantitative genetic variation.

In this paper a method based on maximum likelihood is developed to infer magnitudes of effects of QTL by using information from gene frequency changes at marker loci under artificial selection. The method is applied to data on gene frequencies in the X lines at 16 marker loci, of which 2 were coat colour, 2 were electrophoretic and 12 were proviruses.

2. Materials and methods

(i) *Selection lines*

The base population (generation 0) was the F_2 of the inbred mouse strains C57BL6/J and DBA/2J. Within-family selection was practised on body weight at 6 weeks with a mating scheme that minimizes inbreeding (Falconer, 1973). Six replicates of 8 pairs of parents were selected for increased size (Larges) and 6 replicates were selected for decreased size (Smalls) over 21 generations. The realized heritability for body size was about 25%, and the Large and Small lines differed by between 4 and 5 standard deviations.

Animals. A total of 96 individuals from the 6 Small sublines and 91 individuals from the Large sublines were typed at two electrophoretic loci at generation 10. A total of 93 individuals from the Small sublines and 34 individuals from the Large sublines were typed with a provirus probe at generation 21. The origin with respect to subline of these 34 Larges was unknown, but it is likely that they were a random sample from the 6 sublines.

(ii) *Analysis of molecular markers*

Endogenous retroviruses. The mouse genome contains many tens of copies of various classes of non-ecotropic proviruses. These are stable elements (Frankel *et al.* 1990), and inbred strains differ greatly in their complement of retrovirus insertions. Using oligonucleotide probes Frankel *et al.* (1990) identified 58 proviruses which vary between C57BL/6J and DBA/2J (i.e. are present in one strain, but absent in the other). These are detected as junction fragments between provirus and mouse-flanking DNA. A less specific probe cloned from the envelope glycoprotein of one such endogenous retrovirus (Holland, Wozney & Hopkins, 1983) hybridizes to many of these endogenous retroviruses and generates a very complex pattern of junction fragments in a suitable Southern Blot (Blatt *et al.* 1983). Such a probe was used in the present study, and it was found that 12 loci differing between C57BL/6J could be used as markers. These loci act as dominant markers.

Southern blotting. Genomic DNA was extracted from spleen or from tail tissue by standard methods. Some 10 μg of DNA were digested to completion with *Bam*H I, and electrophoresed at 0·8 V/cm for 48 h in a 20 cm 0·6% agarose–TBE gel containing 1 μg/ml of ethidium bromide. The DNA was partially depurinated by immersing the gel in 0·25 M-HCl for 15 min, then transferred to a charged nylon membrane (Hybond-N+, Amersham Ltd, England) by capillary blotting with 0·4 M-NaOH for 18 h. The 622-base pair *Bam*H I–*Eco*R I fragment of the gp70 gene of murine retrovirus MCF 247 (Holland *et al.*, 1983) (probe designated α_{18} henceforth) was labelled with ^{32}P by random oligonucleotide primer extension (Feinberg & Vogelstein, 1983) to a specific activity of about $1{\cdot}5\times10^9$ dpm/μg and hybridized to the membrane-bound DNA at a concentration of about 5 ng/ml in 1 mM-EDTA, 0·5 M-NaHPO$_4$, pH 7·2, 7% SDS (Church & Gilbert, 1984) for 18 h at 65°. Membranes were washed three times at 65° for 30 min in 1 mM-EDTA, 40 mM-NaHPO$_4$, pH 7·2, 1% SDS, and autoradiographed for 1–5 days at −70° using an intensifying screen. The period of exposure for optimal visualization of bands corresponding to different proviruses varied considerably. Individuals were scored for the presence or absence of 12 provirus junction fragments (see results). Autoradiographs were scored twice by two people independently.

Protein electromorphs. Genotypes of individuals at the *Gpi-1* and *Hbb* loci were determined by isoelectric focusing of blood samples according to the method of Bulfield & Bantin (1981).

(iii) *Inference of selection coefficients associated with marker loci*

A statistical procedure involving Monte Carlo simulation was developed to infer maximum likelihood (ML) estimates of selection coefficients, *s*, associated with marker loci under the assumption of complete linkage between a marker and a QTL. The method uses information from the observed distribution of gene or phenotype frequencies at marker loci in the replicated selection lines.

A replicate of the selection experiment was simulated by initializing a population of *N* diploid parents with one diallelic locus segregating at a frequency of 0·5. The locus altered the relative fertilities of the parents according to an additive model, with *s* the difference in fitness between the homozygotes. The gene frequency was subsequently allowed to change by selection and drift for *t* generations by sampling

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press

parents with replacement, and generating one offspring from each pair. N offspring were regenerated in this way each generation. In order to generate the appropriate amount of random genetic drift, the population size in the simulation was set to 23, equal to the average effective population size, N_E, within the lines of the experiment. This was obtained from the pedigree by first computing the mean relationship, $\bar{\rho}$, among individuals within lines at generation $t = 20$, where the relationship, $\rho_{i,j}$, between two individuals is the (hypothetical) coefficient of inbreeding of an offspring of these individuals, and the effective population size is related to $\bar{\rho}$ according to $N_E = [1-(1-\bar{\rho}/2)^{1/t}]/2$. With no selection ($s = 0$), the effective population size in the simulation was, as expected, close to the actual population size of the simulation, but the former was reduced if selection was included. Although the X-line experiment involved within-family selection, fertility selection was simulated for reasons of computational efficiency.

The likelihood of the data with respect to s was obtained by replicating many sets of selection line simulations for different values of s. The likelihood of the observed phenotype frequency (in the case of proviruses) or gene frequency (in the case of electromorphs) is the proportion of occurrences of such a frequency in the simulation. Log likelihoods were computed independently for each replicate and added together to obtain the overall log likelihood of the data. Divergent selection was implicitly assumed by changing the sign of s when computing likelihoods for the Smalls and Larges. Any asymmetry of the change of marker allele frequency was not accounted for in the model: the ML estimate of s is an average for up and for down selection. In the case of the Larges at generation 21, the origin of animals with respect to subline was unknown, but it was assumed that these were a random sample from all 6 sublines. An overall likelihood for the Larges was computed by randomly sampling from 6 simulated replicates accordingly.

A total of 10^5 selection-line replicates were simulated per selection coefficient value with steps of 0·025. The maximum likelihood with respect to s was obtained by fitting a quadratic to the highest likelihood and the two values on either side. Variances of estimates were obtained from the curvature of the likelihood about the maximum according to

$$var(s_{ML}) = -\left[\frac{d^2 \log(L)}{ds^2}\right]^{-1}. \quad (1)$$

ML estimation of selection with a deterministic model was analysed by DuMouchel & Anderson (1968).

(iv) *Effect of recombination between QTL and marker*

In order to relax the assumption of complete linkage made above, Monte Carlo simulation was used to work out the expected change in gene frequency, Δq, for a marker linked to a QTL of selective advantage, s, under the assumption that recombination can occur. The same population size, structure and selection regime as above were assumed. A QTL was assumed to have additive gene action and to be initially in complete linkage disequilibrium with a marker (as in the experimental F_1). Gametes were generated by allowing recombination to occur with probability c between the marker and the QTL. An F_2 population was generated without selection, then the average change in gene frequency of the marker over t generations of selection in 2000 replicates was computed for a range of s and c.

3. Results

(i) *Marker loci*

Junction fragments generated by the α_{18} *provirus probe.* The patterns of *Bam*H I provirus junction fragments

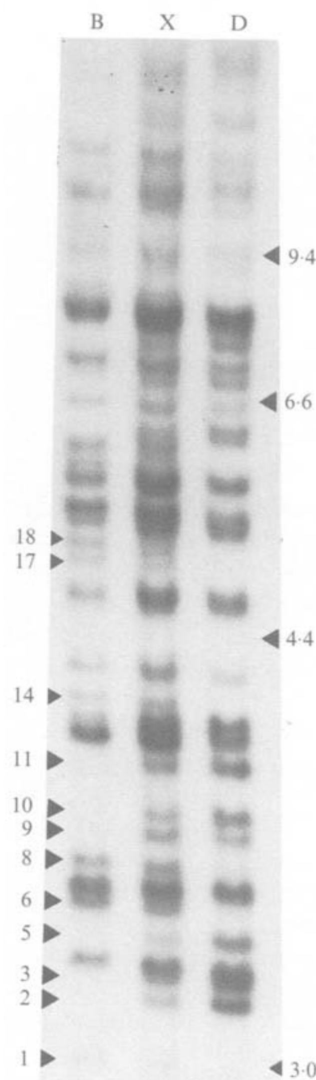

Fig. 1. Autoradiograph showing segregation of α_{18} *Bam*H I junction fragments in C57BL/6J, DBA/2J and the F_1 of these inbred strains. Bands used as markers are numbered down the left hand side, and an approximate scale in thousands of base pairs is shown on the right.

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press

Table 1. *Frequencies at marker loci*

(a) Genotype frequencies at the *Gpi-1* locus

Genotype	A	B	C	D	E	F
Large sublines						
B/B	4	15	0	3	0	0
B/D	2	1	5	6	8	3
D/D	10	0	11	4	5	13
Small sublines						
B/B	10	9	0	5	1	6
B/D	6	5	1	10	2	10
D/D	0	1	15	1	13	0

Numbers in each of three genotypic classes at the *Gpi-1* locus in the X line sublines at generation 10. The B allele refers to $GPi\text{-}1^b$, while the D allele refers to $Gpi\text{-}1^a$.

(b) Genotype frequencies at the *Hbb* locus

Genotype	A	B	C	D	E	F
Large sublines						
B/B	1	2	4	8	0	4
B/D	12	7	4	3	8	11
D/D	3	7	8	3	5	1
Small sublines						
B/B	0	1	8	1	3	2
B/D	5	8	5	9	6	8
D/D	11	6	3	6	7	6

Numbers in each of three genotypic classes at the *Hbb* locus in the X line sublines at generation 10. The B allele refers to Hbb^s, while the D allele refers to Hbb^d.

(c) Phenotype frequencies at retrovirus loci

Locus	Larges Total	Smalls Total	Small sublines A	B	C	D	E	F
Pmv-18	14	81	14	15	14	15	7	16
Pmv-40	12	76	12	12	12	9	16	15
Pmv-41	29	67	11	12	13	15	16	0
P5	27	50	9	4	11	0	16	10
Pmv-16	22	64	10	8	8	15	16	7
Pmv-11	18	74	5	10	15	15	13	16
Mpmv-25	23	42	12	0	7	7	10	6
P10	24	75	15	9	13	15	15	8
Mpmv-24	23	40	3	14	6	9	3	5
Mpmv-2	11	93	15	15	16	15	16	16
Mpmv-4	14	82	9	15	16	12	16	14
Xmv-15	30	40	5	10	14	6	5	0
Number typed	34	93	15	15	16	15	16	16

Phenotype frequencies in Larges and Smalls at retrovirus loci expressed as number of animals having the α_{18} junction fragment. The phenotype frequency within replicates is known for the Smalls only.

generated by the α_{18} non-ecotropic provirus probe in C57B L/6J and DBA/2J are shown in Fig. 1. Many tens of junction fragments are visible in both inbreds, and about one-third appear to be common to the two inbred strains. Many bands could be separated in some autoradiographs, of which 12 could be scored consistently.

Gene/phenotype frequencies. A sample of individuals from the Larges and Smalls from generation 10 was electrophoretically typed at the *Gpi-1* and *Hbb* loci, and at generation 21 a sample was typed using the α_{18} provirus probe. Gene frequencies for the electromorphs are shown in Tables 1 *a* and 1 *b*, and phenotype frequencies at 12 provirus loci are shown in Table 1 *c*. The identities of the provirus loci were obtained by comparing the strain distribution patterns in recombinant inbred lines with those of known non-ecotropic provirus loci (later section). Some loci (e.g. *Mpmv-2*) show obvious divergences in frequency between the Larges and Smalls. The *brown* and *dilute* genes, which segregate in the cross, showed consistent changes in frequency as a consequence of selection. *Brown* increased in frequency in the Larges and decreased in frequency in the Smalls, while *dilute* increased in frequency in the Smalls and decreased in the Larges. The estimated gene frequencies of the coat colour markers at generation 20 are included in Table 2.

Selection coefficients associated with markers. ML estimates of selection coefficients associated with markers computed using the method described above are shown in Table 2. Neither of the electrophoretic loci showed a significant divergence in gene frequency, but both coat colour markers and 4 out of 12 provirus markers show significant associations. Average gene frequencies among lines for the markers were computed under the assumption of Hardy–Weinberg equilibrium within lines (with slight bias because of the small numbers sampled), and are also shown in Table 2.

Gene effects associated with marker loci. For the present case of within-family selection, an expression relating the effect on a quantitative trait to the selection coefficient is,

$$s = i\frac{a}{\sigma_p}(1-r)\left[\frac{n-1}{n(1-t)}\right]^{\frac{1}{2}}, \tag{2}$$

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press

Table 2. *Gene frequencies and effects associated with markers*

	Gene frequency					
Locus	Large	Small	s	S.E.	a/σ_p	h^2 (%)
Pmv-18	0·72	0·23	0·11**	0·03	0·18	0·4
Pmv-40	0·24	0·62	−0·10	0·04	−0·17	0·3
Pmv-41	0·72	0·60	0·04	0·04	0·07	0·0
P5	0·65	0·40	0·07	0·05	0·13	0·1
Pmv-16	0·49	0·45	−0·01	0·07	−0·01	−0·2
Pmv-11	0·62	0·34	0·07	0·03	0·11	0·1
Mpmv-25	0·52	0·28	0·09	0·03	0·16	0·3
P10	0·54	0·66	−0·04	0·03	−0·07	0·0
Mpmv-24	0·52	0·28	0·06	0·04	0·11	0·1
Mpmv-2	0·79	0·00	0·27**	0·04	0·47	2·8
Mpmv-4	0·72	0·33	0·14*	0·04	0·24	0·7
Xmv-15	0·33	0·73	−0·11*	0·03	−0·18	0·4
Gpi-1	0·62	0·49	0·03	0·03	0·06	0·0
Hbb	0·54	0·63	−0·02	0·04	−0·04	−0·0
Brown	0·65	0·09	0·12**	0·06	0·20	0·38
Dilute	0·15	0·84	−0·19**	0·04	−0·32	1·2

Allele frequencies (with respect to the allele of the DBA/2J parent), ML estimates of selection coefficients, effects (a) in standard deviations, and base population heritabilities associated with marker loci. The signs of s and a are with respect to the substitution of the two C57BL/6J alleles with the DBA/2J alleles and refers to selection upwards on body size. The standard errors of a/σ_p are about twice the standard error of the corresponding s. (*$2 < \ln L - \ln L_{s=0} < 3$; **$\ln L - \ln L_{s=0} > 3$)

(Falconer, 1989, chapters 11 and 13), where a is the effect associated with the marker (difference between the homozygotes). Other parameters of this equation together with values (in brackets) estimated from the data are: i, intensity of selection (0·98); σ_p, phenotypic standard deviation (2·1 g); $r = \frac{1}{2}$, correlation of breeding values within families; n, average number in each sex family (3·9); and t, phenotypic correlation of family members (0·50). Effects on body size associated with the markers expressed in grams and in phenotypic standard deviations calculated from equation (2) are shown in Table 2.

Heritability accounted for by the markers. The base population heritability associated with a marker i of effect a_i is $h_i^2 = \frac{1}{2}a_i^2 q_0(1-q_0)/\sigma_p^2$. Because the base population gene frequency, q_0, is known to be $\frac{1}{2}$, this becomes $h_i^2 = a^2/(8\sigma_p^2)$. Estimates of heritability associated with markers are biased downwards because of recombination between the marker and QTL, but biased upwards because of sampling. Without information from several linked markers, it is difficult to account for the bias from recombination, but the sampling bias can be reduced by subtracting the sampling variance associated with the estimate of the effect of each marker. The sampling variance of a is proportional to the sampling variance of s (equation (2)). This variance was obtained from the curvature of the likelihood about the maximum likelihood estimate of s (equation (1)). Estimates of heritabilities associated with the markers are shown in Table 2. Heritability estimates can be negative when the sampling variance exceeds the uncorrected estimate.

Assignment of α_{18} BamH I junction fragments to known non-ecotropic provirus loci using BXD Recombinant Inbred lines. The surprisingly high number of significant selection coefficient estimates suggested that some markers are closely linked. In order to map the provirus junction fragments and to determine if the α_{18} *Bam*HI junction fragments correspond to provirus loci that have already been mapped using more specific probes (Frankel *et al.* 1990), a panel of DNAs from 16 BXD recombinant inbred strains was probed with α_{18} as described above. Table 3 shows the Strain Distribution Patterns (SDPs) for 12 α_{18} junction fragments. In 10 out of 12 cases the α_{18} SDP corresponds uniquely to a published SDP for a non-ecotropic provirus (Frankel *et al.* 1989*a*, *b*; 1990). In two cases, loci *P5* and *P10*, there was no exact match, so these are presumed to be novel proviruses.

In two cases, pairs of loci which showed significant changes of gene frequency are very closely linked to one another. The marker associated with the largest change in gene frequency, *Mpmv-2*, is about 5 cM from *Mpmv-4* on chromosome 11 (Frankel *et al.* 1990), and both are associated with increased size. On chromosome 9, Provirus locus *Xmv-15* is about 5 cM from *dilute*, and both markers are associated with small size. A third pair of loci, *Pmv-18* and *Gpi-1* are within 10 cM of one another on chromosome 7. *Pmv-18* shows a significant association with large size, while *Gpi-1* also shows an association in this direction but not significantly. The direction of the changes of gene frequency associated with these three pairs of closely linked loci therefore accord with one another.

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press

Table 3. *Strain distribution pattern of retroviruses in RI lines*

	BXD line																		
Band	P	1	2	5	6	8	9	11	14	15	18	19	22	24	25	27	30	Known Locus	Chrom
18	B	D	D	D	B	D	D	B	D	B	D	D	B	B	D	B	D	*Xmv-15*	9
17	B	B	D	D	B	B	D	B	D	D	B	D	D	B	B	B	B	*Mpmv-4*	11
14	B	B	D	B	D	B	B	B	D	D	B	D	B	B	B	B	B	*Mpmv-2*	11
11	D	B	B	B	D	B	D	D	B	D	B	B	B	D	D	D	D	*Mpmv-24*	12
10	D	B	B	D	B	B	B	D	B	D	B	B	D	B	D	B	D	*Unknown*	?
9	D	B	B	D	B	D	D	D	B	B	B	B	D	D	D	D	D	*Mpmv-25*	1
8	B	B	D	D	B	B	D	B	D	B	B	D	D	D	B	D	D	*Pmv-11*	5
6	B	B	B	B	B	B	B	D	D	D	B	D	B	B	D	D	B	*Pmv-16*	16
5	D	D	D	D	B	B	D	B	B	D	D	D	B	B	D	D	B	*Unknown*	?
3	D	B	D	B	D	B	D	D	D	B	B	B	B	D	D	D	D	*Pmv-41*	13
2	D	D	D	D	D	B	D	D	B	B	B	D	B	B	B	D	B	*Pmv-40*	5
1	B	B	D	B	D	B	B	B	D	B	B	B	B	B	B	B	B	*Pmv-18*	7

SDP of α_{18} junction fragments in 16 BXD sublines and in the parental strains, *P. Locus* 17 has the same BXD SDP as both *Mpmv-4* and a closely linked provirus, *Xmv-42*. The chromosome assignments are from Frankel *et al.* (1990).

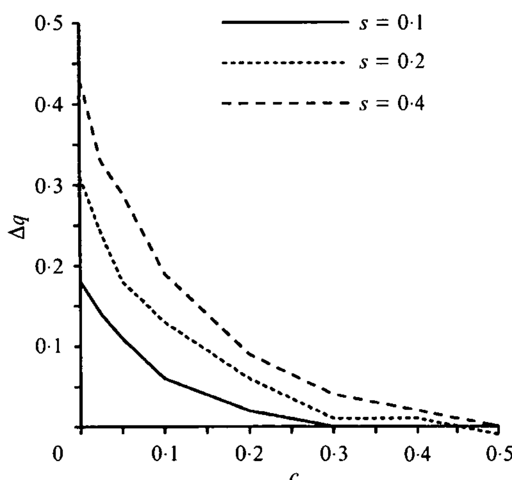

Fig. 2. Change in marker allele frequency (Δq) after 20 generations plotted against recombination fraction, c. The marker is in an initial state of complete linkage disequilibrium with a QTL, and the initial gene frequency is 0·5. Three selective values, s, of the QTL are shown. Additive gene action is assumed.

(ii) *Variable recombination*

The above method, which assumes complete linkage, underestimates the selection coefficient associated with a marker because, if recombination can occur between the marker and the QTL, a higher selection coefficient is required to explain a given change in gene frequency. The effect of varying the amount of recombination on change of gene frequency at a marker was computed by Monte Carlo simulation and is shown in Fig. 2. The expected change in gene frequency decreases by a factor of more than 2 between complete linkage and $c = 0{\cdot}1$.

(iii) *Association between markers and body size in an F_2 population*

In order to check the associations found in the selection lines, a separate experiment involving linkage analysis in an F_2 population was carried out. The size of the experiment implied sufficient power *a priori* to detect the effect associated with *Mpmv-2*, and this proved to be the case. 104 individuals from 12 F_2 litters were weighed at 6 weeks, and typed at non-ecotropic retrovirus loci using the α_{18} probe. In the case of *Mpmv-2*, the three genotypes could be distinguished on the basis of signal intensity by comparison with a nearby junction fragment cor-

Table 4. *Effects on body size associated with markers in an F_2 population*

Locus	Effect (g)	S.E.	$f(-/-)$
Pmv-18	−0·75	1·18	0·12
Pmv-40	−0·45	0·75	0·34
P5	−0·98	0·76	0·31
Pmv-16	0·32	0·84	0·24
Pmv-11	—	—	0·21
Mpmv-25	−0·34	0·76	0·32
P10	−0·70	0·78	0·30
Mpmv-24	−0·14	0·78	0·28
Mpmv-2	1·39*	0·63	0·34
Mpmv-4	—	—	0·34
Xmv-15	—	—	0·24
dilute	0·46	0·92	0·19
brown	−0·12	0·78	0·26

REML estimates of effects on body size associated with provirus and coat colour loci in an F_2 population of the two inbreds. The signs of the effects are with respect to the substitution of C57BL/6J alleles by DBA/2J alleles. Figures given are estimates of the difference between the homozygotes based on an additive model assuming Hardy–Weinberg proportions in the animals having the provirus junction fragment, with the exception of *Mpmv-2* where all three genotypes were included in the model. Three loci, *Mpmv-4*, *Xmv-15* and *Pmv-11* were not included in the model because these are closely linked to *Mpmv-2*, *dilute* and *Pmv-40* respectively. The final column shows the proportion of animals homozygous for the recessive allele, $f(-/-)$, in the sample of 104 animals. * $P < 0{\cdot}05$.

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press

responding to a provirus which does not appear to segregate. The genotypes were scored by eye twice by two people independently, and the agreement was 97%. It was important to be able to distinguish the three genotypes in the F_2 because the expected proportion of homozygous $-/-$ individuals is only $\frac{1}{4}$. In the X line experiment, inbreeding and selection increase the expected proportion of homozygotes. A statistical analysis by Restricted Maximum Likelihood (REML) (Patterson and Thompson, 1971) with litters as a random effect, and sex and genotype fitted as fixed effects provide a ML estimate (± 1 S.E.) for the effect of *Mpmv-2* (difference between the homozygotes relative to DBA/2J) of 1·4 g ($\pm 0·6$ g). The ML estimate of the effect associated with the heterozygote is 0·8 g ($\pm 0·6$ g) implying additive gene action. REML estimates of effects associated with the other retrovirus loci and the two-coat colour genes which segregate in the cross based on an additive model are shown in Table 4.

Segregation ratios of provirus junction fragments in the F_2. Also shown in Table 4 are segregation ratios for absence:presence of the retrovirus junction fragments in the F_2. The expected ratio is 1:3 for an autosomal locus undergoing Mendelian segregation. Only one locus, *Pmv-18* deviates significantly from the Mendelian ratio (χ^2 1 D.F. $= 10·0$, $P = 0·002$), raising the possibility that, for example, there are two junction fragments that migrate to the same position on the gel. In this particular case, however, the BXD strain distribution pattern locus is known for a probe which appears to generate the same 'provirus fingerprint' as α_{18} (Blatt *et al.* 1980), and this pattern was identical in the present BXD mapping experiment.

4. Discussion

Mapping of quantitative trait loci by the present approach is an alternative to more usual methods involving F_2 or backcross populations, but these differ greatly in power of detection. Relatively loosely linked markers are likely to show associations with QTL from analysis of F_2 or backcross populations because only one generation of recombination occurs. With selection over many generations, unless changes of gene frequency occur very rapidly (i.e. the QTL is of large effect and therefore strongly selected), even relatively closely linked markers are not expected to show significant associations. In contrast, the present method of inferring gene effects assumes complete linkage, and therefore under-estimates effects if recombination between markers and QTL can occur. For example, the simulation of the X lines suggests that with a QTL of selective value 0·1, the expected change of gene frequency over 20 generations at a completely linked marker is 0·18, but only 0·04 for a marker 20 cM away. This implies that a relatively dense map of markers will be required to detect the majority of the QTL responsible for the selection response in the X-lines, and contrasts with the conclusion that increasing the density of markers to below 20 cM apart is of little benefit for analysis of backcross populations of reasonable size (Darvasi *et al.* 1993).

There have been several previous reports in the literature concerning associations between body weight and the *dilute* and *brown* coat colour alleles (see references in Hedrick & Comstock, 1968). In most cases the *dilute* and *brown* alleles were associated with large size, but there is a fair degree of inconsistency, and experiments involving the generation of recombinant inbred lines suggest that the associations can be in either direction (Green, 1935). Hedrick & Comstock (1968) performed selection on post-weaning weight gain for 34 generations in the F_2 of inbred lines in which *brown* and *dilute* alleles segregated. The change of gene frequency suggested that the *brown* allele is associated with reduced post-weaning gain (note that *brown* is associated with increased 6 week weight in the present study), while changes of gene frequency of the *dilute* allele were in opposite directions in two replicates. These observations strongly suggest linkage between QTL affecting growth rate and the coat colour genes. In the present study, linkage rather than a direct effect of *dilute* is likely because the within-litter deviation of *dilute* individuals from wild type individuals was about $-0·5$ grams in the F_2, and approached zero after 10 generations. The within litter deviation of *brown* individuals did not show a consistent pattern over generations, but it is possible the effect of *brown* was masked by a difference in the direction of dominance between *brown* and the putative QTL (S. C. Heath, pers. comm.).

In a similar study to the present experiment, Garnett & Falconer (1975) and Simpson *et al.* (1982) measured gene frequencies at a total of 10 markers in a replicated selection experiment on body size (Q-lines). A significant association was found at *Hbb*, but, interestingly, this marker does not show such an association in the present study, implying a different genetic background near this locus. A further interesting non-random association with body size was found later in the Q-lines at the *H-2* locus (Simpson *et al.* 1982). There are two major differences between the Q-line study and the present X-line study which would lead to differences in the outcome of such mapping experiments: (1) In the Q-line experiment the gene frequencies of markers in the base population were unknown, and had to be inferred from the gene frequencies in an unselected control. Genetic drift in the control would lead to a considerable loss in statistical power, as would natural selection. (2) The X-line experiment, which used as its base population the F_2 of two inbreds, in which it is likely that there was much linkage disequilibrium. Although the Q-lines were also derived from a cross, they were mated at random for

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press

at least 15 generations prior to selection, so most of the linkage disequilibrium would have broken down (Falconer, 1973).

How powerful is the present design for detecting the effects of QTL? Assuming complete linkage between a marker and a QTL, expected standard errors (S.E.) of estimates were computed by generating simulated data (using the same subline structure and population size as assumed for the X-lines), and sampling 16 individuals per subline, and performing the maximum likelihood procedure as described above. It was found, for example, that with a codominant marker completely linked to a QTL of effect 0·1 phenotypic standard deviations, the predicted S.E. is about 0·08 standard deviations. The expected S.E. from analysis of an F_2 population using regression to estimate the effect associated with a completely linked marker is S.E. $= [8\sigma_p^2/n]^{\frac{1}{2}}$, where n is the sample size. So to achieve a S.E. of 0·08 standard deviations would require a sample size of 1250. It can be concluded, therefore, that the design of experiment exemplified by the X-lines is extremely powerful for measuring the effects of QTL, but of course depends on the availability of suitable selection lines, and is restricted to one trait only.

High density microsatellite maps of the mouse genome are currently being constructed (Love *et al.* 1990; Cornall *et al.* 1991; Dietrich *et al.* 1992), and it should be possible to map QTL that we have detected more accurately by measuring the effects of small sections of chromosome in suitable crosses (although mapping below 1 cM will become increasingly difficult as recombinants become rare). The use of information from closely linked markers will also permit the simultaneous estimation of QTL location and effect within the present maximum likelihood framework. A detailed identification of map position is a crucial prerequisite to the molecular cloning and functional analysis of these QTL.

We wish to thank W. G. Hill and A. Eyre-Walker for helpful advice, J. P. Stoye for useful discussions and for providing clones, S. C. Heath for his unpublished observations on coat colour markers, R. K. Scougall for providing samples, and L. May for technical assistance. We thank the Agricultural and Food Research Council for support, and P. D. K. acknowledges support from the Royal Society.

References

Blatt, C., Mileham, K., Haas, M., Nesbitt, M. N., Harper, M. E. & Simon, M. I. (1983). Chromosomal mapping of the mink cell focus-inducing and xenotropic *env* gene family in the mouse. *Proceedings of the National Academy of Science USA* **80**, 6298–6302.

Bulfield, G. & Bantin, G. (1981). Genetic monitoring of inbred strains of mice using electrophoresis and electrofocusing. *Laboratory Animals* **15**, 147–149.

Church, G. M. & Gilbert, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Science USA* **81**, 1991–1995.

Cornall, R. J., Aitman, T. J., Hearne, C. M. and Todd, J. A. (1991). The generation of a library of PCR-analyzed microsatellite variants for genetic mapping of the mouse genome. *Genomics* **10**, 874–881.

Darvasi, A., Weinreb, A., Minke, V., Weller, J. I. & Soller, M. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**, 943–951.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H.-S., Friedman, J., Dracopoli, N. & Lander, E. S. (1992). A genetic map of the mouse suitable for typing interspecific crosses. *Genetics* **131**, 423–447.

DuMouchel, W. H. & Anderson W. W. (1968). The analysis of selection in experimental populations. *Genetics* **58**, 435–449.

Falconer, D. S. (1973). Replicated selection for body weight in mice. *Genetical Research* **22**, 291–321.

Falconer, D. S. (1989). *Introduction to Quantitative Genetics*, 3rd edn. London: Longman.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6.

Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. (1989*a*). Genetic analysis of endogenous xenotropic murine leukemia viruses: association with two common mouse mutations and the viral restriction locus *Fv-1*. *Journal of Virology* **63**, 1763–1774.

Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. (1989*b*). Genetic identification of endogenous polytropic proviruses by using recombinant inbred mice. *Journal of Virology* **63**, 3810–3821.

Frankel, W. N., Stoye, J. P., Taylor, B. A. & Coffin, J. M. (1990). A linkage map of endogenous murine leukemia proviruses. *Genetics* **124**, 221–236.

Garnett, I. & Falconer, D. S. (1975). Protein variation in strains of mice differing in body size. *Genetical Research* **25**, 45–57.

Georges, M. (1993). Mapping Quantitative trait loci affecting milk production using a primary bovine DNA marker map. *Proceedings of the Seventeenth International Congress of Genetics*, Birmingham.

Green, C. V. (1935). Apparent changes with age in X-over between colour and size genes in mice. *Journal of Genetics* **30**, 101–106.

Hedrick, P. W. & Comstock, R. E. (1968). Role of linkage in gene frequency changes of coat color alleles in mice. *Genetics* **58**, 297–303.

Holland, C. A., Wozney, J. & Hopkins, N. (1983). Nucleotide sequence of the gp70 gene of murine retrovirus MCF 247. *Journal of Virology* **47**, 413–420.

Jabob, H. J., Lindpainter, K., Lincoln, S. E., Kusumi, K., Bunker, R. K., Mao, Y.-P., Ganten, D., Dzau, V. J. & Lander, E. S. (1991). Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* **67**, 213–224.

Lebowitz, R. J., Soller, M. & Beckmann, J. S. (1987). Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theoretical and Applied Genetics* **73**, 556–562.

Love, J. M., Knight, A. M., McAleer, M. A. & Todd, J. A. (1990). Towards construction of a high resolution map of the mouse genome using PCR analysed microsatellites. *Nucleic Acids Research* **18**, 4123–4130.

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E. & 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.

Patterson, H. D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545.

Sax, K. (1923). The association of size differences with seed

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press

coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**, 552–560.

Simpson, E. Bulfield, G., Brenan, M., Fitzpatrick, W., Hetherington, C. & Blann, A. (1982). *H-2*-associated differences in replicated strains of mice divergently selected for body weight. *Immunogenetics* **15**, 63–70.

Soller, M. & Beckmann, J. S. (1990). Marker-based mapping of quantitative trait loci using replicated progenies. *Theoretical and Applied Genetics* **80**, 205–208.

Stuber, C. W., Lincoln, S. E., Wolff, D. W., Helentjaris, T. & Lander, E. S. (1992). Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* **123**, 823–839.

Thoday, J. M. (1961). Location of polygenes. *Nature* **191**, 368–370.

Todd, J. A., Aitman, T. J., Cornall, R. J., Gosh, S., Hall, J. R. S., Hearne, C. M., Knight, A. M., Love, J. M., McAleer, M. A., Prins, J., Rodrigues, N., Lathrop, M., Pressey, A., DeLarato, N. H., Peterson, L. B. & Wicker, L. S. (1991). Genetic analysis of autoimmune type 1 diabetes mellitus in mice. *Nature* **351**, 542–547.

https://doi.org/10.1017/S0016672300031906 Published online by Cambridge University Press