Theor Appl Genet (1988) 76:228–236

Detection of linkage between marker loci and loci affecting quantitative traits in crosses between segregating populations *

J.S. Beckmann [1] **and M. Soller** [2]

[1] Department of Plant Genetics and Breeding, Agricultural Research Organization, The Volcani Center, Bet Dagan 50250, Israel
[2] Department of Genetics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Received July 6, 1987; Accepted February 9, 1988
Communicated by E.J. Eisen

Summary. By making use of pedigree information and information on marker-genotypes of the parent and F-1 individuals crossed to form an F-2 population, it is possible to carry out a linkage analysis between marker loci and loci affecting quantitative traits in a cross between segregating parent populations that are at fixation for alternative alleles at the QTL, but share the same alleles at the marker loci. For two-allele systems, depending on marker allele frequencies in the parent populations, 2–4 times as many F-2 offspring will have to be raised and scored for markers and quantitative traits in order to provide power equivalent to that obtained in a cross between fully inbred lines. Major savings in number of F-2 offspring raised can be achieved by scoring each parent pair for a large number of markers in each chromosomal region and scoring F-1 and F-2 offspring only for those markers for which the parents were homozygous for alternative alleles. For multiple allele systems, particularly when dealing with hypervariable loci, only 10%–20% additional F-2 offspring will have to be raised and scored to provide power equivalent to that obtained in a cross between inbred lines. When a resource population contains novel favorable alleles at quantitative trait loci that are not present (or rare) in a commercial population, analyses of this sort will enable the loci of interest to be identified, mapped and manipulated effectively in breeding programs.

Key words: Restriction fragment length polymorphisms (RFLPs) – Variable number tandem repeat regions (VNTRs) – Quantitative trait loci (QTL) – Marker-QTL linkage

* Contribution no. 2124-E, 1987 series from The Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel

Introduction

Soller et al. (1976) have described experimental designs for the detection of linkage between marker loci and loci affecting quantitative traits (henceforth: quantitative trait loci, QTL) in crosses between inbred lines. Given sufficient markers, such experiments can enable QTL for which two populations differ to be mapped, and the allelic state at each differentiating QTL to be determined for each population. If more than one QTL is found in the vicinity of the marker locus, these methods provide an estimate of the resultant of their action and location (Mather and Jinks 1971; McMillan and Robertson 1974). In some cases it can be established whether more than one gene is involved (Thoday 1961; Weller 1987).

Implementation of marker-QTL linkage experiments in practice has been limited by a paucity of suitable markers, although a number of laboratories have reported experimental results using morphological and isozyme markers (Edwards et al. 1987; Stuber et al. 1987; Tanksley et al. 1982; Weller 1987; Weller et al. 1988; Zhuchenko et al. 1979). Recently a new class of genetic marker, restriction fragment length polymorphisms (RFLPs), has been uncovered by the use of recombinant DNA methodologies. These have proven to be exceedingly frequent in plant and animal agricultural populations (Beckmann et al. 1986; Bernatzky and Tanksley 1986; Hallerman et al. 1987; Helentjaris et al. 1986). This should allow the widespread comparative mapping of QTL in those agricultural species where inbred lines are available (Nienhuis et al. 1987).

Power of the Soller et al. (1976) design is proportional to $(p_1 - p_2)^2$, where p_1 and p_2 are marker allele frequencies in the populations crossed, roughly proportional to $(a_1 - a_2)^2$, and where a_1 and a_2 are allelic frequencies at the QTL. Thus, power of this design will be essentially nil

for crosses involving populations of agricultural animals, which typically differ in allelic frequencies at marker loci, but are not characterized by fixation of alternative alleles at either marker loci or QTL. Similar considerations apply with respect to fruit trees. We now show that by making use of pedigree information and information on marker-genotypes of the parent and F-1 individuals crossed to form the F-2 population, it is possible to carry out a marker-QTL linkage analysis when parent populations are at, or close to, fixation for alternative alleles at QTL, even when these parent populations do not differ greatly in allelic frequencies at the marker loci monitored. Analyses of this sort will be particularly effective when loci characterized by multiple alleles having roughly equal allelic frequencies are used as markers. The highly polymorphic loci described by Jeffreys et al. (1985) and Nakamura et al. (1987) fill this requirement admirably. Such loci are apparently present in large numbers in the genome of higher plants and animals and can be detected by appropriate techniques.

Theory

Two-allele systems, one marker locus per chromosomal region

Consider two populations: Population 1 at fixation for allele A at a QTL with gene frequencies pM and qm at a linked marker locus; and Population 2 at fixation for allele a at the QTL with gene frequencies rM and sm at the linked marker locus. Genotype effects at the QTL in phenotypic standard deviation units are assumed to be: AA = +d, Aa = h, aa = −d.

Table 1 shows genotype frequencies in the two parental populations and in the F-1 produced by random mating between them. In virtually all cases it is possible to distinguish specific coupling linkage relationships between marker alleles and QTL alleles, among the F-1 offspring of specific matings classified by parental marker genotype. For example, in the mating between a heterozygous marker parent of Population 1 (genotype: MA/mA) and a homozygous marker parent of Population 2 (genotype: Ma/Ma), F-1 offspring, heterozygous at the marker locus, will all have the genotype Ma/mA. Such offspring will be termed "known" heterozygotes. The only mating for which coupling relationships in the F-1 offspring cannot be distinguished unequivocally is the cross between a heterozygous marker parent from Population 1 and a heterozygous marker parent from Population 2. In this case, some of the heterozygous F-1 offspring will have the genotype MA/ma while others will have the genotype Ma/mA. That is, the M marker allele will be in coupling with the A QTL-allele in some heterozygous offspring and with the a QTL-allele in others. Such offspring will be termed "mixed" heterozygotes.

Summing F-1 genotype frequencies from the various matings of Table 1, for the four possible known F-1 genotypes and the mixed heterozygotes, gives the following overall F-1 genotype frequencies:

Known homozygous genotypes:

$$MA/Ma = p^2r^2 + pqr^2 + p^2rs + pqrs = pr$$
$$mA/ma = pqrs + q^2rs + pqs^2 + q^2s^2 = qs$$

Known heterozygous genotypes:

$$MA/ma = pqs^2 + p^2s^2 + p^2rs = ps(p + qs)$$
$$Ma/mA = q^2rs + q^2r^2 + pqr^2 = qr(r + qs)$$

Mixed heterozygous genotypes:

$$MA/ma, Ma/mA = 2pqrs$$

Table 1. Genotypes and genotype frequencies of F-1 individuals produced by a cross between two parental populations homozygous for alternative alleles at a quantitative trait locus (A), but segregating at a single linked diallelic marker locus (M)

Parental genotypes and frequencies							
Population (2)		Population (1)					
		MA/MA (p^2)		MA/mA (2pq)		mA/mA (q^2)	
		Genotype	Frequency	Genotype	Frequency	Genotype	Frequency
Ma/Ma	(r^2)	MA/Ma	p^2r^2	MA/Ma Ma/mA	pqr^2 pqr^2	Ma/mA	q^2r^2
Ma/ma	(2rs)	MA/Ma MA/ma	p^2rs p^2rs	MA/Ma [a] MA/ma [a] [a] Ma/mA [a] mA/ma	pqrs pqrs pqrs pqrs	Ma/mA mA/ma	q^2rs q^2rs
ma/ma	(s^2)	MA/ma	p^2s^2	MA/ma mA/ma	pqs^2 pqs^2	mA/ma	q^2s^2

[a] These genotypes cannot be distinguished on the basis of the F-1 individual marker genotype and parental genotype

In order to be useful for marker-QTL linkage studies, a mating between F-1 individuals must be such that it will be possible to group the F-2 products into at least two marker genotypes that differ in a known manner at their linked QTL. On this basis, when F-1 individuals are mated at random, five classes of matings can be identified with respect to their usefulness for purposes of marker-QTL linkage studies. Two of these classes provide F-2 offspring that are informative with respect to marker-QTL linkage studies, while three do not.

The two informative matings are as follows:

Type A: Known Heterozygous F-1 × Known Heterozygous F-1, producing Known F-2 offspring. For example, in the mating between F-1 genotypes MA/ma × MA/ma, and assuming zero recombination, F-2 offspring will be produced having known genotypes MA/MA, MA/ma and ma/ma. Thus, the difference in mean phenotypic value of MM and mm marker offspring (termed: the "contrast" MM-mm) will have the expectation, 2d. In the presence of some proportion of recombination, t, the expected value of the MM-mm contrast will be $2(1-2t)d$.

Type B: Homozygous F-1 × Known Heterozygous F-1. For example, the mating between known F-1 genotypes MA/Ma × Ma/mA will produce F-2 offspring having known, but mixed, genotypes: MA/Ma, MA/mA, Ma/Ma and Ma/mA. However, informative contrasts can still be made. In particular, the contrast MM-Mm will have the expectation d in the absence of recombination, while in its presence the contrast will have the expectation:

$$\text{MM-Mm} = (1-2t)[(h-d)-(d+h)]/2 = -(1-2t)d\,.$$

The three non-informative matings are:

(i) Homozygous F-1 × Homozygous F-1. All F-2 offspring of any particular mating of this type have the same marker genotype, so informative contrasts cannot be constructed.

(ii) All matings involving Mixed Heterozygous F-1 individuals. The F-2 offspring of such matings will not have known genotypes with respect to the quantitative trait locus, so informative contrasts cannot be constructed.

(iii) Known Heterozygous F-1 × Known Heterozygous F-1, producing uninformative F-2 offspring. For example, in the mating MA/ma × Ma/mA, F-2 offspring will be produced having the genotypes: MA/Ma, MA/mA, Ma/ma and mA/ma, and informative contrasts cannot be formed.

Table 2 shows the various classes of informative matings between F-1 individuals having known or mixed genotypes. Also shown are the expected frequency of each mating class, the informative contrast within each mating class when present and its expected value. Depending on the mating class, the expected value of the informative contrast will be $\pm 2(1-2t)d$ or $\pm(1-2t)d$. The expected frequency of the various mating classes will depend strongly on the allelic frequencies in the two parental populations. For example, when $pM = 1.0$, $rM = 0.0$, the frequency of the MA/ma × MA/ma mating type in the F-1 offspring will be 1.0 (expected value of the informative contrast $= 2d$). When $pM = qm = rM = sm = 0.5$, the frequency of the equivalent MA/ma × MA/ma and Ma/mA × Ma/mA matings (expected value of the informative contrast $= 2d$), will be 0.035 each, or 0.07 together. Similarly, when $pM = 0.9$, $rM = 0.1$, the frequency of the MA/Ma × Ma/mA mating type (expected value of the informative contrast $= d$) will be 0.0003; when $pM = qm = rM = sm = 0.5$, the frequency of this mating type will be 0.094.

Table 2. Informative F-1 matings with a single diallelic marker locus: frequency, informative contrast and main effect of informative contrast

Mating	Frequency	Informative contrast	Main effect
Type A matings:			
MA/ma × MA/ma	$[ps(p+qs)]^2$	MM−mm	$2d(1-2t)$
Ma/mA × Ma/mA	$[qr(r+qs)]^2$	MM−mm	$-2d(1-2t)$
Type B matings:			
MA/Ma × Ma/mA	$2pqr^2(r+qs)$	MM−Mm	$-d(1-2t)$
MA/Ma × MA/ma	$2p^2rs(p+qs)$	MM−Mm	$d(1-2t)$
MA/ma × mA/ma	$2pqs^2(p+qs)$	Mm−mm	$d(1-2t)$
Ma/mA × mA/ma	$2q^2rs(r+qs)$	Mm−mm	$-d(1-2t)$

The total frequency of all mating types with an expected contrast value of 2d (Type A matings) will be

$$A = [ps(p+qs)]^2 + [qr(r+qs)]^2\,.$$

The total frequency of all mating types with an expected contrast value of d (Type B matings) will be

$$B = 2(pr+qs)[qr(r+qs) + ps(p+qs)]\,.$$

In Type B matings, all offspring are included in the contrast evaluation. In Type A matings, only half of the offspring (those that are homozygous for alternative marker alleles) are included in contrast evaluation, but the expected contrast value is twice as large (2d) as in Type B matings (d). Thus, for a given number, T, of offspring produced by each type of mating, the relative ability of a Type B mating to uncover statistically significant marker-QTL linkage (i.e., the power of the mating) as compared to a Type A mating will be in proportion to

$$[d^2/(2/T)]/[4d^2/\{2/(T/2)\}] = 1/2\,,$$

and the overall power of a cross involving two segregating populations as compared to a cross between two inbred lines will be in proportion to $0.5B + A$. The number of offspring required for equivalent power in a cross between two segregating populations as compared to a

cross between two inbred lines will be the reciprocal of this, $1/(0.5B+A)$. For example, if $pM=0.8$, $rM=0.3$, then $A=0.28$, $B=0.42$, the relative power $=0.49$ and 2.04 as many offspring will be required for a power equivalent to that obtained in a cross between two inbred lines.

Two-allele system, many markers per chromosomal region

As is evident from the above argument, offspring of Type A matings are twice as informative as offspring of Type B matings. The overall frequency of Type A matings can be increased if a number of independent markers are available for each chromosomal region. This would make it possible to identify, for each chromosomal region of each parent pair, a marker for which one parent has the MA/MA genotype and the other parent has the ma/ma genotype – producing MA/ma known heterozygotes; or one parent has the mA/mA genotype and the other parent Ma/Ma – producing Ma/mA known heterozygotes. When the parents are mated at random, the expected frequency of such matings will be: p^2s^2 and q^2r^2, respectively. Thus, the overall frequency of such matings out of all matings will be $p^2s^2+q^2r^2$. These matings will provide F-1 offspring having genotypes MA/ma and Ma/mA. The proportion of parental matings for which at least one marker is in the above state will be

$$1-[1-(p^2s^2+q^2r^2)]^n,$$

where n is the number of markers in the chromosomal region, assuming the same allelic frequencies for all markers. The number of markers required to provide a likelihood of 0.90 that at least one parental marker in a given region will be in the most useful state was obtained by numerical substitution in the above expression for various combinations of p, q, r and s.

Given that for all chromosomal regions of all F-1 individuals a marker in the above state is present, then the proportion of Type A matings of MA/ma × MA/ma among all F-1 matings will be

$$[p^2s^2/(p^2s^2+q^2r^2)]^2,$$

the proportion of matings of Ma/mA × Ma/mA will be

$$[q^2r^2/(p^2s^2+q^2r^2)]^2,$$

and the total proportion of Type A matings will be the sum of these two expressions.

Multiple alleles

When there are multiple alleles at the marker locus, expressions for the general case (different frequencies for different alleles within a population, and different frequencies for different alleles between populations) are cumbersome. Consequently, expressions will be derived only for the case of equal frequencies for the various alleles within and between the two parental populations.

Table 3. Frequency of parental and F-1 genotypes, and of informative F-1 mating classes, for a polyallelic marker. n = no. of alleles at the marker locus

Genotypes or matings within each class			
Class	No. of different types	Frequency of each types	Total frequency of all types
Parental genotypes:			
Homozygotes	n	p^2	np^2
Heterozygotes	$n(n-1)/2$	$2p^2$	$n(n-1)p^2$
F-1 genotypes:			
Homozygotes	n	p^2	np^2
Known heterozygotes	$n(n-1)$	$p^2(1-p^2)$	$n(n-1)p^2(1-p^2)$
Mixed heterozygotes	$n(n-1)$	p^4	$n(n-1)p^4$
F-1 informative mating classes:			
Type A	$n(n-1)[n(n-1)-1]$	$[p^2(1-p^2)]^2$	$n(n-1)[n(n-1)-1][p^2(1-p^2)]^2$
Type B	$2n^2(n-1)$	$p^4(1-p^2)$	$2n^2(n-1)p^4(1-p^2)$

This should provide a fair approximation of the general case, since differences in allelic frequencies within populations (e.g., pM1 as compared to pM2) will tend to decrease the power of the cross, while differences in allelic frequencies between populations (e.g., pM1 as compared to rM1) will tend to increase the power of the cross.

Consider the marker locus, M, with multiple alleles, M1, M2, ..., Mi, ..., Mn and allelic frequencies all equal to $p=1/n$. As in the two-allele case, the parental population will be a mixture of homozygotes and heterozygotes and the F-1 will be a mixture of homozygotes, known heterozygotes and mixed heterozygotes. Similarly, crosses between the F-1 individuals will be of the same five types as in the two-allele case. Table 3 shows the frequencies of the various parental individuals, F-1 individuals and informative F-1 mating types. Note that the factor $(1\text{-}p^2)$ that appears in the expressions for the various F-1 mating types tends rapidly to 1 as n increases. With this in mind, the overall frequency of Type A matings will tend to $n^2(n-1)^2p^4$, and the overall frequency of Type B matings will tend to $2n^2(n-1)p^4$. Consideration of these expressions shows that once $n>3$, the majority of F-1 matings will be of Type A.

Numerical results

Two-allele system, one marker per chromosomal region

Table 4 shows the proportion of F-1 matings for a two-allele system, single marker per chromosomal region,

Table 4. Proportion of F-1 matings providing informative F-2 offspring for a single diallelic marker according to mating type and marker allele frequency in the two parental populations. Also shown is the relative number (RN) of F-2 offspring required in order to provide power equivalent to that obtained in a cross between inbred lines. The full table has four-way symmetry (Axes indicated by *bold characters*). Consequently, values for allele frequency combinations not shown can readily be obtained from the values given. See text for details

Population (2)	Mating type/RN	Frequency of marker allele									
		Population (1)									
		0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	B	0.18	0.32	0.42	0.48	0.50	0.48	0.42	0.32	0.18	0.00
	A	0.01	0.04	0.09	0.16	0.25	0.36	0.49	0.64	0.81	1.00
	RN	10.0	5.0	3.3	2.5	2.0	1.7	1.4	1.3	1.1	**1.0**
0.9	B	0.27	0.34	0.40	0.44	0.46	0.45	0.42	0.37	0.29	0.18
	A	0.01	0.03	0.07	0.12	0.18	0.27	0.37	0.50	0.64	0.81
	RN	**6.8**	4.9	3.8	3.0	2.4	2.0	1.7	1.5	**1.3**	1.1
0.8	B		0.37	0.39	0.41	0.42	0.43	0.42	0.40	0.37	0.32
	A		0.04	0.05	0.09	0.13	0.20	0.28	0.38	0.50	0.64
	RN		**4.6**	4.0	3.5	2.9	2.4	2.0	**1.7**	1.5	1.3
0.7	B			0.38	0.39	0.40	0.40	0.41	0.42	0.42	0.42
	A			0.06	0.07	0.10	0.14	0.20	0.28	0.37	0.49
	RN			**4.0**	3.8	3.4	2.9	**2.5**	2.0	1.7	1.4
0.6	B				0.38	0.38	0.39	0.40	0.42	0.45	0.48
	A				0.07	0.08	0.10	0.14	0.20	0.27	0.36
	RN				**3.9**	3.7	**3.4**	2.9	2.4	2.0	1.7
0.5	B					0.38	0.38	0.40	0.42	0.46	0.50
	A					0.07	0.08	0.10	0.13	0.18	0.25
	RN					**3.9**	3.7	3.4	2.9	2.4	2.0

that are of Type A and Type B, according to allelic frequencies in the two parental populations. Because of the double symmetry of allelic frequencies (between Parent Population 1 and Parent Population 2 and between p, q and r, s), the complete Table has a four-fold symmetry, so that only one quarter of the complete table need be shown. That is, the proportion of Type A and Type B matings will be the same, e.g., when $pM = 0.7$, $rM = 0.4$; or $rM = 0.7$, $pM = 0.4$; or $pM = 0.3$, $rM = 0.6$; or $pM = 0.6$, $rM = 0.3$. The proportion of Type B matings does not vary greatly, generally falling in the range 0.3–0.5. The proportion of Type A matings, in contrast, generally falls in the range 0.0–0.30, but can reach much higher values when the difference in allelic frequencies between the two parent populations is great. For a single marker locus, it is possible to screen parents and set up only informative matings. But clearly this cannot be done when many marker loci are followed simultaneously, hence the assumption of random mating of parents and F-1 individuals.

Table 4 also shows the relative number of offspring required in a cross between two segregating populations, as compared with a cross between two inbred lines, according to allelic frequencies in the two populations. It is striking that except for $rM = 1.0$ or 0.9, the values in Table 4 are a fairly constant function of the difference in allelic frequencies between the two parental populations. That is, when the difference in allelic frequency is 1.0 (fixation for alternative alleles in the two populations), the relative number of offspring required is 1.0; when the difference is 0.9, the relative number is 1.1; and, continuing the series for differences of 0.8, relative numbers are 1.3; for 0.7, 1.5; for 0.6, 1.7; for 0.5, 2.0; for 0.4, 2.4; for 0.3, 2.9; for 0.2, 3.3; for 0.1, 3.7; and for 0.0, 3.8. Thus, it will be possible to carry out a linkage analysis even when allelic frequencies are the same in the two parent populations, but about four times as many offspring will be required as for a cross between inbred lines. As seen above, the greater the difference in allelic frequencies between the parent populations the more effective the cross will be.

Two-allele system, many markers per chromosomal region

Table 5 shows the relative number of markers that must be monitored in each chromosomal region in the parent population in order to have a likelihood of 0.90 of identifying in each chromosomal region at least one parental

Table 5. Relative increase in the number of markers that must be scored in the parental populations in a cross between segregating populations in order to provide a likelihood of 0.90 of having, for each chromosomal region, at least one parent marker genotype that provides the most informative F-1 offspring. Given that F-1 offspring are of the most informative type, the relative number of F-2 offspring required in the cross between segregating populations to provide power equivalent to that obtained in a cross between inbred lines according to marker allele frequency in the segregating populations is also shown. The full table has four-way symmetry (Axes indicated by *bold characters*). Consequently values for allele frequency combinations not shown can readily be obtained from the values given. See text for details

Population (2)	Relative increase in	Frequency of marker allele									
		Population (1)									
		0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	Markers	100+	56	24	13	8	5	3	2	1	**1**
	Offspring	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	**1.0**
0.9	Markers	**100+**	58	28	16	10	7	5	3	**2**	1
	Offspring	**2.0**	1.4	1.1	1.1	1.0	1.0	1.0	1.0	**1.0**	1.0
0.8	Markers		**44**	28	18	12	8	6	**4**	3	2
	Offspring		**2.0**	1.6	1.3	1.1	1.1	1.0	**1.0**	1.0	1.0
0.7	Markers			**24**	19	15	11	**8**	6	5	3
	Offspring			**2.0**	1.7	1.4	1.2	**1.1**	1.0	1.0	1.0
0.6	Markers				**18**	17	**14**	11	8	7	5
	Offspring				**2.0**	1.7	**1.4**	1.2	1.1	1.0	1.0
0.5	Markers					**17**	17	15	12	10	8
	Offspring					**2.0**	1.7	1.4	1.1	1.0	1.0

Table 6. Relative proportion of Type A and Type B informative matings among all F-1 matings, and relative increase in number of F-2 offspring required to provide power equivalent to a cross between inbred lines, according to number of alleles per marker locus (n), assuming frequency of all alleles to be equal within and between populations

No. of alleles per locus	Mating type		Relative no. of F-2 offspring required
	B	A	
2	0.375	0.070	3.88
3	0.395	0.293	2.04
4	0.352	0.453	1.59
5	0.307	0.560	1.40
6	0.270	0.635	1.30
7	0.240	0.688	1.24
8	0.215	0.729	1.20
9	0.195	0.760	1.17
10	0.178	0.785	1.14
20	0.095	0.896	1.06

marker that will produce F-1 offspring having the MA/ma or Ma/mA genotypes. A major (5- to 15-fold) increase in number of markers is generally required. However, it should be noted that all of these will have to be monitored simultaneously only in the parental pairs. In the F-1 and F-2 derivatives of each parental pair, only those markers in a favorable state in the parents will need to be followed.

Table 5 also shows the relative number of F-2 offspring required when all F-1 matings involve the most favorable marker genotypes, as described above. Relative numbers reach a maximum of 2.0 and are generally less than 1.5. Thus, this design can offer a real saving in the number of F-2 offspring required for a mapping program in segregating populations at the expense of a major increase in the number of markers followed.

Multiple alleles

Table 6 shows the proportion of Type A and Type B matings and the relative number of F-2 offspring required to give a power equivalent to that obtained in a cross between two inbred lines, according to the number of alleles at the marker locus. The proportion of informative matings increases rapidly with a rise in the number of alleles per marker locus; within this, the proportion of the more informative Type A matings increases at the expense of the less informative Type B matings. It is clear that for highly polymorphic marker loci ($n > 10$), the power of the experiment will be virtually the same as for a cross between inbred lines.

It is striking that most of the advantage is gained in the shift from two alleles to three: with three alleles, the relative number of F-2 offspring required is reduced to half that required for the two-allele case. The important factor here, however, is the assumption that allelic frequencies are equal within populations. Generally, a

three-allele system contains one allele in high frequencies and the remaining two at much lower frequencies. Thus, the primary advantage of the highly polymorphic systems uncovered by Jeffreys et al. (1985) and Nakamura et al. (1987) is the fact that in a highly variable multiple allelic system the frequency of the various alleles tends to be roughly equal.

Discussion

In agricultural species that reproduce by selfing (e.g., wheat, barley, tomato), the various cultivars, landraces and wild progenitors constitute an enormous store of genetic variability potentially available to the breeder. Good use has been made of this resource with respect to traits having a Mendelian mode of inheritance, particularly disease resistance. These resource populations also constitue a potential store of genetic variation with respect to quantitative traits of economic importance. Exploitation of this genetic variation for breeding purposes, however, has been limited by the difficulty of determining whether a given resource population indeed contains novel favorable alleles not present in the commercial population to be improved, although biometrical designs are available that provide at least a partial answer to this question (Dudley 1984a, b; Soller and Beckmann 1987). A marker-QTL linkage analysis provides an alternative approach. Such an analysis gives an indication of whether novel favorable alleles are present in the resource population, and can provide useful genetic markers to monitor the favorable QTL alleles in pedigree breeding or introgression programs (Beckmann and Soller 1986, 1987; Soller and Beckmann 1983, 1987; Soller and Plotkin-Hazan 1977; Tanksley et al. 1981).

When considering outcrossing species (e.g., cattle, poultry), between population genetic variation is a less important resource, since each population, of itself, constitutes a great store of segregating genetic variability. For such species even extreme quantitative differences between populations are likely to be a matter of relative allelic frequencies at QTL, rather than fixation (or near fixation) for alternative alleles. Nevertheless, even in this case there are situations where differences in quantitative trait value may reflect the presence of novel QTL alleles. Hill (1982) suggested that in large populations, under long-term directional selection, mutation may be an important means of maintaining quantitative genetic variation. Consequently, populations under selection, but separated from one another by artificial or natural reproductive barriers, may have accumulated qualitatively different alleles affecting expression of the quantitative traits under selection. Similarly, some resistance traits characterizing particular breeds (e.g., trypanotolerance of the West African N'Dama breed of cattle, tick-fever resistance of Zebu cattle) seem to be best interpreted as polygenic in nature and due to qualitative differences in the kinds of alleles present. In these cases, a marker-QTL linkage analysis would be useful to identify and map potentially valuable QTL alleles for purposes of physiological analysis or introgression.

The experimental design for determination of marker-QTL linkage proposed by Soller et al. (1976) could not be used for these cases, since populations of outcrossers typically differ in allelic frequencies at marker loci, but are not characterized by fixation of alternative marker alleles. The results of the present study, however, show that a marker-QTL linkage analysis can still be carried out in crosses between segregating populations even when the latter share marker alleles. The proposed method requires maintaining complete pedigree records for all F-1 and F-2 animals, and also requires scoring parent and F-1 as well as F-2 individuals for the marker alleles. The total number of F-2 individuals raised would vary for each marker, depending on allelic frequencies in the parental populations at that marker locus. Thus, the total number of F-2 individuals raised would be determined by allelic frequencies at the least favorable locus. This means that generally four times as many F-2 offspring would be required as are needed for equivalent power in a cross between inbred lines, but not all offspring would be scored with respect to all markers.

Setting up and analyzing the experiment

In carrying out an experiment of the sort described in this paper, the procedure is as follows. It is assumed that a detailed marker map of the species genome is available. Using this map, the two parental populations are scored for marker loci in order to provide a set of markers that show maximum differences in allelic frequencies yet provide adequate genome coverage (Beckmann and Soller 1983). Parent individuals of each population are mated at random, producing an F-1 population; the F-1 is mated at random inter se, producing an F-2 population. Parent, F-1 and F-2 individuals are scored for the chosen markers.

Each marker locus is evaluated separately for each pair of F-1 individuals in order to determine whether the mating was of an informative or uninformative type; and, if of an informative type, whether it was Type A or Type B. The informative matings are then grouped by subtypes (i.e., the two sorts of Type A matings and the four sorts of Type B matings shown in Table 2). All of the F-2 offspring within each informative subtype are then grouped separately according to marker genotype classified into two groups, according to whether they were expected to have the higher or lower quantitative trait value. For example, within the MA/ma × MA/ma Type A subtype, MM F-2 individuals are expected to have the higher quantitative

trait value, mm individuals the lower value, while within the Ma/mA × Ma/mA Type A subtype, MM F-2 individuals are expected to have the lower quantitative trait value, mm individuals, the higher. A *t*-test is then carried out comparing mean quantitative trait value of all F-2 offspring (pooled over all mating types and subtypes) having the "higher" expected value with the mean quantitative trait value of all F-2 offspring having the "lower" expected value. A significant difference is taken as an indication of marker-QTL linkage.

Power considerations

In experiments aimed at mapping QTL for purposes of eventual introgression from a resource population to some commercial population, it is important to avoid a Type I error, i.e., falsely identifying a favorable QTL in linkage to a marker when such a QTL is not indeed present. Since experiments of this sort will always involve a large number of markers and sometimes a number of quantitative traits as well, the overall likelihood of a Type I error can be a large multiple of the likelihood of an individual Type I error for any given marker × trait combination tested. Hence the likelihood of an individual Type I error must be kept very low, say $\alpha < 0.001$. The power of an experiment to determine marker-QTL linkage is a very sensitive function of the magnitude of effect at the QTL and of the proportion of recombination between marker and QTL. Following Soller et al. (1976), the number of F-2 offspring required in a cross between inbred lines to provide power ($1-\beta$) against Type II error, and Type I error of α, can be derived from the expression:

$$\text{S.E.(D)} = 2(1-2t)\,d/(z_{\alpha/2}+z_{\beta})$$

where d and t are as previously defined; S.E.(D) = standard error of the difference between the homozygous F-2 marker genotypes, $=[2/(T/4)]^{0.5}$, in units of the phenotypic standard deviation; and T/4 = number of F-2 offspring in a cross between inbred lines homozygous for one of the marker alleles.

The effect of recombination can be limited to some extent by following a large number of markers, so that the average recombination proportion between a QTL and its nearest marker neighbor is small. Assume that the number of markers scored is such that the average proportion of recombination between a QTL and its nearest marker neighbor is 0.05. Then the number of offspring that would have to be scored to provide a power of 0.8 against Type II error ($z_{\beta}=0.84$), while maintaining an individual Type I error of 0.001 ($z_{\alpha/2}=3.32$), can be calculated, for the case of a cross between inbred lines, by substituting appropriate values in the above expression and solving for T. Assuming further that the cross between segregating populations requires scoring on the average three times as many F-2 offspring as would be required in a cross between inbred lines, then multiplying the values so obtained by three and rounding will give the approximate number of F-2 individuals required to provide equivalent power when the cross is carried out between segregating populations. When this is done, values obtained are: T = 10,000 for d = 0.1; T = 2,500 for d = 0.2; T = 1,100 for d = 0.3; T = 600 for d = 0.4; and T = 400 for d = 0.5. Thus, for a two-allele system, experiments of this sort would appear to be feasible only for QTL at which d = 0.2 or more. It should be stressed that the application of this design is limited to situations where the two populations are at or close to fixation for alternative alleles at the QTL. Since power will be proportional to the square of the difference in allelic frequencies at the QTL in the two populations, once this falls below 0.8 power drops dramatically.

Range of applicability

Are such effects likely to be found in practice? Little information is currently available as to the proportional effects of QTL in animal populations, except for biometrical analyses carried out in some selection lines of mouse and *Drosophila* (Falconer 1981). In plants, however, a number of marker-QTL linkage analyses have been carried out (Edwards et al. 1987; Nienhuis et al. 1987; Stuber et al. 1987; Tanksley et al. 1982; Vallejos and Tanksley 1983; Weller et al. 1988), and in these analyses a considerable proportion of markers have been found to have associated effects of magnitude d = 0.2 or greater.

In other situations, a priori considerations suggest the possibility that loci having effects of magnitude d = 0.2 or greater may be present. Consider, for example, the difference between the trypanotolerant N'Dama cattle of West Africa and the sensitive Zebu, mentioned previously. If k loci, all at fixation in the N'Dama and all having equal absolute effect, D, are involved in determining trypanotolerance, genetic variance in the F-2 will equal $2pqkD^2 = kD^2/2$ (Falconer 1981). Considering that the N'Dama has been under challenge by trypanosomiasis for no more than 7,000 years (about 1,000 generations; Murray et al. 1982), and that loci conferring trypanotolerance would have had to arise by mutation and establish themselves by way of selection, it seems reasonable that the number of trypanotolerance loci would not be greater than some small number, say 10. On this assumption, F-2 genetic variance will be $5D^2$; assuming heritability of 0.5, total F-2 phenotypic variance will be $10D^2$. Then the effect of each locus affecting the trait in standard deviation units will equal $D/(10D^2)^{0.5} = 0.3$. Effects of this magnitude could be detected with 80% power in a 1,000 F-2 animal experiment.

The use of the highly polymorphic markers described by Jeffreys et al. (1985) and Nakamura et al. (1987) for experiments of this sort will reduce the required number

of offspring by more than half, bringing it close to that required for crosses between inbred lines. In fact, with the use of highly polymorphic markers, it may be possible to extend experiments with segregating populations to cases where two parental populations differ widely in allelic frequencies at the QTL of interest, but are not at fixation for alternative alleles.

Acknowledgements. This research was supported by a grant from the United States-Israel Binational Agricultural Research and Development Fund (BARD). We acknowledge with thanks helpful comments of the referees.

References

Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in varietal identification and genetic improvement: methodologies, mapping, and costs. Theor Appl Genet 67:35–43

Beckmann JS, Soller M (1986) Restriction fragment length polymorphisms in plant genetic improvement. In: Miflin BJ (ed) Oxford Surv Plant Mol Cell Biol 3:197–250

Beckmann JS, Soller M (1987) Molecular markers in genetic improvement of farm animals. Biotechnology 5:573–576

Beckmann JS, Kashi Y, Hallerman EM, Nave A, Soller M (1986). Restriction fragment length polymorphisms among Israeli Holstein-Friesian dairy bulls. Anim Genet 17:25–38

Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map of tomato based on isozymes and random cDNA sequences. Genetics 112:887–898

Dudley JW (1984a) A method of identifying lines for use in improving parents of a single cross. Crop Sci 24:355–357

Dudley JW (1984b) A method for identifying populations containing favorable alleles not present in elite germplasm. Crop Sci 24:1053–1054

Edwards MD, Stuber CW, Wendel JF (1987) Molecular marker facilitated investigations of quantitative trait loci in maize: I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125

Falconer DS (1981) Introduction to Quantitative Genetics, 2nd edn. Ronald Press, New York, p. 117

Hallerman EM, Nave A, Kashi Y, Beckmann JS, Soller M (1987) Restriction fragment length polymorphism in dairy and beef cattle at the growth hormone and prolactin loci. Anim Genet 18:213–222

Helentjaris R, Slocum M, Wright S, Schaeffer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 72:761–769

Hill WG (1982) Rates of change in quantitative traits from fixation of new mutations. Proc Natl Acad Sci (USA) 79:142–145

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67–73

Mather K, Jinks JL (1971) Biometrical Genetics. Chapman and Hall, London, p. 18

McMillan I, Robertson A (1974) The power of methods for detection of major genes affecting quantitative characters. Heredity 32:349–356

Murray M, Morrison WI, Whitelaw DD (1982) Host susceptibility to African trypanosomiasis: Trypanotolerance. In: Baker JR, Muller R (eds) Advances in Parasitology, vol. 21. Academic Press, London New York pp 1–68

Nakamura Y, Leppert M, O'Connell P, Wolf R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616–1622

Nienhuis J, Helentjaris T, Slocum M, Ruggero B, Schaefer A (1987) Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. Crop Sci 27:797–803

Soller M, Beckmann JS (1983) Genetic polymorphisms in varietal identification and genetic improvement. Theor Appl Genet 67:25–33

Soller M, Beckmann JS (1987) Genomic genetics and the utilization for breeding purposes of genetic variation between populations. 2nd Int Cong Quantitative Genetics. Raleigh NC, June 1987

Soller M, Plotkin-Hazan J (1977) The use of marker alleles for the introgression of linked quantitative alleles. Theor Appl Genet 51:133–137

Soller M, Genizi A, Brody T (1976) On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor Appl Genet 47:35–39

Stuber CW, Edwards MD, Wendel JF (1987) Molecular marker facilitated investigations of quantitative trait loci in maize: II. Factors influencing yield and its component traits. Crop Sci 27:639–648

Tanksley SD, Medina-Filho H, Rick CM (1981). The effect of isozyme selection on metric characters in an interspecific backcross of tomato – basis of an early screening procedure. Theor Appl Genet 60:291–296

Tanksley SD, Medina-Filho H, Rick CM (1982) Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. Heredity 49:11–25

Thoday JM (1961) Location of polygenes. Nature 191:368–370

Vallejos CE, Tanksley SD (1983) Segregation of isozyme markers and cold tolerance in an interspecific backcross of tomato. Theor Appl Genet 66:241–247

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

Weller J, Soller M, Brody T (1988) Linkage analysis of quantitative traits in an interspecific cross of tomato (*L. esculentum* × *L. pimpinellifolium*) by means of genetic markers. Genetics 118:329–339

Zhuchenko AA, Samovol AP, Karol 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