Theor Appl Genet (1983) 67:35–43

Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs *

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

[1] Institute of Field and Garden Crops, Agricultural Research Organization, The Volcani Center, 50250 Bet Dagan, Israel
[2] Department of Genetics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

Received July 14, 1982; Accepted July 3, 1983
Communicated by A. Robertson

Summary. Recently a new class of genetic polymorphism, restriction fragment length polymorphisms (RFLPs), has been uncovered by the use of restriction endonucleases which cleave DNA molecules at specific sites and cloned DNA probes which detect specific homologous DNA fragments. RFLPs promise to be exceedingly numerous and are expected to have genetic characteristics – lack of dominance, multiple allelic forms and absence of pleiotropic effects on economic traits – of particular usefulness in breeding programs. The nature of RFLPs and the methodologies involved in their detection are described and estimated costs per polymorphism determination are derived. The anticipated costs of applying RFLPs to genome mapping are considered in terms of the number of RFLPs required for a given degree of genome coverage, the number of probe x enzyme combinations tested per polymorphism uncovered, and the total number of individuals and polymorphisms scored for mapping purposes. The anticipated costs of applying RFLPs to genetic improvement are considered in terms of the number of individuals and the number of polymorphisms per individual that are scored for the various applications. Applications considered include: varietal identification, identification and mapping of quantitative trait loci, screening genetic resource strains for useful quantitative trait alleles and their marker-assisted introgression from resource strain to commercial variety, and marker-assited early selection of recombinant inbred lines in plant pedigree breeding programs and of young sires in dairy cattle improvement programs. In most cases anticipated costs appear to be commensurate with the scientific or economic value of the application.

Key words: Restriction fragments – Polymorphisms – Genetic markers – Genetic improvement

Introduction

Polymorphic genetic markers have wide potential application in plant and animal breeding practice as a means for varietal and parentage identification and as a means for identification and evaluation of polymorphic genetic loci affecting quantitative economic traits and the subsequent manipulation of these loci in genetic improvement programs.

These applications are reviewed in a companion paper (Soller and Beckmann 1983). The full potential usefulness of genetic markers for these purposes, however, is limited by the paucity of presently available markers. Plant cultivars, in particular selfers, generally differ from one another with respect to only a small number of morphological or pigmentation markers, and have been found to be remarkably uniform with respect to isozyme markers as well (Rick and Fobes 1975; Nevo et al. 1979). A similar, rather low degree of biochemical polymorphism appears to be present in cattle (Baker and Manwell 1980) and poultry (Washburn et al., 1980; C. M. Baker, personal communication; unpublished results from our laboratory). Furthermore, many of the present markers have the disadvantage of being developmentally regulated, i.e., they come to phenotypic expression only at specific stages of development or only in some specific tissue or organ, while morphological and pigmentation markers in particular may have secondary pleiotropic effects on economic characters.

Recently a new class of genetic polymorphism – restriction fragment length polymorphisms (RFLPs) – has been described (Bishop and Skolnick 1980; Botstein et al. 1980; Grodzicker et al. 1974; Solomon and Bodmer 1979) which promises to be exceedingly numerous and is expected to have genetic characteristics of particular usefulness in genetic improvement programs. These include lack of dominance, multiple allelic forms and absence of pleiotropic effect on other characters.

* Contribution from the Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel, No. 769-E, 1983 Series

In this paper the nature of RFLPs and the methodologies involved in their detection are described, and the anticipated costs of applying RFLPs to genome mapping and to a variety of breeding applications are considered.

RFLP methodologies

Restriction fragments

Restriction endonucleases are enzymes that recognize specific nucleic acid sequences in DNA and cleave the DNA at these sites or at adjacent sites (Zabeau and Roberts 1979). The more prevalent the recognition sequence in the DNA, the more frequently the DNA will be cleaved by the enzyme which recognizes the sequence. The DNA fragments so formed can be separated by gel electrophoresis, since smaller fragments migrate more rapidly through the pores of the gel than larger fragments. When the DNA from a higher organism is digested by restriction enzymes, many different-sized fragments are produced and a continuous smear is formed on the gel. Specific fragments are therefore detected by the use of an appropriate probe consisting of a cloned DNA sequence homologous to a particular DNA fragment or some portion of it. The DNA pattern is transferred from the gel to a solid support, such as a nitrocellulose filter (Southern 1975), and exposed to the radioactively labelled probe under conditions that promote DNA-DNA hybridization. The unhybridized ratioactivity is washed away and the filter is dried and placed against photographic film for autoradiographic exposure. After film development the specific DNA fragment that hybridized with the probe will be visualized as a band on the film.

By using specific, unique DNA probes, a DNA fragment that occurs as rarely as once in a million or less can be detected. The DNA probes used in RFLP detection do not have to be homologous with known genes. Any unique DNA sequence will suffice as long as it hybridizes with some part of one of the DNA fragments formed by the restriction process. It is not necessary to isolate specific genes for this method to work.

Restriction fragment length polymorphisms

Base changes can alter the sequences that are recognized by restriction enzymes, abolishing sites or creating new sites for particular enzymes. Deletions or transpositions of large elements will make simultaneous changes in the restriction patterns of a number of enzymes. As a result, a given restriction enzyme will not always cleave a given DNA molecule at the same point in two individuals. Consequently, fragments of different length will be formed when the DNA of the two individuals is digested. The unequal-sized fragments will travel at different rates through the gel and the band formed, following hybridization and autoradiography, will be located at different positions on the film. In this manner a restriction fragment length polymorphism will have been demonstrated. A hypothetical example is illustrated in Fig. 1.

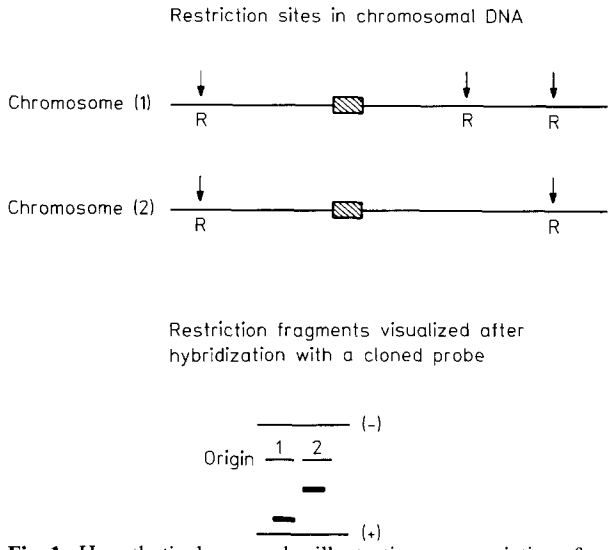

Fig. 1. Hypothetical example illustrating a restriction fragment length polymorphism. A region of homologous chromosomal DNA sequence from two chromosomes is diagrammed. Both segments contain a sequence (indicated by the box with hatched lines) that has been cloned. The site for *enzyme R* on the right side of the box in *chromosome 2* has been abolished by alteration of the DNA sequence: a larger fragment will therefore be produced when DNA from *chromosome 2* is digested with *enzyme R*. Autoradiographic visualization of the fragments containing the cloned sequences is shown in the lower half of the figure

Expected frequency and mode of inheritance of RFLPs

A large variety of restriction enzymes having different specificities are available and RFLPs in both coding and non-coding DNA sequences can be detected with appropriate probes. Thus, the RFLP method can potentially monitor a significant fraction of the total genome and should be virtually unlimited in its ability to detect large numbers of polymorphisms. Preliminary estimates in humans (Jeffreys 1979) based on the degree of polymorphism found in a sample of 60 individuals tested with eight restriction enzymes and a single probe (β-globin), suggest that on the average 1 in 100 base pairs of the human genome may vary polymorphically, and that about 1 in 100 of these can be

detected by combining a suitable restriction enzyme with a suitable probe. This suggests that there may be as many as 3×10^5 detectable polymorphisms in the human genome.

Consideration of Fig. 1 shows that in a heterozygote for both types of DNA molecule, both bands would be present in the autoradiograph. Thus RFLPs can be expected to show a "codominant" mode of inheritance, and this has indeed been found. Furthermore, multiple allelic forms are expected to be common, since independently occurring large substitutions and deletions will generally be different in different individuals, as will independent point mutations that produce a new restriction site in a given DNA fragment. This has also been found (Evola et al., cited in Burr et al. 1983; Rom 1982). Also, since RFLPs will often be located in flanking sequences or introns, one would expect them to lack secondary pleiotropic effects on quantitative economic characters.

Table 1. Number of polymorphic markers required for a given proportion (P) of genome coverage as a function of total genome size (k) and maximum spacing (c), in Morgans, between marker and other loci[a]

Proportion of genome coverage	Total genome size and spacing (in Morgans)					
	k = 10		k = 20		k = 30	
	$c = 0.1$	$c = 0.2$	$c = 0.1$	$c = 0.2$	$c = 0.1$	$c = 0.2$
0.50	41	22	83	45	124	67
0.60	54	29	109	59	164	89
0.70	71	38	144	77	215	117
0.80	96	51	192	104	287	156
0.90	137	73	275	148	411	224
0.95	178	95	358	193	535	291

[a] Values in the Table are those given by the text expression for n, multiplied by 1.2 and 1.3 for $c = 0.1$ and $c = 0.2$, respectively, to take chromosome ends into account. See text for details

Genome mapping

The number of RFLPs required for a given degree of genome coverage

If it is desired that every gene be within c Morgan of a marker, the number of polymorphic markers required for a given degree of genome coverage can be estimated in terms of the proportion, P, of a circular genome of total length k Morgans that will be covered to this degree by a set of n markers occurring randomly over the genome (Lange and Boehnke 1982). This approach ignores the effects of chromosome ends, and yields the expression

$$n = \frac{\log(1-P)}{\log(1-2c/k)}$$

For purposes of genetic analysis, a quantitative trait locus (henceforth, QTL, Geldermann 1975) should be no more than 0.2 Morgan away from a marker locus (Soller et al. 1976), while for purposes of introgression, a marker bracket of no more than 0.2 Morgan enclosing the desired QTL would be desirable (Soller and Plotkin-Hazan 1977). Thus, for these purposes, markers should be spaced every 0.2 to 0.4 Morgan, i.e., c = 0.1 to 0.2 Morgan.

Table 1 gives values of n for selected values of P, as a function of desired spacing (c = 0.1, 0.2) and total genome length (k = 10, 20 or 30 Morgan). From Table 1 it can be seen that the number of polymorphisms required to provide 80% genome coverage ranges from 51 – for c = 0.2 and k = 10, to 287 – for c = 0.1 and k = 30. A genome length of 10 Morgans would correspond to many cultivated plants; a genome length of 30 Morgans to many livestock or poultry species. Since in practice the influence of a marker does not extend past the ends of the chromosome on which it falls, markers near the end of a chromosome will not provide a full 2c of genome coverage. Hence, the values of n in Table 1 are underestimates. A computer simulation by Lange and Boehnke (1982) shows that end effects will add 20–30% to the required number of markers for the range of c values considered in Table 1.

The number of probe x enzyme combinations tested per polymorphism detected

Little information is as yet available on the average number of probes and restriction enzymes that need to be tested per useful polymorphism detected, where a useful polymorphism is defined as one for which allelic frequencies are in the range 0.1 to 0.9. Jeffreys (1979) detected two useful and one rare polymorphism in human DNA using a β-globin cDNA probe and eight restriction enzymes. Wyman and White (1980) tested five randomly chosen unique DNA probes and two restriction enzymes against human DNA. One of the probes uncovered a locus showing a high degree of polymorphism with both enzymes. Evola et al. (cited in Burr et al. 1983) found that 16 of 18 unique random maize-DNA probes showed polymorphism when tested against DNA endonuclease digests from seven maize strains, chosen for diversity on the basis of their isozyme loci. Similarly, in a preliminary study of tomato, Rom (1982) found polymorphism in DNA from four strains digested with a single enzyme and tested against a single random tomato DNA probe. The same probe demonstrated polymorphism in both petunia and tobacco (Rom 1982). Clearly, the fre-

quency of detectable polymorphisms per probe × enzyme combination tested can be expected to vary widely between species and populations. Nevertheless, as an order of magnitude it seems justifiable on the basis of the above results to assume a range of 10–20 probe × enzyme combinations tested, per useful polymorphism detected. On this basis, the number of probe × enzyme combinations that would have to be tested to provide sufficient polymorphisms for 80% genome coverage would range from 500–1,000 for $c=0.2$, $k=10$, to 3,000–6,000 for $c=0.1$, $k=30$.

Total number of polymorphisms scored to map a genome

In selfers or inbred lines, mapping will require a series of crosses between strains that differ with respect to some subset of the loci found to be polymorphic over the entire range of lines or strains. The proportion, Q, of polymorphic loci that would differ in at least one of m crosses would be $(1-2pq)^m$, where p and q are the frequencies of alternative alleles at the polymorphic marker locus. For useful polymorphic loci ($0.1 \leqq p \leqq 0.9$), 2pq will range from 0.18 to 0.5; for a rough average $2\overline{pq}=0.33$. For this value Q will equal 0.94 for $m=7$. Since in practice the various strains would be scored for the polymorphic markers, and cross combinations chosen so as to provide maximum coverage with a minimum number of crosses, the number of crosses required in practice for 95% coverage of the polymorphic loci detected would be less than this, say 5 or 6. About 50 backcross offspring per cross should allow linkage relationships up to a proportion of recombinations, $r=0.2$, to be unequivocally detected. With $2\overline{pq}=0.33$ and six crosses tested, the average polymorphism would be mapped on the basis of two crosses (100 offspring), allowing linkage relationships up to a level of $r=0.3$ to be detected. The offspring of any particular cross, of course, would be scored only for the polymorphic loci differentiating the parents, so that the average number of individuals tested per polymorphic locus would also be 100. In fact, if two crosses only, were tested per polymorphism, the average number of individuals tested per mapped polymorphism would be somewhat less than this (about 80–90). Thus, the total number of polymorphisms scored for 80% genome mapping would range from 5,000 for $c=0.2$, $k=10$ to 30,000 for $c=0.1$, $k=30$.

In the case of outcrossers in linkage equilibrium, mapping would be carried out by analyzing the offspring of individual males heterozygous for a subset of the polymorphic loci in the population as a whole. The average proportion of individuals heterozygous at any given polymorphic locus ($2\overline{pq}$), and the number (m) of males whose offspring are to be scored to allow a given proportion (Q) of the polymorphic loci to be mapped are given by the same expressions as in the case of inbred lines. However, in segregating populations only homozygous offspring genotypes are informative of linkage relationships. Assuming linkage equilibrium between two loci in the female parent, these will range in frequency from 0.25 to 0.35 for $P=0.5$ and $P=0.9$ at the two loci, respectively. Thus, in a segregating population one would have to score three to four times as many offspring per tested male to obtain equivalent power as in a cross between inbred lines, i.e., 150 to 200 offspring per tested male. The total number of polymorphism tests carried out for 80% genome mapping in this case, therefore, would range from 15,000 to 120,000 depending on c and k.

Costs

Breeding applications of polymorphic markers have been reviewed by Soller and Beckmann (1983). The costs of mapping RFLPs and of applying RFLPs to marker-based breeding applications will be discussed in this section. Clearly, these costs will be a function of (i) costs per RFLP scored, and (ii) the number of scored polymorphisms required for the various applications. Mapping requirements were discussed in the previous section. Costs per RFLP scored and breeding requirements will be discussed in the following sections.

Costs per polymorphism determination

The basic operations involved in demonstrating RFLPs include: (i) DNA extraction, (ii) preparation of probes, and (iii) DNA analysis, including restriction, electrophoresis and blotting of DNA, and hybridization and autoradiography of the blots. Table 2 shows estimated costs of materials and labor for each of these activities. The procedure for DNA extraction and purification considered here (Marmur 1963; Davis et al. 1980) consists of rupturing the tissue or cells, purification of the nuclei, proteinase K digestion, phenol and phenol-chloroform extractions, and purification of the DNA through centrifugation in a CsCl density gradient. It is assumed that plasmids harboring the appropriate DNA probes will have been previously cloned and are available. Our estimates for preparation of plasmid DNAs are based on a variation of the "boiling method" (Holmes and Quigley 1981), further modified to suit large-size DNA preparations through equilibrium sedimentation. Purification of internal fragments (purified RFLP probes) can be carried out by restriction enzyme cleavage of the plasmid DNA followed by fragment separation and recovery (Smith 1980; Yang et al. 1979). Labelling of probe DNA is carried out according to the procedure termed "nick translation" (Rigby et al. 1977). The procedure for revealing RFLPs involves: specific

Table 2. Cost estimates for RFLP determinations

Step	Sufficient for	Material per unit	Labor				Costs			
			per run	units per run	per unit	per unit[b]	per unit	per slot	per hybridization	per polymorphism
		($)	(h)[l]	(no.)	(h)	($)	($)	($)	($)	($)
DNA extraction										
per 10 g sample	20 slots[a]	2.00	32	24	1.25	12.50	14.50	0.725[c]	0.181[d]	0.060 – 0.725[e,j]
Probe preparation										
Plasmid DNA	180 hybridizations[f]	4.00	10	12	0.85	8.50	12.50	0.005[c,g]	0.005	0.005
Isolation of internal probe fragment	120 hybridizations[g]	10.00	10	24	0.43	4.30	14.30	0.006[c]	0.006	0.006
Probe labelling	5 hybridizations[h]	5.00[i]	10	20	0.50	5.00	10.00	0.100[c,h]	0.100	0.033[j]
DNA analysis										
Restriction, electrophoresis, and blotting	1 gel	18.00[k]	10	4	2.50	25.00	43.00	2.150[c]	0.540	0.180[j]
Hybridization and autoradiography	1 blot	2.50	10	20	0.50	5.00	7.50	0.375[c]	0.374	0.125[j]
Total										0.409 – 1.074[e]

[a] Based on yield of 100 μg DNA/10 g tissue, and on the use of 5 μg DNA/slot. [b] Based on $ 20,000 per year salary. [c] Based on 20 slots per gel. [d] Based on re-using each Southern blot 3x for a total of 4 hybridizations per blot. [e] Lower figure based on utilization of a single DNA sample for a maximum number (=240) RFLP determinations. Higher figure based on utilization of a single DNA sample for 20 RFLP determinations. [f] Based on yield of 30 μg DNA per plasmid preparation, nick translation in pools of 3 probes, and total of 0.5 μg plasmid DNA per nick translation. [g] Based on 66% recovery of the internal probe fragment. [h] Based on the assumption that 80% of plasmid probe labelled by nick translation does not participate in hybridization, and is lost to medium or may contribute to background. On this assumption, a labelled internal probe fragment should give the same relative intensity of autoradiographic signals with only 20% of the total counts in the hybridization mixture as compared to a whole plasmid probe. [i] Based on labelling in pools of 3 probes, in 10–20 μl reaction containing 20–50 μCi of high specific activity P^{32} labelled deoxytriphosphate precursors. [j] Based on hybridization with a pool of 3 probes. [k] Based on current prices for commercially available endo R Eco RI and endo R Msp I. [l] Based on sum of net labor time, not including waiting times greater than 1 h

cleavage of the various DNAs by restriction enzymes, separation of the resulting DNA fragments by electrophoresis in agarose gels, transferring the DNA fragments from the gel to a nitrocellulose filter or other support (Southern 1975; Alwine et al. 1979), allowing for the annealing of the labelled DNA probe to the complementary DNA sequences present on the filter, and finally, analysis of the hybridization pattern by autoradiography. Further technical details are found in the footnotes to Table 2.

Examination of Table 2 shows that DNA extraction is a major cost, coming to $ 15.00 per sample at current prices. Although a given DNA sample can be used for up to 240 RFLP determinations (20 slots per sample × 4 hybridizations per blot × 3 probes per hybridization), most applications would require far fewer determinations. Preparation of probes is a rather minor cost, coming to about $ 0.05 per RFLP determination. DNA analysis, again, is a major cost, coming to $ 0.30 per RFLP determination. The overall costs of an RFLP application would thus vary from $ 0.41 to $ 1.10 per polymorphism scored, depending on the number of polymorphisms scored per DNA sample.

Number of polymorphisms scored for various breeding applications

The number of polymorphisms scored for various breeding applications are summarized in Table 3. Values for varietal and parentage identification are as given in Soller and Beckmann (1983). Other values are derived as described in the following paragraphs, based on Soller and Beckmann (1983).

Identification, evaluation and mapping of QTL by linkage relationships to polymorphic markers. For this purpose a QTL should be no more than 0.2 Morgan from a marker. In mapping QTL in a cross between two inbred lines, the lines would be scored for all polymorphic

Table 3. Costs for various *RFLP* applications according to genome size (k), as determined by number of individuals tested and number of RFLPs scored per individual tested[a]

Application	Individuals tested	RFLPs scored per individual[b]	Costs		
			DNA extraction[c]	DNA analysis[d]	Total
	(no.)	(no.)	($)	($)	($)
Initial screening of population		500–1,000	600–1,200	3,500– 7,000	4,100– 8,200
for FLPs	20	3,000–6,000	3,600–7,200	21,000–42,000	24,600–49,200
Mapping of RFLPs					
Selfers	300[e]	50	4,500	5,250	9,750
		290	9,000	30,450	39,450
Outcrossers	1,050[f]	50	15,750	18,375	34,125
		290	31,500	106,575	137,075
Strain identification per inbred strain	5	20	75	35	110
Segregating strain	20	20	300	140	440
Parentage identification per individual[g]	1	20	15	7	22
Evaluating and mapping QTL[h]					
Inbred lines	6,000	40	90,000	84,000	174,000
		100	90,000	210,000	300,000
Segregating populations	15,000	40	225,000	210,000	435,000
		100	225,000	525,000	750,000
Screening genetic resources[i]					
Inbred lines	1,000	40	15,000	14,000	29,000
		100	15,000	35,000	50,000
Segregating populations	6,660[l]	40	99,900	93,240	19,314
		100	99,900	233,100	333,000
	1,330[m]	40	19,950	18,620	38,570
		100	19,950	46,550	66,500
Introgression of a useful allele					
Marker bracket only[j]	100	2	1,500	70	1,570
Marker bracket + culling[k]	50	40	750	700	1,450
	150	30	2,250	1,575	3,825
Pedigree breeding	3,000	30	45,000	31,500	176,500
		100	45,000	105,000	150,000
Marker-assisted young sire selection					
per evaluated sire	1,500	90	22,500	47,250	69,750
per young sire	1	90	15	31	46
Identifying a known locus					
per locus	10,000	3	150,000	10,500	160,500
per 20 loci	40,000	60	600,000	840,000	1440,500
Selection on the basis of 20 known loci					
per young sire	1	60	15	21	36

[a] See text for all details not covered in footnotes. [b] Upper value, $k = 10$ Morgans; lower value, $k = 30$ Morgans. [c] Based on $ 15 per DNA extraction and assuming 1 DNA extraction supplies sufficient DNA for 240 RFLP determinations. [d] Based on $ 0.35 per *RFLP* determination. [e] Assuming 6 crosses × 50 offspring per cross. [f] Assuming 6 males tested × 175 offspring per male. [g] Assuming sire and dam already scored. [h] Quantitative trait loci. [i] Per line or strain. [j] Assuming 5 generations × 20 individuals scored per generation. [k] Assuming 1–3 generations × an average of 50 individuals per generation. [l] Total F-2 generation tested, analyzing for all quantitative traits. [m] Top and bottom 10% of F-2 generation tested, analyzing for one quantitative trait only

markers, and a subset of marker differences formed covering as much of the genome as possible. For locating QTL a marker spacing of 0.4 Morgan would be adequate. Considering chromosome ends, and assuming an average chromosome length of 1 Morgan, this works out to 3–4 markers per chromosome, or 30–40 polymorphic markers for k = 10 and 90–100 markers for k = 30. In analyzing a cross between inbred lines, about 1,000 F-2 offspring are required to identify and evaluate QTL differences between the lines, independent of the number of polymorphisms scored. Identifying the bulk of polymorphic QTL across a series of strains would

require five or six such crosses. For segregating populations, mapping of QTL would take place by forming a large F-2 out of a single male parent heterozygous at a large number of polymorphic marker-loci and mated to a number of females, as described in Soller and Beckmann (1983). Three-fourths of the offspring in this F-2 can contribute information on the quantitative value of heterozygous QTL near any particular marker locus. (The remainder can be expected to be uninformative due to overlap of male and female marker genotypes.) Since the marker contrasts in this case include both homozygous and heterozygous types, the power of the design is less than in the F-2 out of inbred lines, and correspondingly more offspring would be required: twice as many for equivalent power in the contrast, and another third to provide sufficient daughters with informative marker contrasts. Thus, this would require about 2,500 offspring per tested male, and again, one would probably test 5 or 6 males to have good genome coverage with respect to marker loci and polymorphic QTL, i.e., some 15,000 F-2 offspring in all, × 30–100 markers per offspring tested, depending on k.

Breeding applications: inbred lines. Screening genetic resources for alleles or chromosome segments showing useful additive effects for introgression into pure lines, or useful dominant or heterotic effects for introgression into inbred lines for crossing, can be carried out by application of the QTL mapping methods described above. Identified segments can then be readily introgressed into the appropriate recipient lines using markers, differentiating the lines, to monitor the introgression of the desired segment and to cull undesired chromosomal material as described by Tanksley et al. (1981) and Soller and Beckmann (1983). About five generations would be required to complete the backcross procedure, if a marker bracket is utilized to ensure retention only, of the desired QTL. One to three backcross generations would be required, depending on k, if markers are used to cull unwanted chromosomal material as well. In the first case 10–20 backcross individuals would be tested each generation × two markers per individual for a total of 100–200 polymorphisms scored per introgressed segment over the entire introgression procedure. In the second case, for $k = 10$, it would be possible to complete the backcross procedure in one generation, scoring about 50–100 BC-1 individuals for 30–40 markers each. For $k = 30$, it would take about three generations to complete the formal introgression procedure, eliminating all donor markers except for the desired bracket. This would require testing about 50 individuals each BC generation × 100 markers the first, 30 markers the second and 10 markers the third BC generation.

In a pedigree breeding program (Simmonds 1979), markers can be used to carry out a preliminary screening of the derived lines indicating a set of lines to be retained for intensive testing (Soller and Beckmann 1983). In this case, the initial cross would be analyzed for QTL in the F-2 generations as described above. This analysis would continue through the next 2 or 3 selfing generations to provide additional information on marker-associated QTL, and on the stability of the effects across years and locations. Thus, the evaluation would require, say, 1,000 individuals per generation × 3 generations × 30–100 polymorphisms scored per individual depending on k. That is, the total number of polymorphisms scored would range from 90,000 to 300,000 depending on k.

Breeding applications: segregating populations. Screening of resource strains and introgression of favorable alleles can be carried out in a manner similar to that for inbred lines. The basic assumption is that the useful QTL allele is more or less absent from the recipient strain, but is found at fairly high frequencies in the resource donor strain. Screening of a resource strain for such alleles would take place by intensive testing of a small number of males from the resource strain in crosses to recipient strain females. Each male and female parent would be scored for as many markers as feasible. For each individual female, chromosome segments for which female and male marker haplotypes differed would be monitored in the F-2 offspring of that female. The proportion of the genome for which female and male marker haplotypes can be expected to differ would depend greatly on the relative frequency of the various marker alleles in donor and recipient populations, on the prevalence of multiple alleles, and on the total density of polymorphisms demonstrated in the population as a whole. Assuming as above that three-fourths of the offspring of any particular male × female combination contribute information as to the quantitative value of the donor QTL near any particular marker, the cross as a whole would be treated as a cross between two inbred lines, and analyzed accordingly. The number of offspring per male would be increased by ⅓ (to 1,333 per male tested) to compensate for offspring not scored for a particular marked segment due to overlap of male and female marker haplotypes. This number of offspring would detect instances in which both of the donor male chromosomes carried the useful QTL allele. Assuming an overall frequency of 0.5 or more for the useful allele in the resource population, testing 4 or 5 males in this manner would give a high probability of detecting a useful QTL allele. The number of polymorphisms tested per F-2 offspring would vary from 40 to 100,

depending on k. If a useful segment were identified, the F-2 offspring themselves would serve as the basis for a marker monitored-introgression procedure as described above for inbred lines. If a resource strain is analyzed for a single quantitative trait only, it would be necessary to score only the top and bottom 10%, say, of the F-2 offspring (according to phenotypic value for the quantitative trait) for RFLPs, in order to identify marker alleles linked to the QTL affecting the trait in question. Thus, it would be possible to screen a resource strain for marker-linked QTL affecting a single trait on the basis of 1,200–1,400 F-2 offspring, instead of the 6,000–7,000 required when analyzing for a number of traits. QTL-linked marker alleles detected in this manner could then be used to monitor the introgression procedure as suggested by Stuber et al. (1982).

Marker-assisted selection of young sires in species, such as dairy cattle ($k = 30$) where very large progeny groups can be obtained from individual males, is discussed in Soller (1978) and Soller and Beckmann (1983). In such a program, alternative homologous chromosomes of a proven sire are evaluated in the offspring of the sire, and young sire selection is based in part on the expected breeding value of the marker assortment they receive from their marker-evaluated sire. It is assumed that initial QTL mapping studies have identified a set of 30–40 marker bracket haplotypes that include polymorphic QTL having significant effects on economic traits. In this case, 1,000–1,500 offspring would be tested per marker-evaluated sire and about 90 polymorphisms scored (defining 30 marker haplotypes) per daughter, for a total of 90,000 to 135,000 polymorphisms scored per evaluated sire. As information accumulates on the economic value of particular marker haplotypes over a number of generations and evaluated sires, some of these marker haplotypes will move into the category of a "known-locus" (Soller and Beckmann 1983) in which case selection of young sires will be based on marker haplotype alone, without preliminary evaluation of haplotype quantitative value. For any particular chromosome region, this would require scoring about 10,000 offspring for alternative marker haplotypes. Considering overlap of sire and dam marker haplotypes at any particular chromosome region, and the likelihood that QTL at the region are homozygous in the sire, each daughter tested would provide information on no more than ¼ of the marker-regions being evaluated in the population as a whole. On this basis, a total of 40,000 scored offspring might suffice to bring some significant number, say 20, of marker bracket haplotypes to a "known-locus" condition. At this point marker haplotype ("known locus") selection of young sires would become a powerful source of genetic progress in the population (Soller and Beckmann 1983).

Costs per breeding application

Table 3 also gives estimated costs for the various RFLP applications described above, as a function of the number of DNA extractions required (equal to the number of individuals tested) and the number of RFLPs scored per individual tested. In most cases costs appear to be commensurate with the scientific or economic value of the application. Costs for evaluation and mapping of QTL are high, particularly in segregating populations, and this information might best be obtained, perhaps, as part of a breeding program, rather than as an independent objective in its own right. Costs for screening segregating resource populations for useful QTL also seem high, except when screening for a single quantitative trait. Thus this could be economically useful if the eventual recipient population were at a selection plateau and the economic value of the trait under improvement was great. Costs for sire selection are high, but in the case of outstanding proven sires, whose sons will enter service as young sires, evaluation of the proven sire and marker-assisted selection among his sons might be economically justified. The investment required to bring marker brackets to the stage of "known loci" seems high, although costs would be expended over a period of 5–10 years. This application might become more attractive, however, as experience accumulated on the value of marker-assisted selection among sons of outstanding proven sires and on the magnitude of marker-linked quantitative effects in dairy cattle populations. Twenty "known loci" might well account, between them, for the bulk of genetic variation in milk production in dairy cattle, and their full exploitation over a relatively short period of time might yield eventual production increases on the order of 2,000 to 3,000 kg per lactation. A return of this magnitude would more than adequately repay the investment involved.

Acknowledgement. This study was supported by a grant from the U.S.-Israel Binational Agricultural Research and Development Fund (BARD).

References

Alwine JC, Kemp DJ, Parker BA, Reiser J, Renart J, Stark GR, Wahl GM (1979) Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. Methods Enzymol 68: 220–242

Baker CM, Manwell C (1980) Chemical classification of cattle. 1. Breed groups. Anim Blood Groups Biochem Genet 11:127–150

Bishop D, Skolnick MH (1980) Numerical considerations for linkage studies using polymorphic DNA markers in humans. Banbury Rep 4: cancer incidence in defined populations. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp 421–433

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314–331

Burr B, Evola SV, Burr FA, Beckmann JS (1983) The application of restriction fragment length polymorphism to plant breeding. In: Setlow JK, Hollaender A (eds) Genetic engineering principles and methods, vol 5. Plenum Press, New York, NY, pp 45–59

Davis RW, Thomas M, Cameron J, St. John TP, Scherer S, Padgett RA (1980) Rapid DNA isolations for enzymatic and hybridization analyses. Methods Enzymol 65:404–411

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

Grodzicker T, Williams J, Sharp P, Sambrook J (1974) Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harbor Symp Quant Biol 39:439–446

Holmes DS, Quigley M (1981) A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem 114: 193–197

Jeffreys AJ (1979) DNA sequence variants in the G_γ-, A_γ-, δ- and β-globin genes of man. Cell 18:1–10

Lange K, Boehnke M (1982) How many polymorphic marker genes will it take to span the human genome? Am J Hum Genet 34:842–845

Marmur J (1963) A procedure for the isolation of deoxyribonucleic acid from microorganisms. Methods Enzymol 6:726–738

Nevo E, Zohary D, Brown AHD, Haber M (1979) Genetic diversity and environmental associations of wild barley *Hordeum spontaneum* in Israel. Evolution 33:815–833

Rick CM, Fobes JF (1975) Allozyme variation in cultivated tomato and closely related species. Bull Torrey Bot Club 102:376–384

Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J Mol Biol 113: 237–251

Rom M (1982) Development of restriction fragment length polymorphism techniques in plants as a tool for strain identification and breeding (in Hebrew, English summary). MSc Thesis, Department of Genetics, The Hebrew University of Jerusalem, Israel

Simmonds NW (1979) Principles of crop improvements. Longman, London

Smith HO (1980) Recovery of DNA from gels. Methods Enzymol 65:371–380

Soller M (1978) The use of loci associated with quantitative traits in dairy cattle improvement. Anim Prod 27:133–139

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

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

Solomon E, Bodmer WF (1979) Evolution of sickle variant gene. Lancet I 8122:923

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Stuber CW, Goodman HH, Moll RH (1982) Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. Crop Sci 22:737–740

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

Washburn KW, Maeda Y, Lanza GM (1980) Protein polymorphism in a random bred chicken population. Anim Blood Groups Biochem Genet 11:261–269

Wyman AR, White R (1980) A highly polymorphic locus in human DNA. Proc Natl Acad Sci USA 77:6754–6758

Yang RCA, Lis J, Wu R (1979) Elution of DNA from agarose gels after electrophoresis. Methods Enzymol 68:176–182

Zabeau M, Roberts RJ (1979) The role of restriction endonucleases in molecular genetics. In: Taylor JH (ed) Molecular genetics, vol III. Academic Press, London New York, pp 1–63