Theor Appl Genet (1983) 67:25–33

Genetic polymorphism in varietal identification and genetic improvement *

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

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

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

Summary. New sources of genetic polymorphisms promise significant additions to the number of useful genetic markers in agricultural plants and animals, and prompt this review of potential applications of polymorphic genetic markers in plant and animal breeding. Two major areas of application can be distinguished. The first is based on the utilization of genetic markers to determine genetic relationships. These applications include varietal identification, protection of breeder's rights, and parentage determination. The second area of application is based on the use of genetic markers to identify and map loci affecting quantitative traits, and to monitor these loci during introgression or selection programs. A variety of breeding applications based on these possibilities can be envisaged for selfers, particularly for those species having a relatively small genome size. These applications include: (i) screening genetic resources for useful quantitative trait alleles, and introgression of chromosome segments containing these alleles from resource strain to commercial variety; (ii) development of improved pure lines out of a cross between two existing commercial varieties; and (iii) development of crosses showing increased hybrid vigor. Breeding applications in segregating populations are more limited, particularly in species with a relatively large genome size. Potential applications, however, include: (i) preliminary selection of young males in dairy cattle on the basis of evaluated chromosomes of their proven sire; (ii) genetic analysis of resource strains characterized by high values for a particular quantitative trait, and introgression of chromosome segments carrying alleles contributing to the high values from resource strain to recipient strain.

Key words: Genetic polymorphisms – Genetic markers – Genetic improvement – Varietal identification

Introduction

Methods have recently been developed, based on the use of restriction enzymes and cloned DNA, that allow a new class of genetic polymorphism – restriction fragment length polymorphisms (henceforth, RFLPs) – to be detected (Bishop and Skolnick 1980; Botstein et al. 1980; Grodzicker et al. 1974; Solomon and Bodmer 1979). It is anticipated that RFLPs will be numerous and developmentally stable, and that they will act as codominant Mendelian markers, free of pleiotropic effects on economically important characters (Beckmann and Soller 1983). The purpose of this paper is to draw attention to the implications of a virtually unlimited source of genetic polymorphisms for breeding practice. Three major areas of potential marker utilization will be reviewed: (i) varietal and parentage identification, (ii) identification of genetic loci affecting quantitative economic traits, and (iii) genetic improvement programs, including screening and evaluation of germplasm resources, introgression, improvement of commercial hybrids and within-population selection. Because of the different genetic architecture of inbred lines and selfers as compared to segregating populations and outcrossers, the discussion of each area will be caried out separately for the two mating types. The nature of RFLPs, the methodologies involved in their detection, and the anticipated costs of applying RFLP methodologies to genome mapping and breeding practice are discussed in a companion paper (Beckmann and Soller 1983).

* Contribution from The Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel, No. 767-E, 1983 Series

Strain and parentage identification

Three situations of interest can be distinguished (i) Varietal identification, requiring a set of markers that can distinguish between n strains or varieties, (ii) Protection of breeder's rights, requiring a set of markers that will allow calculation of the likelihood that an unknown population actually represents a given commercial variety, and (iii) Parentage identification, requiring a set of markers allowing positive parent identification in cases where total control of mating opportunities is difficult for biological or agrotechnical considerations (e.g., honey bees, beef cattle or fruit trees).

Inbred lines and selfers

It is assumed that k independent polymorphic loci are available, two alleles at each locus, and that the frequency of alternative genotypes in the entire population of varieties is 0.5, i.e., half of the varieties are of the genotype "AA" and half of the genotype "aa". This is the worst case with respect to number of alleles per locus and the best case with respect to allelic frequencies. For purposes of varietal identification, the relevant statistic is the probability, P_1, that each of n varieties has a different genotype,

$$P_1 = \frac{2^k!}{2^{kn}\ (2^k - n)!}.$$

For protection of breeder's rights, the relevant statistic is the probability, $P_2 = (\frac{1}{2})^k$ that a variety chosen at random has the same genotype as a given commercial variety. $R_1 = 1/P_2 = 2^k$ will then be the relative likelihood that an individual having a genotype identical to that of the commercial variety, indeed originated from that variety.

Trial calculations using these expressions show that given a set of 20 or more polymorphic loci of this sort, a testing laboratory could confidently expect to distinguish among any two or more varieties submitted; while given a set of 10–15 such polymorphic loci, a breeder could maintain with high plausibility that any individual having a genotype identical to that of his particular variety was indeed a member of that variety. Very high relative likelihoods could be provided if a number of rare marker alleles could be purposely incorporated into new varieties as these are developed.

Segregating populations and outcrossers

Protection of breeders rights in segregating populations requires a set of k polymorphic loci for which allelic frequencies in the commercial variety differ from the values in the overall population. In this case the relative likelihood, R_2, that a sample of n individuals originated from the given commercial variety is calculated by comparing the likelihood that they originated from the commercial variety to the likelihood that they originated from an "average" variety having the overall allelic frequencies of the entire population. R_2 can be calculated by maximum likelihood methods and is generally expressed as the log. likelihood ratio

$$\text{Log } R_2 = \sum_{j=1}^{n} \sum_{i=1}^{k} \text{Log } P(G_{ij}) - \sum_{j=1}^{n} \sum_{i=1}^{k} \text{Log } P'(G_{ij})$$

where,

n and k are defined as above,

G_{ij} is the genotype at the i^{th} polymorphic locus, of the j^{th} individual in the sample,

$P(G_{ij})$ is the frequency of this genotype in the breeder's variety, and

$P'(G_{ij})$ is the frequency of this genotype in the population as a whole or in some specific alternative variety.

A Log R_2 value of 3, for example, would be equivalent to a relative likelihood value, R_2, of 1000 : 1.

Trial calculations show that ten polymorphic loci, and ten individuals sampled, should suffice for relative likelihood values of 1000 : 1 in most instances.

Parentage identification. In random-mating populations, genetic polymorphisms can serve as a means of positive parent identification, for purposes of progeny testing, say, using procedures similar to those employed in cases of disputed parentage in humans (Silver 1978). In this case, the relevant statistic is the "combined probability of exclusion, CPE". This is the a priori probability that a randomly chosen male of a known genotype will be excluded as the parent of a given offspring on the basis of the genotype of both the offspring and its female parent at one or more polymorphic loci.

For multiple polymorphic systems, it can be shown that

$$\text{CPE} = 1 - \prod_{i=1}^{k} (1 - PE_i)$$

where PE_i is the probability of exclusion at the i^{th} polymorphic locus. For loci having two codominant alleles (as in the case of RFLPs) there are six such combinations in which the male parent can be excluded if female parent and offspring are scored, and two combinations in which the male parent can be excluded if only the offspring is scored (Table 1). Based on this table it can be shown that for any given polymorphic locus, with two codominant alleles and allelic frequencies p and q, $PE = pq\,(1 - pq)$ when both female

Table 1. Exclusion combinations and their probabilities

Random male parent	Female parent	Offspring	P
Female parent and offspring scored			
AA	AA	Aa	$p^2 \cdot p^2 \cdot q$
AA	Aa	aa	$p^2 \cdot 2pq \cdot \frac{1}{2}q$
AA	aa	aa	$p^2 \cdot q^2 \cdot q$
aa	aa	Aa	$q^2 \cdot q^2 \cdot p$
aa	Aa	AA	$q^2 \cdot 2pq \cdot \frac{1}{2}p$
aa	AA	AA	$q^2 \cdot p^2 \cdot p$
Total			$pq(1-pq)$
Offspring alone scored			
AA		aa	$p^2 \cdot q^2$
aa		AA	$q^2 \cdot p^2$
Total			$2p^2q^2$

parent and offspring are scored, and $PE = 2p^2q^2$ when the offspring alone is scored.

Trial calculations show that if female parent and offspring are both scored, $k = 15$ polymorphisms will give a CPE of 0.95 when average allelic frequencies are 0.5, while if the female parent is not scored, 22 such polymorphisms would be required for a CPE of 0.95.

In agricultural situations there will often be a fairly large number of possible male parents, and the problem becomes one of identifying the true parent. In this case the probability of identification, PI (the probability that all but the true male parent will be excluded), will be

$$PI = (CPE)^{n-1}$$

and the probability that at least one other male, in addition to the true male parent, will not be excluded, is 1-PI. This probability is not negligible even when the CPE is high. For example, if the $CPE = 0.95$ and $n = 20$, $PI = 0.36$. Thus, in such a case the probability that one would be left with two or more non-excluded males at the end of the scoring procedure is 0.64. In this case the final decision would be made by maximum likelihood "plausibility of paternity" calculations (Silver 1978), carried out as follows. From inspection of the genotypes of female parent and progeny, the set of alleles which must have come from the male parent are identified. For example, if the offspring is Aa, and the female parent is AA, the male parent must have contributed the allele a. The probability that each non-excluded male contributed this set of alleles is then calculated as the product of the probabilities that he contributed each allele in the obligatory set. These individual probabilities will be 1.0 or 0.5, depending on whether the male is homozygous or heterozygous at the locus in question. Since it is assumed that one or other of the non-excluded males is the true parent, the sum of these individual male probabilities for all unexcluded males is taken as unity. The probability of paternity for each male will then be the ratio of his specific probability to the sum of all probabilities.

Strain identification. In segregating populations strain identification can be accomplised in a similar manner by methods analogous to those for establishing plausibility of paternity. This involves calculating the likelihood that a given sample came from each of a series of known strains. Since it is assumed that the sample originates from one or the other of the known strains, the plausibility of origin from some particular strain can be taken as the ratio of specific strain probability to the sum of probabilities of all strains.

Genetic analysis of quantitative traits

Selfers and inbred lines

Methodologies for locating chromosome segments or loci affecting quantitative traits based on linkage to marker loci in backcross populations have been applied in *Drosophila* (Thoday 1961; Spickett and Thoday 1966; Mather and Jinks 1971) and wheat (Law 1966). The adaptation of these methods to agricultural plants and animals has been considered by Soller et al. (1976). When codominant markers are available (i.e., isozymes markers or RFLPs) analyses based on the F-2 generation will be more useful than those based on backcross generations in that they provide information on both dominance and main effects of the identified quantitative trait locus (henceforth, QTL, Geldermann 1975). The basic approach is outlined in Table 2. It can be seen that appropriate functions of the F-2 marker genotype means provide estimates of main effects (2d) and relative dominance (h/d) at the QTL, attenuated by the proportion of recombination between marker locus and QTL. Some notion as to the value of r, and the absolute effects of d and h, can be obtained when the QTL is bracketed between two marker loci (Thoday 1961). In other cases, estimates can be provided by maximum likelihood methods (Weller 1983). As has been pointed out by McMillan and Robertson (1974), a located QTL may actually represent the center of gravity of two or more separated QTLs in the vicinity of the marker. In this case, what is being evaluated is the overall quantitative value of the chromosome segment to both sides of the marker locus. However, for most breeding purposes, this will still be informative and useful. The likelihood of two or more QTLs lying in the vicinity of a single marker may not be great (Soller et al. 1979).

Experimental studies (Weller 1983; Patterson et al. 1968; Zhuschenko et al. 1979; Tanksley et al. 1981; Tanksley et al. 1982; Kluge and Geldermann 1982) show that a wide variety of quantitative effects can in fact be found associated with specific genetic markers. Maximum likelihood and other mapping methods show that in some cases at least the results are due to linkage rather than pleiotropy. These experiments suggest that a significant proportion of marker-linked effects on quantitative traits are of a magnitude that could be readily detected in experiments of about 1000 F-2 or backcross individuals. It is important to note that if two strains differ in

Table 2. Evaluation of main effect and dominance relations of quantitative trait loci (QTL) in the F-2 of a cross between inbred lines differing at both QTL and marker locus

Definitions and symbols

r = proportion of recombination between marker locus and QTL

Genotypes at the QTL and their value: AA = +d
Aa = h
aa = −d

Genotypes at the marker locus: MM, Mm, mm

Mean phenotypic value of marker genotypes in F-2: $\overline{MM}$, $\overline{Mm}$, $\overline{mm}$

Cross between inbred lines differing at marker locus and at QTL

Parental genotype $\frac{MA}{MA} \times \frac{ma}{ma}$

F-1 genotype $\frac{MA}{ma}$

F-2 genotypes	MM	Mm	mm
Value	$(1-2r)\,d + 2r(1-2r)\,h$	$((1-r)^2 + r^2)\,h$	$-(1-2r)\,d + 2r(1-2r)\,h$

Estimates of main effects and dominance at the QTL

Main effects $\overline{MM} - \overline{mm} = 2(1-2r)\,d$

Dominance $\overline{Mm} - \frac{1}{2}(\overline{MM} + \overline{mm}) = (1-2r)^2\,h$

Relative dominance $\frac{\text{Dominance}}{\frac{1}{2}\text{ Main effects}} = (1-2r)\frac{h}{d}$

many markers, suitably spaced along the genome, a single F-2 generation will alow a total evaluation and rough mapping of the entire genome for all quantitative traits scored. Thus the quantity of information yielded by this type of analysis is critically dependent on the number of marker polymorphisms differentiating the two parental lines.

Segregating populations

The use of polymorphic genetic markers for genetic analyses in segregating populations is more problematic than in selfers or inbred lines (Soller and Genizi 1978; Soller 1978), since in a segregating population under linkage equilibrium, the sign of the quantitative allele associated with any particular marker allele will be different in different homologous chromosomes. As a result, linkage of marker and QTL is detected by intensive analysis, in large numbers of F-1 and F-2 offspring, of known homologous chromosome pairs originating from a single parent individual and differing in one or more marker alleles. The procedure would be as follows. Denote the male parent as M_1M_2 and the female parents as M_FM_F, where M_1, M_2, and M_F are all well marked homologous segments. It is assumed that in the male parent, M_1 and M_2 differ at one at least of the marker loci included in the segment. The M_F are all of the corresponding homologous segments in the female parents (not necessarily identical to one another). Then the initial cross will be $M_1M_2 \times M_FM_F$, and the contrast $\overline{M_1M_F} - \overline{M_2M_F}$ in the F-1 will give initial information as to the relative quantitative value of the M_1 and M_2 segments in the male tested. In the F-2, six genotypic classes will be recognized at the marked segment in the following proportions: 1/16 M_1M_1, 1/8 M_1M_2, 1/16 M_2M_2, 1/4 M_1M_F, 1/4 M_2M_F, 1/4 M_FM_F. From among these, appropriate contrasts can readily be formed to evaluate M_1 as compared to M_2, and to evaluate both M_1 and M_2 with respect to the female line homologous segments, M_F. Clearly, in this case the likelihood of having good genome coverage depends critically on the total number of polymorphic markers available. Furthermore, the availability of a large number of markers on the parental chromosomes will facilitate the analyses by increasing the proportion of F-1 and F-2 offspring in which the alternative parental chromosomes can be unequivocally identified. This derives from the fact that a well-marked chromosomal segment (e.g. a 20 cM segment, delimited by specific marker alleles at each end, and containing one or more internal polymorphic marker alleles) serves as a marker haplotype that can be followed from generation to generation almost as readily as a single low frequency allele in a multiple allelic series. For example, assuming $P = 0.5$ at each marker locus, and a segment marked by three heterozygous markers, then 1/8 of female line chromosomes would have the M_1 allelic haplotype and 1/8 the M_2 haplotype. If these are not identified in the F-2 this will reduce the various contrasts by 1/8. This should not have a major effect on the ability of the design to detect QTL having significant effects on the quantitative characters of interest.

When analyzing a resource strain with respect to only one or two quantitative traits, an alternative approach is to score for genetic markers only individuals showing extreme phenotypes for the quantitative trait of interest. In this case, two groups of individuals showing high or low phenotypes for a particular trait, respectively, are selected out of a segregating F-2 population of a cross between resource strain and standard variety, and marker-allele frequencies in the two groups are compared. Such groups will differ in frequency of alleles affecting the quantitative trait, and hence, by hitchhike effects, will also differ in frequency of any linked marker alleles. In this way marker alleles close to QTL affecting the trait of interest can be identified without the necessity for scoring all offspring with respect to marker alleles. This procedure is similar to that suggested by Stuber et al. (1982).

Breeding applications

Selfers

Screening germplasm resources. Marker-based evaluation and mapping of the genome for QTL have the potential to screen germplasm resources (wild progenitor populations, accession lines or cultivars) and identify high frequency chromosomal segments of potential usefulness for the improvement of quantitative traits in commercial varieties. This approach has a number of advantages compared to searching for transgressive variation in the F-2 or subsequent generations of a cross between commercial variety and resource strain as an indication of hidden favorable alleles in the resource strain: (i) for any given magnitude of experiment, the power of the marker analysis will be greater when more than a few loci are involved or when gene effects are relatively small; (ii) information will be available in the F-2 generation, without the necessity of going on to subsequent generations; (iii) it will be possible to distinguish between additive effects and epistatic effects as a source of transgressive variation; (iv) it will be possible to determine whether useful alleles in different resource strains are at the same or different loci; and (v) it will be possible to determine whether the resource strain differs from the commercial variety in one or a number of useful alleles affecting the trait of interest.

Introgression of useful chromosome segments from resource strain to commercial variety. When commercial variety and "resource" strain differ at a QTL and at the adjoining marker locus (or at two nearby marker loci with the QTL bracketed between them), repeated backcrosses to the commercial strain, retaining only backcross progeny carrying the exotic marker allele(s), will allow the effective introgression of the linked QTL from "resource" to commercial variety (Soller and Plotkin-Hazan 1977). Theoretical analyses (Soller et al. 1979) show that under most circumstances the majority of marker-linked quantitative effects should be due to no more than one or two QTL in the near vicinity of the marker locus. Thus, it should be possible to separate the linked favorable allele from any associated linked unfavorable alleles prior to or subsequent to the introgression process (Soller and Brody, to be published). For introgression of useful alleles from wild species Tanksley and Rick (1980) and Tanksley et al. (1981) showed that selection for recurrent parent markers along with the introgressed gene, can result in major savings in the number of generations required for completion of the backcrossing procedure and in the amount of field space required to test mature individuals. A similar procedure could obviously be used with positive selection for a marker-bracket as well, leading to an extremely effective and efficient introgression procedure.

Development of improved pure lines out of a cross between existing pure lines. Genetic improvement in most selfers proceeds by development of improved pure lines out of crosses between different existing pure lines (pedigree, bulk or single-seed descent methods; Simmonds 1979). In such programs the methods for genetic analysis of quantitative traits described above can provide a partial answer to the problem of retaining the best of the families developed in the early generations for further screening in the later generations of the program. This would proceed as follows. Alternative chromosome segments in the vicinity of polymorphic markers differentiating the parental pure lines would be evaluated for QTL in the initial generations of the cross (F-2 to F-5), during the period when homozygosity of the pure line families generated by the cross is established. Repeated evaluations in the F-2 to F-5 generations, over a number of seasons and sites, would also provide information on genotype × environmental interactions affecting the various QTL. At the F-5 or F-6 generations, the many families generated in the initial stages of the program would be subjected to preliminary screening on the basis of the overall value of their QTL as indicated by their marker genotypes. In this way the combined information of a number of seasons and of many individuals and families would be brought to bear on the initial selection of individual F-5 and F-6 families. The selected families would then undergo the usual, more extensive evaluation with respect to quality and yield. This should allow the final stages of the program to start from a significantly superior initial level as compared to present methods, particularly with respect to traits of low heritability.

Development of crosses between inbred lines showing increased hybrid vigor. The methods for genetic analysis described in the previous sections can also be utilized to identify specific QTL showing dominance or overdominance in the F-2 of a cross between two inbred lines. This would enable an existing cross to be analyzed in terms of the number of QTL involved in producing the observed hybrid vigor, the type of genetic effect involved (i.e., complementary dominance or overdominance), and the magnitude of effect of such QTL. Following this, additional inbred lines could be tested against the standard lines, and screened for the presence of additional dominant or overdominant QTL. These could then be incoporated into the alternate standard line by marker-based introgression techniques as described above. In this way OTL contributing to heterotic effects

could be collected from a wide variety of resource lines and introduced into the standard lines, allowing direct and specific construction of crosses showing maximum degrees of hybrid vigor.

Segregating population

Analysis of resource strains and introgression of useful chromosomal segments. In the case of segregating populations, resource strains can be screened for the presence of favorable alleles at QTL by crossing individual males of the resource strain to standard strain females, and intensively analyzing alternative homologous chromosomes of the resource strain, with respect to one another and with respect to the standard strain chromosomes, either over the total F-2 population or over selected extreme phenotypes as described above. Any useful identified alleles (or marker-linked chromosome segments) could then be introgressed into the standard lines using the markers defining the segment to monitor and assist in the introgression procedure.

Marker-based selection in segregating populations

1 Biometric considerations. In species, such as dairy cattle, where very large progeny groups can be obtained from individual males, genetic markers or well marked chromosome segments can also assist in within-population selection. The basic procedure is based on the marker approach to genome evaluation described above, except that the analysis involves a male from the standard population itself, and is aimed at evaluating alternative well marked homologous chromosome segments of the individual male with respect to one another in the F-1 offspring of that male. Selection would then be carried out among future offspring of such marker evaluated males on the basis of the estimated breeding value of the offspring, as indicated by the marked chromosome segments inherited from the evaluated male parent. The procedure proposed here differs from that of Soller (1978) primarily in that it involves the use of multiple rather than single marker alleles to identify chromosomal segments.

Although the contribution in this manner of any single marked segment to genetic progress would be small, the joint contribution of a large number of such segments can be appreciable; also, the procedure becomes more effective with time. At the first stage of its implementation, genetic gains depend on information obtained from the direct progeny of the male being evaluated. As the program proceeds, however, information on a particular marked chromosome segment, with respect to alternative marked segments segregating in the population, is accumulated across generations and from a number of tested males in a particular generation. Eventually, sufficient information on the most useful marked segments should become available as to enable them to be treated as "known loci", markedly increasing the possibilities of genetic improvement (Soller 1978). Thus, the marker-segment procedure contains an internal bootstrap by means of which its effectiveness increases with time. The biometric equations defining expected genetic progress under such a selection program, based on Soller (1978), are given in Table 3.

2 Decay of linkage between QTL and marker allele in evaluated chromosomal segments. Given a sire heterozygous at a QTL and at a nearby marker locus, and assuming that the particular chromosome carrying the favorable QTL and its linked marker allele have been identified by an appropriate evaluation procedure, the proportion, T_m, among the offspring, of like-marked chromosomes derived from the parent, that will retain the original linkage relationship of marker allele and useful QTL allele, can be obtained as follows:

For first generation offspring, since the sire is heterozygous at both marker and QTL, all chromosomes can undergo recombination, so that $T_1 = (1 - r)$, where r is the proportion of recombination between marker and QTL. In all subsequent generations the proportion of chromosomes retaining the original favorable linkage relationship will be given by the recurrence relation

$$T_{t+1} = (1 - rq)\, T_t + r\,(1 - q)\,(1 - T_t) = T_t\,(1 - r) + r\,(1 - q)$$

where,

- q is the population frequency of the unfavorable allele at the QTL,
- rq is the proportion of marked chromosomes that lose the original linkage configuration each generation as a result of recombination with a pairing partner that carries the unfavorable quantitative allele,
- $r\,(1 - q)$ is the proportion of marked chromosomes each generation, having lost the original linkage relationship as a result of recombination, that regain it as a result of reverse recombination involving a pairing partner carrying the favorable quantitative allele.

Along similar lines, for a marker bracket of width r, it can readily be shown that $T_1 = (1 - \frac{1}{4} r^2)$, and

$$T_{t+1} \geqq (1 - \tfrac{1}{4} r^2)\, T_t + \tfrac{1}{4} r^2\,(1 - q)\,(1 - T_t).$$

Table 4 shows loss of favorable linkage relationship for the case $q = 0.5$, and for segments distinguished by a single marker allele, or by a bracket. It is clear that when r is small ($\leqq 0.05$), or for a bracket of size 20–40 cM, loss of linkage relationship is slow. This should enable information on the quantitative value of such a bracket to be accumulated, and effective selection carried out, over a period of at least five generations (or more, for a narrow bracket). One would expect that by this time frequencies for the favorable allele would be high (Soller 1978). At the end of the selection period, some heterogeneity may remain in the marker bracket at a population level as a result of crossing over. This may be desirable as it will allow a certain degree of genetic variation at the chromosomal segment to be retained.

Table 3. Biometric relations defining the increase in breeding value of young sires chosen on the basis of marker-assisted selection or "known locus" selection with respect to quantitative trait loci (QTL)[a]

Definitions and symbols

D = observed phenotypic difference between daugther groups receiving alternative marker alleles or marker brackets from the evaluated sire

σ^2 = phenotypic variance of milk production

n = number of daughters receiving a particular marker allele or bracket from the evaluated sire

p, q = frequency of alternate alleles at the QTL linked to the evaluated marker or enclosed within the marker bracket

d, h = main effect and dominance effect at the QTL, as defined in Table 2

r = proportion of recombination between marker locus and QTL

k = number of evaluated QTL

k′ = number of "known loci"

B = breeding value of dam of young sire

i = intensity of selection

Marker assisted selection

$\frac{2\sigma^2}{n}$ = error variance of D

σ^2_{AD} = genetic variance of D, $\sigma^2_{AD} = 2pq(1-2r)2d^2$ for a sire heterozygous at a single marker, and $2pqd^2$ for a sire heterozygous at both ends of a marker bracket

$\sigma^2_D = \sigma^2_{AD} + \frac{2\sigma}{n}$ = phenotypic variance of D

$b_D = \sigma^2_{AD}/\sigma^2_D$ = regression of breeding value for D on observed value of D

$\frac{1}{2}D$ = mean deviation from overall daughter mean of the evaluated sire, of those daugthers receiving a particular marker or marker bracket

$M = b_D(\frac{1}{2}D)$ = expected breeding value associated with a particular marker or marker bracket

$I = \sum_{i=1}^{k} M_i + B$ = index value of young sire based on breeding value of his dam and expected breeding value of k evaluated markers or marker brackets

$\sigma^2_I = \frac{1}{4}\sum_{i=1}^{k} b^2_{D_i}\sigma^2_{D_i} + \sigma^2_B$ = variance of index values

$S = i(\sigma_I - \sigma_B)$ = increase in expected breeding value of young sires as a result of marker assisted selection

Known locus selection

$\alpha = d + (q-p)h$ = average effect of a gene substitution

K_i = breeding value of i^{th} "known locus"
= $2q\alpha$, $(q-p)\alpha$ and $-2p\alpha$ for the M_1M_1, M_1M_2, and M_2M_2 "known locus" genotypes, respectively.

$I = \sum_{i=1}^{k'} K_i + B$ = index value of young sire based on breeding value of his dam, and expected breeding value for k′ "known loci"

$\sigma^2\left(\sum_{i=1}^{k'} K_i\right) = 2\sum_{i=1}^{k'} p_iq_id_i^2$ = variance of total of breeding values for k′ "known loci"

$\sigma^2_I = \sigma^2\left(\sum_{i=1}^{k'} K_i\right) + \sigma^2_B$ = variance of index values of young sires

$S = i(\sigma_I - \sigma_B)$ = increase in breeding value of "known locus" index selected young sires

[a] For details see text and also Soller 1978

Table 4. Proportion of chromosomes retaining original linkage between favorable quantitative allele and single marker or marker bracket by generation and proportion of recombination[a]

Generation	Proportion of recombination				
	Single marker			Marker bracket	
	0.05	0.10	0.20	0.20	0.40
0	1.00	1.00	1.00	1.00	1.00
1	0.95	0.90	0.80	0.99	0.96
2	0.93	0.86	0.74	0.99	0.94
3	0.91	0.82	0.69	0.98	0.92
4	0.89	0.79	0.65	0.98	0.91
5	0.87	0.76	0.62	0.97	0.89
6	0.85	0.74	0.60	0.97	0.88
7	0.83	0.71	0.58	0.96	0.86
8	0.81	0.69	0.56	0.96	0.85
9	0.80	0.67	0.55	0.95	0.83
10	0.78	0.65	0.54	0.95	0.82

[a] See text for details

3 Numerical example: sire selection in dairy cattle. These considerations can now be applied to the case of young sire selection in dairy cattle. Assume a phenotypic standard deviation of milk production of 1,000 kg and 40 QTL affecting milk production, each with d = 100 kg, h = 0, p = q = 0.5 (see Table 3 for explanation of symbols). This will give a total genetic variance of 200,000, corresponding to the observed heritability of 0.2 for milk production in dairy cattle. It will be further assumed that k = 30 of these QTL are included in marked segments that can be treated in the manner described above, that 500 daughters are initially classified in each marker-genotype, and that intensity of selection of young sires, i, is equal to 2.0. It should be noted that for a marker segment analysis, it will be possible to unequivocally allocate almost all daughters to one or the other marker genotype, so that no more than 1,000–2,000 daughters would be required to give 500 daughters per marker-genotype group.

Variance of breeding value of the young sire as estimated from the breeding value of his dam, σ^2_B, will equal, say, 10,000 in both cases (Soller 1978). On these assumptions additional genetic progress in the offspring of the young sires, as a result of prior selection on the basis of marked segments will be 43 or 76 kg/generation for single marker selection (r = 0.1) or marker-bracket selection, respectively. It should be noted that existing selection procedures along the dam-to-sire path can

be expected to provide 85–130 kg of progress per generation (Owen 1975).

As information accumulates, b_D will increase. When the equivalent of 1,000 offspring are available to evaluate a given marker segment, incremental gains due to single marker or marker bracket selection will be 58 or 90 kg/generation.

If the 30 marker-segments followed eventually yield $k' = 20$ "known loci", incremental progress due to marker-bracket selection will be 231 kg per generation, and marker assisted selection would be the major route for genetic improvement in the population.

Discussion and conclusions

The expressions and numerical examples presented above suggest that given a sufficient number of useful polymorphisms (i.e., $0.1 < P < 0.9$), problems of varietal identification, protection of breeder's rights and parentage identification can receive adequate and general solutions.

With respect to marker applications to genetic improvement programs, two groups of agricultural species can be distinguished. The first group comprises selfers with small genome sizes (~ 10 Morgan). This group includes many important cultivated plants (e.g. tomato, barley, and also corn to the extent that corn improvement is based on crossing inbred lines). In this group marker assisted genetic analyses can be carried out using efficient designs, and relatively few polymorphism need to be scored to cover the genome. For these species genetic markers can make a decisive contribution to breeding practice. The second group comprises outcrossers, with large genome sizes (~ 30 Morgan). This group includes the major agricultural livestock and poultry species. For these species marker assisted breeding methods are not of general applicability, but can make significant contributions in particular instances.

Selfers, small genome size

For such species the methodologies reviewed here allow both marker-assisted genetic analysis of QTL differences between two lines or varieties and marker-assisted introgression or selection. Combinations of these activities allow a variety of breeding applications to be carried out, including:

1. Genetic analysis of a cross between a commercial variety and a resource strain, followed by introgression of useful QTL alleles from resource strain to a commercial variety.

2. Genetic analysis of a cross between two commercial varieties, followed by selfing and selection of recombinant inbreds carrying a maximum number of favorable QTL alleles from both parent lines.

3. Genetic analysis of a cross between two inbred parents of a commercial hybrid, and between the parent inbreds and resource inbreds, with subsequent introgression of dominant or overdominant QTL alleles from resource inbreds to one or another of the alternative parental lines.

These applications seem straightforward and cover the major classes of breeding programs in selfers. Cost estimates of these applications, based on restriction fragment length polymorphisms as genetic markers, are given in a companion paper (Beckmann and Soller 1983) and do not seem excessive. Thus, the availability of a large number of marker polymorphisms in selfer cultivars should allow a major increase in the effectiveness of breeding programs in these species.

Segregating populations, large genome size

The application of genetic markers to segregating populations (outcrossers) with large genome size is more problematic. Nevertheless, two applications can be envisaged. The first, early selection of young males in dairy cattle according to the chromosome evaluation of their proven sires, has been discussed in detail in the preceding sections. Cost evaluations based on restriction fragment length polymorphisms (Beckmann and Soller 1983) indicate that this application may be economically justified. The second potential application involves a genetic analysis of a cross between a resource line characterized by high quantitative value for a low heritability trait of major economic importance, and a standard commercial line. An example of this is the very high incidence of multiple births in Finnsheep, as compared to the usual mutton breeds. In such cases, large differences in gene frequency at the relevant QTL are expected between commercial strain and resource strain enabling relatively low cost analysis to be carried out. In such cases, too, where only a single trait is of interest, a considerable reduction in number of individuals scored for marker polymorphisms can be had by scoring genetic markers only in those individuals showing high or low expression of the quantitative trait, as described above. Once resource strain markers associated with the relevant QTL are identified, they can be used to introgress the desired trait into the commercial strain.

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

References

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

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
Geldermann H (1975) Investigations on inheritance of quantitative characters in animals by gene markers. 1. 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
Kluge R, Geldermann H (1982) Effects of marked chromosome sections on quantitative traits in the mouse. Theor Appl Genet 62:1–4
Law CN (1966) The location of genetic factors affecting a quantitative character in wheat. Genetics 53:487–498
Mather KM, Jinks JL (1971) Biometrical genetics. Chapman and Hall, London
McMillan I, Robertson A (1974) The power of methods for the detection of major genes affecting quantitative characters. Heredity 32:349–356
Owen JB (1975) Selection of dairy bulls on half-sister records. Anim Prod 20:1–10
Patterson FL, Schafer JF, Caldwell RM (1968) Effect of selected linkage blocks on yield and yield components in wheat. In: Finlay KW, Shepard KW (eds) Third Int Wheat Genet Symp. Aust Acad Sci, Butterworths, Canberra, pp 88–92
Silver H (1978) Paternity testing. In: Silver H (ed) Seminar presented by the Committee on Technical Workshops of the Am Ass Blood Banks. New Orleans, Louisiana
Simmonds NW (1979) Principles of crop improvement. Longman Group, London
Soller M (1978) The use of loci associated with quantitative traits in dairy cattle improvement. Anim Prod 27:133–139
Soller M, Brody T, Genizi A (1979) The expected distribution of marker-linked quantitative effects in crosses between inbred lines. Heredity 43:179–190
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, Genizi A (1978) The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34:47–55
Soller M, Plotkin-Hazan J (1977) The use of marker alleles for the introgression of linked quantitative alleles. Theor Appl Genet 51:133–137
Solomon E, Bodmer WF (1979) Evolution of sickle variant gene. Lancet I: 8122: 923
Spickett SG, Thoday JM (1966) Regular responses to selection. 3. Interaction between located polygenes. Genet Res 7:96–121
Stuber CW, Goodman MM, 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, Rick CM (1980) Isozymic gene linkage map of the tomato: applications in genetics and breeding. Theor Appl Genet 57:161–170
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
Weller J (1983) Genetic analysis of quantitative traits in Lycopersicon (tomato) by means of genetic markers, including a comparison of an improved cultivar and a wild variety (in Hebrew, English summary). PhD Thesis, The Hebrew University of Jerusalem, Israel
Zhuschenko AM, Samovol AP, Korol AB, Andryushchenko AB (1979) Linkage between loci of quantitative characters and marker loci. 2. Influence of three tomato chromosomes on variability of five quantitative characters in backcross progenies. Translated from Genetika 15:672–683