Theor Appl Genet (1992) 84: 704–708

Some practical considerations for using RFLP markers to aid in selection during inbreeding of maize

B. E. Zehr [1,*], J. W. Dudley [1], and J. Chojecki [2]

[1] Department of Agronomy, University of Illinois, Urbana, IL 61801, USA
[2] ICI Seeds, Jealott's Hill Research Station, Bracknell, Berkshire, RG12, 6EY, UK

Received December 4, 1991; Accepted January 13, 1992
Communicated by A. L. Kahler

Summary. If molecular markers are to be routinely used in maize (*Zea mays* L.) breeding for selection of quantitative trait loci (QTL), then consistent marker-trait associations across breeding populations are needed, as are efficient methods for weighting information from different markers. Given 15 restriction fragment length polymorphism (RFLP) markers associated with grain yield in testcrosses of 220 [BS11(FR)C7 × FRMo17] F_2 individuals to FRB73, separate weighting schemes were attempted in order to maximize the frequency of favorable marker genotypes associated with increased grain yield in selected F_2 individuals and F_2:S_4 lines. The following principles were apparent: (1) Differential weighting among markers, in addition to weighting individual marker genotypes on the basis of associated mean effects, should be emphasized when using markers to select in breeding populations. This is due to limited population sizes that can readily be handled. (2) Relatively few markers may need to be used to screen segregating populations (e.g., F_2) of limited size for loci affecting complex traits, such as combining ability for grain yield, assuming prior knowledge of marker-QTL associations. Markers given greatest weight (largest estimates of associated effects) will determine most selections. (3) When marker-based selection is among individuals at higher levels of inbreeding (e.g., S_4) within selected families, more markers need to be used in screening because those associated with relatively small effects have an increased chance of affecting selection.

These results suggest a qualitative approach for utilizing RFLP markers to aid in selection of complex traits in commercial hybrid maize breeding programs. Commercial research programs produce thousands of crosses each year aimed at inbred line development. Discovery of molecular markers with consistent QTL associations across breeding populations and close QTL linkages would allow for rapid screening of new F_2 populations at a few key markers. Early elimination of individuals with undesirable genotypes would reduce the extent of hybrid performance testing necessary during later stages of inbreeding.

Key words: *Zea mays* L. – RFLPs – Marker-based selection – RFLP-QTL associations – Favorable marker genotypes

Introduction

Molecular genetic markers have been suggested as a means of indirect selection for traits which have low heritability, are difficult or expensive to measure, or require wide crossing for incorporation (Nienhuis et al. 1987; Soller and Beckmann 1983; Tanksley et al. 1989). Because these markers are easily scored and have complete heritability, their use in screening plant breeding populations may help improve selection efficiency and reduce costs (Tanksley et al. 1981). A number of studies have shown isozyme and restriction fragment length polymorphism (RFLP) markers to be associated with quantitative trait loci (QTL) controlling characters of agronomic importance in plants (see Stuber 1991 for a review). The efficiency of marker-aided selection depends on tight marker-QTL linkage and the development of indices incorporating marker information.

Stuber and Edwards (1986) used genotypes at 15 isozyme loci associated with grain yield to select among

* To whom correspondence should be addressed
Present address: Department of Agronomy, Purdue University, West Lafayette, IN 47907-1150, USA

F_2 plants from two populations of maize (*Zea mays* L.). Homozygous marker genotypes in F_2 individuals were given positive or negative breeding values (one-half the difference between trait means of homozygous genotypes) depending on associated grain yield effect, and those heterozygous were given values of zero. Composite scores were used to rank F_2 individuals. F_2 progeny from divergent marker and phenotypic (mass) selections were evaluated at three locations. In both populations, marker-based selections were as effective as phenotypic selections for increasing grain yield.

Lande and Thompson (1990) presented a theory for maximizing the rate of improvement in quantitative characters by using marker-assisted selection. Selection indices were derived that combined information on marker loci with phenotypic data. For selection among individuals, the optimum index (I) was:

$$I = b_z z + b_m m$$

where z is the phenotypic value of an individual for a trait and m, the net molecular marker score, is the sum of the additive effects for all marker loci associated with the trait. Weight coefficients, b_z and b_m, are dependent on the heritability of the trait (h^2) and the proportion of additive genetic variance explained by marker loci. Marker scores are given more emphasis as h^2 decreases or as the proportion of additive genetic variance explained by marker variation increases. The authors conclude that information provided by molecular genetic markers can substantially increase efficiency for artificial selection.

We examined methods of weighting molecular marker information to enhance the probability of selecting individuals with favorable marker genotypes. The data set used as an example consisted of RFLP genotypes in a maize population of 220 F_2 individuals. Marker genotypes of selected F_2 plants and of plants within heterogeneous S_4 lines derived from F_2 selections were determined. Results and their implications for practical application in commercial hybrid maize breeding programs are discussed.

Materials and methods

Each of the 220 [BS11(FR)C7 × FRMo17] F_2 (S_0) individuals used for QTL analysis by Zehr et al. (1992) was selfed. Within each F_2:S_1 family 5–7 plants were selfed, and equal quantities of seed were bulked to produce the F_2:S_2. This procedure was repeated to produce 220 heterogeneous F_2:S_4 families.

Selection among F_2 individuals was carried out using two approaches. Twenty (9% selection intensity) were selected solely on the basis of highest F_2 × FRB73 testcross performance for grain yield; 20 were also selected solely on the basis of their genotypes at 15 RFLP markers associated with F_2 × FRB73 grain yield at $\alpha \leq 0.05$ (Zehr et al. 1992). Five S_4 individuals per F_2:S_4 family were then selected on the basis of RFLP genotypes of 30 S_4 individuals per family in 16 of the 20 families derived from F_2 RFLP selections. Laboratory methods for RFLP determinations were as indicated by Zehr et al. (1992).

Rankings of F_2 individuals based on RFLP markers were derived using the formula:

$$\mathbf{R}_j = \mathbf{M} \times \mathbf{q}_j$$

where **M** is a matrix weighting F_2 RFLP genotypes within grain yield associated markers, **q** is a column vector weighting the relative importance among markers, and **R** is an index used to determine F_2 rankings. **M** is a 220×15 matrix accounting for mean genotypic effects associated with each marker; each matrix row represents FRB73 testcross grain yield means for RFLP genotypes of an F_2 individual, and each column represents an RFLP marker associated with F_2 × FRB73 grain yield. Genotypic means were calculated as the average grain yield for F_2 individuals with the same RFLP genotype at a marker. The pooled genotypic classes (e.g., AA, AK, KK where K = all RFLP bands at the marker other than A) chosen by Zehr et al. (1992) to represent each grain yield associated RFLP marker were used for calculations. Column vector weights were based on either probability values for associations between RFLP markers and grain yield, or absolute t-values for linear effects of grain yield associated with each marker (Table 1). Whole numbers from 1 to 4 were used to weight markers based on ranges of probability value (see Table 1). Absolute t-values were used to account for differences in the precision of linear effect estimates.

Rankings of F_2 individuals based on their RFLP genotypes were derived using the four column vectors for comparison:

$$\mathbf{R}_1 = \mathbf{M} \times \mathbf{q}_1$$
$$\mathbf{R}_2 = \mathbf{M} \times \mathbf{q}_2$$
$$\mathbf{R}_3 = \mathbf{M} \times \mathbf{q}_3$$
$$\mathbf{R}_4 = \mathbf{M} \times \mathbf{q}_4$$

The relative efficiency of the column vectors was determined by comparing frequencies of 'favorable' RFLP genotypes in selected F_2 individuals (top 20 rankings). Favorable RFLP genotypes in the F_2 were defined as being either homozygous or heterozygous for RFLP marker bands associated with increased grain yield. RFLP bands considered favorable for selection at the 15 grain yield associated markers were as determined by Zehr et al.

Table 1. Column vectors (**q**) for weighting the relative importance of RFLP markers associated with grain yield at $\alpha \leq 0.05$

RFLP marker	Probability > F	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_3$	$\lvert t \rvert = \mathbf{q}_4$[a]
UMC16	0.0001	1	3	4	4.06
NPI234	0.0003	1	3	3	3.65
UMC92	0.0013	1	2	2	3.27
UMC96	0.0063	1	2	2	2.77
UMC38	0.0066	1	2	2	2.76
UMC21	0.0068	1	2	2	2.73
UMC27	0.0116	1	1	1	2.55
NPI391	0.0117	1	1	1	2.55
UMC44-B	0.0127	1	1	1	2.52
UMC81	0.0147	1	1	1	2.46
UMC104	0.0181	1	1	1	2.39
NPI297	0.0182	1	1	1	2.38
NPI114	0.0196	1	1	1	2.36
NPI238	0.0204	1	1	1	2.33
BNL15.07	0.0225	1	1	1	2.30

[a] Absolute t-values for estimated grain yield effects associated with each marker (see Zehr et al. 1992); t-values were used to account for differences in the relative precision of estimates

(1992). Frequencies of favorable RFLP genotypes (di-band combinations) were determined by a comparison of the number of favorable genotypes with the total number of genotypes over all selected F_2 individuals and markers. For example, 20 selected F_2 individuals $\times 15$ RFLP markers = 300 RFLP genotypes. If 250 contain at least one favorable band, the frequency of favorable RFLP genotypes over all markers would be 0.833.

RFLP selections of individuals within S_4 families were based on the among-locus weights given in $\mathbf{q}_3$ (Table 1). S_4 genotypes homozygous for favorable marker bands were given the value of the weight assigned to its marker, heterozygous favorable genotypes were given one-half values, and homozygous unfavorable genotypes were scored as 0. The sum of values across markers was used to rank S_4 individuals within each F_2:S_4 family.

Results and discussion

In attempting to maximize the frequency of favorable RFLP marker genotypes in F_2 and S_4 selections, the following principles were apparent:

Differential weighting among markers, in addition to weighting individual marker genotypes on the basis of associated mean effects, should be emphasized when using markers to select in populations of limited size

In selected F_2 individuals, there was a trade-off between maximizing the frequency of favorable genotypes over all markers under selection and maximizing this frequency at markers considered to be of greatest importance (Table 2). Vector $\mathbf{q}_1$ gave equal inter-marker weight, thus $\mathbf{R}_1$ rankings were based solely on associated mean effects of genotypes at each marker. $\mathbf{R}_1$ selections maximized the frequency of favorable genotypes over all makers, but not at UMC16 and NPI234, the 2 markers with greatest probability of grain-yield association and largest estimated effects as measured by absolute t-values (Table 1). At UMC16 and NPI234, favorable genotypes were fixed only when a vector ($\mathbf{q}_3$) giving these markers preferential weight was applied.

If data from extremely large populations were available for selection, equal marker weights would be able to select individuals with favorable genotypes at all markers. The population size (N) needed to find one individual with favorable genotypes at all markers, with β probability is:

$$N = \log(1-\beta)/\log\left[1-\prod_{i=1}^{n}(p_i^2+2p_i q_i)\right]$$

where p and q are the frequencies of favorable and unfavorable bands at the ith marker ($p+q=1$), and n is the number of markers under selection (derived from Mather 1951). When frequencies of favorable bands at each of the 15 grain yield associated RFLP markers in the unselected F_2 population are used (data not shown), over 7,000 individuals would need to be screened to have a 95% probability of finding 1 with favorable genotypes at all 15 markers. The resources required to score such a large population for RFLP marker genotypes would be prohibitive.

Table 2. Frequency of favorable RFLP genotypes in selected F_2 individuals, over all markers and at the 2 markers (UMC16 and NPI234) with greatest estimate of associated grain yield effect

Index ranking	Vector[a]	Frequency of favorable RFLP genotypes	
		Over all markers	UMC16 and NPI234
$\mathbf{R}_1$	$\mathbf{q}_1$	0.877	0.900
$\mathbf{R}_2$	$\mathbf{q}_2$	0.857	0.925
$\mathbf{R}_3$	$\mathbf{q}_3$	0.867	1.000
$\mathbf{R}_4$	$\mathbf{q}_4$	0.877	0.925
Grain yield selection[b]		0.780	0.725

[a] Vector of relative weights among RFLP markers. $\mathbf{q}_1$ gave equal weight to all markers; $\mathbf{q}_2$, $\mathbf{q}_3$, and $\mathbf{q}_4$ gave differential weighting among markers (see Table 1)

[b] Selection based on grain yield means of $F_2 \times$ FRB73 progeny

The importance of differential weighting depends on the accuracy at which QTL effects can be estimated. An estimated QTL effect is a function of the magnitude of the true effect and the recombination distance between the marker and QTL (Edwards et al. 1987). If recombination distances are unknown, then a marker associated with a large estimated effect has either a moderate-to-strong linkage, with QTL having a true large effect, or a tight linkage, with QTL having intermediate effect. In either case, increased emphasis on large estimated effects in selection is justified.

Relatively few markers may need to be used to screen segregating populations (e.g., F_2) for loci affecting complex traits, such as combining ability for grain yield, assuming prior knowledge of marker-QTL associations

While the average frequency of favorable bands over all 15 RFLP markers increased as a result of both F_2 and S_4 selection, the majority of change came at the 6 markers given greatest weight in vector $\mathbf{q}_3$ (Table 3). At these 6 markers the average frequency of favorable bands increased from 0.34 to 0.62 due to F_2 selection, and to 0.73 after S_4 selection. Changes at the 9 markers given lowest weight were less substantial: 0.71 to 0.77 resulting from F_2 selection, and 0.79 in the S_4. The weights used in $\mathbf{q}_3$ were necessary to maximize favorable genotypes at the 2 most important markers (Table 2); therefore, selection at a limited number of RFLP markers in this population would have been as effective as selection at all associated markers for increasing frequencies of favorable bands over all markers.

The 6 markers with highest weight in $\mathbf{q}_3$ determined almost all F_2 selections. When $\mathbf{R}_3$ selections were compared with those resulting from the use of only the

Table 3. Average frequency of favorable RFLP marker bands at different stages of phenotypic and RFLP-based selection using marker weights in vector $\mathbf{q}_3$

	Average frequency[c]	Vector $\mathbf{q}_3$	
		Six markers (high weight)[d]	Nine markers (low weight)[e]
F_2 population[a]	0.56	0.34	0.71
$\mathbf{R}_3$ – among F_2	0.71	0.62	0.77
– within S_4	0.77	0.73	0.79
Grain yield selection[b]			
– among F_2	0.65	0.47	0.77

[a] Unselected population of 220 F_2 individuals
[b] Selection based on grain yield means of $F_2 \times$ FRB73 progeny
[c] Frequency of favorable RFLP marker bands in selected individuals, averaged over all markers used for selection
[d] Frequency of favorable RFLP marker bands in selected individuals, averaged over the 6 markers given greatest weight in $\mathbf{q}_3$
[e] Frequency of favorable RFLP marker bands in selected individuals, averaged over the 9 markers given least weight in $\mathbf{q}_3$

Table 4. Number of F_2 individuals in common among top 20 (above and right of diagonal) and top 10 (below and left of diagonal) selections

	$\mathbf{R}_3$[a]	Vector $\mathbf{q}_3$ (top 6)[b]	Grain yield selection[c]
$\mathbf{R}_3$[a]	–	19	6
Vector $\mathbf{q}_3$ (top 6)[b]	9	–	6
Grain yield selection[c]	1	1	–

[a] Selection based on all 15 RFLP markers using marker weights in vector $\mathbf{q}_3$
[b] Selection based only on the 6 RFLP markers given greatest weight in $\mathbf{q}_3$
[c] Selection based on grain yield means of $F_2 \times$ FRB73 progeny

6 markers given greatest weight in $\mathbf{q}_3$, 19 of the top 20 selections and 9 of the top 10 were the same (Table 4). Thus, only a limited number of markers were necessary to select most of the same F_2 individuals. This again relates to the number of F_2 individuals from which to choose. When selecting in F_2 populations of limited size, the markers given greater emphasis (largest estimates of associated effects) will have disproportionate influence and, assuming prior knowledge of marker-QTL associations, only a relatively few may need to be used for initial screening.

When marker-based selection is among individuals at higher levels of inbreeding (e.g., S_4) within selected families, an increased number of markers need to be used for screening

RFLP markers given lower weight had greater effect on selection within S_4 families than among F_2 individuals. This can be seen by contrasting S_4 selections based on $\mathbf{q}_3$ weights at all 15 markers with those based solely on the 6 markers given greatest weight in $\mathbf{q}_3$ (Table 5). Five S_4 individuals were to be selected within each F_2:S_4 family. Most families had more than 5 individuals with a top five ranking when selected on the basis of all markers. When selection was based only on the 6 markers with greatest weight, even more individuals had a top five ranking. Thus, at a higher level of inbreeding, the markers given lowest weight had a greater influence in discriminating among individuals. When genotypes at markers given higher weight are fixed, those with lower weights will determine rankings.

Table 5. Comparison of S_4 selections based on all RFLP markers using $\mathbf{q}_3$ weights with S_4 selections based only on the 6 markers given greatest weight in $\mathbf{q}_3$

S_4 family[a]	Number of S_4 individuals with a top 5 ranking[b]	
	$\mathbf{q}_3$ – all 15 markers	$\mathbf{q}_3$ – 6 markers (high weight)
1	7[c]	7
2	5	6
3	10	15
4	8[c]	8
5	6	17
6	14	20
7	7	12
8	8	19
9	8	17
10	17	19
11	6	8
12	10[c]	10
13	5[c]	5
14	5	6
15	7[c]	7
16	8	9

[a] RFLP genotypes of S_4 individuals were available only for 16 of the 20 F_2:S_4 families derived from selections in $\mathbf{R}_3$
[b] Five S_4 individuals were to be selected in each F_2:S_4 family; for most families, more than 5 individuals had a top 5 ranking. Selection of S_4 individuals based on 6 markers was less discriminating than selection based on all 15 markers
[c] The 9 markers given least weight in $\mathbf{q}_3$ showed no segregation of favorable marker bands among selected S_4 individuals

Few conclusions can be drawn by comparing RFLP-based and phenotypic ($F_2 \times$ FRB73 grain yield mean) selections. Frequencies of favorable marker genotypes (Table 2) and favorable marker bands (Table 3) in selected F_2 individuals were greater for RFLP-based selections. The only difference in average frequency of favorable bands between RFLP ($\mathbf{R}_3$) and phenotypic selection in the F_2 was at the top 6 markers in $\mathbf{q}_3$, which may indicate the greater efficiency of RFLP-based selection for critical QTL. However, QTL having substantial cumulative effect on combining ability may not have been accounted for by these markers. Further testing of $S_4 \times$ FRB73 progeny derived from F_2 selections using both phenotyp-

ic and RFLP approaches will be required to determine their relative effectiveness. If greater hybrid performance results from RFLP-based selection, it may be due to the identification of, and improved selection at, specific RFLP markers associated with QTL having large effects on combining ability for grain yield.

Lande and Thompson (1990) proposed the use of marker-assisted selection (indices combining phenotypic and molecular marker data) to maximize genetic gain. Marker-QTL associations would be evaluated one generation after each hybridization event prior to selection; and every evaluation would require scoring genotypes at a few hundred markers in a few hundred to a few thousand individuals. While this approach could be applied to methods of parental line development in hybrid maize breeding, the limitation for its use would be an ability to repeatedly handle the scope of experiments required. From a practical standpoint, analysis of marker-QTL associations prior to and during every pedigree selection procedure is probably unrealistic for most commercial maize breeding programs. Commercial research programs produce thousands of crosses each year aimed at inbred line development. Given the number of markers and population sizes suggested by Lande and Thompson, it would be difficult for a molecular screening laboratory to service all such breeding materials. Also, obtaining hybrid data for QTL analysis in each pedigree procedure would be difficult because individuals in each F_2 population would have to be crossed to a number of inbred testers and grown in field trials over years and locations. Due to resource constraints, only limited population sizes (such as that used for this study) can readily be handled in most breeding programs.

Our results suggest a qualitative approach for utilizing RFLP markers to aid in the selection of complex traits. For routine use in hybrid maize breeding, marker-QTL associations need to have predictive value by showing some degree of consistency across potential populations within a heterotic grouping and across genetic backgrounds for groups of inbred line testers. Databases need to be developed that include a knowledge of markers closely linked with specific QTL, differences in linkage phase among parents, and marker band identity by descent versus identity by molecular weight. This would allow for the initial screening of new breeding populations with markers chosen a priori on the basis of known QTL associations within a specific germplasm group. The goal could be early elimination of individuals with undesirable genotypes at key loci in order to reduce the extent of experimental hybrid performance testing necessary during later stages of selection.

Acknowledgements. All laboratory procedures were performed by ICI-Garst. The assistance of Drs. Ian Bridges, Keith Rufener II, M. Saghai-Maroof, Ron Mowers, Ken Russell and David Foster at the Garst Seed Co., Dr. Steve Barnes and Jacqueline Brannigan at ICI Seeds, and Dr. Samuel Carmer at the University of Illinois is gratefully acknowledged. This research was supported by a grant from ICI Seeds, UK, and Garst Seed Company, USA.

References

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

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Mather K (1951) The measurements of linkage in heredity, 2nd edn. John Wiley & Sons, New York

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 polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:25–33

Stuber CW (1992) Biochemical/molecular markers in plant breeding. In: Janick J (ed) Plant breeding reviews, vol. 9, John Wiley & Son, Inc., New York, pp 37–57

Stuber CW, Edwards MD (1986) Genotypic selection for improvement of quantitative traits in corn using molecular marker loci. In: Dolores Wilkinson (ed) Proc 41st Annu Corn Sorghum Ind Res Conf. American Seed Trade Assoc, Washington, D.C., pp 70–83

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, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/technology 7:257–264

Zehr BE, Dudley JW, Chojecki J, Saghai Maroof MA, Mowers RP (1992) Use of RFLP markers to search for alleles in a maize population for improvement of an elite hybrid. Theor Appl Genet 83:903–911