Theor Appl Genet (1990) 80:273–280

THEORETICAL AND APPLIED GENETICS

Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize*

E. B. Godshalk, M. Lee***, and K. R. Lamkey****, *******

USDA-ARS and Department of Agronomy, Iowa State University, Ames, IA 50011, USA

Received August 28, 1989; Accepted March 6, 1990
Communicated by A. L. Kahler

Summary. Isozymes and restriction fragment length polymorphisms (RFLPs) have been proposed for use in varietal identification and selection for agronomic traits. Although the use of isozymes for these purposes has been well documented, evaluation of the efficacy of RFLP technology as applied to crop improvement is far from complete. This investigation was conducted to study the relationship between RFLP-derived genotypes and heterotic patterns of a group of maize (*Zea mays* L.) inbred lines. A total of 22 inbreds was crossed to four testers (B73, B76, Mo17, and Va26) in combinations that minimized crossing within heterotic groups. Forty-seven single-cross progeny were subsequently evaluated for several agronomic traits (including grain yield and moisture, ear height, and root lodging) over 2–4 consecutive years at two to four Iowa locations in a randomized complete-block design. The inbred lines were subjected to RFLP analysis, which involved 47 genomic clones and the restriction enzymes EcoRI and HindIII. Hybrid RFLP patterns were predicted from their inbred parents. Modified Roger's distances were computed to estimate genetic distance among the inbred lines. Principal component analysis facilitated ascertainment of relative dispersion of the inbreds based on the frequency of variants at specific RFLP loci. Evident associations of variants with genes affecting agronomic traits were identified by principal component regression analysis, in which adjusted hybrid means were regressed on the matrix of hybrid variants frequencies. The hybrid means were adjusted by removing environmental effects, using residuals as dependent variables in the regression analysis. Results from this study suggest that RFLP analysis may be of value in allocating maize inbreds to heterotic groups, but no relationship between RFLP-based genetic distance and hybrid performance was apparent. Principal component regression identified variants potentially linked to genes that control specific agronomic traits.

Key words: Modified Roger's distance – Principal component regression analysis – Hybrid performance – RFLPs

Introduction

A primary goal of most breeding programs is to optimize the effectiveness of selection for quantitatively inherited agronomic traits. Selection for these traits has heretofore been based on phenotype or genetic population parameter estimates. Currently, attempts are being made to integrate molecular biological technology and conventional plant breeding procedures.

Initially, research was done to determine the potential of isozyme-aided selection in maize. Efforts to evaluate the utility of isozymes as criteria for selection and prediction produced varied results. Stuber et al. (1980) reported changes in allozyme frequencies after several cycles of selection for improved maize grain yield, but several other researchers found no discernible relationship between isozyme-based genetic diversity and hybrid vigor (Hadjinov 1982; Price et al. 1986; Lamkey et al. 1987). Frei et al. (1986), on the other hand, found isozyme diversity to be

* Joint contribution: USDA-ARS and Journal Paper No. J-13590 of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA 50011, USA. Projects No. 2818 and 2778

** Present address: DeKalb-Pfizer Genetics, 3100 Sycamore Road, DeKalb, IL 60115, USA
*** Present address: Department of Agronomy, Iowa State University, Ames, IA 50011, USA
**** Present address: USDA-ARS, Department of Agronomy, Iowa State University, Ames, IA 50011, USA
***** To whom correspondence should be addressed

predictive of maize hybrid yield when evaluating lines with similar pedigrees.

The relatively small number of isozyme loci currently available has compelled usage of restriction fragment length polymorphisms (RFLPs) as genetic markers. Helentjaris et al. (1985) and Evola et al. (1986) have discovered a high frequency of polymorphic loci in maize. To date, methods of RFLP data analysis and design of experiments to evaluate the utility of RFLPs for selection purposes remain unexplored. As a preliminary evaluation of the RFLP technology, as applied to agronomic crop breeding, this study was initiated to: (1) investigate the potential for using RFLPs to assign members of a diverse group of maize inbred lines to their correct heterotic groups, and (2) identify associations between RFLPs and genes that control quantitatively inherited traits.

Materials and methods

Forty-seven single-cross hybrids of maize were produced by crossing 22 inbred lines to four testers (B73, B76, Mo17, and Va26) in combinations that minimized crossing within heterotic groups (Table 3). Lines from the Iowa Stiff Stalk Synthetic (BSSS) or the Reid Yellow Dent heterotic group were B14A, B37, B73, B76, B84, B89, H100, Va95, Va96, Va97, and Va98 (Table 1). Lines from the Lancaster Sure Crop heterotic group were H98, H112, Mo17, Pa870, Va22, and Va26. The lines B79, B88, B90, B91, De811, Oh8710, Pa91, Va99, and Va100 were considered to be of diverse origin (Henderson 1984), and were placed in one of the two heterotic groups on the basis of prior single-cross performance.

The hybrids were evaluated over a period ranging from 2 to 4 consecutive years (1984–1987) at two to four Iowa locations (Ames, Clarence, Harlan, and Martinsburg). Each hybrid was evaluated in at least 7 of the 15 year/location combinations. The experimental design was a randomized complete block with three replicates per location. Planting densities ranged from 58,000 to 68,500 plants ha^{-1}. Two-row plots were 5.4 m long, with 0.76 m between rows. Data were collected for grain yield (Mg ha^{-1}), percentage grain moisture at harvest, ear height (1–5 scale), and percentage root lodging.

An RFLP genotype was determined for each of the 26 inbreds. The RFLP analysis procedure consisted of DNA isolation (Saghai-Maroof et al. 1984) and digestion of DNA with the restriction enzymes EcoRI or HindIII. The DNA was subsequently loaded onto neutral agarose gels, electrophoresed, and transferred to nylon membranes according to the Southern blot procedure (Southern 1975). Blots were then treated with prehybridization solution to Helentjaris et al. (1985, 1986). Forty-seven clones were selected from collections of mapped clones provided by B. Burr (Brookhaven National Laboratory), T. Helentjaris (Native Plants, Inc.), D. Grant (Pioneer Hi-Bred International, Inc.), and D. Hoisington (University of Missouri Columbia). The clones were distributed over the ten chromosomes as listed in Table 2. Clones were radiolabelled by random-primer synthesis of isolated inserts (Feinberg and Vogelstein 1983) and hybridized to DNA fragments positioned on the nylon membranes (Helentjaris et al. 1986). For each clone, only data from the restriction enzyme that provided the greatest number of polymorphisms were used in subsequent analyses. This was done to enhance chances of detecting associations between polymorphisms and genes that control agronomic traits. Each unique RFLP pattern at a given clone/enzyme combination (Table 2) was considered a variant. RFLP patterns of the 47 hybrids were inferred from the RFLP patterns of their constituent inbred parents.

Table 1. Pedigrees and heterotic groups of 26 inbred lines of maize

Inbred	Pedigree	Heterotic group
B14A	Cuzco × B14[8]	RYD
B37	BSSS	RYD
B73	BSSS(HT)C5	RYD
B76	CI31A × B37[2]	RYD
B79	BS10	RYD
B84	BSSS(HT)C7	RYD
B88	BS6(RC)C2	LSC
B89	BSSS(R)C7	RYD
B90	BSCB1(R)C7	LSC
B91	BSCB1(R)C8	LSC
De811	B68 × {[B73Ht × (C103 × Mp3204)] − sel}	RYD
H98	Hy × Oh43	LSC
H100	N28 × H91	RYD
H112	(H99 × H98)H98	LSC
Mo17	C103 × CI187−2	LSC
Oh8710[b]		
Pa91	[(WF9 × Oh40B)S4] × {[38−11 × L317)38−11]S4}	LSC
Pa870	75F−5 × Oh43	LSC
Va22	Va17 × C103 Backcross	LSC
Va26	Oh43 × K155	LSC
Va95	(B73 × H84)B73	RYD
Va96	(B73 × H84)B73	RYD
Va97	(B73 × H84)	RYD
Va98	(B73 × H94)	RYD
Va99	Oh7B × Pa91	LSC
Va100	Indiana HCBS A	LSC

[a] RYD = Reid Yellow Dent; LCS = Lancaster Sure Crop
[b] Pedigree and heterotic group of this line are unknown

Statistical analysis of RFLP data consisted of estimating genetic distances between inbreds according to the modified Roger's distance (MRD) equation (Rogers 1972) as presented by Lee et al. (1989). An MRD value of one indicates no variants common to the two inbreds, whereas a value of zero represents complete variant concordance between the two inbreds. Simple correlation coefficients were calculated between MRD values and means of agronomic trait residuals that resulted from removal of environmental effects.

The variant content for each clone/enzyme combination in each hybrid was represented as a numerical value of 0, 1, or 2. These values were converted to corresponding frequency values of 0, 0.5, or 1 and were arrayed in a 47×187 matrix, in which rows corresponded to hybrids and columns represented variants. The model chosen for analysis of the data was $\mathbf{Y} = \mathbf{BX}$, where $\mathbf{Y}$ is a 47×1 vector of agronomic performance data, with each individual element being the mean performance value for one of the 47 hybrids, $\mathbf{X}$ is the aforementioned 47×187 matrix of variant frequencies, with each row being matched by hybrid to the corresponding element of the $\mathbf{Y}$ vector, and $\mathbf{B}$ is the 187×1 vector of unknowns that serve to relate variant content to hybrid performance. The model so constructed clearly takes account of only the additive (or linear) effects between variant content and

Table 2. Probe and number of variants resulting from RFLP analysis of 26 inbred lines of maize

Probe[a]	Chromosome arm[b]	Restriction enzyme	No. of variants
UMC76	1–S	EcoRI	3
UMC29	1–S	EcoRI	3
UMC11	1–S	HindIII	8
UMC86	1–L	HindIII	2
BNL8.10	1–L	EcoRI	2
BNL5.59	1–L	EcoRI	4
UMC34	2–S	HindIII	5
UMC53	2–S	EcoRI	2
UMC49	2–L	HindIII	6
UMC55	2–L	HindIII	2
UMC10	3–S	HindIII	2
UMC18	3–L	HindIII	6
UMC26	3–L	EcoRI	2
UMC39	3–L	HindIII	6
UMC19	4–S	HindIII	3
UMC31	4–S	HindIII	2
UMC52	4–L	EcoRI	3
UMC15	4–L	EcoRI	6
BNL7.65	4–L	EcoRI	8
UMC27	5–S	HindIII	2
UMC72A	5–S	HindIII	3
UMC54	5–L	EcoRI	2
BNL4.36	5–L	EcoRI	4
UMC104	5–L	EcoRI	7
BNL6.29	6–S	EcoRI	2
UMC85	6–S	HindIII	3
UMC65	6–L	EcoRI	3
UMC46	6–L	HindIII	4
UMC71	6–L	HindIII	3
UMC80	7–L	HindIII	3
UMC56	7–L	EcoRI	3
UMC116	7–L	EcoRI	5
BNL13.05	8–S	HindIII	3
BNL9.11	8–L	HindIII	8
BNL9.08	8–L	HindIII	4
NPI268	8–L	HindIII	5
UMC89	8–L	HindIII	4
UMC103	8–L	HindIII	5
BNL5.04	9–S	EcoRI	5
UMC81	9–S	HindIII	2
UMC114	9–S	HindIII	4
BNL5.09	9–L	EcoRI	3
BNL3.04	10–S	HindIII	4
UMC57	10–L	EcoRI	3
PIO10.33	10–L	HindIII	6
BNL7.49	10–L	HindIII	7
BNL10.13	10–L	EcoRI	5

[a] Probe abbreviations represent the following sources: "BNL" = Brookhaven National Laboratory; "NPI" = Native Plants Inc.; "PIO" = Pioneer Hi-Bred International, Inc.; and "UMC" = University of Missouri-Columbia

[b] The number represents a chromosome, whereas "S" refers to the short arm and "L" represents the long arm of that chromosome

hybrid performance. The model $\mathbf{Y}=\mathbf{BX}$ can be regarded as a system of linear equations in which the elements of $\mathbf{Y}$ are the dependent variables, the elements of the columns of $\mathbf{X}$ are the independent variables, and the elements of $\mathbf{B}$ are unknowns. Agronomic data were adjusted for environmental effects, where an environment was a particular year/location combination. Adjustment involved regressing agronomic data on a set of 15 independent variables, representing the 15 year/location combinations. The independent variables had values of 1 and 0, corresponding to the presence or absence of a hybrid in that evironment. Residuals from regression were subsequently used as dependent variables in principal component regression.

Since the rank of the augmented matrix $[\mathbf{X}, \mathbf{Y}]$ is at most only 47, a number less than the number of unknowns (187), the system has infinitely many solutions. Principal component regression has been suggested as a method for analyzing data containing multi-colinearities (Hocking 1976). In this approach, letting $\mathbf{V}$ represent the eigenvector matrix for $\mathbf{X'X}$, the model is rewritten as $\mathbf{Y}=\mathbf{ZG}+\mathbf{E}$, where $\mathbf{Y}$ is a previously described, $\mathbf{Z}=\mathbf{XV}$, $\mathbf{G}=\mathbf{VB}$, and $\mathbf{E}$ is a 47×1 vector of deviations from regression. The variables represented by the remaining zero columns (arising by virtue of linear dependencies among the columns of $\mathbf{X}$) drop out of the model. Further, if any of the other variables represented by the columns of $\mathbf{Z}$ are shown by subsequent regression analysis to account for little variation in the dependent variable, they, too, can be eliminated from the model. Letting $\mathbf{G}$ represent the vector of regression coefficient estimates in the reduced rewritten model and $\mathbf{W}$ the matrix comprised of the corresponding columns of $\mathbf{V}$, the vector of regression coefficient estimates, in terms of the parameters of the original model, is

$$\mathbf{B}_{PC}=\mathbf{VG}\,. \tag{1}$$

If the eigenvalues of $\mathbf{X'X}$ are ordered in a manner such that the first s eigenvalues are those that correspond to the variables removed from the rewritten model, Eq. (1) can be expressed as

$$\mathbf{B}_{PC}=\sum_{j=s+1}^{p} l_j^{-1}\,\mathbf{v}_j\,\mathbf{v}_j{}'\mathbf{X'Y} \tag{2}$$

where p is the total number of independent variables in the original model, l_j is the j^{th} eigenvalue, and $\mathbf{v}_j$ is the corresponding column from $\mathbf{V}$. The regression coefficients so obtained are biased, but reflect the essential sources of variation in the dependent variables. In Eq. (2) it is obvious that the inverse of small eigenvalues have a large effect on the regression coefficient estimates and contribute substantially to their variances.

Principal component estimates were tested for statistical significance by computing the test statistic

$$t=B_{PC(j)}/[\mathrm{MSE}\,(\textstyle\sum l_r^{-1}\,v_{jr}^2)]^{1/2} \tag{3}$$

where $B_{PC(j)}$ is a principal component estimate for the j^{th} predictor variable, MSE is the error mean square, l_r^{-1} is the reciprocal of the r^{th} eigenvalue among those retained, and V_{jr}^2 is the square of the j^{th} element in the r^{th} eigenvector (Gunst and Mason 1980).

Principal component analysis (SAS 1985) was performed on the simple correlation matrix computed from inbred variant frequencies. Plotting of the first three principal components permitted visualization of the dispersion of the inbred lines according to their RFLP patterns.

Results and discussion

The hybrids in Table 3 are arranged according to testcrossed inbred lines. Most inbreds were crossed to at least two testers, and usually the testers represented het-

Table 3. Agronomic trait means and modified Rogers's distance for 47 hybrids of maize evaluated over several years and locations

Hybrid	Modified Rogers' distances	Grain yield Mg ha^{-1}	Grain moisture %	Ear height 1–5	Plants root lodged %
Mo17 × B79	0.80	9.08	22.8	3.9	19.7
B73 × B79	0.79	9.23	24.1	4.1	22.1
Va26 × B79	0.71	8.62	24.1	3.5	23.8
Va26 × B84	0.84	9.41	24.2	3.8	26.4
Mo17 × B84	0.83	9.93	23.5	4.0	26.3
B76 × B84	0.73	8.18	23.1	3.7	23.0
B73 × B88	0.85	9.40	23.2	3.9	20.5
B76 × B88	0.80	8.64	21.4	3.8	7.8
Va26 × B88	0.80	9.19	21.8	3.1	17.3
Mo17 × B88	0.76	9.14	21.4	3.8	15.0
Va26 × B89	0.81	8.10	21.7	3.1	18.3
Mo17 × B89	0.76	8.46	21.8	3.5	18.7
B73 × B89	0.65	8.18	22.0	3.9	29.1
B76 × B90	0.85	8.56	21.6	3.5	13.4
B73 × B90	0.81	9.26	21.1	3.6	22.4
Mo17 × B90	0.73	8.59	20.7	3.5	19.1
B76 × B91	0.79	8.42	20.6	3.5	16.5
B73 × B91	0.74	9.32	21.4	3.6	22.7
B73 × De811	0.73	9.31	24.9	3.7	23.3
Mo17 × De811	0.71	8.63	23.0	3.9	26.3
B73 × Oh8710	0.81	9.28	24.5	4.2	7.9
Mo17 × Oh8710	0.76	8.83	23.9	3.8	4.3
B76 × Pa91	0.81	8.68	23.7	3.6	10.7
B73 × Pa91	0.76	9.58	25.0	3.9	27.9
Mo17 × Pa870	0.80	8.06	24.7	3.0	6.2
B73 × Pa870	0.74	8.75	25.3	3.4	18.1
Va26 × Va95	0.83	9.12	23.8	3.3	17.8
Mo17 × Va95	0.83	8.92	23.6	3.5	25.3
Mo17 × Va96	0.81	8.52	22.4	3.6	22.0
Va26 × Va96	0.77	9.21	22.8	3.3	22.9
Va26 × Va97	0.83	8.89	23.0	3.1	14.8
Mo17 × Va97	0.83	9.24	22.9	3.4	16.0
Va26 × Va98	0.73	8.62	21.2	3.3	14.6
Mo17 × Va98	0.81	8.06	20.9	3.4	16.2
B73 × Va26	0.71	9.18	23.2	3.7	22.0
B76 × Mo17	0.81	8.54	21.8	3.9	18.8
B73 × Mo17	0.81	9.04	23.0	4.2	25.5
B76 × B73	0.71	8.18	22.4	3.6	15.1
B73 × Va100	0.86	8.13	21.8	4.1	52.5
B73 × Va22	0.84	9.16	24.3	4.3	31.4
B73 × H98	0.80	8.75	24.2	3.9	19.4
B73 × H112	0.80	8.50	20.2	3.6	28.1
B73 × Va99	0.77	10.04	26.4	4.1	31.6
Mo17 × B37	0.81	8.68	24.1	3.9	13.1
Mo17 × B14A	0.77	8.08	22.1	4.1	18.4
Mo17 × B91	0.77	8.43	21.4	3.5	20.6
Mo17 × H100	0.73	9.26	24.5	3.9	18.0

erotic groups that differed from that of the testcrossed inbred. Unadjusted mean grain yield computed according to tester were: B73–9.02 Mg ha^{-1}, Va96–8.93 Mg ha^{-1}, Mo17–8.75 Mg ha^{-1}, and B76–8.46 Mg ha^{-1}. For a given inbred line, the hybrids in Table 3 are ranked according to MRD, from highest to lowest.

Generally, the lowest MRD values were computed between BSSS inbreds (Table 4). These genetic distance estimates suggest that B37 and B76 are similar among the lines included in this study, with the lowest MRD value of 0.25. Other similar lines include B73-Va95, B73-Va96, Va95-Va96, Va95-Va97, and H98-H112. MRD values for

Table 4. Modified Roger's distances among 26 maize inbred lines

	B14A	B37	B73	B76	B84	B89	H100	Va95	Va96	Va97	Va98	H98	H112	Mo17	Pa870	Va22	Va26	B79	Oh8710	B88	B90	B91	De811	Pa91	Va99
B37	0.73																								
B73	0.67	0.70																							
B76	0.73	0.25	0.71																						
B84	0.77	0.73	0.55	0.73																					
B89	0.77	0.71	0.65	0.74	0.67																				
H100	0.76	0.84	0.84	0.84	0.84	0.89																			
Va95	0.70	0.70	0.46	0.71	0.58	0.67	0.83																		
Va96	0.71	0.65	0.51	0.67	0.62	0.62	0.85	0.41																	
Va97	0.71	0.64	0.58	0.60	0.60	0.68	0.80	0.48	0.60																
Va98	0.71	0.65	0.60	0.65	0.67	0.74	0.83	0.56	0.60	0.62															
H98	0.83	0.80	0.80	0.80	0.74	0.80	0.71	0.80	0.80	0.80	0.74														
H112	0.83	0.77	0.80	0.79	0.73	0.84	0.74	0.81	0.80	0.80	0.74	0.44													
Mo17	0.77	0.81	0.80	0.81	0.83	0.76	0.73	0.83	0.81	0.83	0.79	0.81	0.84												
Pa870	0.74	0.80	0.74	0.81	0.79	0.83	0.73	0.70	0.76	0.73	0.80	0.79	0.84	0.80											
Va22	0.83	0.84	0.84	0.83	0.83	0.80	0.76	0.80	0.84	0.76	0.76	0.79	0.79	0.74	0.74										
Va26	0.81	0.80	0.84	0.81	0.84	0.81	0.74	0.83	0.77	0.83	0.81	0.73	0.80	0.80	0.68	0.77									
B79	0.74	0.83	0.79	0.83	0.80	0.77	0.74	0.79	0.80	0.83	0.73	0.70	0.77	0.80	0.79	0.80	0.71								
Oh8710	0.76	0.71	0.81	0.76	0.83	0.81	0.67	0.83	0.81	0.79	0.83	0.77	0.76	0.76	0.74	0.74	0.74	0.74							
B88	0.81	0.81	0.85	0.80	0.81	0.84	0.70	0.83	0.85	0.79	0.77	0.71	0.76	0.76	0.79	0.73	0.80	0.77	0.70						
B90	0.74	0.85	0.81	0.85	0.83	0.81	0.76	0.74	0.79	0.83	0.77	0.80	0.85	0.73	0.81	0.76	0.83	0.71	0.80	0.73					
B91	0.70	0.79	0.74	0.79	0.81	0.80	0.79	0.70	0.76	0.73	0.74	0.74	0.80	0.77	0.77	0.74	0.77	0.67	0.76	0.73	0.62				
De811	0.68	0.73	0.73	0.74	0.74	0.74	0.79	0.74	0.70	0.80	0.70	0.77	0.79	0.71	0.80	0.77	0.76	0.76	0.79	0.84	0.79	0.77			
Pa91	0.70	0.80	0.76	0.81	0.74	0.76	0.65	0.70	0.74	0.76	0.73	0.76	0.79	0.65	0.62	0.68	0.67	0.70	0.74	0.74	0.71	0.74	0.68		
Va99	0.71	0.84	0.77	0.85	0.74	0.80	0.73	0.79	0.79	0.80	0.74	0.77	0.77	0.70	0.71	0.79	0.73	0.71	0.73	0.76	0.80	0.81	0.71	0.55	
Va100	0.89	0.81	0.86	0.83	0.88	0.83	0.81	0.86	0.85	0.79	0.83	0.83	0.84	0.74	0.76	0.73	0.76	0.79	0.79	0.81	0.85	0.84	0.80	0.76	0.77

non-BSSS pairs tended to be larger than those of their BSSS counterparts, implying relatively greater genetic diversity among the non-BSSS lines.

No clear relationship was evident between agronomic traits (Table 3) and MRD values (Table 4). In some cases, hybrids with low MRD values actually had relatively high yields. The highest yield was obtained from B73 × Va99 (10.04 Mg ha^{-1}), a hybrid with a relatively low MRD value (0.77). Other traits exhibited a similarly random association with genetic distances. Correlations between MRD and the traits confirmed the absence of a relationship between RFLP diversity and agronomic performance. MRD was not significantly correlated with any of the traits (MRD versus yield, $r=0.09$; moisture, $r=-0.03$; root lodging, $r=0.03$; ear height, $r=0.004$). Given the narrow range of MRD values and the minimal crossing within heterotic groups, these results may indicate only that RFLP information adds little to prediction of performance of single-cross hybrids between unrelated inbreds beyond that grained from prior experience.

The molecular data also were used to estimate the potential utility of RFLP analysis in allocating maize inbred lines to heterotic groups. A similar analysis was reported earlier by Lee et al. (1989). However, in our study, the group of inbred lines was much more diverse than that evaluated by Lee et al. (1989). In both studies, principal component analysis was performed on the correlation matrix obtained from the matrix of inbred variant frequenies. The first, second, and third principal components, in this study, accounted for 12.1%, 8.1%, and 7.2%, respectively, of the total variation present among variants. The plot of principal components one, two, and three (Fig. 1) shows a tendency for the inbred lines to aggregate according to the heterotic groups identified in

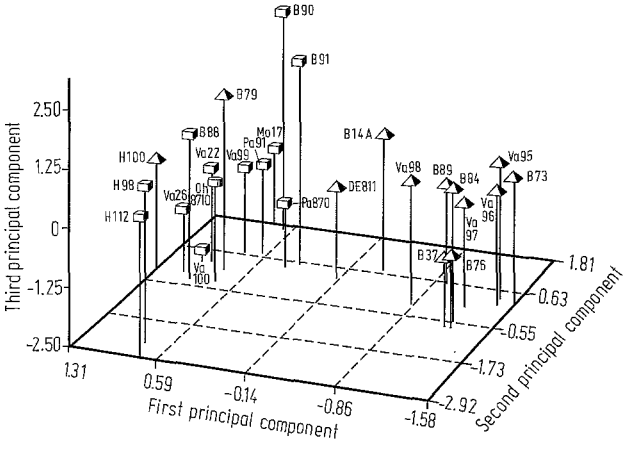

Fig. 1. Plot of the first three principal components from the simple correlation matrix computed from the inbred variant frequencies. *Pyramids* designate lines from the Reid Yellow Dent heterotic group; *cubes* designate lines from the Lancaster Sure Crop heterotic group; and a *cylinder* designates a line with unknown heterotic grouping

Table 5. Principal component regression analysis of adjusted means vs. RFLP variant frequencies

Trait	RFLP variant[a]	Regression coefficient	Prob > |T|
Grain yield	UMC11-8	−0.096	0.009886
	BNL5.59-4	0.113	0.003324
	UMC49-3	−0.163	0.008446
	UMC49-4	0.155	0.002719
	BNL4.36-2	−0.182	0.007186
	UMC104-2	−0.167	0.007664
	BNL6.29-1	0.138	0.008927
	BNL6.29-2	−0.138	0.008927
	UMC80-3	−0.126	0.006551
	BNL9.11-4	0.142	0.005281
	BNL5.09-3	0.160	0.007983
	BNL3.04-4	−0.147	0.007995
Grain moisture	UMC11-8	−0.294	0.000029
	BNL5.59-4	0.174	0.000057
	UMC52-2	−0.505	0.000006
	UMC52-3	0.453	0.000046
	BNL7.65-1	0.134	0.000078
	BNL7.65-4	−0.469	0.000017
	BNL7.65-6	0.663	0.000038
	BNL4.36-1	0.746	0.000064
	BNL4.36-3	−0.399	0.000053
	UMC104-3	0.497	0.000033
	BNL6.29-1	0.360	0.000098
	BNL6.29-2	−0.360	0.000098
	UMC65-3	0.134	0.000078
	UMC89-2	−0.398	0.000027
	UMC80-3	−0.239	0.000029
	UMC114-2	0.134	0.000078
Ear height	UMC86-1	0.021	0.000090
	UMC86-2	−0.021	0.000090
	BNL8.10-1	−0.031	0.000099
	BNL8.10-2	0.031	0.000099
	UMC34-2	0.023	0.000093
	UMC26-1	−0.019	0.000092
	UMC26-2	0.019	0.000092
	UMC39-2	−0.004	0.000093
	BNL6.29-1	−0.017	0.000087
	BNL6.29-2	0.017	0.000087
	UMC80-1	0.004	0.000098
	BNL9.11-4	−0.024	0.000093
	UMC103-2	−0.021	0.000088
	BNL5.09-1	−0.026	0.000088
Root lodging	UMC76-1	−0.810	0.000046
	UMC76-2	2.437	0.000010
	UMC29-1	1.254	0.000083
	UMC11-5	−1.913	0.000057
	UMC11-8	1.178	0.000084
	UMC86-1	2.301	0.000070
	UMC86-2	−2.301	0.000070
	UMC34-3	−0.958	0.000049
	UMC34-4	−2.281	0.000046
	UMC34-5	2.433	0.000005
	UMC49-1	1.961	0.000007
	UMC39-4	−0.734	0.000064
	UMC15-6	−0.783	0.000027
	BNL7.65-5	1.120	0.000001
	BNL7.65-7	1.961	0.000007
	BNL4.36-3	2.088	0.000013
	UMC89-2	2.009	0.000017

Table 5. (continued)

Trait	RFLP variant[a]	Regression coefficient	Prob $>\|T\|$
	UMC80-1	2.306	0.000008
	UMC56-3	−0.863	0.000014
	UMC116-2	−0.783	0.000027
	BNL13.05-1	−1.721	0.000095
	BNL13.05-3	1.936	0.000026
	BNL9.11-2	−2.672	0.000053
	BNL9.11-5	−0.958	0.000049
	NPI268-1	−1.631	0.000011
	UMC103-2	2.329	0.000029
	UMC81-1	0.781	0.000027
	UMC81-2	−0.781	0.000027
	BNL5.04-1	−0.783	0.000027
	BNL7.49-1	1.848	0.000016
	BNL10.13-3	1.120	0.000001

[a] Numbers affixed to clone designation indicate assigned variant number

Table 1. Among the Reid Yellow Dent inbred lines, only H100 and B79 were markedly different from the remaining lines in that heterotic group (Fig. 1). Variant frequencies were more variable in Lancaster Sure Crop lines than in the Reid Yellow Dent lines. Based on their positions in Fig. 1, six Lancaster Sure Crop lines (B90, B91, De811, H98, H112, and Va100) had variant frequencies noticeably distinct from those of the remaining lines in that heterotic group. This result indicates that significant genetic differences can exist among lines within heterotic groups. Despite this within-group variation, however, the general dispersion pattern suggests the potential for using RFLP data to identify combinations of inbred lines from different heterotic groups.

Several variants had principal component regression coefficients that were statistically significant at the 0.01% probability level for grain moisture, root lodging, and ear height, and the 1% level for grain yield (Table 5): 12 variants were associated with genes controlling maize grain yield, 16 variants with grain moisture, 14 with ear height, and 32 with root lodging. Clone UMC80 had variants significantly associated with all four traits, whereas UMC 11, BNL4.36, BNL6.29, and BNL9.11 had variants that were significant for three of the traits. Statistitical significance suggests that, when attempting to increase grain yield, one could select for the presence of variants that have positive regression coefficients and for the absence of variants with negative coefficients. To obtain reductions in grain moisture, ear height, or root lodging, favorable variants would be those with negative principal component regression coefficients.

In research that involves RFLP analysis, it may not be possible to include an adequate number of observations so as to reduce or eliminate linear dependencies. Optimally, the number of traits should be greater than the total number of RFLPs variants. Frequently, however, the primary goal is to saturate the genome with as many markers as possible. If the number of observations is not correspondingly increased, statistical problems, such as those described in this study, will ensue. Biased regression procedures, such as principal component regression, are suggested as statistical tools for reducing the impact of highly correlated predictor variables when attempting to relate RFLP patterns to agronomic data.

Provided that the variants in Table 5 are verified to be of agronomic importance, an effort to use plant transformation techniques to transfer the chromosomal segments to designated recipients may be in order. Furthermore, once important segments of chromosomes are identified, they may be saturated with RFLP loci to more precisely elucidate linkages between polymorphisms and quantitative trait loci. A logical extension of this research is to predict, from RFLPs and linkages of polymorphisms with quantitative trait loci, single crosses that optimize performance for a given trait.

Acknowledgements. The authors acknowledge Dr. W.A. Russell professor in the Department of Agronomy at Iowa State University, who produced the single crosses and made available the hybrid agronomic data analyzed in this study.

References

Evola SV, Burr FA, Burr B (1986) The suitability of restriction fragment length polymorphisms as genetic markers in maize. Theor Appl Genet 71:765–771

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction fragment length polymorphisms to hight specific activity. Anal Biochem 132:6–13

Frei OM, Stuber CW, Goodman MM (1986) Uses of allozymes as genetic markers for predicting performance in maize single-cross hybrids. Crop Sci 26:37–42

Gunst RF, Mason RL (1980) Regression analysis and its application. Marcel Dekker, New York

Hadjinov MI (1982) Interrelationships between isozymic diversity and combining ability in maize lines. Maydica 27:135–149

Helentjaris T, King G, Slocum M, Siedenstrang C, Wegman S (1985) Restriction fragment length polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol Biol 5:109–118

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

Henderson CB (1984) Maize research and breeders's manual, No. X. Illinois Foundation Seeds, Champaign, FL

Hocking RR (1976) The analysis and selection of variables in linear regression. Biometrics 32:1–49

Lamkey KR, Hallauer AR, Kahler AL (1987) Allelic differences at enzyme loci and hybrid performance in maize. J Hered 78:231–234

Lee M, Godshalk EB, Lamkey KR, Woodman WW (1989) Association of restriction fragment length polymorphism with

agronomic performance of maize inbreds and hybrids. Crop Sci 29:1067–1071

Price SC, Kahler AL, Hallauer AR, Charmley P, Giegel P (1986) Relationships between performance and multilocus heterozygosity at enzyme loci in single-cross hybrids of maize. J Hered 77:341–344

Rogers JS (1972) IV. Measures of genetic similarity and genetic distance. Studies in Genetics VII. Univ Tex Publ 7213:145–153

Saghai-Maroof MA, Soliman KM, Jorgenson R, Allard RA (1984) Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

SAS Institute (1985) SAS user's guide: statistics. SAS Institute, Cary/NC

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

Stuber CW, Moll RH, Goodman MM, Schaeffer HE, Weir BS (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). Genetics 95:225–236