Theor Appl Genet (1988) 76:39–44

Associations among inbred lines of maize using electrophoretic, chromatographic, and pedigree data

2. Multivariate and cluster analysis of data from Iowa Stiff Stalk Synthetic derived lines

J. S. C. Smith and O. S. Smith

Departments of Research Specialists and Data Management, respectively, Pioneer Hi-Bred International, Inc., 7301 NW 62nd Avenue, Johnston, IA 50131, USA

Received August 17, 1987; Accepted January 23, 1988
Communicated by A. R. Hallauer

Summary.

Associations among 17 "Iowa Stiff Stalk Synthetic" derived inbred lines of maize (*Zea mays* L.) were determined using multivariate and cluster analysis. Objectives were to assess the level of unique characterization among lines afforded by reversed-phase high-performance liquid chromatography (RP-HPLC) of zeins and starch gel electrophoresis of isozymes and to compare associations among lines revealed by biochemical and pedigree data. Isozymic data for 33 loci provided unique discrimination among 88% of the lines; 2 closely related lines were indistinguishable. Seventy-one percent of the lines could be uniquely and unambiguously identified by RP-HPLC. Biochemical data showed associations between lines that would be expected on the basis of pedigree. Nevertheless, different associations were revealed by allozymic and chromatographic data. Although these data permitted a high degree of unique identification, additional markers, covering a larger proportion of the genome, are needed to more adequately monitor similarities among genes that respond to selection during plant breeding.

Key words: High performance liquid chromatography – Zeins – Allozymes – Plant variety protection – Germplasm security – Heterosis

Introduction

Isozyme and zein data allowed the unique identification of important and widely available Lancaster Sure Crop inbred lines of maize (*Zea mays* L.) that were more than isogenically different (Smith and Smith 1987). Multivariate analyses showed the lines could be assigned to 3 groups that corresponded to those expected on the basis of pedigrees.

Iowa Stiff Stalk Synthetic inbreds provide a similar opportunity to determine the potential of biochemical data to characterize and to reveal associations between lines for a second important pool of elite maize germplasm. The Iowa Stiff Stalk Synthetic was put together from 16 lines that had strong stalks (Sprague 1946). Development of lines from Iowa Stiff Stalk Synthetic has either been through direct selfing out of the original population, selfing out of advanced cycles, or by multiple backcrosses to lines so developed (Hallauer et al. 1983). The diverse origins of inbred lines making up the Iowa Stiff Stalk Synthetic germplasm pool (Sprague 1946; Henderson 1976; Baker 1984) means that the present study places particular emphasis on first, the potential of biochemical data to reveal associations between lines, many of which have no connection through direct pedigree breeding and second, on the potential to differentiate lines, many of which have a very close pedigree connection through multiple backcrosses. For a similar study of Lancaster Sure Crop derived lines (Smith and Smith 1987) only Oh43 and C103 or Oh43 and Mo17 were unrelated by pedigree breeding, and there were fewer examples of lines with multiple backcross pedigrees.

Materials and Methods

Seventeen publicly available lines of current and historic importance in U.S. maize breeding that have been either selected directly or were derived (Henderson 1976) from the Iowa Stiff Stalk Synthetic (Sprague 1946) were analyzed (Table 1). Chromatographic and isozymic data for 22 isozymic loci were collected and associations among lines were revealed as described previously (Smith and Smith 1987). Data from an additional 11 isozymic loci (Table 2) were collected using gel, electrode, and staining conditions described by Stuber et al. (1988).

Results

Fifty-one isozyme alleles were revealed; of the 33 loci that were examined, 15 were found to be monomorphic and 18 were polymorphic (Table 2). Fifteen (88%) of the Iowa Stiff Stalk Synthetic lines had unique profiles with a common isozymic constitution shown by B14 and B14A. Reversed-phase high-performance liquid chromatography (RP-HPLC) revealed 49 hydrophobically different zein maize endosperm protein peaks. Fourteen different zein profiles, each of which was comprised of from 10–14 peaks, were found across the 17 lines; some examples are presented in Fig. 1. Twelve (71%) of the lines could be uniquely identified; the remaining five lines could be separated into two groups within which chromatographic profiles were qualitatively identical: *viz*, 1) B14, B14A, A635, (Fig. 2a) and 2) H84, H100 (Fig. 2b). One major and one minor peak were quantitatively different for B14 versus B14A and for B14A versus A635; the minor peak was quantitatively different for B14 in comparison to A635. H84 and H100 showed a quantitative difference for one minor peak.

Table 1. Lines used in the analysis of pedigree, electrophoretic, and chromatographic data

Lines selected directly from a synthetic (Group DIR)	
B14	Iowa Stiff Stalk Synthetic
A657	Iowa Stiff Stalk Synthetic
N28	Nebraska Stiff Stalk Synthetic (ISSS with added material)
B37	Iowa Stiff Stalk Synthetic
B73	Iowa Stiff Stalk Synthetic
B79	Iowa 2-Ear synthetic [see Russell et al. (1971)]
B84	Iowa Stiff Stalk Synthetic
Derived lines with no immediate backcrossing (Group DERBC−)	
H100	N28 × H91 [H91 = (B14 × GE440) $B14^4$]
Derived lines with immediate backcrossing (Group DERBC+)	
B14A	(Cuzco × $B14^8$) rust res. sel.
A632	(Mt42 × B14) $B14^3$
A634	(Mt42 × B14) B14 × $B14^2$
A635	(ND203 × B14) $B14^2$
A636	(ND203 × B14) B14 × B14
B68	(41.2504B × $B14^3$) sel.
N7A	$Oh7^2$/Stiff Stalk Synthetic [Oh7 = Illinois 2 ear × ILL.L.]
B76	(CI.31A × $B37^2$) B37
H84	(B37 × GE440) Ht

Pedigree data

Associations among lines on the basis of pedigree data are given in Fig. 3. The first three vectors accounted for 60%, 30%, 9%, and 42%, 13%, and 9% of total variation in multivariate and cluster analysis, respectively. Both cluster analysis (Fig. 3a) and principal coordinate analysis (Fig. 3b) showed associations between B14 and lines derived by backcrossing with B14 as the recurrent parent. Cluster analysis indicated an association of H100 with B14 and B14 backcross derived lines (Fig. 3a). However, principal coordinate analysis (Fig. 3b) placed H100 intermediate to B14 and N28. Both multivariate analyses indicated a close association of N7A with B79. Principal coordinate analysis (Fig. 3b) indicated similarities between B37, B76, and H84 that was not shown by cluster analysis (Fig. 3a). Individual lines selected directly from the Iowa Stiff Stalk Synthetic (Table 1) were not

Table 2. Isozymic genotypes of inbred lines used in the present study. In addition to the loci that are tabulated, all lines were fixed for *Adh1-4, Adk1-4, Cat3-9, Dia2-4, Got1-4, Got2-4, Got3-4, Idh1-4, Mdh1-6, Mdh3-16, Mdh4-12, Mdh5-12, Mmm-Mmm, Pgm1-9, TPI1-4, TPI2-4, TPI3-4,* and *TPI4-4*

Inbred	*Aco1*	*Acp1*	*Amp1*	*Amp3*	*Dia1*	*E8*	*Glu1*	*Hex2*	*Idh2*	*Mdh2*	*Pgd1*	*Pgd2*	*Pgm2*	*Phi1*	*Sad1*
A632	1	4	4	4	8	4.5	7	2	6	6	3.8	2.8	4	5	4.5
A634	4	4	4	4	8	4.5	7	2	6	6	3.8	5	4	5	4.5
A635	4	4	4	5	8	5	7	2	6	6	3.8	2.8	4	5	4
A636	4	4	4	4	8	4.5	6	2	6	6	3.8	5	4	5	4
A657	4	4	5	4	8	4.5	7	4	6	6	2	2.8	4	4	4
B14	4	4	4	4	8	5	7	2	6	6	3.8	2.8	4	5	4.5
B14A	4	4	4	4	8	5	7	2	6	6	3.8	2.8	4	5	4.5
B37	4	2	4	5	12	5	7	2	6	6	2	5	4	4	4
B68	4	3	4	5	8	5	7	2	6	6	2	5	4	5	4.5
B73	4	2	4	5	8	5	7	2	4	3.5	3.8	5	4	4	4
B76	4	2	4	5	12	5	7	2	6	3.5	2	5	4	4	4
B79	4	4	4	5	12	4.5	7	2	6	6	3.8	5	4	4	4
B84	4	4	4	4	8	5	7	4	4	3.5	2	5	3	4	4
H84	4	2	4	4	8	5	6	1	4	6	2	5	4	4	4
H100	4	2	4	4	8	5	7	2	6	6	3.8	2.8	4	5	4.5
N74	1	4	4	4	12	4.5	1	4	6	6	2	2.8	4	4	4
N28	4	2	4	4	12	4.5	7	4	6	6	2	5	4	4	4

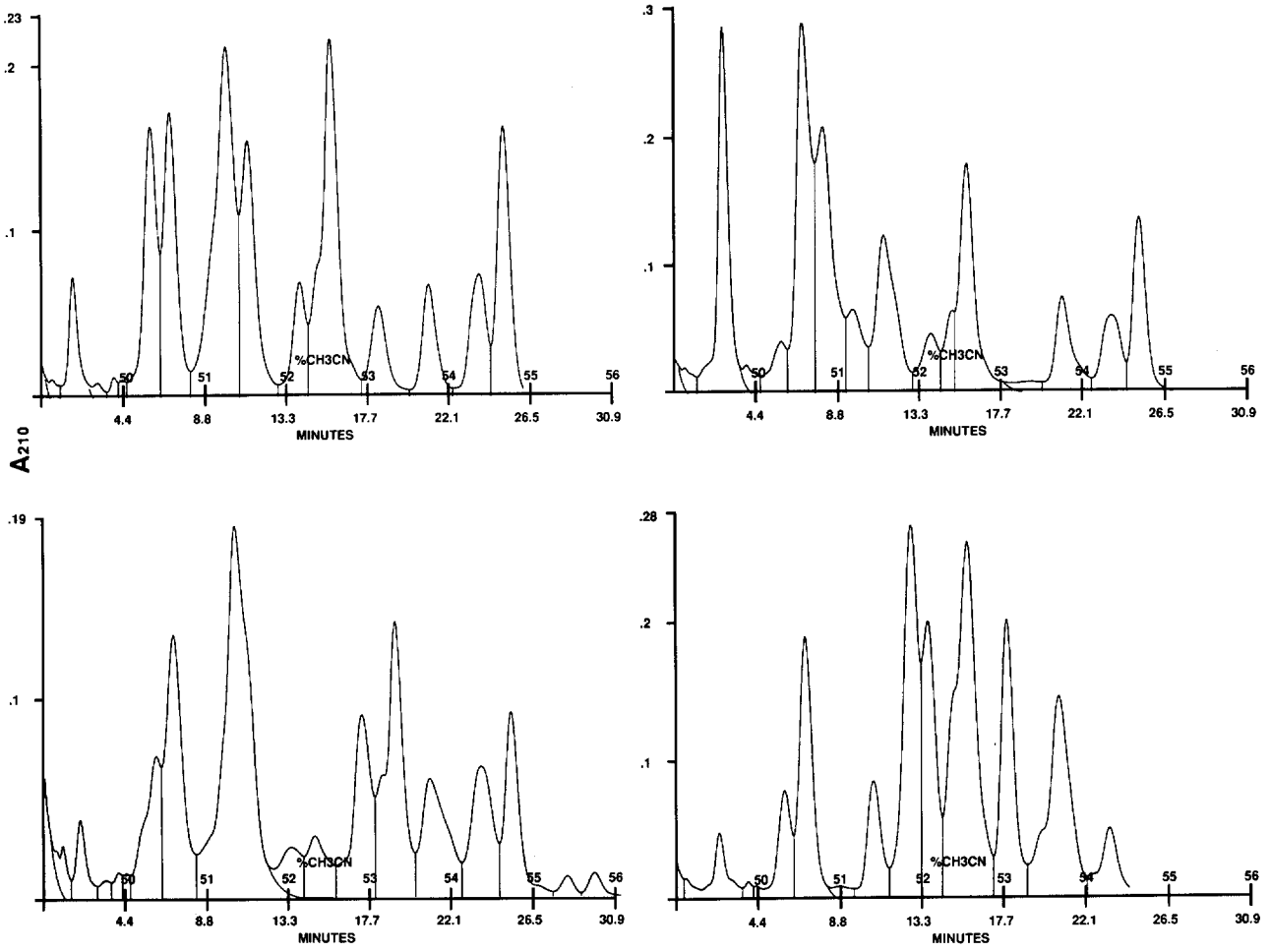

Fig. 1. Zein chromatograms (*top to bottom*) of B14, B37, B73, and B79

closely associated by cluster analysis (Fig. 3a). This was also true of principal coordinate analysis, except for associations revealed between 1) A657 and N28 and 2) B73 and B84. This is, in part, due to the lack of knowledge of how similar the lines derived directly from Iowa Stiff Stalk Synthetic actually are and thus, highlights the need for data which can reveal associations among lines that reflect phylogeny and selection.

Isozyme data

Both cluster analysis (not shown) and principal coordinate analysis of isozymic data (Fig. 4) revealed similar associations among lines to that of the analysis based on pedigree data. The first three vectors accounted for 36%, 21%, 11% and 43%, 19%, and 10% of the total variation for multivariate and cluster analysis, respectively. B14 and all lines derived from B14 by both immediate and non-immediate backcrossing were associated. Among these lines, B68, A636, and A632 were the least closely associated with B14. Remaining lines were associated into three groups, *viz*; 1) B84, B73, H84; 2) N7A, A657; and 3) N28, B76, B37, B79. Of these lines, principal coordinate analysis (Fig. 4) showed B79, N7A, A657 to be more closely associated with B14 and B14 derived lines.

RP-HPLC unweighted data

The first three vectors of cluster analysis (Fig. 5) encompassed 51%, 16% and 11% of the variation. Likewise, principal component analysis (not shown) encompassed 77%, 5%, and 4% of the variation. Both techniques revealed similar associations between lines. B14, B14 backcross derived lines, and H100 were associated. B68 was the least closely associated among these lines. Identical profiles were revealed for two pairs of lines; 1) B14 and B14A and 2) H84 and H100. The remaining lines were grouped into three clusters; *viz*, 1) B37 and B76 and 2) B73, N7A, and N28, and 3) a more distant cluster formed by B84, A657, and B79.

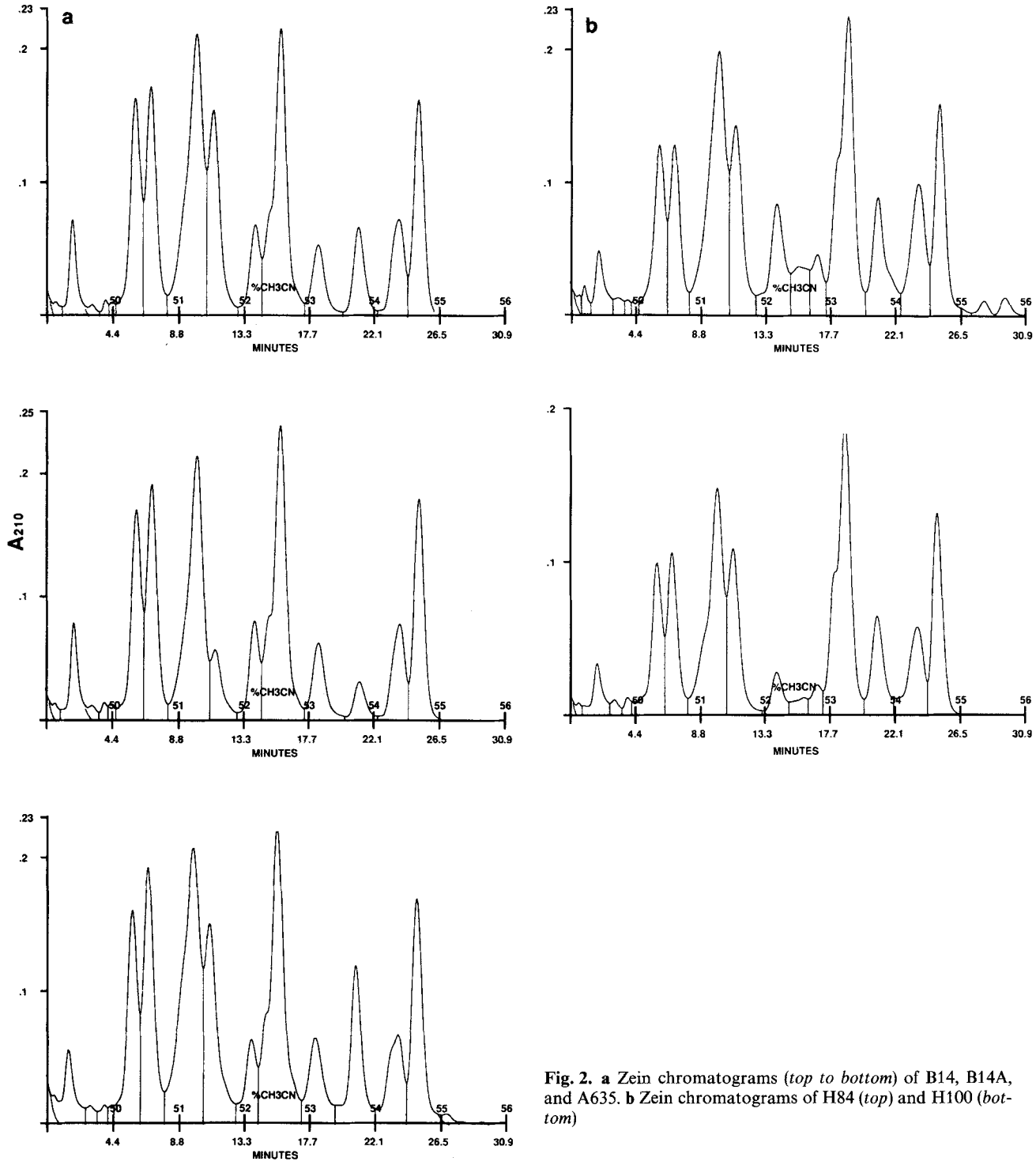

Fig. 2. **a** Zein chromatograms (*top to bottom*) of B14, B14A, and A635. **b** Zein chromatograms of H84 (*top*) and H100 (*bottom*)

RP-HPLC weighted data

Cluster analysis (Fig. 6), encompassing 55%, 19%, and 10% of the variation for the first three vectors, and principal component analysis (not shown), which encompassed 53%, 15% and 8% of the total variation for the first three vectors, respectively, revealed similar associations among lines that were nearly identical to those shown by unweighted chromatographic data (Fig. 5). Weighted chromatographic data showed less associations between the individual pairs of lines, B14, B14A and H84, H100. However, B14 and lines with immediate backcrossing to B14, with the sole exception of B68, were

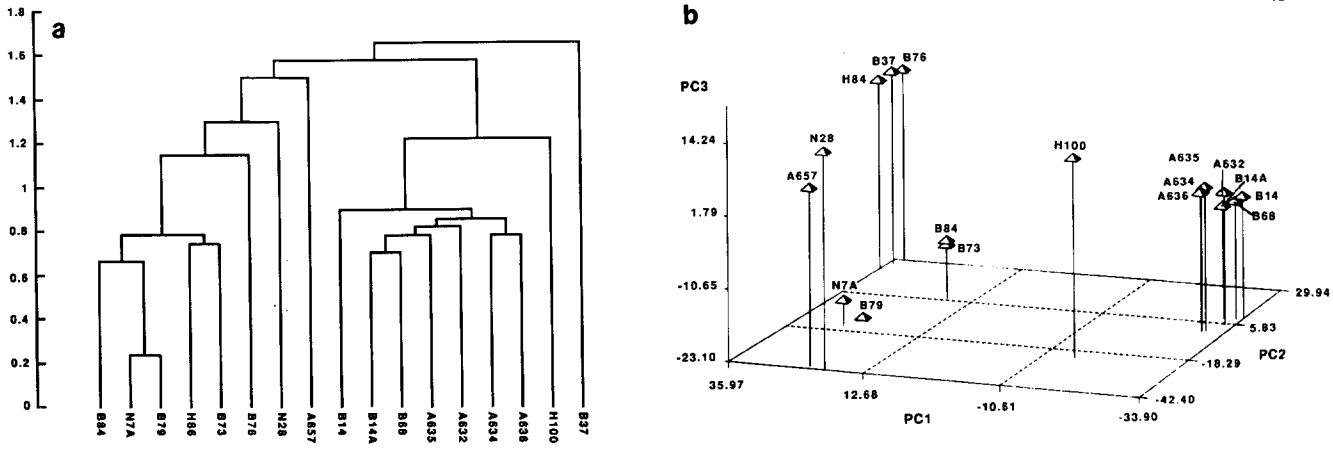

Fig. 3. **a** Associations of inbred lines revealed by cluster analysis of pedigree data. **b** Associations of inbred lines revealed by principal coordinate analysis of pedigree data

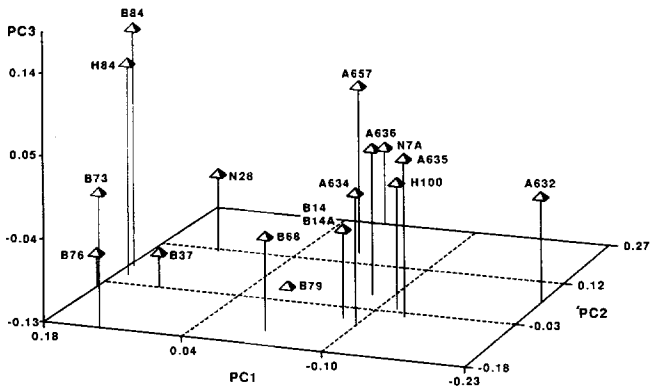

Fig. 4. Associations of inbred lines revealed by principal coordinate analysis of isozymic data

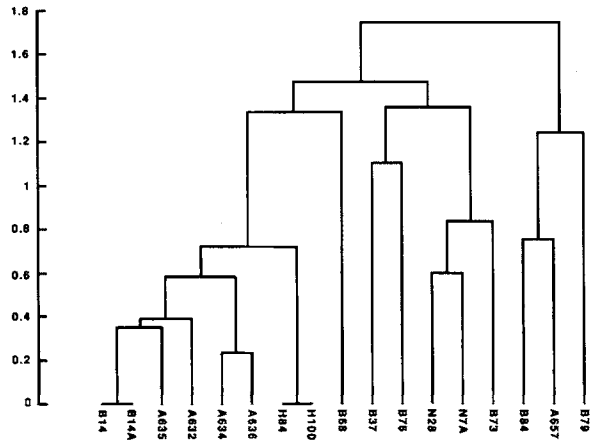

Fig. 5. Associations of inbred lines revealed by cluster analysis of unweighted zein chromatographic data

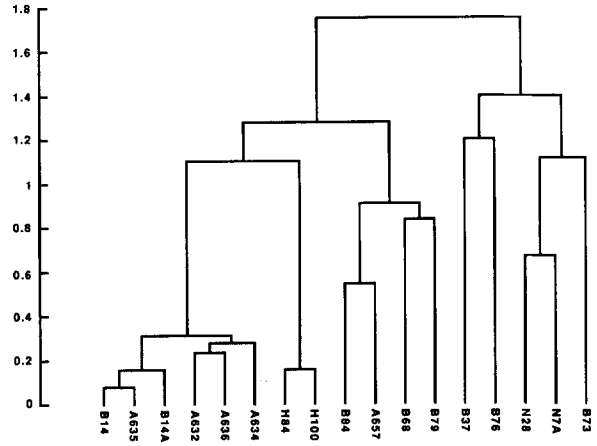

Fig. 6. Associations of inbred lines revealed by cluster analysis of weighted zein chromatographic data

more closely associated on the basis of weighted chromatographic data. Instead of the loose association of B68 to B14 and other B14 derived lines that was shown by unweighted chromatographic data (Fig. 5), the weighted data associated B68 with B79, A657, and B84 (Fig. 6).

Discussion

The 16 inbred lines that were used to constitute the Iowa Stiff Stalk Synthetic appear to have encompassed a broad range of germplasm (Sprague 1946; Wallace and Brown 1956; Smith et al. 1985). Therefore, individual selections made from this synthetic could be very different, and biochemical data should be expected to reveal any such differences. For those lines selected directly from the synthetic, there are no precise pedigree data. Furthermore, the very breadth of diversity of Iowa Stiff Stalk Synthetic germplasm increases the chances that estimates of relatedness based solely on pedigree will be inaccurate. Thus, data that can reveal associations between lines on the basis of comparisons among gene loci could be especially useful in providing estimates of relationships. There are also Iowa Stiff Stalk Synthetic lines that have been selected following two or more backcrosses to either B14 or B37 as the recurrent parent (Hender-

son 1976). These backcross and recurrent parent lines present a more stringent test of the ability of biochemical, or indeed of any data set, to provide unique line identification.

The level of discrimination afforded by isozymic data permitted discrimination among all Iowa Stiff Stalk Synthetic lines with the exception of one backcross derived line (B14A) and its respective recurrent parent line (B14). RP-HPLC provided lesser discrimination because two groups of lines (B14, B14A, and A635 and H84 and H100) had shared profiles. These analyses provided a high level of unique identification among inbred Iowa Stiff Stalk Synthetic lines and derived lines, including those that are closely related by pedigree.

Lines derived by self-pollination from the synthetic could be expected to be associated by biochemical data if genes upon which selection operated during breeding were also those assayed by the biochemical tests or were instead closely linked to the isozymic and zein loci. However, the biochemical loci were most probably very few in comparison with the number of loci that responded to selection. Thus linkage disequilibrium caused through breeding would not be measured by the zein or isozyme loci. It is, therefore, hardly surprising that lines directly selected from the Iowa Stiff Stalk Synthetic were not associated by biochemical data thus supporting the hypothesis that none is closely related. Of these lines only B37 – N28 and A657 – B84 were joined by cluster analysis of isozymic and chromatographic data, respectively, but then only at an average cluster distance of ≥ 0.5. For lines derived from B14 and from B37, allozymic and zein data revealed close associations between 1) B14, B14A, A632, A634, A635, and A636 and 2) B37 and B76. H84, less related than B76 to B37 by pedigree, revealed a lesser association with B37 on the basis of chromatographic and isozymic data.

Sometimes, however, isozymic and chromatographic data also revealed associations of lines that were not in agreement with expectations based upon known pedigree. Both isozymic and chromatographic data showed B68 to be the least associated among B14 derived lines; chromatographic data excised B68 from the B14 group of lines. Isozymic data, in contrast to zein chromatographic data, separated A632 from B14. Yet, B68 and A632, on the basis of pedigree data, would not be expected to be less closely associated with B14 than other B14 derived lines. Isozymic and zein data revealed dissimilar associations for H100. Isozymic data showed that H84 and H100 differed at seven loci, apparently because of the B14 parentage of H100 (Table 1). In contrast, a close association of H100 with H84 was shown by chromatographic data. H100 and H84 have GE440 in their pedigrees.

Additional laboratory and field data will be necessary to test the validity and significance of differences in associations shown by isozymic, chromatographic, and pedigree data. It is understood that predictions of relationship based upon pedigree data alone may not be entirely accurate due to the inadequacies of a model that assumes equal parental contribution and no selection. Furthermore, for selections made from this synthetic, or indeed for any lines for which pedigree data are unreliable or unavailable, then other data must be sought. These additional data [e.g., restriction fragment length polymorphisms, 2-dimensional gel electrophoresis profiles or heterosis data] that survey a greater sample of the total genome could allow associations between inbred lines to be investigated in sufficient detail to provide data that might then have descriptive or predictive value for traits of agronomic interest.

Acknowledgements. Isozymic analyses were conducted by D. Phillips and R. Tenborg. Assistance with multivariate and cluster analysis was provided by S. Wall. Useful suggestions on a preliminary draft of the manuscript were given by Drs. A. Cavalieri, R. Fincher, and W. Kuhn of Pioneer Hi-Bred International, Inc.

References

Baker R (1984) Some of the open pollinated varieties that contributed the most to modern hybrid corn. In: Proc 20th Annu Illinois Corn Breeders School, Champaign, May 6–8, 1984. University of Illinois, Urbana-Champaign IL, pp 1–19

Hallauer AR, Russell WA, Smith OS (1983) Quantitative analysis of Iowa Stiff Stalk Synthetic. Stadler Symp 15:83–104

Henderson CB (1976) Maize research and breeders manual No 8. Illinois Foundation Seeds Inc, Champaign IL

Russell WA, Penny LH, Hallauer AR, Eberhart SA, Scott GE, Guthrie WD, Dicke FF (1971) Registration of maize germplasm synthetics. Crop Sci 11:140–141

Smith JSC, Smith OS (1987) Associations among inbred lines of maize using electrophoretic, chromatographic, and pedigree data. Part 1. Multivariate and cluster analysis of data from 'Lancaster Sure Crop' derived lines. Theor Appl Genet 73:654–664

Smith JSC, Goodman MM, Stuber CW (1985) Genetic variability within U.S. maize germplasm. I. Historically important lines. Crop Sci 25:550–555

Sprague GF (1946) Early testing of inbred lines of corn. J Am Soc Agron 38:108–117

Stuber CW, Wendel JF, Goodman MM, Smith JSC (1988) Techniques and scoring procedures for starch gel electrophoresis of enzymes from maize (*Zea mays* L.). University of North Carolina, Raleigh NC

Wallace HA, Brown WL (1956) Corn and its early fathers. Chicago: Michigan State University Press