Theor Appl Genet (1994) 89:943–950

J. Burstin · D. de Vienne · P. Dubreuil · C. Damerval

Molecular markers and protein quantities as genetic descriptors in maize. I. Genetic diversity among 21 inbred lines

Received: 22 February 1994 / Accepted: 6 July 1994

Abstract Twenty-one maize (*Zea mays* L.) inbred lines were analysed using isozyme electrophoresis, restriction fragment length polymorphism (RFLP), and two-dimensional electrophoresis of denatured proteins (2-D PAGE). Our goal was (1) to assess the genetic variability among these lines which are potential progenitors for the development of forage maize hybrids in Europe, and (2) to compare the relationship pattern revealed by the polymorphism at marker loci with the one derived from the amount of protein variability assessed by computer-assisted analysis of the 2-D electrophoregrams. Fourteen markers were obtained from isozyme polymorphism, 84 from the restriction fragment length polymorphism, and 70 from protein shifts revealed by 2-D PAGE. The Rogers' distance computed on the set of molecular markers was the most efficient to describe the pedigree relationships between lines. Quantitative protein data gave a picture of relationships between lines clearly different from the monogenic markers. When unrelated pairs of lines were considered, the Rogers' distance was weakly correlated to distances based on quantitative variations in the amount of protein which may be consistent with their polygenic control and the occurrence of gene interactions.

Key words *Zea mays* L. · Genetic distances
Protein polymorphism · Two-dimensional electrophoresis
Restriction fragment length polymorphism

Introduction

Accurate evaluation of genetic diversity is one of the main objectives of maize (*Zea mays* L.) breeding programs. Different descriptors have been used to characterize inbred lines of maize. Pedigree information provides a global estimate of the expected genetic relatedness among lines, but relies on the assumption of the absence of gametic and zygotic selection, which is often not the case. Morphological and biochemical data have provided an increasing number of genetic markers. Since 1980, measurements of genetic relatedness based on DNA restriction fragment length polymorphism (RFLP) have progressively replaced those based on isozyme variation because the RFLP technique provides a nearly unlimited number of polymorphic markers, distributed over the entire genome (Botstein et al. 1980). Two-dimensional electrophoresis of denatured proteins (2-D PAGE, O'Farrell 1975), also reveals a large number of genetic variants. This technique allows the detection of two types of polymorphism: (1) structural variations of proteins, and (2) variations in the amount of the proteins, quantified through variations in spot intensity.

In the present study, we have analysed the genetic diversity among 21 elite maize inbreds by 2-D PAGE of total proteins, RFLPs and isozymes, based on available pedigree information. Our objectives were (1) to investigate the genetic relationships among lines of different origins and different maturities that could be useful for the development of silage maize hybrids, and (2) to compare the relationship patterns obtained from marker loci and from data on the amount of proteins. The heterotic pattern between early European flint lines and U.S. Corn Belt dent lines has been widely used for the development of elite grain maize hybrids as well as for silage maize hybrids. However, the definition of heterotic groups has become less clear because of lines with unknown pedigree or lines from mixed origins, such as CO255 which has been widely used for silage hybrid development (Y. Barrière, personnal communication). Silage maize breeding also could benefit from a re-definition of the heterotic groups for biomass yield improvement, and from a broadening of the genetic basis of breeding germplasm (Dhillon et al. 1990). Structural protein variations and RFLPs display discrete variations which are determined by one polymorphic gene. Isozymes and 2-D PAGE address the variability expressed at the protein level while the RFLPs focus on the variability at the

Communicated by P. L. Pfahler

J. Burstin (✉) · D. de Vienne · P. Dubreuil · C. Damerval
INRA-UPS-CNRS-URA1492, Station de Génétique Végétale, Ferme du Moulon, F-91190 Gif-sur-Yvette, France

DNA level itself. Most of these variations are assumed to be selectively and phenotypically neutral (Kimura 1985). Two-dimensional PAGE also makes it possible to analyse quantitative variations of proteins, which are related to polymorphism in gene expression. These variations are, in most cases, polygenically inherited and reveal frequent gene interactions such as dominance and epistasis (Damerval and de Vienne 1993; Damerval et al. 1994). Therefore, they represent an integrated level of genetic variation as compared with the DNA and protein structure variation level and may constitute an intermediary between the gene and the macroscopic levels. Leonardi et al. (1991) have pointed out the original feature of these quantitative 2-D PAGE variations as compared with structural protein variations. In their study, the 2-D patterns were visually inspected, and few discrete levels of quantitative variation could be defined. Automatic systems of image analysis now provide quantitative measurements of protein amounts. In our study, we extended the number of lines from eight to 21 and considered the protein quantitative variations as continuous characters.

Materials and methods

Maize inbred lines

Twenty-one inbreds representative of potential progenitors for the development of elite silage maize hybrids in Europe were analysed. These lines have different maturities and come from various germplasm sources. Three originated from the French population Lacaune, three from the Iowa Stiff Stalk Synthetic population (BSSS), and four from the line CO125. Two lines from an unknown pedigree come from the Iodent group, one from the Lancaster Sure Crop group, while the others have miscellaneous or unknown origins (Table 1).

Allozyme analyses

Twenty-one enzyme loci were analysed (Table 2). Most of the data were obtained from GEVES (Groupement d'Etude des Variétés et des Semences, France), following the procedure of Stuber et al. (1988).

RFLP analyses

The 21 lines were assayed for their RFLP profiles according to the procedure described in Damerval et al. (1994). The RFLP assays were performed using two restriction enzymes (*Eco*RI and *Hin*dIII) with 59 probes including 57 genomic DNA clones from the maize core map (Coe 1993) and two cDNAs kindly provided by C. Baysdorfer and mapped by M. Causse (personal communication). Only autoradiograms with clear single-copy signals or those with multicopy patterns with a clear major band were scored. In all, 84 probe-enzyme combinations were examined. The chromosomal locations of the different marker loci are listed in Table 2.

2-D PAGE analyses

The proteins were extracted in denaturing conditions from etiolated coleoptiles of 8-day old seedlings grown under controlled conditions, and were separated in the first dimension according to their isoelectric point, and in the second dimension according to their molecular masses. High resolution was allowed by large-size gels (24×23 cm^2) and improved protocols (Damerval et al. 1986, 1987). For each genotype, at least four replicate gels were done according to Burstin et al. (1993).

Seventy protein shifts (PS) were scored visually among the 21 lines. Each PS included a series of 2–6 spots of similar aspect, only one of the series being found in any inbred. Thirty-four of these variants were found in an F_2 progeny between the two lines Io1 and F2 (Damerval et al. 1994) and revealed a monogenic determinism. For the 36 other PS, monogenic determinism was assumed. The chromosomal locations of 26 PS loci were determined (Table 2).

The gels were automatically analyzed with the Kepler 2-D Gel Analysis Software (LSB Corp., Rockville, Maryland, USA) as described in Burstin et al. (1993). We focused on 190 well quantified

Table 1 Maize inbred lines

Line	Kernel type	Developer	Origin
F2	Flint	INRA	Lacaune pop.
F268	Flint	INRA	Lacaune pop. (37.5% F2, 12.5% F7)
F283	Flint	INRA	Lacaune pop. (50% F7), Lizargarote pop. (22% Ep1)
Co255	Flint-dent	Canada Ottawa	Lacaune pop. (25% F7), Lizargarote pop. (25% Ep1), Wisonsin
F476	Flint-dent	INRA	Pouyastruc pop. (25% F47), Chalosse pop. (25% F49), BSSS (25% B14)
F7001	Flint-dent	INRA	Yugoslavian line, pop corn, BSSS (12,5% B37)
F1852	Flint-dent	INRA	F32 × F19, eastern European germplasm
BSSS1	Dent	Holden	BSSS
BSSS2	Dent	Holden	BSSS
F618	Dent	INRA	BSSS (75% B37)
F584	Dent	INRA	50%, A636, 50% A434
F608	Dent	INRA	25% A434, 75% A239
Io1	Dent	Mike Brayton	Iodent commercial hybrid
Io2	Dent	Mike Brayton	Iodent commercial hybrid
F252	Dent	INRA	50% Co125, 25% W325A
F292	Dent	INRA	50% F252, 25% W401
F271	Dent	INRA	50% Co125, 50% W103
F284	Dent	INRA	25% Co125, 25% W401, 18.5% W325A
M1	Dent	Holden	Minn 13, Lancaster Sure Crop
M2	Dent	Holden	Related to M1
F113	Dent	INRA	Spooner (Wisconsin)

Table 2 Chromosomal location of the isozyme loci, protein position shift loci and RFLP loci, and list of unlocalised marker loci

Chromosome	Loci
Chromosome 1	*adh1 phi1 mdh4* *psl25 psl18 psl6 psl24 psl44 psl13* UMC11 UMC157 BNL5.59 BNL5.62 BNL6.32
Chromosome 2	*dia1* *psl11 psl32 psl1* UMC6 UMC34 UMC36 UMC49 UMC55 BA40
Chromosome 3	*pgd2 mdh3* *psl5 psl10 psl28 psl16* UMC10 UMC16 UMC60 UMC63 UMC102
Chromosome 4	UMC19 UMC47 UMC66 UMC111 BNL8.23
Chromosome 5	*mdh5 pgm2 got2* *psl21 psl39 psl43 psl20 psl17* UMC27 UMC51 UMC54 UMC68 UMC90 BNL6.25 BNL7.71
Chromosome 6	*idh2 mdh2 pgd1* UMC21 UMC38 UMC59 UMC62 UMC65 UMC85
Chromosome 7	*psl27 psl23* UMC35 UMC168 BNL16.06 BA81
Chromosome 8	*idh1 mdh1* *psl42 psl38 psl19* UMC30 UMC89 BNL9.11
Chromosome 9	*acp1* *psl3 psl22 psl46* UMC109 UMC113 UMC114 BNL5.09 BNL5.10 BNL14.28
Chromosome 10	*βglu1* BNL7.49
Unlocalised	*cat3 got1 got3 Mmm pgm1* 44 PS loci UMC103 UMC104 UMC107 UMC140 UMC15 UMC31 UMC39 UMC4 UMC44 UMC55

spots which were distributed throughout the gel. Their intensity was estimated as the volume of a two-dimensional Gaussian fitted on the digitized image and corrected for the gel effect (Burstin et al. 1993). Protein spot intensity has been shown to be related, linearly in most cases to protein amount (Damerval and de Vienne 1993).

Statistical analyses

The coancestry coefficient, f, defined by Malecot (1948) was calculated for each pair of related lines with known pedigree, and a pedigree distance d was derived (d=1–f). Lines without known common parentage are supposed to be unrelated (f=0). Two lines derived from a commercial hybrid have f=0.5. We assumed that two lines derived from the same population, have f=0, and that two lines derived from the BSSS synthetic have a f=1/16 since it was founded from 16 progenitors.

Genetic distances between all possible pairs of inbred lines were estimated through isozyme, RFLP and PS data, separately and then pooled, by the multilocus Rogers' distance (RD, Rogers 1972). For inbred lines, the RD is the ratio of the number of polymorphic loci between two lines to the total number of considered loci. The standard errors (SE) of RD estimates were: $SE=[RD\ (1-RD)/(n-1)]^{0.5}$ where n=number of markers (Bar-Hen and Charcosset, personal communication).

Associations among lines were investigated by a principal component analysis (PCA) on the binary data matrix of the global set of markers. This set included the 70 PS loci, the 58 RFLP loci corresponding to the most polymorphic probe-enzyme combination per locus, and the 14 polymorphic isozyme loci. For all loci, each allele was considered as one variable, taking the values 0 for the absence and 1 for the presence of the allele. The principal components were calculated from the covariance matrix between these binary variables.

The amount of protein variability among lines was tested using analyses of variance with the volume of each spot as the dependent variable and the genotype as the factor.

Associations among lines were investigated by a PCA on the volumes of the 64 significantly variable spots among the 21 lines ($P< 0.01$) and present in at least 20 lines ; the PCA was performed on the correlation matrix between the spot volumes. Two distances, whose genetic interpretation is discussed later, were computed between pairs of lines: (1) Stud was calculated as the number of spots with significant intensity differences between two lines (with $P<0.01$ for the t-test), and (2) Maha was a Mahalanobis distance computed from the coordinates of the lines on the 20 first axes of the PCA analysis on the spot volumes. Pearson's correlations between RDs for the different markers, the pedigree distance d, and the distances based on quantitative variations were computed. Since the distance values for the 210 pairs of lines are not independent, the degrees of freedom to test the significance of the correlation coefficient are not known. Thus, the significance of the correlation was tested by a Monte Carlo permutation procedure (Brunel and Rodolphe 1985).

Unrooted trees were constructed from the global RD and stud, using the Fitch and then the Drawtree programs of the PHYLIP package (Felsenstein 1989). All statistical computations, except for the construction of unrooted trees, were performed using SAS procedures (SAS Institute 1988).

Results

Genetic variation for PS, RFLPs, isozymes and protein amounts

Fourteen of the twenty-one enzyme loci showed polymorphism among the 21 inbreds, with 2–4 alleles per locus. The average number of alleles per polymorphic locus was 2.2. On average, 2.4 alleles per locus were found for the 70 PS loci scored, with 2–6 alleles per locus. All but 2 of the 84 probe-enzyme combinations revealed polymorphism among the lines. When only one probe-enzyme combination per locus was considered and the most polymorphic probe-enzyme combination selected, the average number of variants per locus was 4.6 for 58 combinations, with up to eight alleles per locus. Only three enzyme alleles (10%) against 33 protein shift alleles (20%) and 104 RFLP variants (30%) were line-specific. The inbred line F2 had the largest number (13) of unique RFLP alleles (the average was five).

Out of the 190 spots quantified, 129 (68%) were significantly variable among the lines at the 5% level, and 103 (54%) at the 1% level. The ratio of the highest to the lowest intensity in the 21 lines ranged from 1.4 to 26. Some spots displayed continuous variation among the 21 lines, while other spots grouped lines into discrete classes of intensity (Fig. 1a, b).

Genetic distances among inbred lines

Among the 210 possible pairs of lines, 28 involved related lines. The coefficients of coancestry ranged from 0.0019

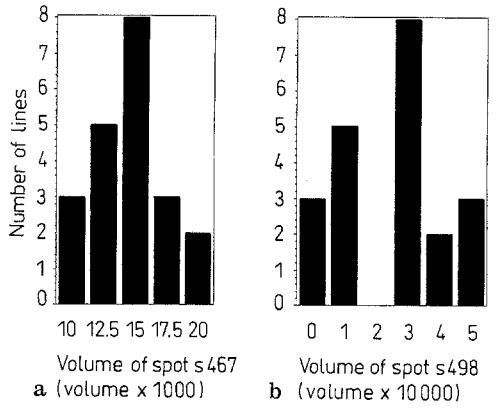

Fig. 1 Histogram of the volumes of spot S467 (**a**) and spot S498 (**b**) in the 21 lines

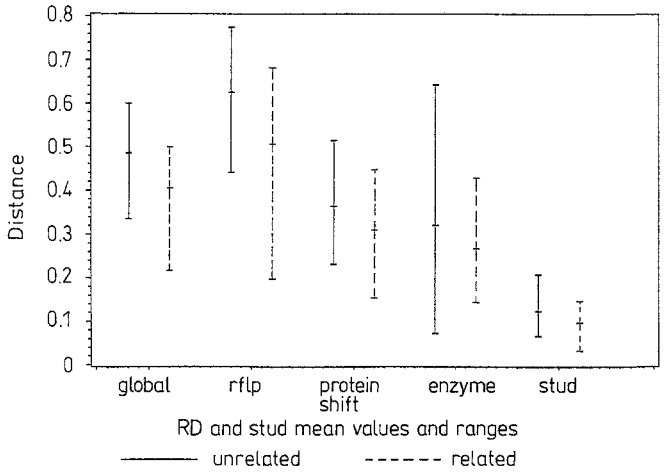

Fig. 2 Means and ranges of stud and RDs for the different markers and for the global set of markers, computed on related and unrelated pairs of lines

for pairs of lines with miscellaneous origins (F476-F7001 and F284-CO255) to 0.5 for the Io1 and Io2 lines, which are derived from the same commercial hybrid.

The RD values and ranges for the different markers and for unrelated and related pairs of lines are shown in Fig. 2. As expected, for any type of marker, the mean RD between unrelated inbred lines was greater than between related ones, and for any given pair of lines, the RDs computed on protein markers were less than computed on RFLP markers. Among unrelated pairs of inbreds, from 44 to 77% of RFLP loci differed between two lines compared to 23–51% of PS loci and 7–64% of isozyme loci; RD ranges were quite similar between RFLP and PS markers, and both were lower than the isozyme RD range; though significant, the correlations between RDs were small (Table 3). Among related pairs of inbreds, the PS RD was significantly correlated with the RFLP RD (r=0.53, $P<0.05$), but not with the isozyme RD. The correlations between the pedigree distance d and the RD values were significant for the PS RD (r=0.83, P<0.001) and the RFLP RD (r=0.63, P<0.05) but not for the isozyme RD. The global RD computed on the complete marker data set was highly correlated to d (r=0.85, $P<0.001$). RD has already been found to be a direct function of the parentage coefficient provided that markers sufficiently represent the genome (Messmer et al. 1993). The smallest global RD values corresponded to the related pairs of lines Io1-Io2, BSSS1-BSSS2, F252-F292, and F268-F2. Seven out of the ten highest global RDs involved the line F2 with late dent lines (M2, BSSS1, Io1, Io2), early dent lines (F252, F292), and the line F1852. The line F2, which is widely used in maize breeding programs for its earliness, has, on average, 68% of RFLP loci, 43% of PS loci, and 40% of isozyme loci differing from the other lines. The RD standard errors ranged from 0.07 to 0.14 for isozymes, from 0.04 to 0.06 for the PS, from 0.04 to 0.05 for the RFLP, and from 0.03 to 0.04 for the global RD.

Six to thirty-nine spots varied quantitatively between two lines at the 1% level (average=23). The stud mean values and ranges for related and unrelated pairs of lines are shown in Fig. 2. The mean value between related lines was not significantly lower than that between unrelated lines. The smallest stud values corresponded to the pairs of related lines F2-F268, F252-F292, and BSSS1-BSSS2, and the seven highest values involved either F1852 or F7001. Maha and stud were significantly correlated (r=0.65, $P<0.001$). When considering only related pairs of lines, stud and maha were highly correlated ($P< 0.001$) to d (r=0.66 and r=0.54, respectively), and to the global RD (r=0.62 and r=0.53, respectively). When considering unrelated pairs of lines, the correlation coefficient between the protein-

Table 3 Correlation matrix between RDs computed on the different marker datasets and the global marker dataset, stud and maha, and the pedigree distance d, for unrelated (above the diagonal) and related (below the diagonal) pairs of lines

	Global RD	PS RD	RFLP RD	Isozyme RD	Stud	Maha
Global *RD*		0.78***	0.71***	0.49***	0.14*	0.01
PS RD	0.88***		0.29**	0.23**	0.24**	0.15
RFLP RD	0.84***	0.53*		0.21**	0.09	–0.03
Isozyme RD	0.23	0.17	0.03		–0.20	–0.28
Stud	0.62***	0.57**	0.55***	0.18		0.45***
Maha	0.53***	0.59***	0.47**	0.04	0.66***	
d	0.85***	0.83***	0.63*	0.23	0.66***	0.54***

***: $P<0.001$ **: $P<0.01$ *: $P<0.05$

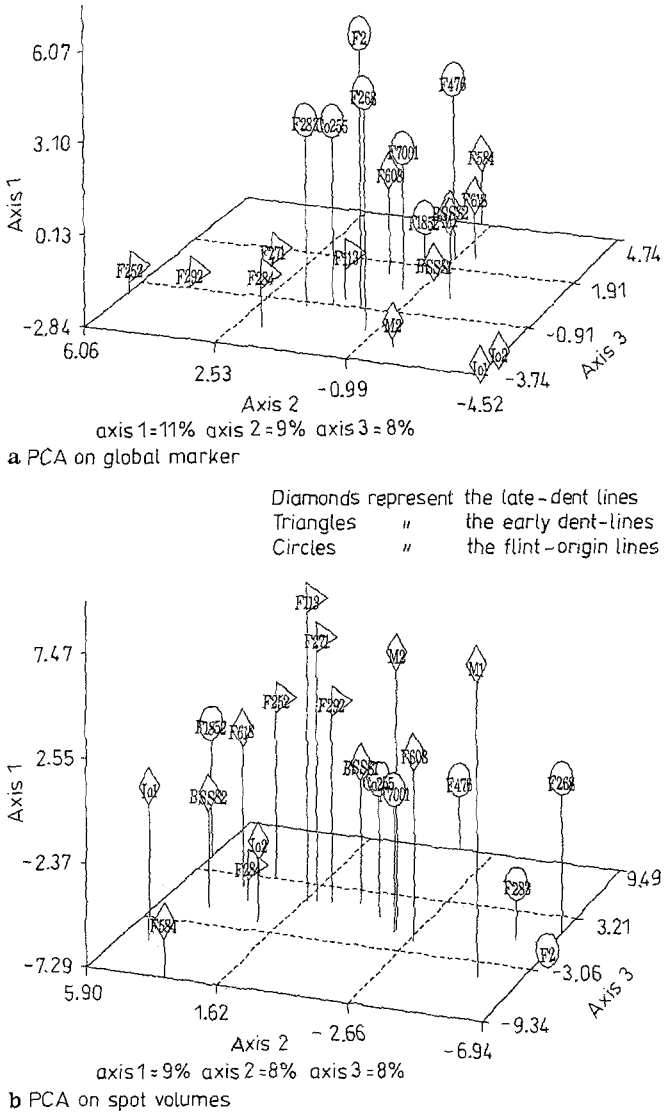

Fig. 3 Associations between the 21 lines revealed by principal component analyses on the global marker dataset (**a**) and on the spot volumes (**b**)

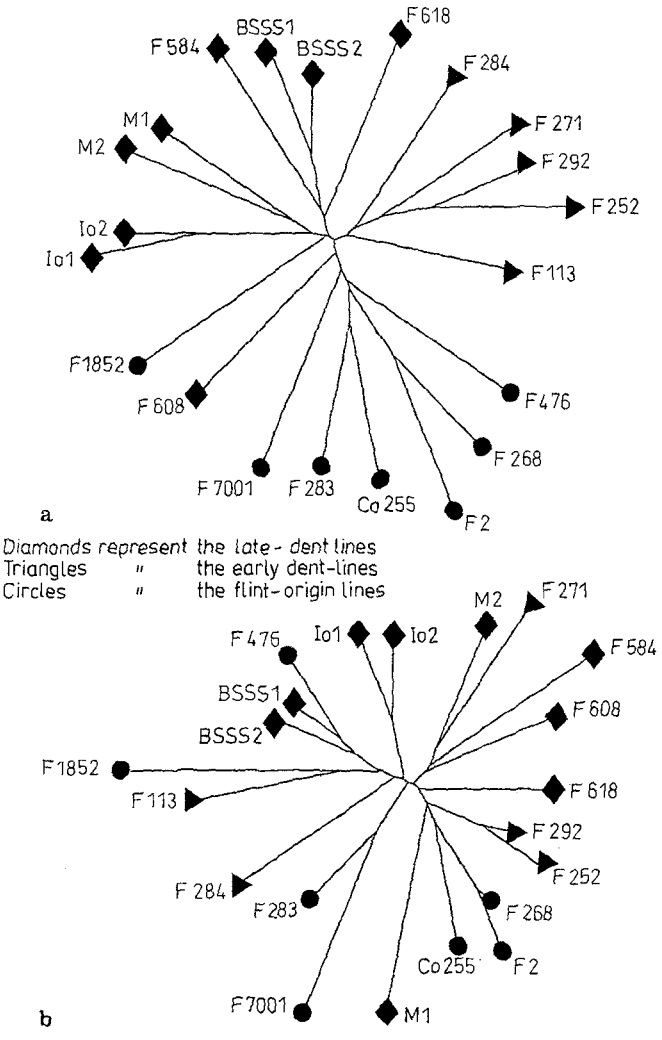

Fig. 4 Unrooted trees from RD on the global marker dataset (**a**) and on stud (**b**)

amount distance and the global RD was not significant for maha and was very small for stud ($r=0.14$, $P< 0.05$).

Description of the genetic variability

Associations among the 21 inbred lines revealed by the PCA on global marker data agreed well with the pedigree information (Fig. 3 a). The three first axes explained 11%, 9%, and 8% of the total variation. The first axis separated the lines with flint origins from the dent lines. The flint lines have three subgroups, according to their genetic background: F2 with F268, F283 with CO255, and F476 with F7001, which mapped near to the BSSS lines. Along the second axis, CO125-related lines were on the opposite side to the two Io1 and Io2 sister lines. The other lines spread along the third axis from a medium position between the Iodent lines and the CO125-related lines. The PCA on the spot volumes revealed a less clear separation of the lines (Fig. 3 b). The first axis separated F284 from the other CO125-related lines, Io1 from Io2, and F268 from F2. However, all but one of the early dent lines were mapped at one side of the second axis, and F2, F283, and F268 at the other, according to their kernel type. F1852 was still adjacent to late dent lines. Lines with mixed flint-dent germplasm (F476, F7001, CO255) had an intermediate position. The three first axes represented 9%, 8%, and 8% of the total variation.

The unrooted tree from the global RD consistently represented the origins of the lines, confirming the results from the PCA analysis (Fig. 4 a). Three principal clusters occurred: (1) flint lines and mixed lines with flint origin, except for the eastern European line F1852, (2) early dent lines (i.e., CO125-related lines) and F113, and (3) BSSS lines with Iodent lines. The two related lines M1 and M2 clustered together and F1852 and F608 were distant from

the others. The tree from stud (Fig. 4 b) was different. Even if the most related lines were clustered together (Io1 and Io2, F252 and F292, and F2 and F268), the grouping did not reflect the origins of the other lines.

Discussion

Genetic diversity among the 21 lines

The set of lines analysed was genetically diverse and represented a wide sample of the variability used in maize breeding programs in France. As often reported in studies on genetic diversity in maize, the level of polymorphism was significant for any type of marker. All but two (98%) of the probe-enzyme combinations were polymorphic and the number of alleles was as high as eight among the 21 lines assayed. The average number of RFLP variants per probe-enzyme combination was similar to that observed in other studies, provided these focused only on single-banded patterns ; with an average of 68% of differing RFLP loci in all possible pairs of lines, the level of variability achieved was amongst the highest reported (Smith et al. 1990; Messmer et al. 1992). Although parentage relationships exist between the lines, the percentage of unique RFLP pattern was similar to the one observed by Messmer et al. (1991) in the 16 unrelated progenitors of BSSS. As expected, polymorphism was lower at the protein level than at the DNA level. Only 67% of isozyme loci revealed polymorphism. The average numbers of isozyme (2.2) and PS (2.4) variants per polymorphic locus and their maximum numbers of alleles (four and six respectively) were lower than observed with RFLPs. Our results indicated that the ability of 2-D PAGE markers, i.e., of total proteins, to reveal polymorphism was similar to that of isozymes. The variability of protein spot intensity was also significant. The automatic analysis substantially improves the quantitative exploitation of the 2-D gels, allowing the detection of small variations. More than two spots out of three were quantitatively variable in our subset, and the intensity of several spots varied in a ten-fold range.

The line F2 remains an important source of variability at the marker loci. Compared with the two other Lacaune flint lines (F268 and F283), the mean RD was greatest between F2 and the other lines . When considering the protein-amount variability, F1852 and F7001, which are particularly adapted for silage maize breeding, were the most distant to the other lines, and to each other. The line F1852, with eastern European origin, was more similar to dent lines than to flint lines in spite of its flint origin, whatever the marker.

Related versus unrelated lines

The correlations between distances computed on the different data were small when considering unrelated pairs of lines, but increased when considering related lines (except for isozymes). This discrepancy may be due to differences in the level of linkage disequilibrium in the two sets of lines. Indeed, linkage disequilibrium is necessarily high among the related pairs of lines, because the common ancestor line is relatively close to the descendant lines, and the number of recombinations which could occur in the identical segments is limited. When no parentage relationship exist between two lines, the same gametic associations are not likely to be observed in the two lines, even if they are derived by one cycle of selection from the same population (Charcosset and Essioux 1994). In that case, markers only measure the long-term divergence between the two lines.

Relationships between the lines according to the neutral markers

Both RFLP and protein structure polymorphisms (including enzymes) are monogenic. Most have been found to be neutral from a selective point of view (Kimura 1985). Thus, they are expected to give similar patterns of relationship between genotypes. Among the numerous studies on genetic diversity, only a few compare the results from enzyme vs RFLP markers (Landry et al. 1987; Messmer et al. 1991; McGrath and Quiros 1992; Smith and Smith 1992; Qifa Zhang et al. 1993). In these studies, the authors found discrepancies between results from RFLP data and those from isozyme data, but could not determine whether these differences were due to a sampling bias, or to the fact that these two types of markers do not reveal genetic variability at the same level. Messmer et al. (1991) reported poor correlations between Rogers' distances computed for 22 isozyme loci and 144 probe-enzyme combinations, and also between distances computed for two different samples of 27 probes in combination with two restriction enzymes. They emphasized the importance of the number of marker loci on the reliability of the RD estimates. In our study, RFLP, PS, and isozyme RDs were poorly correlated when only unrelated pairs of lines were considered. Poor correlations between distances computed on two independent sets of markers can result from two error levels: (1) the inaccuracy on the estimate of the distance, illustrated by large confidence intervals and (2) a sampling effect, because markers give only a partial view of the genome, depending on the sample of markers chosen. These biases are more important when the number of markers is reduced. In our case, the mean standard error was twice as large for isozyme RD as for PS and RFLP RDs, and decreased to 0.04 for the global RD. The different classes of markers did not equally cover the genome ; no isozyme markers appeared on chromosomes 4 and 7, and only one RFLP probe was mapped to chromosome 10. In any event, the global RD seemed the best estimate of the relatedness between lines (Table 3) and the PCA on the whole set of markers revealed associations among lines consistent with available information on their origins (Fig. 3).

Quantitative polymorphism of proteins

Most quantitative variations of proteins have been found to be under polygenic control (Damerval et al. 1994). Because stud counts the number of quantitative variants, it was expected not to be independent of distances computed from neutral markers, which measure genome divergence. However, when unrelated pairs of lines were considered, the variability of protein amounts was not related to the global RD. The determination coefficient between the global RD and stud was low (R^2=0.02). The global RD relies on the polymorphism at 142 loci. Stud takes into account the quantitative variation of 129 variable proteins, but the number of polymorphic loci involved is difficult to deduce. Damerval et al. (1994), in an F_2 progeny, reported that a minimum of 70 genetic factors controlled the amounts of 42 proteins. Both stud and the global RD probably sampled more than 100 loci: therefore the bias due to sampling may be limited. The low correlation values could indicate that (1) the relationship between the two parameters is not linear; a more complexe relationship may result from the polygenic control of protein amount and from possible epistatic/pleiotropic interactions, and (2) the evolutionary forces act differently upon genes controlling protein amounts and neutral polymorphisms. Some observations suggested that the factors controlling the activity/amount of some enzymes do not respond to selection in the same way as structural genes (Matsuo and Yamazaki 1984; Bush and Paigen 1992). The magnitude of quantity variation did not seem to be correlated with the number of factors controlling this variation (Damerval, unpublished). Maha takes into account the magnitude of quantitative variation which could be physiologically pertinent. Although stud and maha were significantly correlated, the coefficient of determination (R^2) between them was only 0.4 confirming that the two distances provide partly non-redundant information.

The different origins of the lines were clearly defined by their marker patterns, as illustrated by the unrooted tree from the global RD. The two lines F1852 and F7001 were found to be distant from the different groups, especially according to protein quantitative polymorphism. The discriminatory power of RFLP has already been reported (Smith and Smith 1991). Two-dimensional PAGE also provides a powerful tool for investigating genotype relationships: Higginbotham et al. (1991) had earlier mentioned the usefulness of qualitative and quantitative variation of protein spots in order to distinguish between very closely related genotypes. Since the protein quantitative variation is often under polygenic control, and subject to epistasis and pleiotropy, two genotypes with similar quantitative patterns are unlikely to occur by chance. As compared to neutral polymorphisms, the predictive power of this kind of polymorphism towards agronomical performances is currently being examined.

Acknowledgements We thank M. Zivy for valuable comments and considerable help in computer analysis and programming, A. Charcosset for helpful discussions and anonymous reviewers for useful remarks on the manuscript.

References

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

Brunel D, Rodolphe F (1985) Genetic neighbourhood structure in a population of *Picea abies* L. Theor Appl Genet 71:101–110

Burstin J, Zivy M, de Vienne D, Damerval C (1993) Analysis of scaling methods to minimize experimental variations in two-dimensional electrophoresis quantitative data. Applications to the comparison of maize inbred lines. Electrophoresis 14:1067–1073

Bush RM, Paigen K (1992) Evolution of β-glucuronidase regulation in the genus Mus. Evolution 46:1–15

Coe EH (1993) Gene list and working maps. Maize Genet Coop Newslett 67:133–166

Charcosset A, Essioux L (1994) The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theor Appl Genet (in press)

Damerval C, de Vienne D (1993) Quantification of dominance for proteins pleitropically affected by *opaque-2* in maize. Heredity 70:38–51

Damerval C, de Vienne D, Zivy M, Thiellement H (1986) Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. Electrophoresis 7:52–54

Damerval C, Le Guilloux M, Blaisonneau J, de Vienne D (1987) A simplification of Heukeshoven and Dernick's silver staining of proteins. Electrophoresis 8:158–159

Damerval C, Maurice A, Josse JM, de Vienne D (1994) Quantitative trait loci underlying gene product variation – a novel perspective for analysing regulation of genome expression. Genetics 137:289–301

Dhillon BS, Gurrath PA, Zimmer E, Wermke M, Pollmer WG, Klein D (1990) Analysis of diallel crosses of maize for variation and covariation in agronomic traits at silage and grain harvests. Maydica 35:297–302

Felsenstein (1989) PHYLIP Phylogeny Inference Package (version 3.2). Cladistics 5:164–166

Higginbotham JW, Smith JSC, Smith OS (1991) Quantitative analysis of two-dimensional protein profiles of inbred lines of maize (*Zea mays* L.). Electrophoresis 12:425–431

Kimura M (1985) The neutral theory of molecular evolution. New Sci 107:41–46

Landry BS, Kesseli R, Leung H, Michelmore RW (1987) Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphism in lettuce (*Lactuca sativa* L.) Theor Appl Genet 74:646–653

Leonardi A, Damerval C, Hébert Y, Gallais A, de Vienne D (1991) Association of protein amount polymorphism (PAP) among maize lines with performances of their hybrids. Theor Appl Genet 82:552–560

Malecot G (1948) Les mathématiques de l'hérédité. Masson et Cie, Paris

McGrath JM, Quiros CF (1992) Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. Theor Appl Genet 83:783–790

Matsuo Y, Yamazaki T (1984) Genetic analysis of natural populations of *Drosophila melanogaster* in Japan. IV. Natural selection on the inducibility, but not on the structural genes, of amylase loci. Genetics 108: 879–896

Messmer MM, Melchinger AE, Lee M, Woodman WL, Lee EA, Lamkey KR (1991) Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor Appl Genet 83:97–107

Messmer MM, Melchinger AE, Boppenmaier J, Brunclaus-Jung E (1992) Relationships among early european maize inbreds. I. Genetic diversity among flint and dent lines revealed by RFLPs. Crop Sci 32:1301–1309

Messmer MM, Melchinger AE, Hermann RG, Boppenmaier J (1993) Relationships among early european maize inbreds. II. Comparison of pedigree and RFLP data. Crop Sci 33:944–950

O'Farrell PH (1975) High resolution two-dimensional gel electrophoresis of proteins. J Biol Chem 250:4007–4021

Qifa Zhang, Saghai Maroof MA, Kleihofs A (1993) Comparative diversity analysis of RFLPs and isozymes within and among populations of Hordeum vulgare ssp. spontaneum. Genetics 134: 909–916

Rogers JS (1972) IV. Measures of genetic similarity and genetic distance. Studies in genetics VII. University of Texas Publishers, Austin, Texas 7213:145–153

SAS/STAT User's Guide, Release 603 Edition Cary (1988) SAS Institute Inc., Cary, North Carolina

Smith JSC, Smith OS (1991). Restriction fragment length polymorphism can differentiate among U.S. maize hybrids. Crop Sci 31:893–899

Smith JSC, Smith OS (1992) Measurement of genetic diversity among maize hybrids; a comparison of isozymic, RFLP, pedigree, and heterosis data. Maydica 37:53–60

Smith OS, Smith JSC, Bowen SL, Tenborg RA, Wall SJ (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, heterosis, and RFLPs. Theor Appl Genet 80:833–840

Stuber CW, Wendel JF, Goodman MM, Smith JSC (1988) Techniques and scoring procedures for starch gel electrophoresis of enzymes from maize (*Zea mays* L.). Tech Bull 286. Res Service, North Carolina State University