Original Paper

Open Access

Indel and single nucleotide variations of zeins generate unique 2D-zein patterns and molecular markers useful in maize (*Zea mays*) genotyping

Sara Piccinini[1], Alfonso S Negri[2], Massimiliano Lauria[1], Angelo Viotti[1]*

[1]Istituto di Biologia e Biotecnologia Agraria, CNR, Via Bassini 15, 20133 Milano, Italy
[2]Dipartimento di Scienze Agrarie ed Ambientali, Università di Milano, Via Celoria 2, 20133 Milano, Italy
*Corresponding author: E-mail: viotti@ibba.cnr.it

Abstract

In this study, we investigated the inter- and intra-genomic sequence variation of alpha-zein genes and their polypeptide expression in different maize genotypes, i.e. inbreds and a set of Lombardy open pollinated varieties, by analyzing their RFLP, coding nucleotides and 2-dimensional (2D) protein fractionation profiles. An extensive analysis of coding capacity of alpha-zein sequences in various genotypes and in the B73 reference inbred allowed us to assign 2D-spots to specific zein sequences. Moreover, we found that some genes reported to contain in frame stop codons are very likely expressed. Collectively these data allowed us to constitute two barcodes respectively based on nucleotide variation and on 2-D protein patterns that identify univocally each genotype.

Keywords: zein gene RFLP, 2D zein polypeptides, stop codons, barcoding, genotyping

Introduction

In the maize kernel, storage protein genes are expressed only in the maize endosperm where their products are exclusively accumulated during grain filling. Those genes encoding for those polypeptides belong to two biochemical classes, i.e. glutelin and prolamine, differing for their peculiar solubility in alcoholic solvents in presence or absence of S-S bridges reducing agent (Soave et al, 1975; Landry and Moureaux, 1980). Glutelins were further subdivided in G1L plus G1S, G2 and G3 sub-classes currently termed gamma- beta- and delta-zeins, respectively, while zeins in sensu stricto i.e. the prolamine class was termed alpha-zein (see Lopes and Larkins, 1998, for a review).

The alpha-zeins were one of the first cloned and molecularly characterized genes and polypeptides in plants (Righetti et al, 1977; Burr et al, 1982; Rubenstein et al, 1982; Viotti et al, 1982. Furthermore, data on nucleotide hybridization and polypeptide analyses clearly indicated the occurrence of a multigene family divided into various subfamilies. (Righetti et al, 1977; Viotti et al, 1978, 1979). Additional stidies about their accumulation in sub-cellular storage organelles surrounded by membranes of rough endoplasmic reticulum (RER) suggested in the zein polypeptides the presence of a signal peptide responsible for their targeting to the endomembranes (Burr et al, 1978). This observation was further confirmed by nucleotide sequence and amino acid comparison analyses to occur at the amino terminus with a constant length of 21 amino acids for all the alpha-zeins (Nien-Tai et al, 1982; Viotti et al, 1982). Similarly, the coding capacities of the mature alpha-zeins were shown to range between 212 and 246 amino acids (Burr et al, 1982; Spena et al, 1983; Larkins et al; 1985, Viotti et al, 1985).

Their wide spectrum in size, charge and their different relative abundance of zeins among genotypes, have indicated a complex system of transcriptional/expression control in time and space (Dolfini et al, 1992; Ciceri et al, 2000; Song and Messing, 2003). Furthermore, detailed nucleotide analyses of both genomic and cDNA sequences of inbreds also revealed that the zein genes are clustered in few chromosomal regions and that a collinear relationships did not always exist among genotypes (Song and Messing, 2002, 2003). In addition, it was found that the occurrence of single nucleotide variations and in frame short- or long-indel within each sub-group modify their coding capacities and charge heterogeneity. In this context, it was also noted that the nucleotide sequence identity, rather than the length in their coding capacity, had permitted to classify zeins into four subfamilies SF1, SF2, SF3, and SF4 (Larkins et al, 1985; Viotti et al, 1985). Among these subfamilies the latter one was the most abundant with about 20 copies per haploid genome (Song and Messing, 2003). However, because of partial chromosomal duplication events within the alpha-zein cluster, the total copy number of each subfamily was shown to vary among inbreds (Song and Messing, 2003). Studies, by comparing genes and alleles among inbreds and collecting data on cDNA libraries, indicated that the zein system shows the occurrence of in frame pre-stop codons in sequences from genomic fragments (Song and Messing, 2002, 2003). Therefore, this amplifies, by consequence, the heterogeneity of the

RECEIVED 07/06/2014

nucleotide sequences, the polypeptide expression and the 2D zein patterns. Accordingly, in this study we have investigated the zein sequences and polypeptides heterogeneities among a set of maize inbred lines and open pollinated varieties of maize. Our results indicated that the zein system is a valuable tool useful for maize genotype identification in maize. In addition, the analyses on the 2D fractionations of alpha-zein polypeptides compared to genomic sequences of two elite inbred lines, indicate that some genes characterized by the presence of stop codons in the coding sequence are expressed.

Materials and Methods

Plant Material and Growth Conditions

For the different genotypes used in this study, plants were grown in the field or in the greenhouse during 2009 and 2010. Few plants of each genotype were used to obtain immature ears, 4-5 cm long, to extract DNA. Seeds were harvested at maturity. A list of the genotypes used in this study is reported in Supplementary Tables 2 and 3.

Nucleic Acid Isolation and DNA Blot Analysis

DNA was extracted and purified as described (Dolfini et al, 1992; Bernard et al, 1994). Purified DNA was digested to completion (Bianchi and Viotti, 1986) and DNA gel blot analysis were carried out as previously reported (Bernard et al, 1994). Fractionation of digested DNAs was in 0.8-1.0% agarose gel in TAE buffer. Probes were labelled with the Rediprime Kit according to the manufacturer procedure (GE Healthcare) by addition of [α-32P]dCTP. The zein cDNA (M6, E19, and M1) probes were those already described (Viotti et al, 1985; Dolfini et al, 1992). The D3 cDNA probe was obtained from cloning and analyses of cDNA from endosperm RNA of W64a as given by Viotti et al (1982). Hybridization of the various genotypes occurred with either the prototype sequences or the amplicons of the different subfamilies obtained and characterized as described in the following section. Washing conditions of the nucleic acid filters were performed at high stringency (Viotti et al, 1982, 1985), which discriminates identity lower than 96%. Filters were exposed to X-ray film with or without intensifying screens or by their scanning with Starion (Elite Healthcare).

DNA amplification and analysis of amplicon specificity

Amplified zein sequences of the four subfamilies, i.e. SF1, SF2, SF3, and SF4, were obtained using DNA of the W64a inbred and purified as described above. Specific primers of each subfamily are reported in Supplementary Table 1. Fractionation occurred in 2.0 NuSeive agarose gel (FMC) in TBE buffer. The amplified fragments were checked for their group identity using digestion of part of the total amplified fragments with specific restriction enzymes. For instance, all the SF4 have only one *HincII* site that is absent in all the other three subfamilies. Similarly, all the SF2 have only one *BamHI* site absent in all the other three subfamilies. For SF1 or SF3 we used *PstI* and *VspI*, respectively. In all instances, we made a quadruple check of digestions for each subfamily with the above-mentioned restriction enzymes to ascertain presence or absence of digestion or even generation of different fragment sizes, as it occurs for the SF1 and SF4 when using *PstI* that is present only in some SF4 sequences while occurring in all the SF1.

PCR polymorphism of B73 accessions

Four forward primers spanning the 5' coding re-

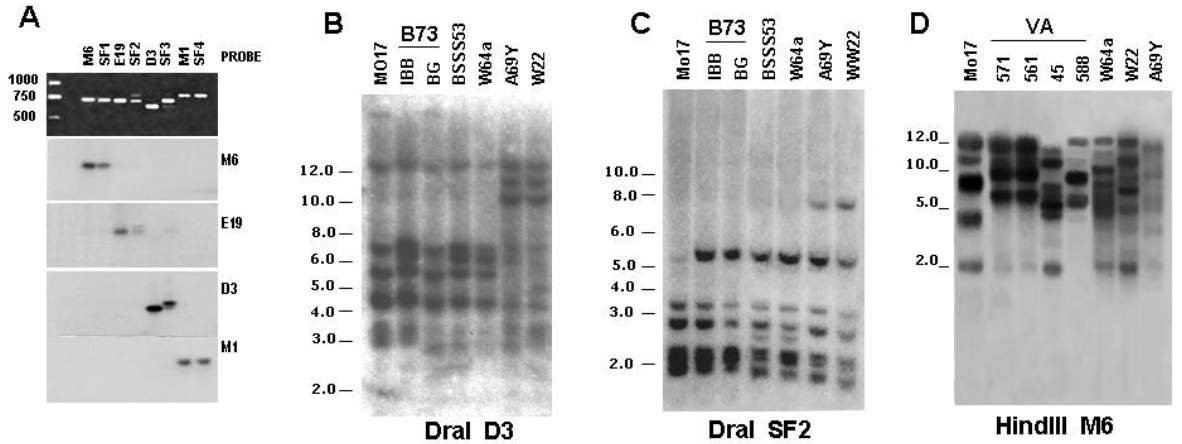


Figure 1 - Analysis of probe specificity and RFLP. **A**: Four gels were run in parallel. Each gel was loaded with the four prototype sequences M6, E19, D3, and M1 together with the amplified fragments that identify each subfamily, respectively, SF1, SF2, SF3, and SF4, as indicated on the top. DNAs were transferred to filters and each filter was hybridized to probes as indicated to the right. Markers in bp are reported on the left. **B**, **C**, and **D**: DNA gel blot analysis of genomic DNA of different genotypes digested with different restriction enzymes and probed as indicated at the bottom of each panel. Varieties are indicated with their number as reported in Supplementary Table 1. Size markers in kb are indicated on the left.

gion of different alpha-zein genes were constructed together with two as reverse primers designed in the mature coding part considering published sequences of both the B73 and BSSS53 inbreds (Woo et al 2001; Song and Messing, 2003). They are reported in Supplementary Table 1. Two Taq systems were used: GoTaq (M7112, Promega) and 5' Hot start (5Prime). In both a final concentration of 2.5 mM Mg ion was used. PCR reactions were carried out in the presence of 2% deionised formamide with annealing temperature at 58°C for 33 cycles. Other conditions were according to manufacturer protocols. Fragments were resolved in 2.2% Metaphore agarose gel (FMC) in TBE buffer.

Storage protein extraction and electrophoresis analysis

Extraction of glutelin and prolamine from mature seeds (five-six per each genotype from the central part of the ear) was carried out as described in Lund et al, (1995) in the presence of 2% beta-mercaptoethanol in 70% ethanolic solution. Mono-dimensional (1D) fractionation by SDS-PAGE and two-dimensional (2D) by IEF/SDS-PAGE were carried out essentially as described (Lund et al, 1995). IEF markers used in 2D fractionation were purchased from GE-Healthcare. Gels were stained with Comassie Brilliant Blue R250.

Nucleic acid and protein Barcodes

The obtained DNA banding pattern for presence or absence of each fragment was converted in a binary code and then into a barcode using the free online software barcode.tec-it/com. The 2D gels were scanned in an Epson Expression 1680 Pro Scanner and analyzed with ImageMaster 2-D Platinum Software v6.0 (GE Healthcare Life Sciences, USA). Automatic matching was complemented by manual matching. Molecular weights of the spots were estimated using the migration range standard as reported in Viotti et al (1982), while pI was determined as described in Righetti et al (1977).

Results

Zein probe recovery and specificity

In our analysis we refer to zein prototypes the M6, E19 and M1cDNA sequences, belonging to alpha-zein genes encoding for polypeptides of the subfamilies SF1, SF2 and SF4, respectively. (Viotti et al, 1982, 1985). Recently, during a screening of clones

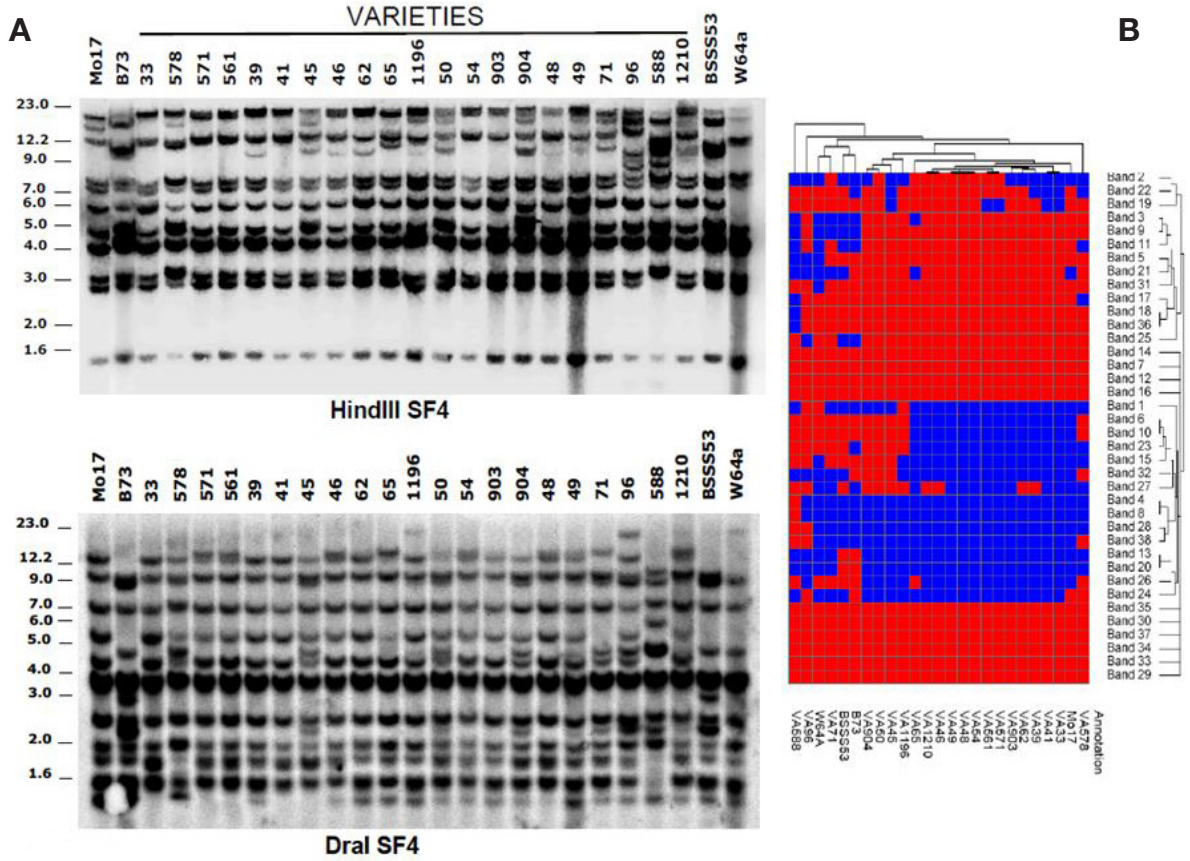


Figure 2 - RFLP of maize varieties and inbreds. **A**: DNA gel blot analysis of genomic DNA of the varieties and four inbreds as reference. DNAs were digested and probed as indicated at the bottom. Varieties are indicated with their number as reported in Supplementary Table 1. Size markers in kb are indicated on the left. **B**: Heatmap of the presence/absence of the RFLP bands obtained from the blots in **A**.

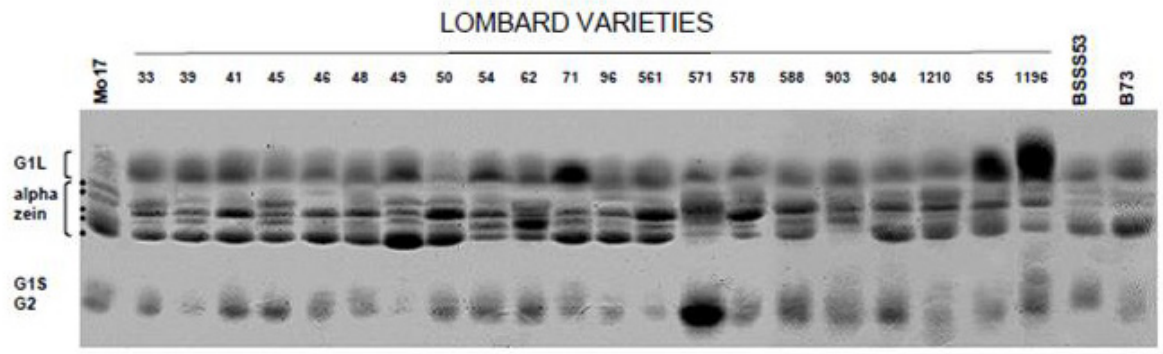


Figure 3 - Mono-dimensional (1D) fractionation of ethanol extracts. Zeins were from mature seeds of maize varieties and of three inbreds. On the left relative mobility of mature gamma-zein (long type, G1L), alpha-zein, gamma-zein (short type, G1S), and G2 are indicated. Varieties are indicated with their number as reported in Supplementary Table 1. Dots on the left indicate the position of the five major size classes of alpha-zeins.

from the cDNA library of the W64A inbred (Viotti et al, 1982), we identified a clone, D3, belonging to the SF3. This clone and a cDNA clone from the inbred Ohio43 (GeneBank accession AI677029.1) represent variants of the SF3 subgroup, containing a deletion of 93 nucleotides when compared with other sequences of the same subfamily (Supplementary Figure1). Apart from the internal deletion, D3 has an identity of 99.6% to the az19D1 cDNA (GeneBank accession NM001111586.1).

We designed SF-specific primers for each zein prototype at the 5' and at the 3' ends of the coding sequences (Supplementary Table1) and used them to amplify, form genomic DNA, PCR products that represent the entire set of sequences of each subfamily. Amplification results are reported in Figure 1A together with the results of their hybridization specificity as determined by DNA gel blot analysis. The specificity and uniqueness of sequences is further supported by results of specific restriction enzyme analysis as reported in Materials and Methods.

Genomic analysis of alpha zein genes in inbreds and varieties

The PCR products of each SF together with the four zein prototypes were used as probes in DNA blot experiments aimed to characterize the extent of DNA polymorphisms at the alpha zein loci among a set of Lombard open pollinated varieties. A preliminary analysis was performed on a subset of these varieties and elite inbred lines, representing different maize breeding group, to identify the more informative probe-restriction enzyme combination for fingerprinting purpose. The list of the various genotypes is reported in Supplementary Tables 2 and 3. DNAs was extracted from sporophytic tissue and digested with the *HindIII* and *DraI* restriction enzymes. These enzymes were chosen for the analysis as they do not digest DNA within the zein coding region and are insensitive to DNA methylation (Bianchi and Viotti, 1988). Although each probe-restriction enzymes combination produces DNA blot patterns that are specific for each zein SF (Figure 1B, C, D and data not shown), we found that only the SF4-HindIII and SF4-DraI combinations were the more informative (Figure 2A). Indeed, the analysis of the entire set of Lombard varieties with these two probe-restriction enzymes combinations permitted to generate on the basis of presence/absence of the various bands, a binary code, this was transformed into a barcode that identified univocally the twenty-five genotypes investigated herein (Supplementary Table 4). The heatmap diagram depicts the uniqueness and creates a relationship among the various genotypes (Figure 2B).

Zein polypeptide analysis by 1D and 2D fractionations

The set of the Lombard varieties and three or four inbred lines were fractionated by mono-dimensional (1D) (SDS-PAGE) and two-dimensional (2D, IEF/SDS-PAGE) analyses. As expected on the basis of the previous results (Viotti et al, 1982; Ciceri et al, 2000) each genotype showed different and specific 1D pattern of the five size classes of alpha-zein (Figure 3). Some genotype as Va46 misses the H2 and few other genotypes some bands in the light size classes. The G1L or gamma-zein-27 migrates as a broad band in 1D and as a number of smeared spots in 2D (Figures 3, 4A and Supplementary Figure 2). The two other glutelins, G1S and G2, migrate faster and generate different spots. The fractionations of the Lombard varieties show that none of the 1D patterns (Figure 3) or the 2D patterns (Figure 4A and Supplementary Figure 2) have either identical relative abundance of the five size classes or 2D-spots distribution, with a number of spots ranging from eight to thirteen for VA54 and W64a, respectively.

In the 2D analysis five IEF markers were properly chosen and added to each extract (roman numbers). The 2D spot distribution of each genotype was compared to that of the W64a considering the position of the five IEF markers and that of the two spots of G1S (A and B as in the panels of Figure 4A and Supplementary Figure 2). The software program, as reported

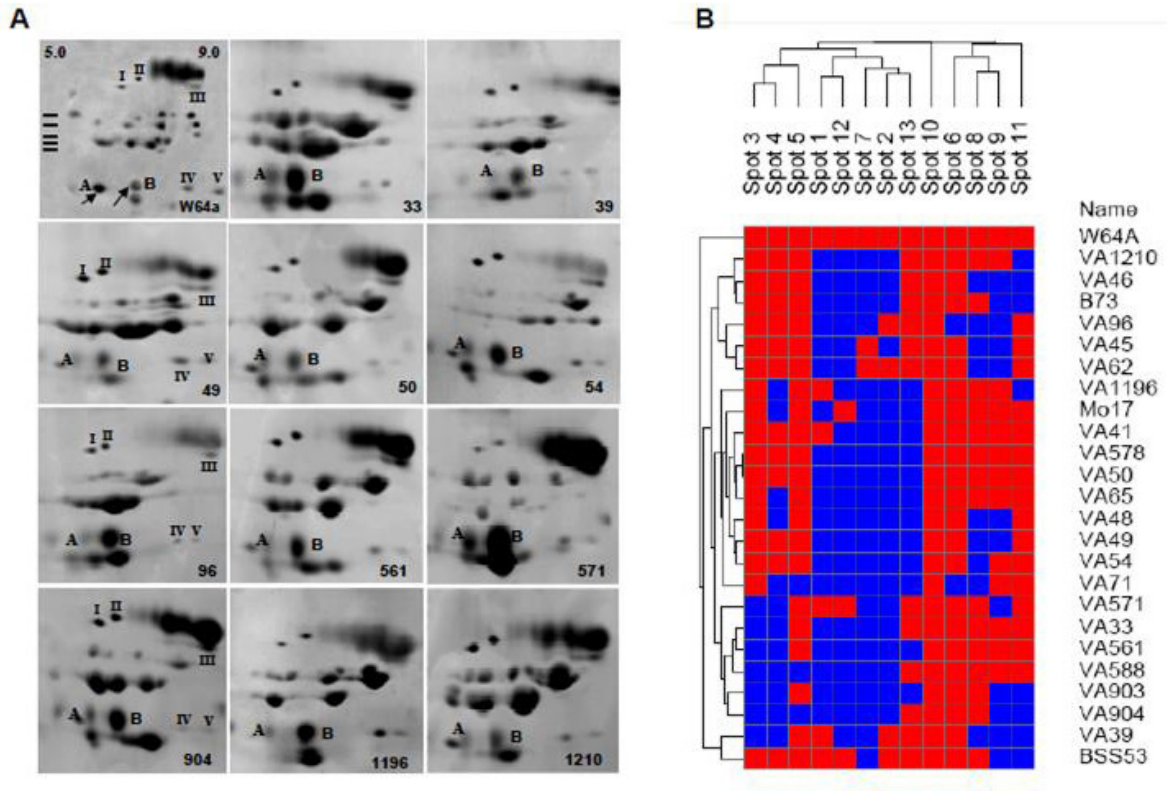


Figure 4 - Two-dimensional (2D) fractionation of storage protein polypeptides.
A: Total ethanol extracts from mature seeds of maize varieties and the W64a inbred were subjected to isoelectrofocusing in the first dimension and then to SDS-PAGE in the second dimension. Loading was about 130 μg for each genotype. The pH ranges are reported on the top of the upper left panel and apply for all the panels. Roman numbers in the four left panels identify IEF markers. In each panel the position of the two polypeptides of the short gamma-zein (G1S) is also indicated (A and B) that together with the IEF markers are utilised for proper positioning of the alpha-zein spots. Dashes in the upper right panel identifies from top to bottom the five major size classes of alpha-zein polypeptides, H1, H2, L1, L2 and L3, respectively. **B**: Heatmap of the 2D fractionations of the twenty-five genotypes reported in **A** and Supplementary Figure 3.

in Materials and Methods, applied to the twenty-five panels of the two figures allowed us to obtain a binary code (Supplementary Table 5) and a cluster reported in Figure 4B. As for the DNA RFLP each genotype can be univocally identified.

Correlation of 2D spots to sequenced alpha-zein genes of heavy type and SNP analysis in B73 accessions

This broad variation in 2D patterns prompted us to verify in different accessions and year harvesting of B73 the expression stability of zein genes. We considered eight different accessions of B73 inbred line because zein sequences of this inbred were reported from different laboratories (Woo et al, 2002; Song and Messing, 2003). Moreover, for this line a high quality reference genome is available. DNA gel blot analyses (Figure 5A) of the eight accession yielded identical patterns, with the exception of the BG accession that showed in the HindIII-D3 combination an additional band of about 18 kb and a missing band around 4 kb. Similarly, the same accession showed an additional band of about 10.5 kb in the HindIII-SF2 combination. However, in the HindIII-SF4 combination the RFLP was identical for all the eight accessions (data not shown).

2D analysis of the eight accessions revealed a constant pattern in term of spot distribution of the H1 and H2 size classes, although, some differences in their intensities were observed. Six spots are present in the H1 region and three spots in the H2 region (Figure 5B). This pattern was compared to that of the BSSS53 inbred, whose entire gene-cluster coding for H1 and H2 was sequenced (Song and Messing, 2003). The spots of the two inbreds were compared for their relative position and for the identification of genes responsible for their expression (Table 1). Some of those genes were reported to contain in frame stop codons (Song and Messing, 2003), however, their substitution with a coding codon generates polypeptides that are present with a good correspondence in the 2D patter, moreover, the intact gene 22.12 does not generate a transcript in the reported expression analysis (Song and Messing, 2003). For instance, the 22.20 gene present in both the inbreds BSSS53 and B73 corresponds as coding characteristics to spot -1 in the former and spot 1 in the latter (Figure 5B

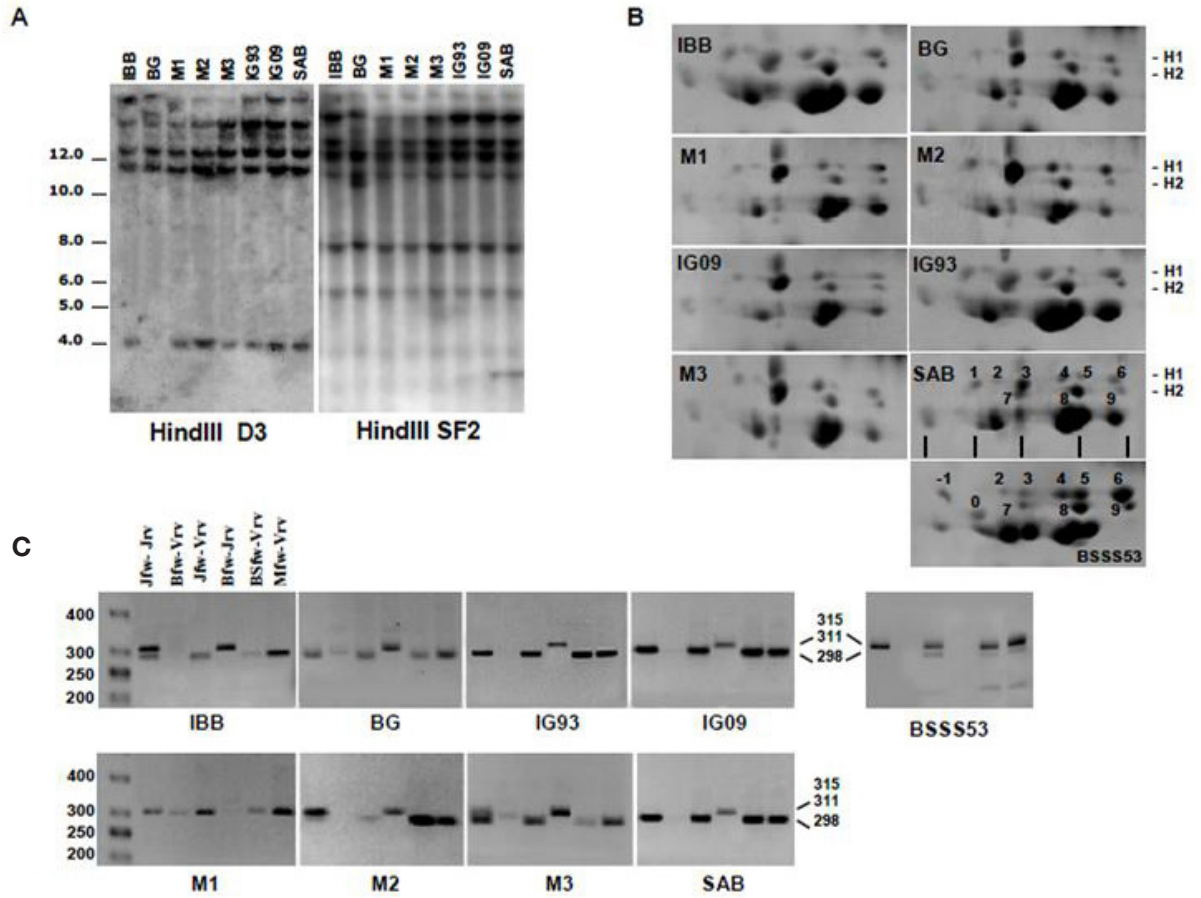


Figure 5 - Genomic DNA and alpha-zein analyses of B73 accessions. **A**: Genomic DNA of the eight accessions was digested and probed as indicated at the bottom of each panel. Size markers in kb are indicated on the left. **B**: Fractionation by 2D of alpha-zein of the eight B73 accessions and of the Bsss53 as comparison. The two size classes of heavy type of alpha-zein are indicated on the right. In the SAB and BSSS53 panels the spots of H1 and H2 are numbered. These two panels were used for comparison and correlation analyses of spots and coding capacity of genes, see Table 1. In SAB panel the five vertical bars indicate from left to right pH 5.0, 6.0, 7.0, 8.0, and 9.0, respectively. **C**: The amplified fragment from the B73 accessions and BSSS53 was obtained and fractionated as reported in Materials and Methods. On the upper left panel combinations of primers, as listed in Supplementary Table 1, are indicated and apply for all the panels. On the right or on the left of the panels size markers are indicated in bp. The abbreviated name of the eight B73 accessions is reported in SupplementaryTable 3.

and Table 1, see note and comments). In analyzing the entire set of genes we succeeded to assign most of the spots to the published sequences. This type of analysis indicates: i) that the substitution of single base in generating in frame stop codons may have derived during bacterial cloning or ii) the absence of transcript for intact genes or alleles, as in the case of the *22.12* (spot 2 in both inbreds), may reflect mRNA paucity at the developmental stage tested, or even error in their identification (Woo et al, 2001; Song and Messing 2003; Luo et al, 2008; Feng et al, 2009).

In comparing alpha-zein sequenced genes of B73 present in the MaizeGDB and NCBI data Banks we noticed several nucleotide variation of the same allele. For instance the *floury2* locus has been reported to code for a mature zein polypeptide of 241 aa (Woo et al, 2001) or 242 aa (Song and Messing, 2003; MaizeGDB). Similarly, the *az22z1* gene (GenBank Accession AF371274; Woo et al, 2001) codes for a mature polypeptide of 242 aa, a unique sequence among the other annotated genes of the two B73 (Song and Messing, 2003; MaizeGDB). Comparison of the 5' nucleotide end of their coding region showed several SNPs for the same allele from the various data banks and even between alleles of the B73 and BSSS53 lines, which allowed the construction of different primers (Supplementary Table 1) with specific SNPs used in PCR reaction and challenged against the DNAs of the eight B73 accessions (Figure 5C). As expected, the patterns of amplification of the eight accessions had two main bands, one around 311-315 nt and the other at 298 nt, respectively. By comparison also the BSSS53 was considered. This line manifested an additional band of 227 nt that derives from the amplification of the *Zp22/D87* gene, absent in B73 (Song and Messing, 2003), and occured only with two combinations of primers. In comparing the patterns of the eight B73 and starting from the left

couple of primers (Figure 5C) it was noted: i) the first couple generates in some cases a double band (IBB, M2 and M3) or a single band (the fast migrating) in all the others; ii) the second couple a faint band in BG and M3, or its absence in all the others apart M1; iii) in M1 only the upper band in all the primer combination; iv) the last four couples of primers generate an identical pattern for the eight B73 apart M1.

Discussion

Our data from RFLP and 2D fractionation assayes, with the generation of the corresponding barcode, identify each genotype as unique. These findings indicate that the alpha-zein system has acquired during generations a large heterogeneity generating a broad diversification and therefore yelding specific haplotype (Song and Messing, 2003). The D3 clone represents such an example as it is recovered into different inbreds, namely W64a and Ohio43. A further indication of the broad diversification occurring among genotypes derives from the analysis of two cDNA clones azs1-42 and azs2-2 (GeneBank accession AJ491308 and AJ491309, respectively) that were previously identified in the NYRo2 inbred line (Ciceri et al, 2000). These sequences have a 98% identity to those of the BSSS53 Zp22/6 and Zp22/D87, respectively. The azs1-42 produces a polypeptide in the position of spot 6 (as Zp22/6 does, Figure 5B) and the azs2-2 a spot toward more acidic pH, pI to 5.44, in respect of Zp22/D87 pI 6.74, because azs2-2 contains the aa substitution 42A>E (Lauria and Viotti, unpublished). This single-nucleotide mutation, as for B73 and BSSS53 genes, may derive from the proposed mutational mechanism associated to insertions/deletions occurring close to genes (Tian et al, 2008; Hollister et al, 2010), as was showed for absence of collinearity and frequent indels in zein gene clusters (Song and Messing, 2002, 2003).

The dendograms of the RFLP bands and 2D patterns (Figures 4B and 5B) generate a different relationship among genotypes because they are based onto different types of data that clearly discriminate zein alleles: inter genic variation and gene expression. In the eight accessions of B73 the inter genic sequences, at least for the three restriction enzymes used (Figure 4A, for *EcoRI* data not shown), were almost consistent in their RFLPs, whereas differences were observed in the coding sequences by PCR analysis (Figure 5C). This suggests that zein intra genic sequences may vary with higher frequency in respect of inter genic regions generating transcript isoforms specific for each genotype. Accordingly, 2D fractionation (Figure 4A and Supplementary Figure 2) of the twenty-five genotypes showed a number of alpha-

Table 1 - Correlation of 2D alpha zein spots and gene coding capacity and characteristics.

BSSS53 spot*	pI/MW[&]	gene/ transcript*	note	B73 spot*	pI/MW[&]	gene/ transcript*	note[$]
-1	5.27/26740	22.20/no	stop[a]	1	5.99/26740	22.20/no[a]	stop[a]
2	6.91/26709	22.12/no	intact[b]	2	6.91/26709	22.12/no	intact[b]
3	7.01/26852	22.21/no	stop	3	7.01/26760	22.9/yes	intact
4		NI[c]		4	8.09/26877	22.8/yes	intact
5	8.11/26891 8.14/26700	22.8/yes 22.10/yes	intact intact	5	8.11/26838	22.19/yes	intact
6	~8.94/~26800	22.14+22.4 +Zp22/6[e]/yes	intact	6	8.95/26760	22.4/yes	intact
7	7.01/~26500	NI		7	7.01/~26500	NI	
8	8.11/26358	fl2/yes	intact	8[d]	8.11/26358	fl2/yes	intact
9	8.95/26500	NI		9	8.95/26500	NI	
10	6.74/23527	Zp22/D87[e]/yes	intact		//	Zp22/D87[e]	absent

*The gene name and the transcript occurrence were derived from Song and Messing, 2003. It should be noted that transcript level was carried out on endosperm tissue at 18 DAP, while spots are from mature endosperm and are the result of an accumulation process. On the other hand transcript level does not necessarily reflect a proportional amount of polypeptide as translational control of alpha-zein mRNA has been reported Spena et al (1985).

[&]The theoretical Isoelectric point (pI) and the MW were calculated with ExPASy program (http://web.expasy.org/compute_pi/)

[$]Reports the annotations of Song and Messing (2003).

[a]The correction of the stop codons in the BSSS53 and B73 alleles generate in the former a polypeptides with a pI of 5.27 that in respect of the 22.20 allele of B73 contains an aa substitution 157Q>E that modifies its focalization towards more acidic pH (spot -1).

[b] The two alleles are intact, but their transcripts were not detected in the reported analysis (Song and Messing, 2003). In any case their theoretical pI and MW fit spot 2.

[c] Not Identified.

[d]The gene *az22z1* of the B73 reported by Woo et al (2001) has a pI of 8.13 and MW of 26358 that fit spot 8. The other four genes reported in this article, az22-3, -4, -5 and –fl2 fit spots, 6, 3, 2 and 8, respectively (data not shown).

[e]Genes absent in the B73. The Zp22/D87 in BSSS53 contains an internal deletion that generates the 227 nt band reported in PCR analysis of Figure 5B.

zein spots that vary from about eight up to thirteen, with the H1 and H2 size classes showing the most heterogeneous pattern in respect of the group of light size classes. Moreover, so far none of the 2D patterns presently analyzed in our laboratory for more than thirty-five genotypes were shown to be identical to each others (present data, Lund et al, 1995; Ciceri et al, 2000). In addition, the 2D analysis indicates that single aa substitution that involves neutral to charge aa, as in the case of the alleles of the 22.20 gene of BSSS53 and B73, results into a distinct spot position (Figure 5B and Table 1). This is also the case for the genes *Zp22/D87* and *azs2-2* as discussed above.

A further consideration suggested us that some spots of major intensity, as spot 6 in BSSS53 in respect of the same spot in B73, may reflect the expression of three genes versus only one in B73. Moreover, on the basis of notes and considerations reported in Table 1, spot 3 in the BSSS53 and B73 appeared result of the expression of two different genes, *22.21* in the former and *22.9* in the latter: note that the *22.21* is absent in the cluster of B73 and the only one that fit spot 3 has to be the *22.9*. Moreover, the 22.9 sequence of BSSS53 has several stop codons and a short rearrangement at the 3' end that strongly may hamper its expression as polypeptide. On the other hand, some genes of both B73 and BSSS53 were reported to contain only one in frame stop codon that always involve the transition C>T of the two triplets CAA and CAG coding for glutamine (Song and Messing, 2002, 2003). In this analysis we have found that some of them are expressed indicating that base transition may occur during cloning and amplification in the bacterial cell. Collectively, alleles may generate two spots with different pI, but on the other hand one spot of different genotypes with the same pI may be the result of the expression of two or more different genes.

The overall data from the current study clearly indicate that among genotypes the entire set of genes of the alpha zeins has a high degree of variation in the coding region and that even genotypes can accumulate dominant negative mutations (Coleman et al, 1995; Kim et al, 2004; Kim et al, 2006) for these polypeptides important for feed utilization that are, however, not essential for maize reproduction.

Acknowledgements

We thank Dr Mario Motto for helpful comments on the manuscript, Dr Raul Pirona for help in manuscript editing and the MGCSC for providing seed stocks. This research was supported by Regione Lombardia (Italy), agreement Regione/CNR, project 5: ''Nuovi strumenti per la selezione di varietà di mais con migliore valore proteico".

References

Burr B, Burr FA, Rubenstein I, Simon MN, 1978. Purification and translation of zein messenger RNA from maize endosperm protein bodies. Proc Natl Acad Sci USA 75: 696-700

Burr B, Burr FA, St John TP, Thomas M, Davis RW, 1982. Zein storage protein gene family of maize. An assessment of heterogeneity with cloned messenger RNA sequences. J Mol Biol 154: 33-49

Burr FA, Burr B, 1981. In vitro uptake and processing of prezein and other maize preproteins by maize membranes. J Cell Biol 90: 427-434

Burr FA, Burr B, 1982. Three mutations in Zea mays affecting zein accumulation: a comparison of zein polypeptides, in vitro synthesis and processing, mRNA levels, and genomic organization. J Cell Biol 94: 201-206

Ciceri P, Castelli S, Lauria M, Lazzari B, Genga A, Bernard L, Sturaro M, Viotti A, 2000. Specific combinations of zein genes and genetic backgrounds influence the transcription of the heavy-chain zein genes in maize *opaque-2* endosperms. Plant Physiol 124: 451-460

Clark RM, Linton E, Messing J, Doebley JF, 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proc Natl Acad Sci USA 101: 700-707

Coleman CE, Clore AM, Ranch JP, Higgins R, Lopes MA, Larkins BA, 1997. Expression of a mutant alpha-zein creates the floury2 phenotype in transgenic maize. Proc Natl Acad Sci USA 94: 7094-7097

Dolfini S, Landoni M, Tonelli C, Bernard L, Viotti A, 1992. Spatial regulation in the expression of structural and regulatory storage-protein genes in Zea mays endosperm. Dev Gen 13: 264-276

Geraghty DE, Messing J, Rubenstein I, 1982. Sequence analysis and comparison of cDNAs of the zein multigene family. EMBO J 1: 1329-1335

Kim CS, Gibbon BC, Gillikin JW, Larkins BA, Boston RS, Jung R, 2006. The maize Mucronate mutation is a deletion in the 16-kDa gamma-zein gene that induces the unfolded protein response. Plant J 48: 440-451

Kim CS, Hunter BG, Kraft J, Boston RS, Yans S, Jung R, Larkins BA, 2004. A defective signal peptide in a 19-kD alpha-zein protein causes the unfolded protein response and an opaque endosperm phenotype in the maize *De*-B30* mutant. Plant Physiol 134: 380-387

Kirihara JA, Petri JB, Messing J, 1988. Isolation and sequence of a gene encoding a methionine-rich 10-kDa zein protein from maize. Gene 71: 359-70

Landry J, Moureaux T, 1980. Distribution and Amino-Acid-Composition of Protein-Fractions in Opaque-2 Maize Grains. J Agric Food Chemistry 21: 1865-1869

Lopes MA and Larkins BA, 1993. Endosperm Origin, development, and function. Plant Cell 5: 1383-

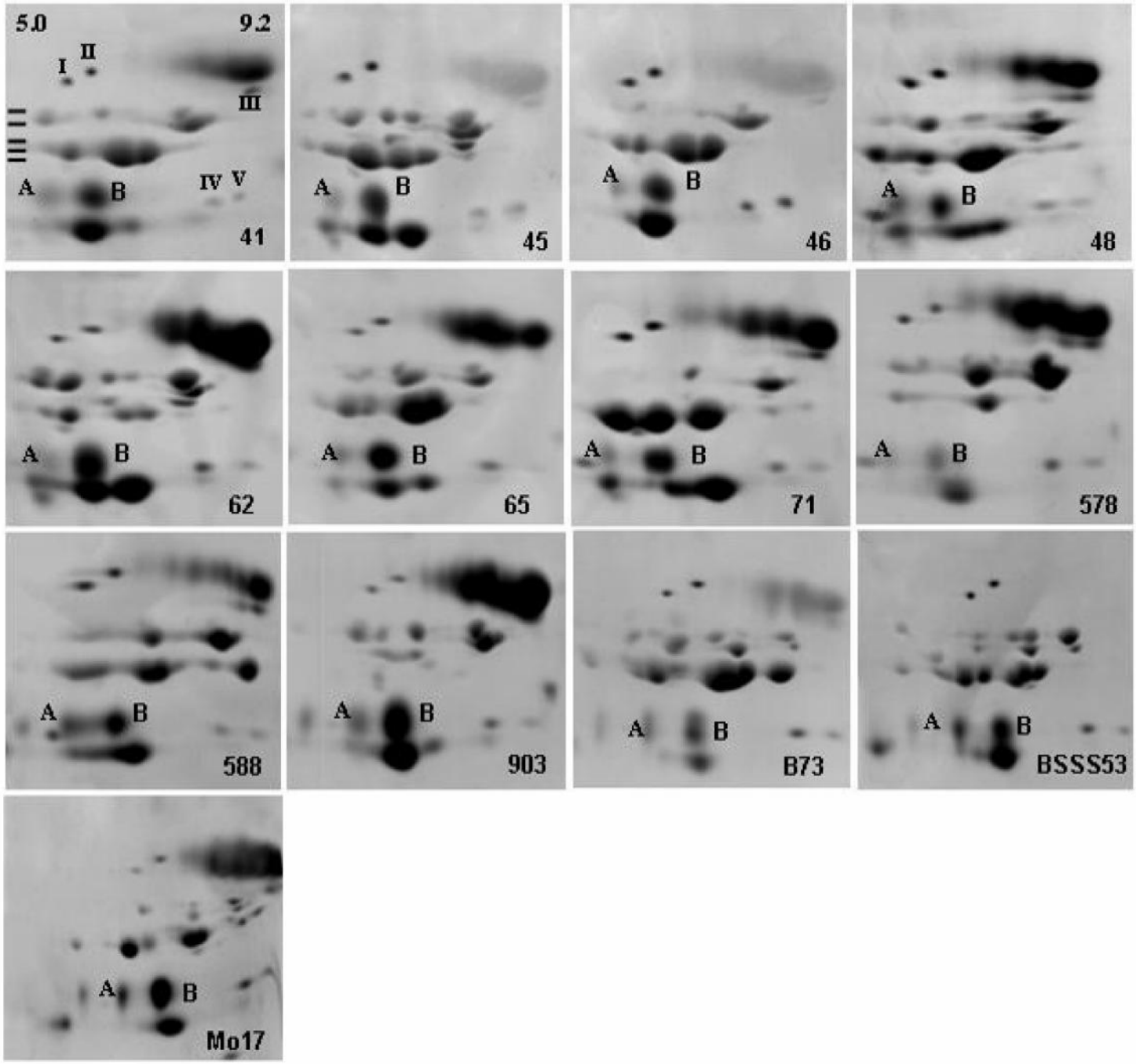
1399
Lund G, Ciceri P, Viotti A, 1995. Maternal-specific demethylation and expression of specific alleles of zein genes in the endosperm of *Zea mays* L. Plant J 8: 571-581
Luo M, Liu J, Lee RD, Guo BZ, 2008. Characterization of gene expression profiles in developing kernels of maize (*Zea mays*) inbred Tex6. Plant Breeding 127: 569-578
Marks MD, Lindell JS, Larkins BA, 1985. Nucleotide sequence analysis of zein mRNAs from maize endosperm. J Biol Chem 260: 16451-16459
Nien-Tai H, Peifer MA, Hidecker G, Messing J, Rubenstein I, 1982. Primary structure of a genomic zein sequence of maize. EMBO J 1: 1337-1342
Righetti PG, Gianazza E, Viotti A, Soave C, 1977. Heterogeneity of storage proteins in maize. Planta 136: 115-123
Soave C, Pioli F, Viotti A, Salamini F, Righetti PG, 1975. Synthesis and heterogeneity of endosperm proteins in normal and *opaque-2* maize. Maydica XX: 83-94
Soave C, Suman N, Viotti A, Salamini F, 1978. Linkage relationships between regulatory and structural gene loci involved in zein synthesis in maize. Theor Appl Genet 52: 263-267
Soderlund C, Descour A, Kudrna D, Bomhoff M, Boyd L, Currie J, Angelova A, Collura K, Wissotski M, Ashley E, Morrow D, Fernandes J, Walbot V, Yu Y, 2009. Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs. PLoS Genet 5: e1000740
Song R, Llaca V, Linton E, Messing J, 2001. Sequence, regulation, and evolution of the maize 22-kD alpha zein gene family. Genome Res 11: 1817-1825
Song R, Messing J, 2002. Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. Plant Physiol 130: 1626-1635
Song R, Messing J, 2003. Gene expression of a gene family in maize based on noncollinear haplotypes. Proc Natl Acad Sci U S A 100: 9055-9060
Spena A, Krause E, Dobberstein B, 1985. Translation efficiency of zein mRNA is reduced by hybrid formation between the 5'- and 3'-untranslated region. EMBO J 4: 2153-2158
Spena A, Viotti A, Pirrotta V, 1983. Two adjacent genomic zein sequences: structure, organization and tissue-specific restriction pattern. J Mol Biol 169: 799-811
Tian D, Wang Q, Zhang P, Araki H, Yang S, Kreitman M, Nagylaki T, Hudson R, Bergelson J, Chen JQ, 2008. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature 455: 105-108
Viotti A, Balducci C, Weil JH, 1978. Adaptation of the tRNA population of maize endosperm for zein synthesis. Biochim Biophys Acta 517: 125-132
Viotti A, Cairo G, Vitale A, Sala E, 1985. Each zein gene class can produce polypeptides of different sizes. Embo J 4: 1103-1110
Viotti A, Sala E, Marotta R, Alberi P, Balducci C, Soave C, 1979. Genes and mRNAs coding for zein polypeptides in *Zea mays*. Eur J Biochem 102: 211-222
Woo YM, Hu DW, Larkins BA, Jung R, 2001. Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. Plant Cell 13: 2297-2317
Wu Y, Goettel W, Messing J, 2009. Non-Mendelian regulation and allelic variation of methionine-rich delta-zein genes in maize. Theor Appl Genet 119: 721-731

S_figure 1- Sequence of clone D3 and main features.

Coding region is in capital letter, 627 nt. Letters in italic identify the coding of the signal peptide: mature polypeptide 188 aa, 21145 Daltons. Stop codon in bold. Double slash indicate the occurrence of the 93 nt deletion in respect to the az19D1 (NM_001111586.1). Underlined bases indicate the five nucleotide changes, in bold those that generate the three aa substitutions. The aa sequence of the mature D3 polypeptide has a 99% identity to the GeneBank accession AI677029.1 cDNA clone of the Ohio43 inbred.

acatagtgaagtacatcagcaacatcttagcacca*ATGGGAGCCAAGATTTTTGCCCTCCTTGCCCTCCTTGCTCTTTCAGCAAGCG*
CTGCTACCTCGACTTTTATTCCACAATGCTCACAACAATACCTCTCTCCGGCGACAGCCGTGGGATTTCAATACCCAACTATACAAT
CCTACATGGTACAAGAGGCCATCCAAGCAAGCATCTTACGGTCATTAGCATTAACCCTCCAACAACCATATGCTCTATTGCAACAGC
CATCCTTAGTGCATCTGTATCTCCAAAGAATCGCGGCACAACAACTACAACAACAGTTGCTACCAACAATCAATCAA//GCACAACA
GCTACTACCATTTAACCAACTTGTCGGGAGCCCTTATGCCTTCTTACTGCAACAACAGCTTCTACCATTCCATCTGCAAGCTGTGGC
AAACATTGCTGCTTTCTTGAGACAACAACATTTGTTGCCATTTTACCCACAGGTTGTGGGAAACATTAATGCCTTCTTGCAACAGCA
ACAATTGCTACCATTCTACCCACAGAATGTGGCAAACATTGTTGCCTTCTTACAACAACAACAATTGTTGCCATTTAGCCAACATGC
TTTGACGAATCCTACCACCTTATTGCAACCGCCCACCATTGGTGGTGCCATCTTC**tag**attttttatgatttatactgtaataataa
agttctcatgctgatatgtgcgacctctcagtaataaagtattagagatctatattttaaaaaaaaaaaaaaaaa

S_figure 2. Two-dimensional (2D) fractionation of storage protein polypeptides. Fractionation was carried out as describe in the legend of Fig. 4.



S_Table 1. Primers used in the amplification of alpha-zein subfamilies and analysis of B73 accessions

	Forward 5'-3'	Reverse 5'-3'
SF1	CAATGGCGACCAAGATATTTTCC	CAATCTAAAAGAGGGCACCACC
SF2	ACAATGGCAGCCAAAATATTTTGC	TAAGAAATCTAAAAGAGGGCACC
SF3	CACCAATGGCAGCCAAGATTTTTG	AAATCTAGAAGATGGCACCACCA
SF4	CAATGGCTACCAAGRTATTAKCC	ATGTAATCTAAAAGATGGCACCTC
Jfw	CGCTTCTTGCCCTTTTAGTG	
Bfw	GCTCCTTTCCCTTTCAGTG	
BSfw	CGCTTCTTGCCCTTTTTGTG	
Mfw	GCGATTCTTGCCCTTTTAGTGA	
Jrv		GCTGCTGTTGCAAGTAGGTG
Vrv		CTGCTGTTGCAAGTAGGC

In the SF primers the start codons and the reverse complementary stop codons are underlined.

S_table 2. Maize genotypes (inbreds and varieties)

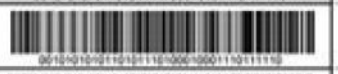
Genotypes	Origin	Name	Abbreviated Name
W64a	MGCSC[1]		
BSSS53	Messing lab.		
B73	Different accessions, see Table S2		See S_table 3
Mo17	MGCSC[1]		
Variety 33	Clusone (Bergamo)	Locale Fiorine	VA 33
Variety 39	Buffalora (Brescia)	Quarantino Nostrano	VA 39
Variety 41	Paderno Franciacorta (Brescia)	Quarantino Nostrano	VA 41
Variety 45	Motta Baluffi (Cremona)	Ottofile Mantovano	VA 45
Variety 46	Stagno Lombardo (Cremona)	Quarantino S. Famiglia	VA 46
Variety 48	Gaidella di Quistello (Mantova)	Quartino Giallo	VA 48
Variety 49w	S Benedetto Po (Mantova)	Cinquantino Bianco	VA 49
Variety 50	Passirana (Milano)	Locale di Passirana	VA 50
Variety 54	Isola Melzese (Milano)	Agostinello	VA 54
Variety 62	Pala (Sondrio)	Nostrano dell'Isola	VA 62
Variety 65	Verceia (Sondrio)	Locale Chiavenna	VA 65
Variety 71	Lonate Pozzolo (Varese)	Agostanello	VA 71
Variety 96	Marano Vicentino (Vicenza)	Marano Vicentino	VA 96
Variety 561	Fontanella Sotto il Monte (Bergamo)	Locale Rostrato	VA 561
Variety 571	Stezzano (Bergamo)	Sintetico Zanchi	VA 571
Variety 578	Torre Boldone (Bergamo)	Rostrano	VA 578
Variety 588	Stezzano (Bergamo)	Microsperma-Pignoletto	VA 588
Variety 903	Alto milanese (Milano)	Cinquantino 2° raccolto	VA 903
Variety 904	Alto milanese (Milano)	Cinquantino 2° raccolto	VA 904
Variety 1196	Chiavenna (Sondrio)	Rostrato di Valchiavenna	VA 1196
Variety 1210	Carenno (Lecco)	Rostrato	VA 1210

[1]Maize Genetics Cooperation Stock Center

S_table 3 Accessions of B73 inbreds

Year harvest	Origin	Abbreviated Name
1999	Ist. Biologia Biotechnologia agraria	IBB
2008	Centro Ricerche Cerealicoltura, Bergamo	BG
2005	MGCSC[1], C736G	M1
1985	MGCSC[1], 3409-95	M2
2006	MGCSC[1], C736G	M3
1993	Ist. Genetica, Milan University	IG93
2009	Ist. Genetica, Milan University	IG09
2002	Dip. Scienze Agrarie, Bologna University	SAB

[1]Maize Genetics Cooperation Stock Center

Inbred lines and Lombard varieties (VA)	Binary Code DraSF4+H3SF4	Barcode
Mo17	00101010101101010111101011000111011111	
B73	00001110010111111111100101001111111110	
VA 33	00101010101101011101000100011101111	
VA 578	00101110110101010110100111001111111111	
VA 571	01101010101101011101100100011101111	
VA 561	01101010101101011111100100011101111	
VA 39	00101010101101011111000101011101111	
VA 41	00101010101101011101000100011101111	
VA 45	00101110111101111101010101011111111110	
VA 46	01101010101100101111111001010111011111	
VA 62	00101010101101011111110010101110111110	
VA 65	01001010101101011110100110011101111	
VA 1196	10101110111101011111111010101110111110	

Inbred lines and Lombard varieties (VA)	Binary Code DraSF4+H3SF4	Barcode
VA 50	01101110111101111111111010101111111110	
VA 54	01101010101101011111100100011101111	
VA 903	00101010101101011111100100011101111	
VA 904	00101110111101111111111010101111111110	
VA 48	01101010101101011101100100011101111	
VA 49	01101010101101011111100100011101111	
VA 71	01001110011101111111011011001110111110	
VA 96	10100110111101111111011100111110111111	
VA 588	00010111010101110010110111111110111011	
VA 1210	01101010101101011111100101011101111	
BSSS53	00001110010111111111011001101110111110	
W64A	10000110010101011110110110011001111110	

S_table 5. Binary code of spots from 2D gels reported in Figures 4 and S2

Genotype	Spot number 1 2 3 4 5 6 7 8 9 10 11 12 13	Genotype	Spot number 1 2 3 4 5 6 7 8 9 10 11 12 13
W64a	1111111111111	48	0010110001100
1210	0011110111001	49	0011110001100
46	0011110001001	54	0011110011100
B73	0011110101001	71	0010000011100
96	0111100001101	571	1000110101111
45	0011111001101	33	0000110111101
62	0111111001101	561	0000110111100
1196	1010110111000	588	0000010111101
Mo17	0010110111110	903	0000110101000
41	1011110111100	904	0000010101001
578	0011110111100	39	1100110001001
50	0011110111100	BSSS53	1111110101011
65	0010110111100		