

# Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize $\alpha$ -zein super gene family

Lingna Feng · Jia Zhu · Gang Wang · Yuanping Tang · Hanjun Chen · Weibo Jin · Fei Wang · Bing Mei · Zhengkai Xu · Rentao Song

Received: 12 February 2008 / Accepted: 5 December 2008 / Published online: 28 December 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** The  $\alpha$ -zein super gene family encodes the most predominant storage protein in maize (*Zea mays*) endosperm. In maize inbred line B73, it consists of four gene families with 41 member genes. In this study, we combined quantitative real-time PCR and random clone sequencing to successfully profile the expression of  $\alpha$ -zein super gene family during endosperm development. We found that only 18 of the 41 member genes were expressed, and their expression levels diverge greatly. At the gene family level, all families had characteristic “up-and-down” oscillating expressional patterns that diverged into two major groups. At the individual gene level, member genes showed dramatic divergence of expression patterns, indicating fast differentiation of their expression regulation. A comparison study among different inbred lines revealed significantly different expressed gene sets, indicating the existence of highly diverged haplotypes. Large gene families containing long gene clusters, e.g. z1A or z1C, mainly contributed the highly divergent haplotypes. In addition, allelic genes also showed significant divergence in their expressional levels. These results indicated a highly dynamic and fast evolving nature to the maize  $\alpha$ -zein super gene family, which might be a common feature for other large gene families.

**Keywords** *Zea mays* ·  $\alpha$ -zein · Expression profile · Super gene family · Endosperm development

## Abbreviations

DAP Days after pollination

O2 Opaque2, transcriptional factor for z1C gene family

## Introduction

A gene family is a group of genes that has descended from a common ancestral gene and therefore inherits similar DNA sequences and similar functions. In *Escherichia coli*, >50% of the genes are members of identified gene families (Koonin et al. 1998), and the proportion of gene family members in eukaryotes may be in the same range or even higher (Chervitz et al. 1998; Semple and Wolfe 1999). Multi-gene families are a distinctive feature of plant genomes. For instance, 65% of *Arabidopsis thaliana* genes belong to families, with half of them containing at least five members (see The Arabidopsis Genome Initiative 2000). In the rice genome, about 29% of the genes belong to gene families (International Rice Genome Sequencing Project, 2005), and in maize, it was estimated that 1/3 of the genes were members of gene families (Messing et al. 2004). The evolution of gene families has been the subject of controversy for many years (Nei and Rooney 2005). Therefore, expression divergence between member genes may be useful evidence in determining how gene families evolved. Such evidence requires extensive expressional analysis of large gene families.

The relative expression and amino acid composition of seed storage proteins significantly impact the nutritional value of maize as animal feed (Ueda and Messing 1993).

---

Lingna Feng and Jia Zhu contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11103-008-9444-z) contains supplementary material, which is available to authorized users.

---

L. Feng · J. Zhu · G. Wang · Y. Tang · H. Chen · W. Jin · F. Wang · B. Mei · Z. Xu · R. Song (✉)  
Shanghai Key Laboratory of Bio-energy Crop, School of Life Sciences, Shanghai University, 99 Shangda Road, Shanghai 200444, People's Republic of China  
e-mail: rentaosong@staff.shu.edu.cn

**Table 1** Classification of major zein proteins in maize

| Zein classes <sup>a</sup>   | Molecular mass (SDS-PAGE) (kD) | Gene copy number <sup>a</sup> | Map location      |
|-----------------------------|--------------------------------|-------------------------------|-------------------|
| $\alpha$ -zein <sup>b</sup> | z1A                            | 9                             | 4.02 <sup>b</sup> |
|                             |                                | 3                             | 4.04              |
|                             | z1B                            | 8                             | 7.01              |
|                             |                                | 15                            | 4.02              |
|                             | z1C                            | 1                             | 4.03              |
| z1D                         | 5                              | 1.06                          |                   |
| $\beta$ -zein               | 15                             | 1                             | 6.01              |
| $\gamma$ -zein              | 16                             | 1                             | 2.07              |
|                             |                                | 1                             | 7.05              |
|                             | 50                             | 1                             | 7.05              |
| $\delta$ -zein              | 10                             | 1                             | 9.03              |
|                             | 18                             | 1                             | 6.04              |

<sup>a</sup> Based on B73 inbred line

<sup>b</sup> 4.02: second bin on chromosome 4

Maize endosperm, the major site of storage protein accumulation, contains predominantly prolamins, the so-called zeins (Landry and Moureaux 1970; Wilson 1983). Zeins are one of the best-characterized sets of storage proteins specifically expressed during seed development and act as a reservoir for free amino acids. Zeins are about 50% of the total proteins in mature seeds (Soave et al. 1981) and 62–74% of the endosperm proteins (Hamaker et al. 1995; Landry et al. 2000). Classification of various zein proteins was on the basis of solubility and structural relationships. Zein proteins were classified as  $\alpha$ -,  $\beta$ -,  $\gamma$ - and  $\delta$ -zeins (Table 1, Coleman and Larkins 1998; Esen 1987).

The  $\alpha$ -zeins are the major fraction of maize zeins, and were among the first storage protein genes to be described (Pedersen et al. 1982; Wienand et al. 1981). Early studies showed that  $\alpha$ -zeins are encoded by a large super gene family, with estimates ranging from 75 to 100 coding sequences (Hagen and Rubenstein 1981; Wilson and Larkins 1984). Determined by sequence similarity and copy number,  $\alpha$ -zein gene family could be further divided in to four gene families: z1A, z1B, z1C, and z1D (Table 1, Heidecker and Messing 1986). Three of the gene families, z1A, z1B, and z1D, have apparent molecular mass on SDS polyacrylamide gels of 19 kD, whereas the z1C gene family has a molecular mass of 22 kD. Based on genomic DNA sequence analysis, it was found that the entire z1C gene family contained a total of 23 gene copies in BSSS53 inbred and 16 in B73 inbred (Song et al. 2001; Song and Messing 2003). While the three 19 kD  $\alpha$ -zein gene families in B73 contained 25 gene copies in total and were composed of five different genomic regions covering a total length of over 1 Mb (Song and Messing 2002). The  $\alpha$ -zein

gene family is a typical super gene family that could be used as a model system for super gene family analysis.

The  $\alpha$ -zein gene family history shows a recent gene amplification. Most of the 22 and 19 kD zein genes arose within the last 3–5 mya (Song et al. 2001; Song and Messing 2002). Such a short evolution history has resulted in high sequence similarity among gene members. In the z1A gene family, 2/3 of the member genes are >95% similar, with 1/3 > 99% similar. In the z1B gene family, the similarities among all the member genes are >90% and three of them are >95%. In the z1D gene family, 40% of the member genes have >90% similarity. The member genes of the z1C gene family are approximately 90% similar (Llaca and Messing 1998; Wilson and Larkins 1984). In an extreme case, two member genes of the z1C gene family—*azs22.10* and *zp22/6* are 99.3% similar with only three nucleotide differences in the BSSS53 inbred line (Song et al. 2001). With such high sequence similarity, even under high stringent conditions, cross-hybridization of cloned zein sequences are reported in dot blot hybridization (Burr et al. 1982). This may explain why few expression analyses could be linked to individual  $\alpha$ -zein genes on the basis of hybridization.

Earlier studies using RNA hybridization-based approaches revealed that there is a temporal regulation of  $\alpha$ -zein gene family expression (Kodrzycki et al. 1989; Marks et al. 1985). Woo et al. (2001) found similar temporal pattern of expression among some members of z1B, z1D and z1C gene families by in situ hybridization at 5, 10, 15, 20 and 25 DAP. Later Song and Messing (2002) found that two highly expressed 19 kD zein genes in the Woo et al. (2001) study (*az19B1* and *az19B3*) were actually a mixture of different expressed genes. Therefore more sensitive methods would be required to distinguish member genes with high sequence similarity. After genomic DNA sequencing, Song and Messing investigated the expression status of each member gene of the 19 kD  $\alpha$ -zein gene family by matching the EST database derived from different inbred lines, other than B73, and showed there were 15 of 25 genes expressed (Song and Messing 2002). The expression of z1C genes at 18 DAP was studied by random sequencing of RT-PCR clones in B73, BSSS53 inbred lines and their hybrids (Song and Messing 2003). There were 7 out of 23 member genes from the z1C gene family expressed in BSSS53 and 6 out of 16 in B73 (Song et al. 2001; Song and Messing 2003). However, for the purpose of gene expression analysis a sequencing approach appeared to be not only less efficient, but also expensive. A better method, more efficient and less expensive, needed to be developed for large gene family analysis.

To provide a comprehensive temporal expression profile of the entire  $\alpha$ -zein super gene family at the individual gene level, we developed a method combining quantitative real-time

PCR with random RT-PCR clone sequencing, which enabled us to study  $\alpha$ -zein gene expression pattern and level during different endosperm development stages (10–34 DAP), in a single maize plant.

## Material and methods

### Plant growth and endosperm collection

Maize plants were grown in a greenhouse facility at Shanghai University, Shanghai, China. According to a previous study (Kiesselbach 1949) and our pilot experiment for zein expression (data not shown), at 10 days after pollination (DAP), the endosperm tissue was apparently visible and zein genes started to express. At 34 DAP, the kernel was hard to dissect due to dehydration. Therefore endosperms were collected from kernels of inbred B73 plants from 10 DAP till 34 DAP. In order to minimize biological variance between plants, each complete set of samples was collected from a single plant sequentially from 10 to 34 DAP. Each sample contained endosperm tissues from about 3 to 5 kernels. Since a previous study showed Opaque2 (O2) was regulated diurnally (Ciceri et al. 1999), all tissues were collected during midday (11:00 am to 14:00 pm) of the indicated day, flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

### RNA isolation and cDNA synthesis

RNAs were extracted using Trizol reagent (BioDev, Cat. No. MK072). To remove genomic DNA contamination, RNAs were subsequently subjected to DNase I treatment (TaKaRa, Cat. No. D2215). A 1  $\mu\text{g}$  aliquot of purified RNA was then reverse transcribed in a 20- $\mu\text{l}$  reaction volume using ReverTra Ace reverse transcriptase (TOYOBO, Cat. No. TRT-101).

### Random sequencing

NetPrimer (<http://www.premierbiosoft.com>) was used to design gene family-specific primers (Primer I), which were universal for member genes within a gene family. Such primers were used for RT-PCR cloning for random sequencing purpose (Supplementary Table 1). All primers were validated by testing with genomic DNA as template to insure no biases among the member genes (data not shown). RT-PCR of cDNAs was carried out with the optimal cycle number within linear amplification stage, and PCR products were subsequently cloned into pMD-18T vector (TaKaRa, Cat. No. D101A). Clones were randomly picked for sequencing analysis using the DYEnamic ET dye terminator sequencing kit (Pharmacia, Cat. No.

US81095), and analyzed on a MEGABACE 4500 DNA analysis system (Amersham). Base-calling and quality assessment of sequences were based on the phred/phrap program (Ewing et al. 1998). In order to have sufficient data for relative abundance calculation, for each developmental sample point, at least 192 random clones were sequenced for z1A or z1C gene family, and at least 96 random clones for z1B or z1D gene family.

### Real-time quantitative PCR

Using the sequences of expressed member genes obtained through random sequencing, the gene family-specific but expressed member genes universal primers (PrimeII) were designed for Quantitative Real time PCR. Besides the common parameters for regular PCR primers the design followed three rules of Wang et al. (2006) (Supplementary Table 1).

SYBR Green I real time PCR was conducted on a DNA Engine Opticon 2 continuous fluorescence detection system (MJ Research, Inc.). The reactions were performed using SYBR Green Premix (TOYOBO, Cat. No. QPK-211). The annealing temperature and plate read temperature of each amplicon were listed in Supplementary Table 1. Quantitative Real Time PCR test included: (1) a standard curve of 8 serial 10-fold dilution points of cloned cDNA plasmids (ranging from  $10^2$  to  $10^9$  copies  $\mu\text{l}^{-1}$ ); (2) all reactions were performed in triplicate; and (3) a mock reaction lacking template as a control. Housekeeping gene,  $\beta$ -Actin, was used as an endogenous reference. For each sample point, three biological replicas (three sets of collected tissues) were used. The final data represented the mean of data from three biological replicas, with SD represented standard deviation from the three biological replicas. The MJ Opticon Monitor version 3.1.32 software (MJ Research, Inc.) collected and analyzed fluorescence data, and estimated the number of cDNA copies relative to  $\beta$ -Actin.

## Results

### Expression of $\alpha$ -zein super gene family members in B73 inbred

According to previous studies only a fraction of genes in z1C gene family were expressed in endosperm (Song et al. 2001; Song and Messing 2003). Because only the expressed genes would be of interest in a profiling study, we first determined how many genes were expressed in the  $\alpha$ -zein gene family. Twenty two DAP was within the window of peak expression of zein genes (Marks et al. 1985). Using a universal primer (PrimeI) for each gene family, cDNAs of expressed genes from 4 gene families were cloned by RT-PCR. To

reconstruct the pool of transcripts from each gene family in endosperm tissue, random cDNA clones were sequenced (see “Materials and methods”). Because there are no introns in  $\alpha$ -zein genes, sequencing data was matched with each of the genomic sequences separately to determine their relative level of mRNA accumulation, according to their occurrence among sequencing samples.

This analysis revealed the expressed member genes from each gene family (Table 2). For the z1C gene family, there were 6 out of 16 member genes expressed. The result was consistent with the previous study by Song and Messing (2003). This study also revealed information about expressed gene members in the 19 kD gene families. There were 8 out of 12 member genes of the z1A gene

**Table 2** Summary of  $\alpha$ -zein genes in B73 and their expression

| Gene family      | Gene                 | Status               | Expression | Abundance ratio within gene family (%) |
|------------------|----------------------|----------------------|------------|--|
| z1A              | <i>Z350D07-1</i>     | Intact               | Yes        | 20.71                                  |
|                  | <i>Z350D07-2</i>     | Intact               | Yes        | 18.34                                  |
|                  | <i>Z448F14-2</i>     | Intact               | Yes        | 1.78                                   |
|                  | <i>Z448F14-3</i>     | Intact               | Yes        | 14.20                                  |
|                  | <i>Z448F14-4</i>     | Intact               | Yes        | 15.38                                  |
|                  | <i>Z448F14-5</i>     | Intact               | Yes        | 13.61                                  |
|                  | <i>Z448F14-6</i>     | Intact               | Yes        | 8.28                                   |
|                  | <i>Z448F14-7</i>     | Intact               | Yes        | 7.69                                   |
|                  | <i>Z448F14-1</i>     | Internal del         | No         | –                                      |
|                  | <i>Z448F14-8</i>     | Pre-stop             | No         | –                                      |
|                  | <i>Z448F14-9</i>     | Truncated (3')       | No         | –                                      |
|                  | <i>Z350D07-3</i>     | Pre-stop             | No         | –                                      |
|                  | z1B                  | <i>Z492M16-4</i>     | Intact     | Yes                                    |
| <i>Z492M16-6</i> |                      | Intact               | Yes        | 49.45                                  |
| <i>Z492M16-1</i> |                      | Pre-stop             | No         | –                                      |
| <i>Z492M16-2</i> |                      | Pre-stop             | No         | –                                      |
| <i>Z492M16-3</i> |                      | Pre-stop             | No         | –                                      |
| <i>Z492M16-5</i> |                      | Pre-stop             | No         | –                                      |
| <i>Z531H07-1</i> |                      | Pre-stop             | No         | –                                      |
| <i>Z531H07-2</i> |                      | Pre-stop             | No         | –                                      |
| z1D              | <i>Z513H09-1</i>     | Intact               | Yes        | 40                                     |
|                  | <i>Z576A02-2</i>     | Intact               | Yes        | 60                                     |
|                  | <i>Z410H16-1</i>     | Truncated (5')       | No         | –                                      |
|                  | <i>Z576A02-1</i>     | Truncated (5')       | No         | –                                      |
|                  | <i>Z576A02-3</i>     | Truncated (3')       | No         | –                                      |
| z1C              | <i>fl2(azs22.16)</i> | Intact               | Yes        | 18.09                                  |
|                  | <i>azs22.4</i>       | Intact               | Yes        | 33.67                                  |
|                  | <i>azs22.7</i>       | Intact               | Yes        | 12.56                                  |
|                  | <i>azs22.8</i>       | Intact               | Yes        | 13.07                                  |
|                  | <i>azs22.9</i>       | Intact               | Yes        | 12.06                                  |
|                  | <i>azs22.19</i>      | Intact               | Yes        | 10.55                                  |
|                  | <i>azs22.1</i>       | Premature stop       | No         | –                                      |
|                  | <i>azs22.2</i>       | Insertion, 11,426 bp | No         | –                                      |
|                  | <i>azs22.3</i>       | Premature stop       | No         | –                                      |
|                  | <i>azs22.5</i>       | Premature stop       | No         | –                                      |
|                  | <i>azs22.6</i>       | Premature stop       | No         | –                                      |
|                  | <i>azs22.10</i>      | Truncation (5')      | No         | –                                      |
|                  | <i>azs22.11</i>      | Premature stop       | No         | –                                      |
|                  | <i>azs22.12</i>      | Intact               | No         | –                                      |
|                  | <i>azs22.13</i>      | Premature stop       | No         | –                                      |
|                  | <i>azs22.20</i>      | Premature stop       | No         | –                                      |

family, 2 out of 8 member genes of the z1B gene family, and 2 out of 5 member genes of the z1D gene family expressed in endosperm. We found that the proportion of expressed genes among different gene families was uneven. Z1A had the highest ratio of expressed genes (3/4), while z1B had the lowest (1/4).

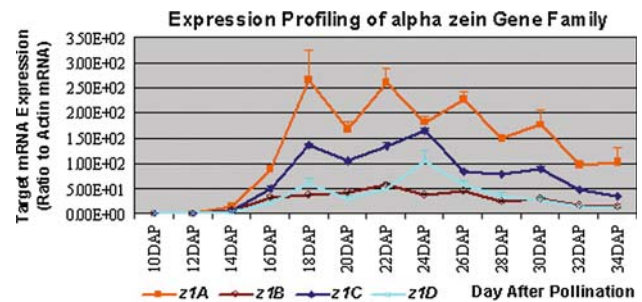
This analysis also revealed the relative transcript abundance of expressed member genes in each gene family (Table 2). Z1A gene family had the most expressed genes among  $\alpha$ -zein gene families. But the contributions of these expressed genes were highly variable. One gene copy, *Z350D07-1*, contributed over 20% of the entire z1A gene family. On the other end, *Z448F14-2* contributed only 1.78%. The rest of the member genes contributed between 8 and 18%. For z1C gene family, *azs22.4* contributed nearly 1/3 of the expression level of the entire 22-kD gene family (33.67%), while the expression levels of the other expressed member genes were lower (10–18%). This was different from a previous study that showed *azs22.4* and *azs22.9* were equally large contributors to the expression level at 18 DAP (Song and Messing 2003). For the z1B gene family, we found that the expression levels of the two expressed gene copies were nearly the same. A similar case was found for z1D gene family (Table 2).

We also noticed that all expressed genes appeared to have intact coding regions. All member genes of the  $\alpha$ -zein gene family with truncations appear to be incapable of accumulating mRNA at detectable level (Table 2). There was one gene (*azs22.12*) of z1C, which had an intact coding region but had no detectable expression level.

#### Expression profiling of $\alpha$ -zein super gene family in B73 inbred

After defining the expressed gene members of each gene family, primers (PrimerII) were designed for real-time PCR to profile the expression of each gene family separately. Using the common endogenous *actin* as a reference, the levels of gene expression among the different gene families were compared. Interestingly, an apparent expression rhythm was observed with all  $\alpha$  zein gene families (Fig. 1). The expression rhythm was in an “up-and-down” oscillating pattern, indicating there was regulatory feedback for  $\alpha$  zein synthesis during endosperm development. There were two groups of patterns. In the first group, represented by z1A and z1B, the expression pattern fluctuated with an interval of 4 days; in the second group, represented by z1C and z1D, the fluctuation interval was 6 days.

Although there were two groups of expression patterns, different gene families showed differences in expression patterns. For the z1A gene family, expression levels reached their peak at 18 DAP, and then decreased with fluctuation until 34 DAP without cease. For the z1B gene

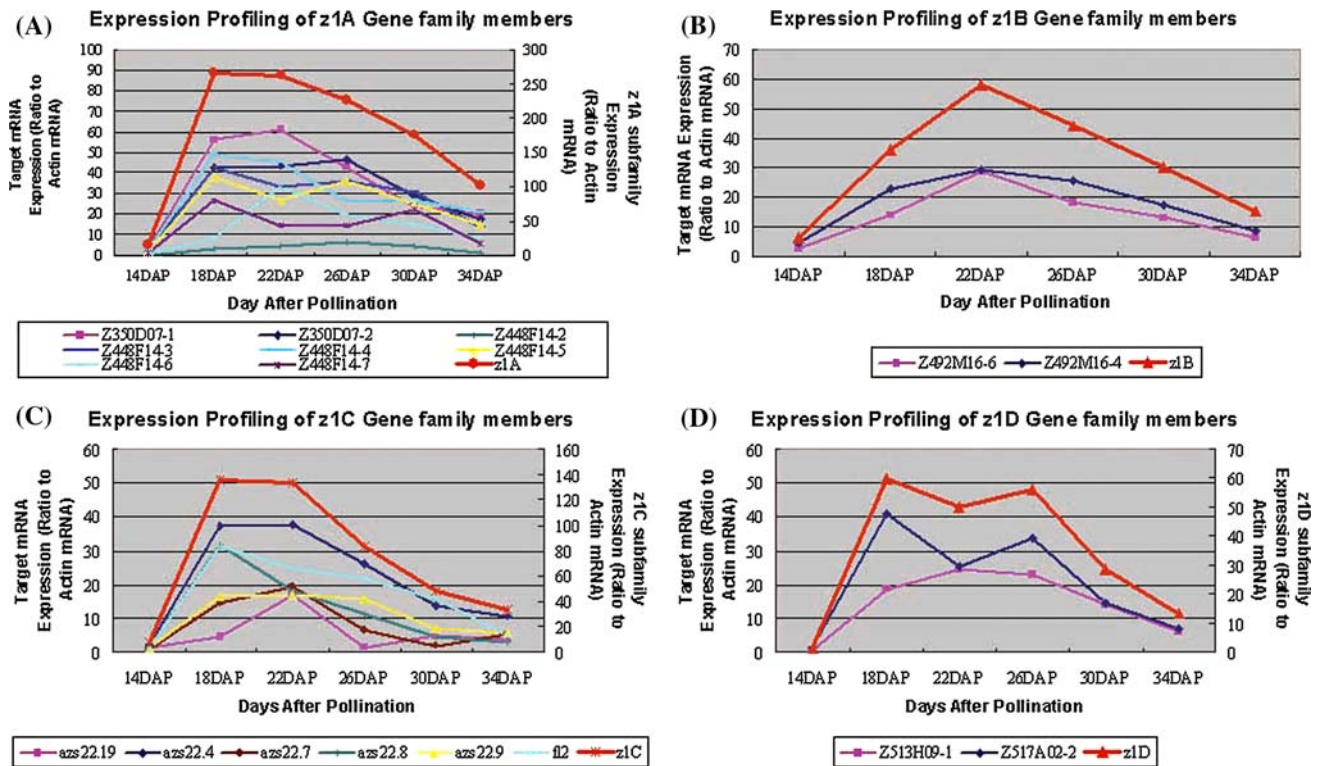


**Fig. 1** Quantitative real time PCR analysis of the expression of four  $\alpha$ -zein gene families in endosperm of B73 inbred. The expression profiling of the four gene families was represented by color-coded lines. The longitudinal axis shows the mRNA level. The scale on left indicates the ratio to actin mRNA level. Data are presented as means  $\pm$  SD ( $n = 3$  biological replicates). The horizontal axis shows the day of endosperm tissue collection after pollination (DAP)

family, expression levels peaked at 22 DAP, and then decreased with a fluctuation until 34 DAP without cease. The expression pattern of the z1C and z1D gene families are similar, both of which have a peak at 24 DAP (Fig. 1). In general, z1A had the highest expression level, followed by z1C, and then z1D and z1B (Fig. 1). Comparing each gene families' peak expression level, z1A had the highest expression level, about 4.50-fold to z1B, which had the lowest expression level. Z1C had the second highest expression level, about 2.81-fold to that of z1B. And z1D was about 1.77-fold to that of z1B. However, due to the non-synchronized oscillating patterns, the relative expression level among different gene families was variable during different endosperm developmental stages (Fig. 1). For example, at 18 DAP and at 24 DAP, the relative expression level between z1A and z1C was significantly different.

#### Expression profiling of member genes in each gene family

In order to reveal the expression pattern for individual genes within each gene family, we used random sequencing of RT-PCR clones to quantify the relative abundance of individual member genes (see “Materials and methods”). Based on sequence polymorphisms, all member genes could be distinguished. The relative expression of a particular gene member in its gene family would be shown by the occurrence of its sequence among the total samples sequenced. Combined with the quantitative expression level of gene families obtained by real time PCR as described above, we were able to obtain the quantitative expression level of individual member genes. Different endosperm developmental stages were analyzed (14, 18, 22, 26, 30, and 34 DAP, see also in “Materials and methods”), and the developmental expression pattern of



**Fig. 2** Random sequencing and quantitative real time PCR analysis of member genes of  $\alpha$ -zein gene families in endosperm of B73 inbred. The longitudinal axis shows the mRNA level. The scale on left indicates the expression level of individual member genes as the ratio to actin mRNA level. The scale on right indicates the expression level of the entire gene family (red line) as the ratio to actin mRNA level.

individual genes could be drawn and compared with each other. By combining the result of random sequencing with the expression level of the corresponding gene family obtained through quantitative real time PCR analysis, the expression patterns of all gene members were obtained (Fig. 2).

Individual genes within gene families were found to have very different expression patterns during endosperm development. In large gene families like z1A and z1C, very complicated expression patterns were observed, and no two genes shared an identical expression pattern in the same gene family. However, genes with similar patterns could be found. For example, if grouped by the overall trends of expression, z1C genes formed two groups. One group containing *azs22.4*, *azs22.9* and *fl2*, showed a relative steady expression level across different developmental stages. The other group containing *azs22.7*, *azs22.8* and *azs22.9*, showed a sharp decline after the expression peak. However, if grouped by the time to reach the expression peak, z1C formed three groups. One group with *azs22.8* and *fl2*, had an apparent expression peak at 18 DAP. The second group with *azs22.7* and *azs22.19* had an apparent expression peak at 22 DAP. The third group with *azs22.4*

The expression level of each gene member is obtained by multiplying its relative expressional percentage to the overall expression level of its gene family. The horizontal axis represents the day of endosperm tissues collection after pollination. Color-coded lines represent different member genes are shown

and *azs22.9* had a flat expression peak. The z1A gene family had even more expressed gene members. There were two major pattern groups, one with a single expression peak, as with *Z350D07-1*, *Z448F14-4* and *Z448F14-6*; the other group had two expression peaks, as with *Z350D07-2*, *Z448F14-3*, *Z448F14-5* and *Z448F14-7*. Within the single expression peak group, *Z350D07-1* and *Z448F14-6* had their expression peak at 22 DAP, while *Z448F14-4* had its expression peak at 18 DAP. For the group with two expression peaks, *Z350D07-2*, *Z448F14-3* and *Z448F14-5* had their expression peaks at 18 and 26 DAP; while *Z448F14-7* had its expression peaks at 18 and 30 DAP. Therefore with different grouping standards, member genes could form into different groups, indicating dramatic divergence of expression patterns of individual genes within a gene family.

Compared with z1A and z1C gene families, the z1B and z1D gene families were smaller, i.e. fewer members. Both had only two expressed genes in the B73 inbred line. Accordingly, their expression patterns were also relatively simple. The two expressed member genes of z1B gene family had almost identical expression patterns. However, the two expressed member genes of z1D had divergent

expression patterns. One gene had two peaks at 18 and 26 DAP (*Z517A02-2*), and the other only has one peak at 22 DAP (*Z513H09-1*) (Fig. 2).

#### Expression of gene families among different maize inbreds

A previous study indicated that dramatically different z1C haplotypes existed among different maize inbred lines resulting in significantly different gene sets (Song and Messing 2003). In order to have a more comprehensive view of the distribution of the zein super gene family among maize inbred lines, we carried out a comparison study of  $\alpha$  zein gene family expression among three maize inbred lines, i.e. B73, BSSS53 and W22. Similar quantitative real time PCR and RT-PCR random clone sequencing was carried out for  $\alpha$  zein gene expression at 24 DAP.

We found that significantly different expressed gene sets existed among these inbred lines. Between B73 and BSSS53, there were 16 non-allelic genes, and 11 allelic genes. Between BSSS53 and W22, 18 non-allelic genes and 10 allelic genes were found. Between B73 and W22, 8 non-allelic genes and 14 allelic genes were found (Table 3). We found that all non-allelic genes were contributed by large families, i.e. z1A and z1C. All the expressed genes in small families, i.e. z1B and z1D, appeared to be allelic. With large family like z1A, non-allelic genes were mostly contributed by a locus with a large gene cluster. The z1A gene family fell into two BAC contigs—Z350D07 and Z448F14 (Song and Messing 2002), that were mapped on two noncontiguous chromosome 4 loci (Song et al. 2003). Z350D07 was a locus with only three copies of z1A genes, and Z448F14 was a locus with a large gene cluster of nine copies of z1A genes. We determined that the two expressed genes from Z350D07 all appeared to be allelic among the three inbred lines, while all non-allelic genes were contributed by the Z448F14 locus.

We also discovered major haplotypes existed among these inbred lines, particularly in large gene families like z1A and z1C. For the z1A gene family, at least two major haplotypes were found, all contributed by the Z448F14 locus. One represented by BSSS53, the other one represented by B73 and W22. However, there was a minor difference between B73 and W22; an extra unknown z1A gene was found in W22. Between the two z1A haplotypes, only two genes appeared to be allelic, while at least eight genes appeared to be non-allelic. In BSSS53 haplotype, we found one z1A gene (*Z350D07-2*) had a premature stop codon (Table 3). This was the only gene in this study that had a premature stop codon, but was still expressed. This result confirmed an earlier observation that some 19 kD

zein ESTs in maize EST database have premature stop codons (Song and Messing 2002).

For the z1C gene family, a previous study indicated that there were four major haplotypes (Song and Messing 2003). However, our study found that the three inbred lines represented three major haplotypes. There was only one gene (*azs22.4*) that appeared to be allelic among all three lines. The results for haplotypes in B73 and BSSS53 lines were inconsistent with the previous study (Song and Messing 2003). It was interesting that the third major haplotype was represented by W22. In this haplotype, the *azs22.16* gene, i.e. *floury2*, was found to be unexpressed. Also, the *azs22.12* gene, which was found to be unexpressed in B73 (see “Results”), was expressed at a low level in W22. Also there was an unknown z1C gene that was expressed at a high level (30.57%).

Despite the dramatic difference of expressed gene sets among these inbred lines, it was interesting that the overall expressed gene numbers appeared to be surprisingly consistent among these lines. In B73, a total of 18 expressed genes were detected, while there were 18 and 19 expressed genes detected from BSSS53 and W22, respectively (Table 3). Not only the number of total expressed genes, but also the number of expressed genes within each family was highly consistent. For example, z1B and z1D all had two expressed genes among these inbred lines. Z1A had eight, eight and nine expressed genes in B73, BSSS53 and W22, respectively. While z1C had six, seven and six expressed genes in B73, BSSS53 and W22, respectively. A selective pressure for maintaining the expressed gene number of  $\alpha$ -zein genes appeared to be present.

In addition to the large amount of non-allelic genes, the differentiation among allelic genes was also significant. For example, the two z1A allelic genes from Z350D07 were diverged so much, that their relative abundance in three inbred lines was completely different (Table 3). A similar divergence was found with the two allelic genes in the z1D gene family (Table 3). For the z1B gene family, although the two allelic genes expressed at similar level between B73 and BSSS53, significant expression divergence was found in W22. Similar expression divergence could also be found from allelic genes in z1A and z1C gene families (Table 3).

## Discussion

An effective strategy for expression profiling analysis of the maize  $\alpha$  zein super gene family

The  $\alpha$ -zein gene family encodes the most prominent storage protein in maize seed. It represents a typical gene super family that contains four small but related gene families,

**Table 3** Expressed  $\alpha$ -zein genes and their expression levels in maize inbred lines B73, BSSS53 and W22 at 24 DAP

| Gene family | Gene                  | B73 inbred                      | BSSS53 inbred     | W22 inbred        |
|-------------|-----------------------|---------------------------------|-------------------|-------------------|
| z1A         | <i>Z350D07-1</i>      | 3.77E+01 (20.71% <sup>a</sup> ) | 7.55E+00 (11.48%) | 4.60E+01 (24.06%) |
|             | <i>Z350D07-2</i>      | 3.34E+01 (18.34%)               | 1.26E+01 (19.13%) | 1.02E+01 (5.35%)  |
|             | <i>Z448F14-2</i>      | 3.23E+00 (1.78%)                | 2.16E+00 (3.28%)  | 4.09E+00 (2.14%)  |
|             | <i>Z448F14-3</i>      | 2.58E+01 (14.20%)               |                   | 3.37E+01 (17.65%) |
|             | <i>Z448F14-4</i>      | 2.80E+01 (15.38%)               |                   | 4.19E+01 (21.93%) |
|             | <i>Z448F14-5</i>      | 2.48E+01 (13.61%)               |                   | 2.76E+01 (14.44%) |
|             | <i>Z448F14-6</i>      | 1.51E+01 (8.28%)                |                   | 8.17E+00 (4.28%)  |
|             | <i>Z448F14-7</i>      | 1.40E+01 (7.69%)                | 1.04E+01 (15.85%) | 1.33E+01 (6.95%)  |
|             | Unknown-1             |                                 | 3.95E+00 (6.01%)  |                   |
|             | Unknown-2             |                                 | 6.11E+00 (9.29%)  |                   |
|             | Unknown-3             |                                 | 1.08E+01 (16.39%) |                   |
|             | Unknown-4             |                                 | 1.22E+01 (18.58%) |                   |
|             | Unknown-5             |                                 |                   |                   |
|             | Total <sup>b</sup>    | 1.82E+02                        | 6.58E+01          | 1.91E+02          |
| z1B         | <i>Z492M16-4</i>      | 1.94E+01 (50.55%)               | 1.29E+01 (52.75%) | 2.15E+01 (77.92%) |
|             | <i>Z492M16-6</i>      | 1.90E+01 (49.45%)               | 1.44E+01 (47.25%) | 6.01E+00 (22.08%) |
|             | Total                 | 3.84E+01                        | 2.73E+01          | 2.76E+01          |
| z1C         | <i>azs22.19</i>       | 1.72E+01 (10.55%)               |                   | 1.73E+01 (9.55%)  |
|             | <i>azs22.4</i>        | 5.49E+01 (33.67%)               | 1.74E+01 (15.17%) | 3.69E+01 (20.38%) |
|             | <i>azs22.7</i>        | 2.05E+01 (12.56%)               |                   |                   |
|             | <i>azs22.8</i>        | 2.13E+01 (13.07%)               | 6.46E+00 (5.62%)  |                   |
|             | <i>azs22.9</i>        | 1.97E+01 (12.06%)               |                   | 3.23E+01 (17.83%) |
|             | <i>azs22.16 (fl2)</i> | 2.95E+01 (18.09%)               | 1.87E+01 (16.29%) |                   |
|             | <i>azs22.14</i>       |                                 | 1.03E+01 (8.99%)  | 3.46E+01 (19.11%) |
|             | <i>azs22.10</i>       |                                 | 1.68E+01 (14.61%) |                   |
|             | <i>zp22/6</i>         |                                 | 1.81E+01 (15.73%) |                   |
|             | <i>zP22/d87</i>       |                                 | 2.71E+01 (23.60%) |                   |
|             | <i>azs22.12</i>       |                                 |                   | 4.61E+00 (2.55%)  |
|             | Unknown               |                                 |                   | 5.53E+01 (30.57%) |
|             | Total                 | 1.63E+02                        | 1.15E+02          | 1.81E+02          |
| z1D         | <i>Z531H09-1</i>      | 4.12E+01 (40%)                  | 5.01E+01 (80.65%) | 1.65E+01 (48.89%) |
|             | <i>Z517A02-2</i>      | 6.18E+01 (60%)                  | 1.20E+01 (19.35%) | 1.72E+01 (51.11%) |
|             | Total                 | 1.03E+02                        | 6.21E+01          | 3.37E+01          |

Relative expression levels of each gene are expressed as the percentages of sequenced clones (see “Materials and methods”). Total expression of the gene families is from quantitative real time analysis of 24 DAP in each inbred

<sup>a</sup> Expression abundance of gene family by random sequencing

<sup>b</sup> Expression relative to actin by real-time PCR

namely z1A, z1B, z1C and z1D. Due to the complexity of the super gene family and high sequence similarity among member genes, it was technically challenging to carry out a quantitative expression profiling analysis of the  $\alpha$ -zein gene family during endosperm development.

Previous studies relied on RNA hybridization, EST data analysis or random RT-PCR clone sequencing (Dolfini et al. 1992; Kodrzycki et al. 1989; Marks et al. 1985; Woo et al. 2001; Song and Messing 2003). Although these studies yielded valuable information about  $\alpha$ -zein gene expression during endosperm development, they either

failed to distinguish individual member genes, or could not provide accurate quantitative data. In this study, we attempted a different strategy to profile the expression of maize  $\alpha$ -zein gene family during endosperm development. We divided the analysis into two parts. In part one, universal primers were designed so that the family members could be amplified without bias. Quantitative real-time PCR analysis was carried out with these primers, and accurate quantitative data was obtained for the entire gene family at multiple time points during endosperm development. In part two, transcripts of expressed member genes



from each family were recovered by RT-PCR and sequenced randomly. Different member genes were distinguished by their sequence polymorphism, and their relative abundance could be counted as their occurrence among the total samples sequenced. When combined with the quantitative data achieved before, the quantitative data for individual member genes could be deduced.

Because a PCR technique was used for both steps (real-time PCR and RT-PCR cloning), small tissue samples were required for the analysis. This enabled a complete set of experiments to be conducted on a single plant, providing much better data quality. This appeared to be quite critical for gene profiling analysis. Since in our pilot experiments we noticed that even with the same genetic background (same inbred line), growing in the same greenhouse conditions (e.g. pollinated on same day, tissue harvested on same day), different individual plants still showed significant variation in gene expression level (data not shown). Therefore an experiment based on a single plant provided the best control for a multi-point expression profiling study. Using this strategy, we were able to carry out an extensive expression profiling study of the maize  $\alpha$ -zein gene super family. This included all 4 gene families, composed of 18 active gene members at 13 different developmental points covering most of the endosperm development. This same approach should be effective for expression profiling studies of other large gene super families as well.

#### Characteristic expressional patterns revealed regulatory feedback of maize $\alpha$ zein genes during endosperm development

For the first time, the expression profile of the  $\alpha$ -zein gene super family during endosperm development has been obtained. Due to dense sampling (every 2 day interval) during endosperm development, this study revealed new findings of  $\alpha$ -zein developmental expression patterns. Characteristic “up-and-down” oscillating expression patterns were discovered with all four gene families (Fig. 1). If less dense sampling points (e.g. every 4 day interval) were collected, such characteristic patterns might be missed. The “up-and-down” oscillating patterns clearly indicate that there was feedback regulation of  $\alpha$  zein genes expression during endosperm development. Two distinct groups of “up-and-down” oscillating patterns indicated two different feedback regulation pathways. Based on this, we postulate that z1A and z1B might share one regulation pathway, while z1C and z1D might share another regulation pathway. Despite the four gene families being derived from a single ancestor gene, their regulation diverged during evolution.

A previous study had demonstrated that 22 kD zein (z1C) gene family was under the control of transcriptional

factor O2 (Kodrzycki et al. 1989), but the transcriptional factor(s) responsible for expression of the 19 kD zein gene families is unknown. Because the O2 mutation affected some of the 19 kD zein proteins and based on the similar expression pattern z1D had to that of z1C, we predict O2 could also be the transcriptional factor for the z1D gene family. However, the z1A and z1B gene families might have a completely different, i.e. non-O2, transcriptional factor, as suggested by their distinct expressional patterns. Then how could such feedback regulation be carried out? Previous studies found that the activity of O2 was regulated diurnally, and phosphorylation was found to play an important role in diurnal regulation and DNA binding activity (Ciceri et al. 1997, 1999). These data suggest that there is an unknown signaling pathway with a Kinase component involved in O2 regulation. This could provide a possible answer for the “up-and-down” oscillating pattern found with z1C genes, the direct target genes of O2. A similar mechanism might also occur with transcriptional factor(s) for the 19 kD gene families.

Dramatically divergent member genes indicate a fast evolution rate of the maize  $\alpha$  zein gene family

A striking finding from this study was the high degree of expression divergence among the  $\alpha$ -zein gene family members. A large number of the member genes were found to be pseudogenes. We also noticed that nearly all of the expressed genes had intact coding sequences, while almost all unexpressed genes had mutations in the coding sequences (Tables 2, 3). It has been shown that premature stop codons decrease mRNA stability (Van Hoof and Green 1996) and that a single in-frame stop codon reduces the mRNA concentration significantly (Liu and Rubenstein 1993). Mutations that cause premature stop codons might be an early step in creating pseudogenes. Moreover, we found a large number of non-allelic expressed genes among different inbred lines (Table 3). The non-allelic expressed genes found were mainly contributed by pseudogenes of allelic gene copies from different inbred lines. This result indicates that the  $\alpha$ -zein gene family is still creating pseudogenes. This provides further evidence that the  $\alpha$ -zein gene family is subject to the “birth-and-death” evolution model (Nei and Hughes 1992).

Besides the divergence of expression status, the expression level of member genes also diverged greatly. A comparison among different inbred lines yielded more information about expression level divergence of allelic genes (Table 3). The differentiation of gene expressional patterns reflected the divergence of gene regulation. It's surprising that closely related member genes within the same gene family could have such diverged expression patterns. Although the transcriptional factor(s) for the

19 kD zein genes is unknown, the 22 kD zein (z1C) genes are under the control of Opaque2 (Rubenstein and Gerahty 1986; Schmidt et al. 1987). However, in BSSS53 (*o2*) line, two member genes of the z1C gene family, *zp22/6* and *zp22/D87*, are expressed in the absence of O2 (Song et al. 2001). Also some other transcription factors binding the P-box may interact with O2 to activate 22 kD zein gene expression, like a maize zinc-finger protein (Vicente-Carbajosa et al. 1997). Previous studies suggested that factor(s) other than O2 might activate the 22 kD zein (z1C) gene transcription (Ciceri et al. 2000; Muller et al. 1997). All of this suggests that the expressional divergences of the  $\alpha$ -zein super gene family observed in this study might result from differential transcriptional regulation due to dissimilar transcription factors. Because the maize  $\alpha$ -zein gene super family had a relatively recent evolutionary history, most of the gene copies arose from gene amplifications within the last 5 million years. Many member genes have very high sequence similarity. The dramatic divergence of these member genes, particularly in their expression level, could be explained by the entire  $\alpha$ -zein gene super family having had a fast rate of evolution. This fast evolutionary rate was possibly due to the coexistence of many redundant gene copies that had similar biological functions, as compared to single copy genes. On the other hand, co-evolution of  $\alpha$ -zein genes and their transcriptional regulators may have further accelerated their evolution at the gene expression and regulation level.

#### High diversity of $\alpha$ -zein genes among maize inbred lines contributed by highly diverged haplotypes

Previous work revealed highly diverged haplotypes of the z1C gene family in the maize genome (Song and Messing 2003). This would result from very different gene sets among the various inbred lines harboring different haplotypes. In this study, the expression of the  $\alpha$ -zein gene super family was compared in 3 maize inbred lines. We found that these inbred lines have drastically different expressed gene sets. Large numbers of non-allelic genes were contributed by the large families (e.g. z1A or z1C), or by a chromosome locus with a large gene cluster (e.g. 448F14 locus). It appears that large gene families contribute more genetic variation than small gene families and single genes.

Based on these expressed gene sets, we predict different haplotypes of  $\alpha$ -zein gene families (two for z1A, three for z1C). Crosses between these inbred lines would result in very different expressed genes sets in their hybrids. It's interesting to note that the total expressed  $\alpha$ -zein gene numbers among these inbred lines were highly consistent, not only for the entire  $\alpha$ -zein gene super family but also for the four gene families (Table 3). We postulate that there

might be a selection pressure for minimal expressed gene number in each haplotype. Above this number, there would be no selective advantage. However, less than this number might cause selective disadvantage. Besides the great diversity of expressed gene sets among inbred lines, the allelic genes also showed dramatic divergence in expression level. This further demonstrates the fast evolving nature of maize  $\alpha$  zein gene family.

As a consequence of the combination of different haplotypes with different expression capacity, the overall expression level of  $\alpha$ -zein genes could be significantly different. For breeding purposes, rational combination of different  $\alpha$ -zein haplotypes would provide an effective way for manipulating the seed protein content in maize seeds.

**Acknowledgments** We would like to thank Dr. Eric Linton for critical reading of the manuscript, Xiang-zong Meng, Jian-ping Xu for their help with the plant materials, Liang-liang Zhou, Bing-bing Huang for their help with DNA sequencing analysis. This work was supported by NSFC (30671303), Ministry of Science and Technology of China (2006AA10Z148, 2006AA10A107), Ministry of Education of China (NCET-06-0435, 206048), Fok Ying Tung Education Foundation (101024), Shanghai Education Foundation (04SG40), Education Commission, Science and Technology Commission of Shanghai Municipality (05ZZ02, 05QMX1424).

#### References

- Burr B, Burr FA, 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. doi:10.1016/0022-2836(82)90415-6
- Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T et al (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* 282:2022–2028. doi:10.1126/science.282.5396.2022
- Ciceri P, Gianazza E, Lazzari B, Lippoli G, Genga A, Hoscheck G, Schmidt RJ, Viotti A (1997) Phosphorylation of Opaque2 changes diurnally and impacts its DNA binding activity. *Plant Cell* 9:97–108
- Ciceri P, Locatelli F, Genga A, Viotti A, Schmidt RJ (1999) The activity of the maize Opaque2 transcriptional activator is regulated diurnally. *Plant Physiol* 121:1321–1328. doi:10.1104/pp.121.4.1321
- 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. doi:10.1104/pp.124.1.451
- Coleman CE, Larkins BA (1998) The prolamins of maize. In: Shewry PR, Casey R (eds) Seed proteins. Kluwer, Dordrecht, pp 109–139
- Dolfini SF, Landoni M, Tonelli C, Bemard L, Viotti A (1992) Spatial regulation in the expression of structural and regulatory storage protein genes in *Zea mays* endosperm. *Dev Genet* 13:264–276. doi:10.1002/dvg.1020130404
- Esen A (1987) Proposed nomenclature for the alcohol-soluble proteins (zeins) of maize (*Zea mays* L.). *J Cereal Sci* 5:117–128

- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185
- Hagen G, Rubenstein I (1981) Complex organization of zein genes in maize. *Gene* 13:239–249. doi:[10.1016/0378-1119\(81\)90029-9](https://doi.org/10.1016/0378-1119(81)90029-9)
- Hamaker BR, Mohamed AA, Habben JE, Huang CP, Larkins BA (1995) Efficient procedure for extracting maize and sorghum kernel proteins reveals higher rolinin contents than the conventional method. *Cereal Chem* 72:583–588
- Heidecker G, Messing J (1986) Structural analysis of plant genes. *Annu Rev Plant Physiol* 37:439–446. doi:[10.1146/annurev.pp.37.060186.002255](https://doi.org/10.1146/annurev.pp.37.060186.002255)
- Kiesselbach TA (1949) The structure and reproduction of corn. University of Nebraska College of Agriculture. *Agric Exp Stn Res Bull* 161:1–96
- Kodrzycki R, Boston RS, Larkins BA (1989) The opaque-2 mutation of maize differentially reduces zein gene transcription. *Plant Cell* 1:105–114
- Koonin EV, Tatusov RL, Galperin MY (1998) Beyond complete genomes: from sequence to structure and function. *Curr Opin Struct Biol* 8:355–363. doi:[10.1016/S0959-440X\(98\)80070-5](https://doi.org/10.1016/S0959-440X(98)80070-5)
- Landry J, Moureaux T (1970) Heterogeneity of corn seed glutelin: selective extraction and amino acid composition of the 3 isolated fractions. *Bull Soc Chim Biol (Paris)* 52:1021–1037
- Landry J, Delhaye S, Damerval C (2000) Improved method for isolating and quantitating  $\alpha$ -amino nitrogen as nonprotein, true protein, salt-soluble proteins, zeins, and true glutelins in maize endosperm. *Cereal Chem* 77:620–626. doi:[10.1094/CCHEM.2000.77.5.620](https://doi.org/10.1094/CCHEM.2000.77.5.620)
- Liu CN, Rubenstein I (1993) Transcriptional characterization of an alpha-zein gene cluster in maize. *Plant Mol Biol* 22:323–336. doi:[10.1007/BF00014939](https://doi.org/10.1007/BF00014939)
- Llaca V, Messing J (1998) Amplicons of maize zein genes are conserved within genic but expanded and constricted in intergenic regions. *Plant J* 15:211–220. doi:[10.1046/j.1365-3113X.1998.00200.x](https://doi.org/10.1046/j.1365-3113X.1998.00200.x)
- Marks MD, Lindell JS, Larkins BA (1985) Quantitative analysis of the accumulation of Zein mRNA during maize endosperm development. *J Biol Chem* 260:16445–16450
- Messing J, Bharti AK, Karlowski WM, Gundlach H, Kim HR, Yu Y, Wei F, Fuks G, Soderlund CA, Mayer KF et al (2004) Sequence composition and genome organization of maize. *Proc Natl Acad Sci USA* 101:14349–14354. doi:[10.1073/pnas.0406163101](https://doi.org/10.1073/pnas.0406163101)
- Muller M, Dues G, Balconi C, Salamini F, Thompson RD (1997) Nitrogen and hormonal responsiveness of the 22 kD alpha-zein and b-32 genes in maize endosperm is displayed in the absence of the transcriptional regulator Opaque-2. *Plant J* 12:281–291. doi:[10.1046/j.1365-3113X.1997.12020281.x](https://doi.org/10.1046/j.1365-3113X.1997.12020281.x)
- Nei M, Hughes A (1992) Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In: Tsuji K, Aizawa M, Sasazuki T (eds) 11th Histocompatibility workshop and conference. Oxford University Press, Oxford, pp 27–38
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121–152. doi:[10.1146/annurev.genet.39.073003.112240](https://doi.org/10.1146/annurev.genet.39.073003.112240)
- Pedersen K, Devereux J, Wilson DR, Sheldon E, Larkins BA (1982) Cloning and sequence analysis reveal structural variation among related zein genes in maize. *Cell* 29:1015–1026. doi:[10.1016/0092-8674\(82\)90465-2](https://doi.org/10.1016/0092-8674(82)90465-2)
- Rubenstein I, Geraghty DE (1986) Genetic organization of zeins. In: Pomeranz Y (ed) *Advances in cereal science and technology*, vol VIII. American Association of Cereal Chemists Press, St. Paul, pp 297–315
- Schmidt RJ, Burr FA, Burr B (1987) Transposon tagging and molecular analysis of the maize regulatory locus opaque-2. *Science* 238:960–963. doi:[10.1126/science.2823388](https://doi.org/10.1126/science.2823388)
- Semple C, Wolfe KH (1999) Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. *J Mol Evol* 48:555–564. doi:[10.1007/PL00006498](https://doi.org/10.1007/PL00006498)
- Soave C, Tardani L, Di Fonzo N, Salamini F (1981) Zein level in maize endosperm depends on a protein under control of the opaque-2 and opaque-6 loci. *Cell* 27:403–410. doi:[10.1016/0092-8674\(81\)90423-2](https://doi.org/10.1016/0092-8674(81)90423-2)
- Song R, Messing J (2002) Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. *Plant Physiol* 130:1626–1635. doi:[10.1104/pp.012179](https://doi.org/10.1104/pp.012179)
- Song R, Messing J (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. *Proc Natl Acad Sci USA* 100:9055–9060. doi:[10.1073/pnas.1032999100](https://doi.org/10.1073/pnas.1032999100)
- 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, Ward K, Messing J (2003) Maize alpha zein genes map to seven positions in the maize genome. *Maize Genet Coop News Lett* 77:59–60
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815. doi:[10.1038/35048692](https://doi.org/10.1038/35048692)
- Ueda T, Messing J (1993) Manipulation of amino acid balance in maize seeds. *Genet Eng (N Y)* 15:109–130
- Van Hoof A, Green P (1996) Premature nonsense codons decrease the stability of phytohemagglutinin mRNA in a position-dependent manner. *Plant J* 10:415–424. doi:[10.1046/j.1365-3113x.1996.10030415.x](https://doi.org/10.1046/j.1365-3113x.1996.10030415.x)
- Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ (1997) A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc Natl Acad Sci USA* 94:7685–7690. doi:[10.1073/pnas.94.14.7685](https://doi.org/10.1073/pnas.94.14.7685)
- Wang Y, Zhu W, Levy DE (2006) Nuclear and cytoplasmic mRNA quantification by SYBR green based real-time RT-PCR. *Methods* 39:356–362. doi:[10.1016/j.ymeth.2006.06.010](https://doi.org/10.1016/j.ymeth.2006.06.010)
- Wienand U, Langridge PL, Feix G (1981) Isolation and characterization of a genomic sequence of maize coding for a zein gene. *Mol Gen Genet* 182:440–444. doi:[10.1007/BF00293933](https://doi.org/10.1007/BF00293933)
- Wilson CM (1983) Seed protein fractions of maize, sorghum and related cereals. In: Gottschalk W, Muller HP (eds) *Seed proteins: biochemistry, genetics, nutritive value*. Martinus Nijhoff/Dr. W. Junk Publishers, The Hague, pp 271–307
- Wilson DR, Larkins BA (1984) Zein gene organization in maize and related grasses. *J Mol Evol* 20:330–340. doi:[10.1007/BF02104739](https://doi.org/10.1007/BF02104739)
- 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