Evolution of Gene Expression after Gene Amplification

Nelson Garcia[1], Wei Zhang[1], Yongrui Wu[1,2], and Joachim Messing[1,*]

[1]Waksman Institute of Microbiology, Rutgers University

[2]Present address: National Key Laboratory of Plant Molecular Genetics, Institute of Plant Physiology & Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

*Corresponding author: E-mail: messing@waksman.rutgers.edu.

Accepted: April 17, 2015

Data deposition: This project has been deposited at NCBI Genbank under the accessions KR105945 and KR105946.

Abstract

We took a rather unique approach to investigate the conservation of gene expression of prolamin storage protein genes across two different subfamilies of the Poaceae. We took advantage of oat plants carrying single maize chromosomes in different cultivars, called oat–maize addition (OMA) lines, which permitted us to determine whether regulation of gene expression was conserved between the two species. We found that *γ-zeins* are expressed in OMA7.06, which carries maize chromosome 7 even in the absence of the trans-acting maize prolamin-box-binding factor (PBF), which regulates their expression. This is likely because oat PBF can substitute for the function of maize PBF as shown in our transient expression data, using a *γ-zein* promoter fused to green fluorescent protein (GFP). Despite this conservation, the younger, recently amplified prolamin genes in maize, absent in oat, are not expressed in the corresponding OMAs. However, maize can express the oldest prolamin gene, the wheat high-molecular weight glutenin *Dx5* gene, even when maize *Pbf* is knocked down (through *PbfRNAi*), and/or another maize transcription factor, *Opaque-2* (O2) is knocked out (in maize *o2* mutant). Therefore, older genes are conserved in their regulation, whereas younger ones diverged during evolution and eventually acquired a new repertoire of suitable transcriptional activators.

Key words: oat–maize addition (OMA) lines, prolamin-box-binding factor, Opaque2, zeins, high-molecular weight glutenin *Dx5*.

Introduction

Cereal species are among the most studied plants because of their agro-economic importance. They are currently categorized into three subfamilies—Ehrhartoideae (rice), Pooideae (oat, wheat, and barley), and Panicoideae (maize, sorghum, and millets) (Kellogg 2001). Relative to other cereals, oat (*Avena sativa*, 2n = 6x = 42) and maize (*Zea mays*, 2n = 2x = 20) are evolutionarily distant, having shared a common ancestor with all the rest of the cereals about 50 Ma (Gaut 2002; Salse and Feuillet 2007). Despite being remotely related, maize pollen can be used to pollinate oat flowers and can sometimes produce fertile partial hybrids containing single maize chromosome additions. The complete set of ten maize chromosomes has been added into oat plants as single chromosome additions and is collectively called oat–maize addition (OMA) lines (Kynast et al. 2001; Rines et al. 2009). Wheat and maize are also evolutionarily distant, diverging from each other at about the same time when oat and maize formed separate lineages about 50 Ma (Gaut 2002; Salse and Feuillet 2007).

Aside from being a major source of carbohydrates, cereals can also be a source of protein for both human and livestock nutrition. During seed development, they accumulate storage proteins mainly in the form of alcohol-soluble prolamins with the exception of rice and oat, whose grains mainly accumulate saline-soluble globulins (Shewry and Halford 2002). The prolamins are deposited into endoplasmic reticulum-associated protein bodies (PBs) in maize, but oat and wheat have protein storage vacuoles in addition to PBs (Lending et al. 1989; Herman and Larkins 1999).

In maize, prolamins are called zeins and are encoded by a medium-sized gene family. Based on structural differences, *zeins* are subdivided into α, β, γ, and δ groups. The α-zeins are encoded by five multicopy gene clusters and a single gene locus (*19-* and *22-kDa α-zeins*) located on chromosomes 1, 4, and 7 (Song and Messing 2003; Miclaus et al. 2011),

whereas γ-zeins are encoded by three genes each as single copies (27- and 50-kDa *γ-zeins* on chromosome 7, 16-kDa *γ-zein* on chromosome 2). The δ-zeins are encoded by two single copy genes (*18-kDa* on chromosome 6 and *10-kDa* on chromosome 9) and the *β-zein* is encoded by one single copy gene (*15-kDa β-zein* on chromosome 6) (Xu and Messing 2008; Feng et al. 2009).

Based on syntenic alignments and nucleotide substitution rates, it has been proposed that prolamins arose from the duplication of a globulin gene and evolved into a high-molecular weight (HMW) glutenin. Thus, HMW-glutenin genes likely represent the founding gene of this family within the grasses (Xu and Messing 2009). During evolution of the species of this family, the HMW-glutenin gene got amplified in a tandem fashion, but more importantly in dispersed chromosomal location. Dispersal led to greater divergence so that today there appeared to be three groups of related prolamins. Group III is comprised by the HMW-glutenin, present only in the subfamily Pooideae. Interestingly, they are critical for the baking quality and must have been lost in the other two subfamilies of the Poaceae after copies dispersed and diverged. Therefore, we need wheat today for baking. The other wheat prolamins, called gliadins, are also important for baking and belong to group II, which includes the maize γ-zeins and β-zeins. Group I is absent in wheat and represents the most recent amplification and dispersal of prolamin genes.

Dispersal of copies also led to tandem amplification resulting in gene clusters, but those can vary in maize haplotypes (Song et al. 2001). There are two transcription factors—prolamin-box-binding factor (PBF) and Opaque2 (O2) that have been shown so far to regulate the expression of a significant number of zein genes. *Pbf* is located on maize chromosome 2 and encodes a *Dof* class of zinc finger transcription factor that engages in protein–protein interaction with O2 (Vicente-Carbajosa et al. 1997). The role of PBF in regulating zein gene expression was elucidated in *PbfRNAi* transgenic lines, showing that knocking down expression of *Pbf* reduces the accumulation of only the 22-kDa α- and 27-kDa γ-zeins (Wu and Messing 2012). The *O2* gene on the other hand is located on chromosome 7 and encodes a leucine zipper-type of transcription factor (Schmidt et al. 1990) that has been shown to affect the accumulation of most 22-kDa α-zeins and the 15-kDa β-zein (Schmidt et al. 1992; Cord-Neto et al. 1995; Song and Messing 2003). In maize, PBF has been shown to bind a TGTAAAG-conserved motif called the P-box, whereas O2 binds a CACGTA/C (also called A/C-box) motif downstream to the P-box (Schmidt et al. 1992; Vicente-Carbajosa et al. 1997). In wheat, the *Dx5* promoter has also been shown to contain P-box motifs, but has also a G-box-like motif (TTACGTGG) located upstream of the P-box. Evidence through gel shift assays has shown that the G-box-like motif has similar affinity to the A-box in maize, and therefore might be another binding site for O2 (Norre et al. 2002).

Although DNA-binding studies have been useful in characterizing transcriptional activators, they are not conclusive for in vivo interactions. A different approach is facilitated in the OMA, where *cis*- and *trans*-acting sites can be combined from evolutionary diverged regulatory processes. Using this path, we show evidence that oat PBF can substitute for the function of maize PBF (ZmPBF) in OMA lines, whereas ZmPBF and maize O2 (ZmO2) do not seem to be major factors for the expression of the wheat HMW-glutenin *Dx5* (group III prolamin) in transgenic maize. However, group I prolamins, present in species of the Panicoideae, have diverged to a degree that distantly related transcription factors can no longer activate their expression. The resulting knowledge not only provides us with an understanding of the evolution of gene amplification and expression but also will be useful for translational genetics/breeding of agronomic traits.

Materials and Methods

Plant Materials

Dr Ronald Phillips and Dr Howard Rines of the University of Minnesota kindly provided the set of OMA lines. These lines represent single disomic maize chromosome additions (either from maize B73, Mo17, or Seneca 60) in the background of either Starter or Sun2 oat cultivars. For nomenclature purposes, these materials are named in the format OMA*y.z*, wherein *y* is the maize chromosome number added, and *z* is used to trace its original chromosome recovery event (Kynast et al. 2001; Rines et al. 2009).

Transgenic maize expressing *Dx5* controlled by its native promoter was obtained from the Scott Lab at Iowa State University (Sangtong et al. 2002) and was backcrossed several times to B73 (Zhang et al. 2013). A maize transgenic line with a knockdown of *ZmPbf* by *PbfRNAi* was generated in the Messing Lab (Wu and Messing 2012). The *o2* mutant used was also a Messing Lab stock. A small section was cut from the top of the seed from the cross *Dx5/-, O2/o2* × *pbfRNAi/-, O2/o2* and saved for protein extraction, and the rest of the seed containing the embryo was saved for germination and genotyping. Primers used to genotype *Dx5, O2,* and *Pbf* were described previously (Sangtong et al. 2002; Wu and Messing 2012). Individuals with *Dx5/-*, *o2/o2*, *PbfRNAi/-*; *Dx5/-*, *o2/o2*, *-/-*; and *Dx5/-*, *O2/o2*, *PbfRNAi/-* genotypes were selected for prolamin extraction and Dx5 phenotyping.

Dx5 Phenotyping

The seed section saved for protein extraction was ground into powder and protein was extracted as described (Zhang et al. 2013) with some modifications. Briefly, 600 μl of borate extraction buffer was added in the powder and left overnight in a shaker. The mixture was then centrifuged and 200 μl of

the supernatant was mixed with equal amount of 100% ethanol and shaken for over 2 h to extract the prolamins, which include the wheat Dx5 protein. The prolamin extract was then run in a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel and blotted on a polyvinylidene difluoride (PVDF) membrane. Western blot using anti-Dx5 antibodies was then performed to examine the accumulation of the Dx5 protein (Zhang et al. 2014).

Polymerase Chain Reaction Assays

Genomic DNA was extracted using the CTAB method, whereas mRNA was extracted using Qiagen RNEAsy kit. Maize simple sequence repeat (SSR) markers, one on each chromosome arm, were used to verify the presence of intact maize chromosomes in the OMA lines (Rines et al. 2009). Polymerase chain reaction (PCR) primers to amplify various *zein* genes as well as *ZmO2* and *ZmPbf* are described elsewhere (Song and Messing 2003; Wu and Messing 2009; Miclaus et al. 2011). We also used a previously described primer pair for gene expression of actin as a control (Brautigam et al. 2005). Total RNA was extracted from leaf (30 days after sowing), stem (30 days after sowing), seedling (14 days after sowing), and immature endosperm (milk grain stage, 7–18 days postanthesis) using Qiagen RNEasy Kit. The same amount of RNA for each sample was then treated with DNAse and reverse transcribed to cDNA using SuperScript III First Strand Synthesis Kit (LifeTechnologies) using Oligo-dT primers. The cDNA was then used for PCR using Platinum *Pfx* polymerase (LifeTechnologies).

Expression studies of *AsPbf* copies were based on the protocol used by Song and Messing (2003) and Miclaus et al. (2011). Briefly, PCR primers were designed to amplify a 300-bp region that contains sequence polymorphisms between the two copies. The reverse transcriptase (RT)-PCR products were then ligated into pGEM-Teasy vector and transformed in *Escherichia coli* DH10B. Plasmid DNA was then extracted from 96 colonies and the clones were sequenced using M13 primers. The proportion of clones that correspond to each copy was taken as an indication of their relative expression levels.

To monitor the expression of *AsPbf* and 27-kDa *γ-zein* during OMA 7.06 endosperm development, PCR was performed on cDNA from developing endosperm at 7, 9, 11, and 13 days after anthesis (DAA) using *AsPbf* and 27 kDa *γ-zein* primers.

Cloning of Oat Pbf cDNA

To clone the oat *Pbf* (*AsPbf*), we designed PCR primers from *Pbf* sequence of barley (GenBank accession number AJ000991.1). Primers were selected as close to start or stop codons as possible. The primer sequences used were HvPBF-L (GAGGAAGTGTTTTCGTCCAA) and HvPBF-R (CATCAGGGAG GTGCTGTTGA). Total RNA was extracted from immature seeds (15 DAA) of oat (cultivar Starter) and was used for reverse transcription using RT-PCR. The resulting cDNA was used as template for PCR amplification using the barley *Pbf* primers. Amplicons were cloned in pGEM Teasy vector (Promega) and transformed into *E. coli* DH10B, after which we picked single colonies for plasmid extraction and DNA sequencing using M13 primers. The sequences were then BLASTed against the NCBI (*National Center for Biotechnology Information*) nonredundant nucleotide database to see whether they have significant hits against known *Pbf* sequences. The 5′ and 3′ rapid amplification of cDNA ends (5′- and 3′-RACE) were then conducted using the GeneRacer Kit (LifeTechnologies) to reveal the full coding sequence (CDS) of the candidate oat *Pbf*.

Plasmid Constructs and Transactivation Assay

To use GFP as a reporter of PBF activity, we modified an existing plasmid in our lab (PTF102-P4) that has a GFP gene driven by a *27-kDa γ-zein* promoter (P27:GFP) (Wu and Messing 2012) (fig. 5). We then removed the *Bar* gene in the T-DNA of pTF102-P4 using *Bgl*II and *Xho*1 restriction enzymes, and replaced it with our candidate *AsPbf1* gene such that it is now under the control of the 35S promoter (35S:AsPbf1). The resulting plasmid now contains a T-DNA with P27:GFP and 35S:AsPbf1 expression cassettes. To clone *AsPbf1* into the plasmid, it was amplified using primers with ends that overlap the target insertion site sequence (AsPbf-F AGCTAGATTGTTGAGCATTACATCAGGGAGCTGCT and AsPbf-R TCATTTGGAGAGGACCATGGAGGAAGTGTTGTCAT). This enabled us to use a Gibson reaction kit (NEB) to piece the linearized PTF102-P4 without the *Bar* gene and the candidate *AsPbf* together. The ligation ends were then sequenced with the ABI3730XL platform to verify the insertion, and to identify which copy was inserted. This new plasmid was then renamed pTF102-P6. Another plasmid also available in our lab, pTF102-P5, was used as a positive control. This plasmid contains P27:GFP and *ZmPbf* driven by 35S promoter (35S: ZmPBF) (Wu and Messing 2012). The original pTF102-P4 with only the P27:GFP on the other hand served as the negative control.

The plasmid constructs described above were used in a transient expression experiment by Agro-injection method in tobacco (*Nicotiana benthamiana*) leaves as previously described (Gallavotti et al. 2011). Briefly, the plasmid constructs were transformed into *Agrobacterium tumefaciens* strain EHA101 and plated on YEP-agar media. Single colonies were then picked and grown in liquid YEP media for about 18–20 h, aiming for an optical density (OD_{600}) measurement of 1. The cultures were then centrifuged and the bacterial pellets were resuspended and washed with 10 mM $MgCl_2$. This cycle of centrifugation and washing was done twice, until the pellet was finally resuspended in 5 ml of 10 mM MgCl2. This was mixed with an equal amount of *Agrobacterium* suspension containing a plasmid expressing

p19 protein to increase the GFP signal. Tobacco leaves were then injected on the abaxial side using a syringe. At least three leaves were injected for each treatment. Imaging of the leaves was done 5 days after injection using a fluorescence microscope with an appropriate filter. The same camera setting was used for all samples.

GFP Quantification

Western blot was used to measure the GFP signal from the transactivation assay in a semiquantitative way. Using a leaf punch, leaf tissues from Agro-injected leaves were collected in 2-ml tubes and ground in liquid nitrogen. Total protein was then extracted using Tris-beta mercaptoethanol buffer. To ensure equal loading, the protein extracts were quantified using the DC Protein Quantification kit (BioRad). One microgram of the protein extract (three biological replicates per treatment) was loaded onto a 15% SDS-PAGE for electrophoresis, and then blotted on a PVDF membrane (BioRad). The membrane was then blocked with 5% nonfat dry milk in Tris-Buffered Saline and Tween 20 (TBST) buffer. Rabbit GFP antibody (LifeTechnologies) was then used to probe the membrane, followed by goat antirabbit immunoglobulin G (IgG) secondary antibody with horseradish peroxidase (HRP) conjugate (LifeTechnologies). The membrane was then incubated for 5 min in HRP substrate (BioRad Western ECL) and exposed on X-ray film.

To quantitate the GFP bands from the Western blot, densitograms were made using ImageJ software (http://imagej.nih.gov/ij/, last accessed April 30, 2015). First, Western blot images were saved as TIFF files and converted to 8-bit image. Bands were selected using the rectangle tool and the background was cropped. The saved band images were then lined up vertically on PowerPoint, allowing sufficient white space in between, and saved as TIFF file. The new TIFF file was opened in ImageJ and converted back to an 8-bit image. The rectangle tool was used to drag a box surrounding the bands to define the lane. The densitogram for the bands was then created using "Plot lanes" command. The area inside each peak in the densitogram, which corresponds to the band density, was then calculated using the "Wand" tool.

GFP fluorescence was also measured directly on the image files using ImageJ. The pictures were opened in ImageJ and then converted to 16-bit image files. For each image, the whole field was selected using the rectangle tool and measurements were set to take integrated density, area measured, and mean gray value (integrated density over area measured). The mean gray values were then compared between treatments.

Sequence Analysis

DNA and protein sequence alignment was done in MegAlign program of DNASTAR Lasergene 10 software package. The maximum-likelihood phylogenetic tree of predicted protein sequences of previously described PBF from maize (ZmPBF, GenBank accession number NP_001105400.1), rice (OsPBF, GenBank accession number BAA78574.2), wheat (TaPBF, GenBank accession number CAA09976.1), and barley (HvPBF, GenBank accession number CAA04440.1), and a putative PBF sequence from *Brachypodium* (BdPBF, GenBank accession number XP_010239574.1) was created using the JTT model (Jones et al. 1992) as implemented in software package MEGA (Tamura et al. 2007).

Results

Expression of ZmPbf, O2, and Zeins in OMA Lines

OMA lines represent single disomic maize chromosome additions. The presence of intact corresponding maize chromosomes on the OMA lines was verified by PCR assay using SSR markers (fig. 1*A*). Both SSR primer pairs on the short and long chromosome arms successfully amplified, indicating that the maize chromosomes were present. The *zeins* and their regulator genes (*ZmO2* and *ZmPbf*) were also PCR-amplified and mapped correctly to the OMA lines with the corresponding maize chromosome addition (fig. 1*B*). *Zein* regulators *ZmO2* and *ZmPbf* were PCR-amplified on OMA7.06 and OMA2.51, respectively, whereas the *zeins* were PCR-amplified on OMA2.51, 4.42, 6.33, 7.06, and 9.41. The oat parent used in creating the OMA lines did not give any amplification.

For the gene expression assay, we found that both *zein* regulators *ZmO2* and *ZmPbf* were expressed in immature endosperm of OMA7.06 and OMA2.51, respectively (fig. 1*B*). For the *zeins*, we were only able to detect the expression of two γ-*zein* genes (*27-* and *50-kDa*), both of which are on OMA7.06. No expression was detected for the rest of the *zeins*. The expressed *zeins* also maintained their tissue specificity as they were expressed only in endosperm tissues but not in leaves. Therefore, regulation of gene expression must have diverged between group I and group II prolamin genes between these species.

Cloning and Characterization of Oat Pbf (AsPbf)

As previous studies have shown that ZmPBF is the main transcriptional regulator of the *27-kDa γ-zein* (Marzabal et al. 2008; Wu and Messing 2012), its expression in the absence of *ZmPbf* in OMA7.06 led us to hypothesize that the ZmPbf homolog in oat can substitute for it. Because barley is more closely related to oat than maize, we cloned *AsPbf* by PCR using primers designed from a previously cloned barley *Pbf* (*HvPBF*) gene (Mena et al. 1998) to validate our assumption. Indeed, PCR of oat immature endosperm cDNA using the barley primers gave us an amplicon of approximately 1 kb in size. Sequencing and BLAST (Basic Local Alignment Search Tool) analysis of the amplicon showed that our sequence is

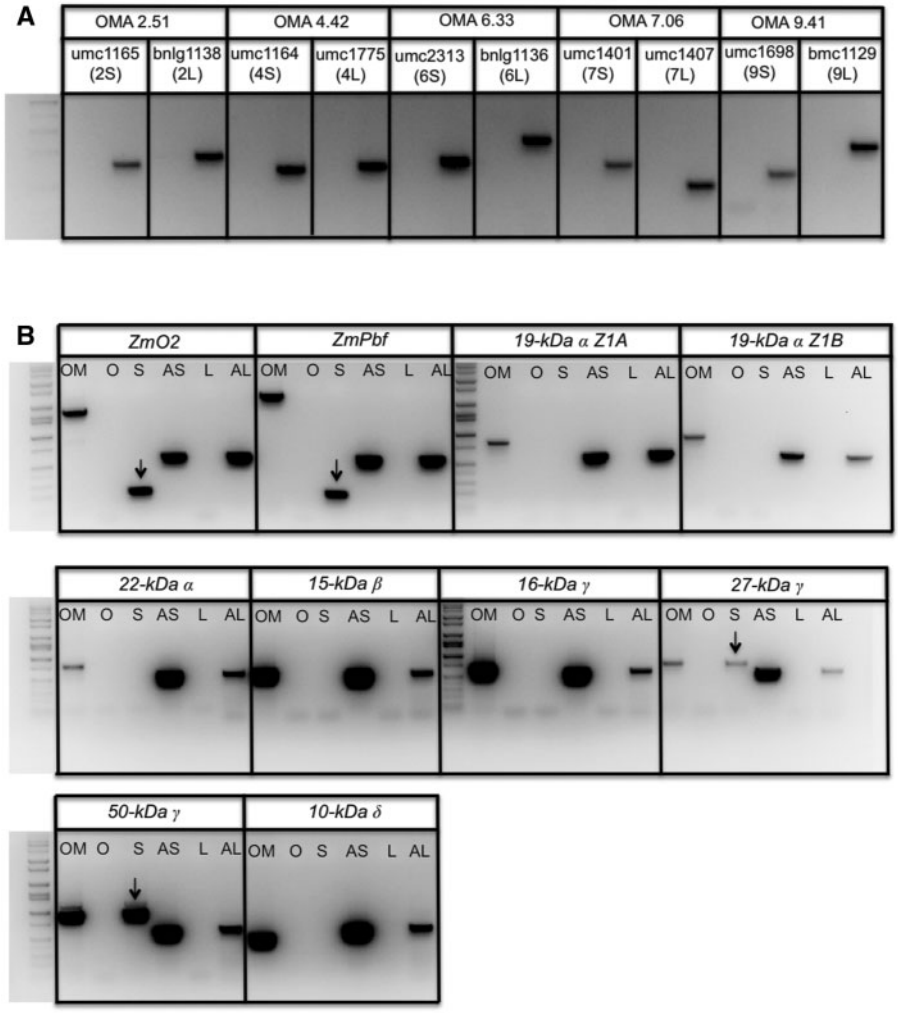


Fig. 1.—(*A*) Verification of presence of maize chromosomes in OMA lines. For each SSR marker, the first lane is the oat parent. The second lane is the OMA line. (*B*) Expression of maize *Pbf*, maize *O2*, and zeins in OMA lines. OM, PCR of OMA line where the gene being assayed is present; O, PCR of oat parent control; S, RT-PCR of seed mRNA; AS, actin control for RT-PCR of seed mRNA; L, RT-PCR of leaf mRNA; AL, actin control for RT-PCR of leaf mRNA. Arrows indicate genes that are expressed in the seed. OMA 2.51 was used to assay for *ZmPbf* and *16-kDa γ-zein*, OMA 4.42 for 19-kDa *α-zein* Z1A and *22-kDa α-zein*, OMA 6.33 for *16-kDa β-zein,* OMA 7.06 for 19-kDa *α-zein* Z1B, *27-kDa γ-zein*, *50-kDa γ-zein*, and *ZmO2*, and OMA 9.41 for *10-kDa δ-zein*.

very similar to barley and wheat *Pbf* genes, as well as other Dof domain-containing genes from other grasses.

The 5′-RACE revealed sequence variations upstream of the start codon, which indicates variation in 5′-UTR sequences and the presence of at least two copies of *AsPbf*. The presence of small indels and single nucleotide polymorphisms (SNPs) within the CDS of these two genes also supports the presence of two copies. We named these two copies *AsPbf1* and *AsPbf2*. AsPBF1 (325-amino acid long) is slightly shorter than AsPBF2 (328-amino acid long) but they are highly similar, sharing 87.7% predicted protein sequence identity (fig. 2*B*).

The predicted protein sequences of *AsPbf1* and *AsPbf2* were compared with the Pfam database (http://pfam.xfam.org/search, last accessed April 30, 2015); indeed both possess a Dof domain, a defining characteristic of PBF proteins. The Dof domains are close to the amino terminal end, consistent with PBF proteins from other grasses (fig. 3*A*). The Dof domain is highly conserved between oat, barley, and wheat PBF proteins, which is located between the 31st and 93rd amino acid residues. The two AsPBF sequences also contain an asparagine-rich C-terminus like the other PBF proteins.

The result of the 5′-RACE indicated that the two *Pbf* copies are not expressed equally because of the difference in the number of reads of the two copies. To verify this, we cloned the RT-PCR products in pGEM-Teasy vector, transformed them in *E. coli* DH10B, and selected 96 white bacterial colonies for sequencing with the ABI3730XL platform. The

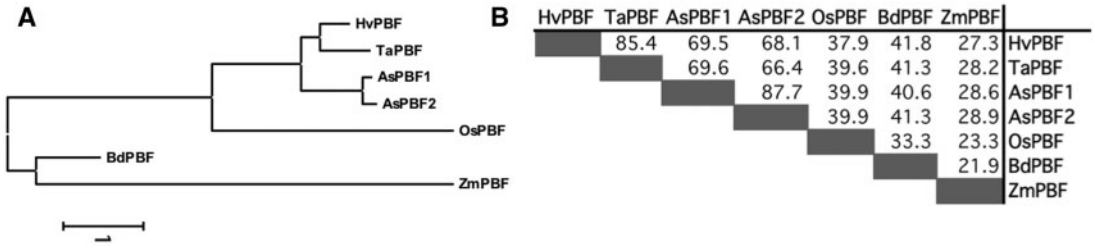


Fig. 2.—(*A*) Phylogenetic tree of PBF proteins from several grass species. The bar below is the distance scale in amino acid substitutions per site. (*B*) Pairwise percent identity of the protein sequences. HvPBF, barley PBF; TaPBF, wheat PBF; AsPBF, oat PBF; OsPBF, rice PBF; BdPBF, *Brachypodium* PBF; ZmPBF, maize PBF.

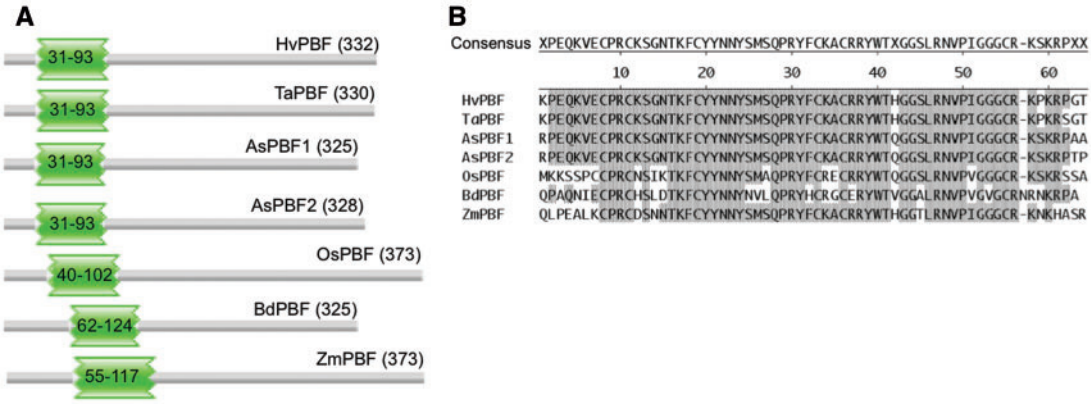


Fig. 3.—(*A*) Comparison of PBF proteins from different grass species. The green boxes with numbers indicate the position of the Dof domain. The number after the protein name is the length of the protein. (*B*) Alignment of Dof domain of PBF proteins from different grasses.

presence of two small indels and SNPs between the two *AsPbf* copies allowed us to make a quantitative comparison of the expression level of the two copies. The result of the sequencing showed that the *AsPbf1* copy is present in 87 out of the 96 clones, whereas sequence reads that correspond to *AsPbf2* showed up in only nine clones. This is evidence that *AsPbf1* is more strongly expressed than *AsPbf2*.

The phylogenetic tree based on predicted protein sequences of PBF sequences from different grasses agrees with their known phylogenetic relationships (fig. 2*A*). HvPBF and TaPBF are grouped together as members of the *Triticeae* tribe, and the two AsPBF sequences are more closely related to this tribe than to the PBF sequences from other species. It is interesting to see that the putative BdPBF did not show a close relationship to oat, barley, and wheat given that they all belong to the *Pooideae* subfamily. Because of the absence of the oat genome sequence, it is currently not possible to investigate whether the two *AsPbf* copies are being contributed by two of the three subgenomes of oat, or are duplicated genes within one subgenome. However, the presence of only one putative *Pbf* in *Brachypodium* indicates that the two copies are more likely contributed by two of the three oat subgenomes.

We then examined the spatial expression pattern of our candidate oat *Pbf* genes by investigating their expression on leaf, stem, seedling, and root tissues in addition to seed endosperm (fig. 4*A*). Like other *Pbf* genes, our candidate *AsPbf* genes are only expressed in the endosperm. In addition, the timing of expression of the *27-kDa γ-zeins* coincides with the expression of the *AsPbf* in OMA 7.06 (fig. 4*B*).

AsPBF Can Trans-Activate 27-kDa γ-Zein

To investigate whether the AsPBF that we cloned can indeed substitute for ZmPBF's function in OMA 7.06, a transient expression assay in tobacco leaves was performed using plasmids shown in figure 5. To monitor the activation of *27-kDa γ-zein* promoter by oat PBF, *AsPbf1* was selected as it has higher expression compared with *AsPbf2* based on our previous gene expression analysis. The fluorescences of GFP were compared among treatments in absence of PBF, presence of ZmPBF, and presence of AsPBF1. As seen in figure 6*A*, the fluorescence of GFP in the treatment containing *ZmPbf* and *AsPbf1* was more intense than the control without any *Pbf* for all of the three replicates, indicating that AsPBF1 activates the *27-kDa γ-zein* promoter just like ZmPBF. Quantification of GFP

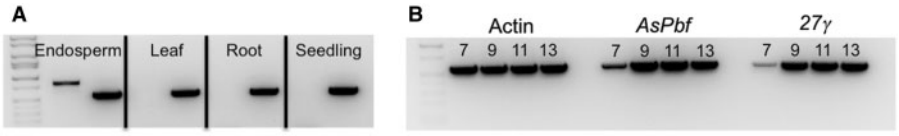


Fig. 4.—(*A*) Expression of *AsPbf* from different oat tissues. For each tissue, the first lane is the RT-PCR for *AsPbf* whereas the second lane is the actin control. (*B*) Expression of *AsPbf* and *27-kDa γ-zein* in OMA 7.06 at 7, 9, 11, and 13 DAA.

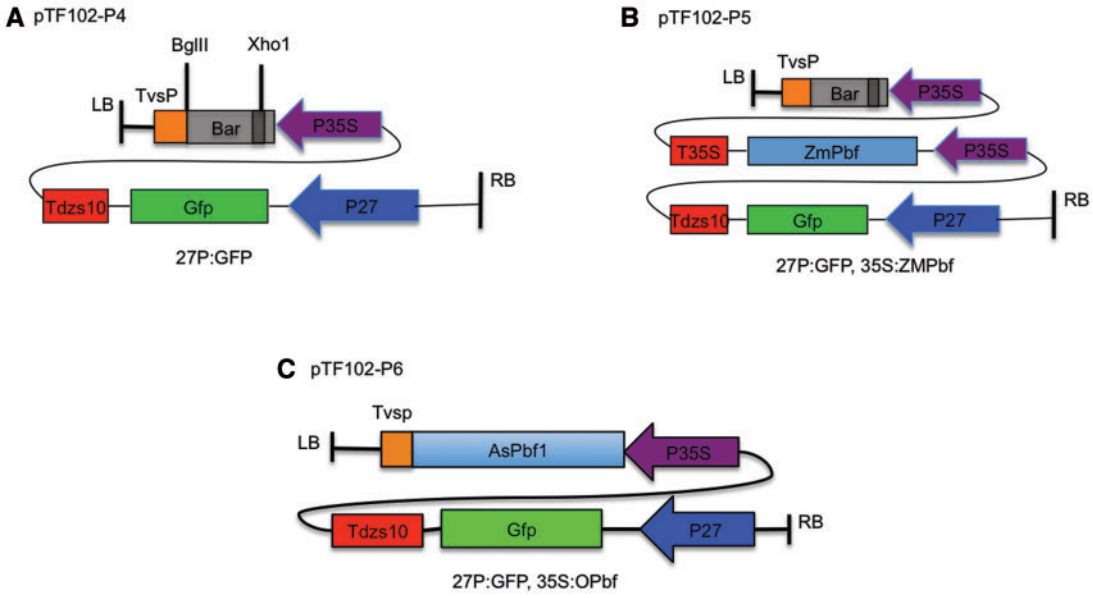


Fig. 5.—Plasmid constructs used for transient expression assay. (*A*) pTF102-P4 contains a *Gfp* gene driven by *27-kDa γ-zein* promoter. (*B*) pTF102-P5 contains maize *Pbf* (*ZmPbf*) driven by 35S promoter in addition to 27P:GFP. (*C*) pTF102-P6 is a modified pTF102-P4, with the *Bar* gene replaced by *AsPbf1*.

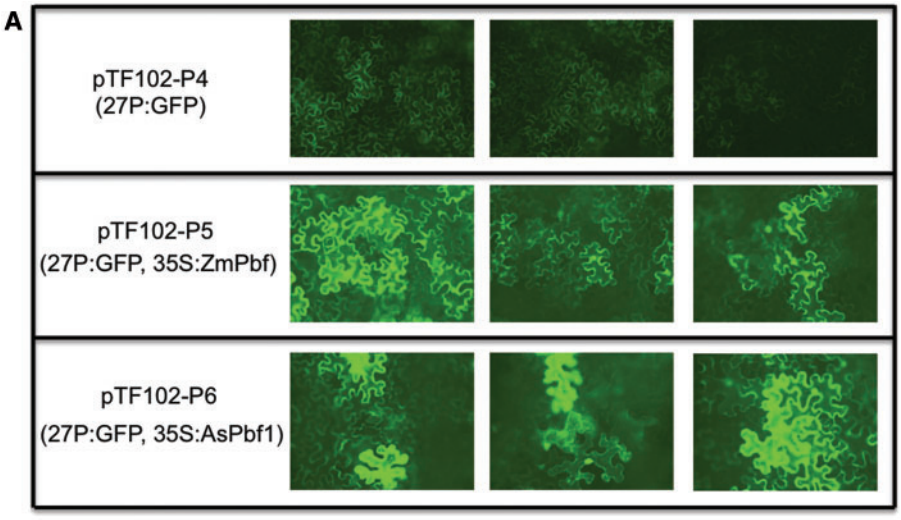
from the image files indicates that the presence of *AsPbf1* resulted in a 2-fold increase in fluorescence on average over the control without any *Pbf*, and that this is statistically significant ($P < 0.05$) based on a *T*-test. The densitograms of the Western blot also showed that the GFP band is twice more intense on average in the treatment with *AsPbf1* than without *AsPbf1* and the difference is also statistically significant (fig. 6*B*).

Wheat Dx5 Is Not Controlled by Pbf or O2

If the expression of the *γ-zeins* is conserved between the two subfamilies of the grasses, the *Pooideae* and the *Panicoideae*, one could ask whether the reciprocal is true as well. The closest prolamins to the *γ-zeins* in the Panicoideae subfamily are the glutenins and the gliadins (Xu and Messing 2009). Indeed, transformation of wheat prolamin *Dx5*, an HMW-glutenin, with its native promoter into maize has led to its successful expression in maize endosperm (Sangtong et al. 2002), but it was unclear which maize trans-acting factor activated the expression of the wheat glutenin gene. Surprisingly, Dx5 protein accumulation was not reduced in the *PbfRNAi* or *o2* backgrounds. There is also no visible reduction of Dx5 in the *PbfRNAi;o2* background (fig. 7). These results suggest that the *Dx5* gene was differently regulated than the *γ-zeins*. Given that *Dx5* is older (group III) than the *γ-zein* genes, it appears that the group II prolamin genes have acquired *cis*-acting elements for their regulation by PBF and O2 after amplification from group I.

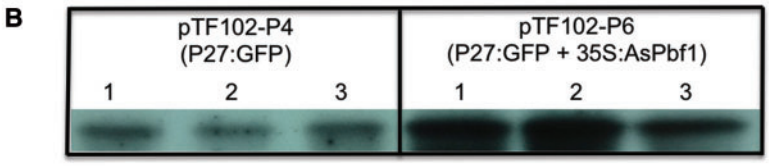
Discussion

The control of seed protein gene expression in cereals has been studied extensively due to their importance in determining the grain's nutritional and physical properties. Cereal species are highly conserved in terms of gene content and synteny, but their seed protein composition and properties can vary significantly. Although most cereals have prolamins as their main seed storage proteins, oat and rice are exceptions in that they contain mostly globulins (Shewry et al. 1995). Also, although the majority of both wheat and maize



Construct	Replicate	Area	Integrated density	Mean Gray Value
pTF102-P4*	1	153114	3338176	21.80
	2	153340	3128425	20.40
	3	153567	2617827	17.05
			Average	19.75
pTF102-P6*	1	153340	5845432	38.12
	2	153567	5624795	36.63
	3	153340	7514108	49.00
			Average	41.25

*p-value = 0.003



Construct	Replicate	Densitogram value
pTF102-P4*	1	6953.7
	2	5561.4
	3	5608.6
	Average	6041.2
pTF102-P6*	1	12076.8
	2	12938.6
	3	10812.7
	Average	11942.7

*p-value = 0.0008

Fig. 6.—(*A*) Tobacco leaf transient expression experiment to assay the activity of PBF using GFP as a reporter. Software quantification of fluorescence from the image files using ImageJ is shown at the right. (*B*) Quantification of GFP signal in Western blot using anti-GFP antibody. The densitogram values for the bands are shown below.

seed storage proteins are prolamins, maize does not contain the wheat prolamin variants (gluten proteins) that are needed for bread making. The characterization and cloning of many cereal seed protein genes with different properties open up the prospect of engineering their seed protein compositions and expanding their utilization. Thus, studying conservation or divergence of genetic control is important for heterologous expression of these genes among cereals.

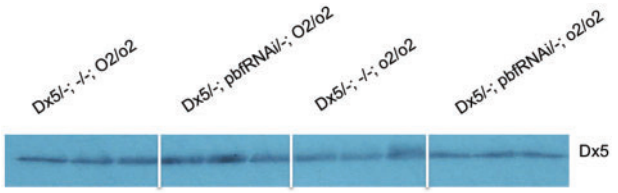


Fig. 7.—Western blot showing the accumulation of Dx5 protein in different *O2/Pbf* backgrounds.

Previous gene regulation studies of *22-kDa α-zeins* showed that their tissue specificity and temporal expression are regulated epigenetically. This occurs at the DNA level through methylation, and at the chromatin level through histone methylation and acetylation marks (Locatelli et al. 2009). The presence of individual maize chromosomes in the OMA lines separates it from regulatory elements present in other chromosomes. In our RT-PCR experiments, we did not see any expression of *ZmO2*, *ZmPbf*, or any of the *zeins* in the leaf. However, the coordinated temporal expression of *ZmPbf* and the *27-kDa γ-zein* was well preserved. These results indicate that the epigenetic states of these genes were preserved, and that the oat host genome was able to correctly assign and maintain appropriate epigenetic marks on *ZmO2, ZmPbf,* and *27-kDa γ-zein*. On the other hand, failure of some *zeins* to be expressed maybe due to either silencing by epigenetic modification or the absence/incompatibility of the host regulators to trans-activate them.

PBF and O2 are the two major *zein* regulators in maize. Previous reports have provided evidence for the role of *O2* in *zein* gene expression. In an *o2* mutant, both the *22-kDa α-* and *15-kDa β-*zein proteins were reduced (Vicente-Carbajosa et al. 1997). We did not detect expression of both genes in the OMA lines where they are located, which can be due to two reasons. First, OMA 4.42 and OMA 6.01 (which contain the *22-kDa α-* and *15-kDa β-zein* genes, respectively) do not contain its native *ZmO2* regulator because it is present on another OMA line (OMA 7.06). Second, in the absence of *ZmO2*, the *AsO2* was not able to trans-activate them. Whether this is due to failure of AsO2 to recognize the *22-kDa α-* and *15-kDa β-zein* promoters or the silencing of these genes through epigenetic modifications is unknown at this time. The *19-kDa α-zein* cluster on OMA 7.06 is also not expressed even when *ZmO2* is expressed in this OMA line. Interestingly, this cluster arose from one of the progenitor of maize before allotetraploidization, whereas the *19-kDa α-zein* cluster on chromosome 1 and 4 arose from the other homeolog (Xu and Messing 2008).

Previous studies have shown that PBF binds to the P-box motif (TGTAAAG) (Vicente-Carbajosa et al. 1997). In *ZmPbfRNAi* transgenic lines, only the *22-kDa α-* and 27-kDa γ-zeins are reduced but other zeins accumulate normally. This is surprising after analysis of *zein* promoters revealed that most have P-box motifs (Wu and Messing 2012). Although ZmPbf might not be necessary for expression of all *zeins*, it is clearly required for the expression of the *27-kDa γ-zeins* (Marzabal et al. 2008). Thus, the expression of *γ-zeins* in OMA 7.06 even without *ZmPBF* points to a trans-acting action of AsPBF on maize *γ-zeins*. Our transient expression data are evidence that this might indeed be the case.

All PBFs have a Dof domain near the N-terminus, which is highly conserved across the cereals (fig. 3*B*). The high degree of sequence conservation in the position and alignment of the Dof domain might explain the ability of AsPBF to trans-activate the *γ-zeins* in OMA 7.06. This trans-activating property was also observed for ZmPBF, which has been shown to recognize and activate glutelin promoters in transient assay using rice endosperm tissue culture (Hwang et al. 2004). Based on these evidences from oat and rice, it is very likely that TaPBF, HvPBF, or OsPBF can also bind other prolamins that contain P-box in their promoters.

Oat has an allohexaploid genome consisting of A, C, and D homeologous chromosome sets. Species that underwent polyploidization often show bias in gene expression between homeoloci. Studies in maize, an ancient tetraploid, have shown that gene expression can change between homeologous copies, with one homeolocus expressed significantly higher than the other (Schnable et al. 2011; Jiang et al. 2013). A recent genome-wide study in wheat has also shown that 73–76% of expressed homeoloci show significant gene expression difference, and that one of the three homeoloci dominate the gene expression (Leach et al. 2014). The divergence in the regulation of homeoloci after polyploidization may explain the significant difference in gene expression between the two copies of *AsPbf* that we detected.

A previous study based on *Dx5* promoter segment deletions has shown that the G-box-like motif is important and fundamental to its expression (Norre et al. 2002). The deletion of this G-box-like motif resulted in a significant reduction of promoter activity. On the other hand, the deletion of a known enhancer element (Thomas and Flavell 1990) that contains a P-box motif resulted in a nonsignificant reduction of promoter activity. The study also showed that stronger expression can be achieved when the G-box-like motif is paired with the P-Box and if these two are repeated in tandem. The same study also found that the G-box-like motif can bind at least two types of transcription factors—bZip factors such as O2, as well as those belonging to the ASF-1 family—due to the similarity of their binding patterns in gel-shift assays. In the absence of any reduction in Dx5 protein in *PbfRNAi*, *o2*, and *PbfRNAi-o2* backgrounds, our data are in agreement with the finding that other transcription factors can bind to the G-box-like motif. Therefore, there seems to be a genetic redundant mechanism that controls the expression of *Dx5* (and perhaps other G-box containing prolamin promoters) that enables them to be expressed in the absence of one of the *Pbfs* and, or *O2*.

Apparently, the redundancy of transcription factors in these species appears to be a prerequisite for recently amplified gene copies to acquire the capacity to switch to different regulators, as long as they achieve a suitable expression pattern. This is particularly evident from the group I prolamins. Even, when both PBF and O2 are knocked down, there are still α-zeins being expressed. Moreover, there are even maize haplotypes of the *z1C1* locus that can differ in copy number variation and their transcriptional activation by O2 despite the conservation of the *cis*-acting elements defined by DNA-protein-binding studies and promoter truncation experiments. Interestingly, this change affects the two youngest, 3′-terminal copies in the tandem gene clusters (Song et al. 2001). Our studies in oats are consistent with this trend, where promoter regions are acquiring mosaics of elements, when copied.

Acknowledgment

This work was supported with the Selman A. Waksman Chair in Molecular Genetics to J.M.

Literature Cited

Brautigam M, et al. 2005. Generation and analysis of 9792 EST sequences from cold acclimated oat, *Avena sativa*. BMC Plant Biol. 5:18.

Cord-Neto G, et al. 1995. The involvement of Opaque2 on beta-prolamin gene regulation in maize and coix suggests a more general role for this transcriptional activator. Plant Mol Biol. 27:1015–1029.

Feng L, et al. 2009. Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize a-zein super gene family. Plant Mol Biol. 69:649–659.

Gallavotti A, et al. 2011. BARREN STALK FASTIGIATE1 is an AT-hook protein required for the formation of maize ears. Plant Cell 23: 1756–1771.

Gaut B. 2002. Evolutionary dynamics of grass genomes. New Phytol. 154: 15–28.

Herman EM, Larkins BA. 1999. Protein storage bodies and vacuoles. Plant Cell 11:601–613.

Hwang Y, et al. 2004. The maize O2 and PBF proteins act additively to promote transcription from storage protein gene promoters in rice endosperm cells. Plant Cell Physiol. 45:1509–1518.

Jiang WK, Liu YL, Xia EH, Gao LZ. 2013. Prevalent role of gene features in determining evolutionary fates of whole-genome duplication duplicated genes in flowering plants. Plant Physiol. 161:1844–1861.

Jones D, Taylor W, Thornton J. 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8:275–282.

Kellogg EA. 2001. Evolutionary history of the grasses. Plant Physiol. 125: 1198–1205.

Kynast RG, et al. 2001. A complete set of maize individual chromosome additions to the oat genome. Plant Physiol. 125:1216–1227.

Leach L, et al. 2014. Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. BMC Genomics 15:276.

Lending CR, Chesnut RS, Shaw KL, Larkins BA. 1989. Immunolocalization of avenin and globulin storage proteins in developing endosperm of *Avena sativa* L. Planta 178:315–324.

Locatelli S, Piatti P, Motto M, Rossi V. 2009. Chromatin and DNA modifications in the Opaque2-mediated regulation of gene transcription during maize endosperm development. Plant Cell 21:1410–1427.

Marzabal P, et al. 2008. The maize Dof protein PBF activates transcription of gamma-zein during maize seed development. Plant Mol Biol. 67: 441–454.

Mena M, Vicente-Carbajosa J, Schmidt RJ, Carbonero P. 1998. An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. Plant J. 16:53–62.

Miclaus M, Xu J-H, Messing J. 2011. Differential gene expression and epiregulation of alpha zein gene copies in maize haplotypes. PLoS Genet. 7:e1002131.

Norre F, et al. 2002. Powerful effect of an atypical bifactorial endosperm box from wheat HMWG-Dx5 promoter in maize endosperm. Plant Mol Biol. 50:699–712.

Rines HW, et al. 2009. Addition of individual chromosomes of maize inbreds B73 and Mo17 to oat cultivars Starter and Sun II: maize chromosome retention, transmission, and plant phenotype. Theor Appl Genet. 119:1255–1264.

Salse J, Feuillet C. 2007. Comparative genomics of cereals. In: Varshney RK, Tuberosa R, editors. Genomics-assisted crop improvement. Dordrecht (The Netherlands): Springer. p. 177–205.

Sangtong V, et al. 2002. Expression and inheritance of the wheat Glu-1DX5 gene in transgenic maize. Theor Appl Genet. 105:937–945.

Schmidt RJ, Burr FA, Aukerman MJ, Burr B. 1990. Maize regulatory gene Opaque-2 encodes a protein with a leucine-zipper motif that binds to zein DNA. Proc Natl Acad Sci U S A. 47:46–50.

Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G. 1992. Opaque2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. Plant Cell 4:689–700.

Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci U S A. 108:4069–4074.

Shewry PR, Halford NG. 2002. Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot. 53:947–958.

Shewry PR, Napier JA, Tatham AS. 1995. Seed storage proteins: structures and biosynthesis. Plant Cell 7:945–956.

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. 2003. Gene expression of a gene family based on non-collinear haplotypes. Proc Natl Acad Sci U S A. 100:9055–9060.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.

Thomas MS, Flavell RB. 1990. Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell 2:1171–1180.

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 basic leucine zipper transcriptional activator Opaque2. Proc Natl Acad Sci U S A. 94:7685–7690.

Wu Y, Messing J. 2009. Tissue-specificity of storage protein genes has evolved with younger gene copies. Maydica 54:409–415.

Wu Y, Messing J. 2012. Rapid divergence of prolamin gene promoters of maize after gene amplification and dispersal. Genetics 192:507–519.

Xu J, Messing J. 2008. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. Proc Natl Acad Sci U S A. 105:14330–14335.

Xu JH, Messing J. 2009. Amplification of prolamin storage protein genes in different subfamilies of the Poaceae. Theor Appl Genet. 119: 1397–1412.

Zhang W, Ciclitira P, Messing J. 2014. PacBio sequencing of gene families—a case study with wheat gluten genes. Gene 533:541–546.

Zhang W, Sangtong V, Peterson J, Scott MP, Messing J. 2013. Divergent properties of prolamins in wheat and maize. Planta 237:1465–1473.

Associate editor: Yves Van De Peer