

Expression of the sorghum 10-member kafirin gene cluster in maize endosperm

Rentao Song, Gregorio Segal and Joachim Messing*

Waksman Institute, Rutgers University, 190 Frelinghuysen Road, Piscataway, NJ 08854-8020, USA

Received August 16, 2004; Revised and Accepted November 29, 2004

ABSTRACT

Functional analysis of chromosomal segments containing linked genes requires the insertion of contiguous genomic sequences from bacterial artificial chromosomes (BACs) into the genome. Therefore, we introduced a 90-kb large BAC clone carrying a 10-copy tandem array of kafirin storage protein genes from sorghum linkage group J, mixed with a selectable marker gene, directly into maize cells using the particle bombardment method. Transgenic plants were regenerated and seeds from eight different transgenic lines were produced. One such transgenic plant was selected that had the entire kafirin gene cluster on a single continuous DNA fragment spanning more than 45 kb integrated into its genome. When alcohol-soluble proteins from individual T2 and T3 seeds of this event were analyzed, significant levels of kafirin were found in addition to the endogenous zein storage proteins, demonstrating that the large exogenous DNA segment is stably integrated into the maize genome and expressed at high levels in subsequent generations. Therefore, we could provide a new utility of plant transformation by the particle bombardment method for functional genomics of multigene families and the modification of the nutritive quality of cereal grains. Despite a tandem array of highly homologous sequences at the transgenic locus, no gene silencing was observed, probably owing to the effects of co-transformed flanking sequences. The expression studies of the transgenic locus also revealed new features of storage protein gene promoters that differed from previous transient gene expression studies, thereby illustrating the significance of the concentration and configuration of DNA–protein interactions in the regulation of gene expression.

INTRODUCTION

Plant transformation of large DNA fragments would greatly facilitate map-based cloning, functional analysis of large

genes and distantly located regulatory elements, and the genetic engineering of major crops. Instead of combining transgenes by crossing multiple integration events, the time of introgressing genes into the same background could be accelerated and multiple genes can become genetically linked, thus simplifying the subsequent breeding manipulations. More importantly, it has been shown that stable expression of transgenes frequently require flanking sequences that act as insulators (1). Yet another reason is that many proteins are produced from gene families whose members are tandemly arrayed and thereby occupy large DNA segments in their chromosomal location. In those cases, it is difficult to determine the contribution of individual gene copies in gene expression without analyzing the intact gene cluster. Clusters like those of disease resistance genes can provide broader spectra of resistance to different pathogens if introduced into a new genetic background in their entirety (2). Another typical example is the clustered storage protein gene families of cereals, the activity of which can result in extremely high gene product accumulations, not achievable by any individual gene member (3). Precise quantification of gene activities and allelic interactions provided by the naturally occurring genomic environment justifies the fact that gene clusters are better studied as a whole.

One possible way to accomplish the introduction of large segments of genomic DNA into plants by transformation has been achieved by converting bacterial artificial chromosome (BAC) vectors into a binary vector system for *Agrobacterium*-mediated transformation capabilities. Such binary BAC vectors containing large fragments of human DNA (BIBAC) and *Arabidopsis* DNA (TAC) have been used to transform tobacco and *Arabidopsis*, respectively (4,5). However, in neither case the expression of multiple genes was intended. Moreover, some studies have shown transferred DNA instability occurring in instances when propagating such vectors in *Agrobacterium* (6). Such instability could be significant if the BAC clone contained a plant genomic DNA segment with tandemly arrayed, highly conserved genes copies. Although such clones are known to be stable in *E.coli*, such plant transformation protocols require the mobilization of the BAC clone by conjugation into *Agrobacterium*, where the cloned-DNA could become unstable.

Therefore, it would be attractive to avoid the additional step of mobilizing gene clusters into *Agrobacterium* and transform plants directly with DNA from *E.coli*. This would have the

*To whom correspondence should be addressed. Tel: +1 732 445 4256; Fax: +1 732 445 0072; Email: messing@mbcl.rutgers.edu
Present address:

Rentao Song, School of Life Science, Shanghai University, 99 Shang Da Road, Shanghai 200436, Republic of China

additional attraction that any BAC vector could be used because many plant genomic libraries are not made with vectors suitable for *Agrobacterium*-mediated transformation methods. To pursue direct DNA uptake, methods would require co-transformation with a plasmid containing a suitable selectable marker gene and the particle bombardment process. One interesting idea is to aid this process with a site-directed recombination system (7). In such a case, the BAC vector contains a *lox* site that can recombine with a *lox* site in the plant genome, when the *cre* gene is expressed. Therefore, only plant cells that have such a *lox* site already engineered in their genome can be used for such transformation methods. However, after transformation of such a tobacco/*lox* system with a BAC clone containing cotton DNA, integrity and stability of the transgenic DNA remained unclear because the gene content of the cotton DNA and its expression in subsequent generations still needed to be determined. In another case, BAC fragments were directly transformed into potato to identify a disease resistance gene against late blight by map-based cloning (8). Since such complementation requires only the portion of the BAC DNA containing the candidate gene, transgenic plants exhibiting a hypersensitive response do not confirm the integration of large DNA segments. One would need in addition the analysis of the integrated DNA by pulsed-field gel electrophoresis to determine its size from a single restriction fragment.

Nevertheless, these approaches illustrate that the co-transformation method using the particle bombardment process is very promising and a useful alternative to the transformation of large DNA segments by *Agrobacterium*-mediated DNA transfer. What was needed in our case was the transformation of a DNA segment containing an entire gene cluster, which upon transformation could be examined for gene expression in subsequent generations. Such an example of a gene cluster in plants is seed storage protein genes. These genes are expressed either in the embryo or endosperm of the seed and their products are an important source for essential amino acids in the human and animal diet. In maize, the majority of seed storage proteins are the prolamins, which constitute half of the proteins in endosperm and are also called zeins (9). For instance, in maize inbred B73 the gene family that codes for the 19 and 22 kDa α -zeins consists of 41 members distributed over seven locations in the genome (10,11). In many of these genomic locations, members of this gene family are arranged in tandem arrays of highly homologous sequences, indicating that their amplification has occurred recently. Interestingly, tandem amplification of such a gene has also occurred in a close relative of maize, sorghum (12), where they are called kafirin genes (13). When the progenitors of sorghum and maize split ~ 11.9 mya (14), each of them contained only one or two copies of such a gene (3). Given this divergence and independent amplification the question arose whether gene regulation has also changed. Therefore, the entire kafirin cluster would have to be inserted into the maize genome and stably transmitted to determine which regulatory elements are conserved in their function over such a period of time. The sorghum gene cluster contains 10 tandem copies of the 22 kDa α -kafirin class, with a repeat unit of ~ 3.5 kb. In contrast to the orthologous 22 kDa zein cluster, the kafirin cluster is not interrupted by transposable elements and does not contain any

pseudogenes. It is therefore only ~ 35 kb in size, and is contained within a single 90 kb long BAC clone (12). Therefore, we co-transform this BAC clone into maize cells by the particle bombardment process as described previously (15). We could show that the kafirin cluster was integrated into the maize genome as a single contiguous fragment, segregating as a Mendelian trait by expressing kafirins in maize endosperm in subsequent generations. Tissue-specific expression and processing of kafirins in maize indicated that regulatory elements between these two species were conserved both at the transcriptional and post-transcriptional level. In contrast to transient expression studies of reporter genes under the control of over-expressed transcription factors, promoters in their normal chromosomal context provided us a different picture of gene regulation than previously believed. Demonstration of the expression of linked genes as a transgenic locus should therefore enhance future functional genomics studies in plants.

MATERIALS AND METHODS

Plant material

High Type II maize lines A and B from our own seed stocks were grown under greenhouse conditions. Immature embryos derived from B \times A hybrid ears were excised and subcultured to induce highly embryogenic callus formation as described previously (16).

DNA clones

Sorghum BAC SB40L16 containing the 22 kDa kafirin gene cluster in the vector BeloBAC II and 7.4 kb in size (17), has been sequenced [(12); GenBank accession no. AF527808]. BAC DNA was purified using a Large Construction Kit (Qiagen, Valencia, CA), according to the manufacturer's instructions. The maize expression cassette pUbi-BAR (18) was amplified from a plasmid by the PCR and gel-purified using a Gel Purification Kit (Qiagen).

Plant transformation by particle bombardment

The bacterial phosphinothricin acetylase (*BAR*) gene, under the control of the maize ubiquitin promoter and its first intron and the *A.tumefaciens* nopaline synthase polyadenylation signal, provides tolerance of transgenic plants to the herbicide bialaphos. BAC and selectable marker DNAs were mixed in a 3:1 molar ratio and used to coat gold particles. Coating of gold particles with DNA and particle bombardment of maize callus plates has been described recently (15), except that vortexing was avoided, and mixing was achieved by gentle finger tapping for 1 min periods. Another difference to the previous protocol was the use of pre-embryogenic callus (sponge-like callus), which improved the transformation efficiency by $\sim 50\%$. Maize High Type II embryogenic callus pieces were densely arrayed in the center of Petri dishes containing osmotic medium and bombarded with vacuum-accelerated DNA-gold particles preparations. A total of eight plates were prepared and, after a one-week recovery period, their callus pieces were transferred onto bialaphos selection medium to allow transgenic callus proliferation based on the expression of the co-transformed *BAR* gene. Selection was carried out on medium containing bialaphos at 2 mg l^{-1}

for a period of 12 weeks. Established callus lines were regenerated into plants and grown to maturity in the greenhouse.

Molecular analysis

Two primers designed from the kafirin gene-coding region: 5' primer ATGGCTACCAAGATATTTGTC and 3' primer CT-AGAAGATGGCACTTCCAAC, were used to identify co-transformed BAC DNA by the Polymerase Chain Reaction (PCR). A piece of resistant callus tissue was ground with extraction buffer [100 mM Tris-HCl (pH 8.5), 100 mM NaCl, 20 mM EDTA (pH 8.0) and 1% *N*-lauroyl-sarcosine], using a disposable pestle in a 1.5 ml tube. The ground tissue was phenol extracted, and precipitated by ethanol. DNA pellet was resuspended in 1× TE buffer and an aliquot used for PCR reaction. PCR products were analyzed on agarose gels. Leaf tissue was collected from the regenerated plants and ground with liquid nitrogen. DNA was extracted using phenol/chloroform extraction, and ethanol precipitation. DNA was resuspended in 1× TE, and stored at 4°C. DNA was digested with restriction enzymes, and separated on agarose gels. DNA was blotted onto nylon membrane, and then subjected to Southern-blot hybridization using ³²P-dCTP-labeled probes.

Pulsed field gel analysis

Pulsed field gel electrophoresis was carried out on a CHEF III apparatus (Bio-Rad, Hercules, CA) in 0.5× TBE buffer, with the following setting: 0.1–40 s ramping, 6 V cm⁻¹, 14 h at 14°C. The gel was treated briefly with 0.25 N HCl, and blotted onto a positively charged nylon membrane by 0.4 N NaOH overnight. The membrane was then used for Southern-blot analysis with the respective radioactively labeled probes.

Protein analysis

Zein proteins were extracted by the method of Das *et al.* (19). Protein was quantified using a protein assay (Bio-Rad) according to the manufacturer's protocol. Approximately 5 µg of proteins were subject to 15% SDS-PAGE analysis using a protein mini apparatus (Bio-Rad). The gel was fixed by

TCA and stained by EzBlue (Sigma-Aldrich, St Louis, MO) before being digitally scanned. About 5 µg of proteins were also subjected to iso-electro-focusing (IEF) gel analysis using a LKB 2117 Multiphor II apparatus (LKB Produkter AB, Bromma, Sweden) as described previously (3). The IEF gel, with a pH range of pH 3–10, was run with 15 W constant power for 3 h at 10°C. The proteins on the gel were then fixed by TCA and stained with Ez-Blue (Sigma-Aldrich).

RESULTS

Maize transformation with sorghum BAC clone SB40L16

BAC clone SB40L16 contains a 90 923 bp insert from the genomic DNA of *Sorghum bicolor*, cv Btx623 (12). As shown in Figure 1, the entire kafirin gene cluster is ~35 kb in length with a regular repetition of 10 gene copies and is located in the downstream region of the cloned DNA. The rest of the cloned DNA fragment contains seven unrelated genes, such as a putative *Mla1*-like disease resistance gene, a disease response gene, a gene with homology to a receptor kinase, a fatty acid elongase-2 gene homolog and other hypothetical genes with no similarity in the databases. Embedded into those is a single mutated copy of a 22 kDa kafirin gene ~45 kb upstream of the kafirin gene cluster. Isolation and sequence comparison of cDNA clones derived from kafirin specific RT-PCR products of immature sorghum seeds allowed us to conclude that the upstream mutated kafirin gene is not expressed, but that all members of the 22-kDa kafirin gene cluster in BAC clone SB40L16 are expressed (data not shown).

In order to obtain transgenic maize plants with the intact sorghum kafirin gene cluster, we co-transformed the sorghum BAC DNA with a DNA fragment comprising a herbicide tolerance gene as a selectable marker as described under Materials and Methods. After ~12 weeks of selection, 17 resistant callus lines were produced. Callus DNA was tested for the presence of kafirin-specific sequences by PCR as described

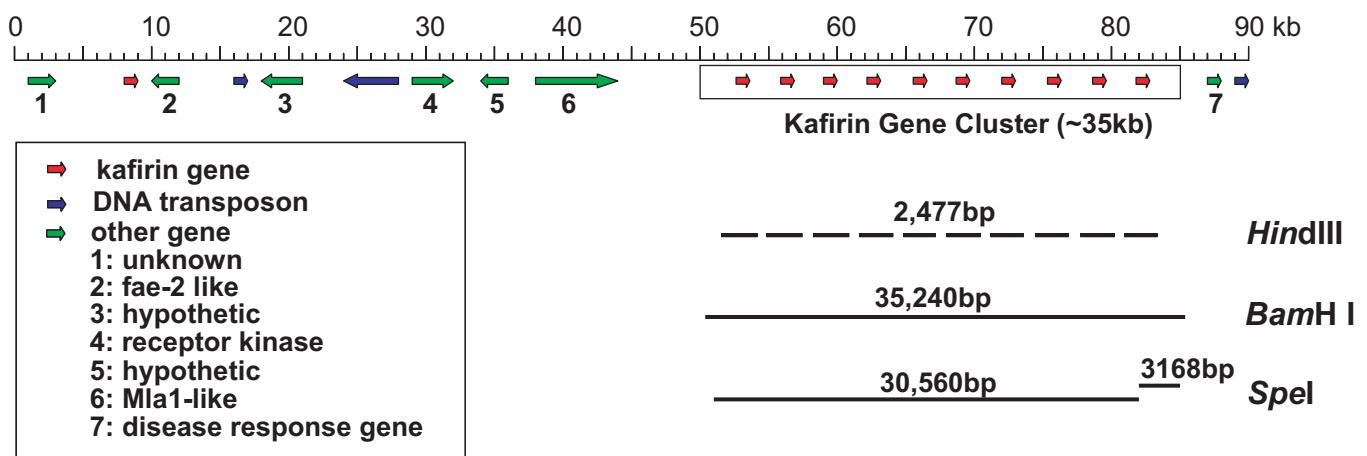


Figure 1. Diagram of sorghum BAC SB40L16. The insert of sorghum BAC 40L16 is indicated along with a ruler with the size in kb on top. The red arrows indicate kafirin genes, green arrows other genes, and blue arrows DNA transposons. A key is inserted to indicate the putative function of genes in this BAC clone. The kafirin gene cluster is highlighted by a box. The bars below the gene cluster represent different size of fragments containing the kafirin genes, those released by different restriction enzymes such as HindIII, BamHI and SpeI, respectively. These three enzymes were used for genomic or pulsed-field gel electrophoresis and Southern-blot analysis.

under Materials and Methods. Thus, we confirmed 12 positive calli with successful co-transformation of both kafirin and BAR genes. The positive calli were regenerated into plants (T1 plants), which were either self-pollinated or out-crossed to non-transformed parental lines to produce transgenic seeds. A total of eight different transformation events led to viable seeds (T2 seeds).

Molecular analysis of transgenic events

Genomic DNA was extracted from leaf tissue of the regenerated T1 plants from the greenhouse. Southern-blot analysis of these DNAs was carried out using HindIII endonuclease, which cuts once into the vector backbone (7.4 kb) and between the kafirin gene copies in the cluster (Figure 1). Two different probes, BAC vector DNA and the 22 kDa kafirin gene-coding region, were used. The result of the analysis corresponding to seven transgenic events is shown in Figure 2. When using the

BAC vector as the probe (Figure 2A), the number of bands observed represent the insertion copy number in the maize genome. Three events (TG4, TG13 and TG14) had low copy numbers of BAC vector sequences; one event (TG10) showed no BAC vector signal and the remaining events (TG7, TG8 and TG9) had medium to high copy numbers of BAC vector (estimated from 10 to >100 copies). A combination of HindIII-restricted DNA with a probe derived from the coding region of a kafirin gene results in the detection of a 2.5 kb DNA fragment (Figure 1), which corresponds to the 10 co-migrating kafirin gene fragments from the cluster and the purified sorghum BAC DNA (Figure 2B). In addition, a non-specific 4.0 kb band also present in non-transformed plant tissue, is present in all lanes. Event TG14 alone appeared to have a banding pattern consistent with a single copy insert of the intact kafirin cluster. According to the pattern and intensity of the bands visualized, other events either contained only part of the kafirin cluster (events TG4, TG10 and TG13) or had medium to high copy number (events TG7, TG8 and TG9) or rearranged copies (TG13) of kafirin genes. Based on the genomic arrangement of the transgenes uncovered by these two probes, we selected event TG14 for further characterization.

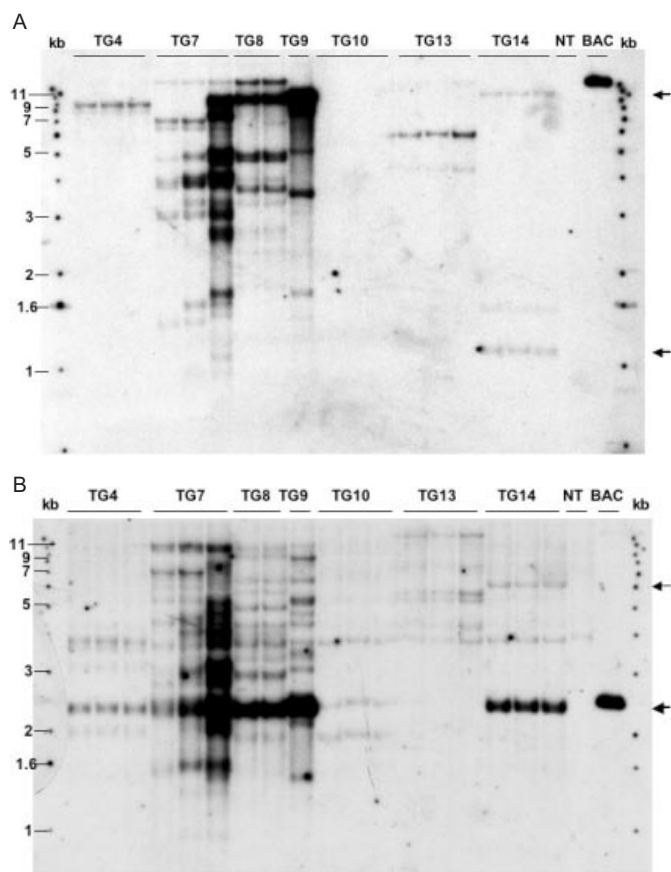


Figure 2. Genomic Southern-blot analysis of transgenic plants. Genomic DNAs extracted from individual regenerated plants of independent events were digested by HindIII, and separated on 1% agarose gel. Different lanes under the same transgenic event (named as TG) represent individual plants of this event. A non-transgenic plant (NT lane) and BAC SB40L16 DNA (BAC lane) was included as negative and positive control, respectively. A 1 kb ladder was used as a size reference on both sides of the gel. Fractionated DNA was transferred onto a nylon membrane, and hybridized with ³²P-labeled probes derived from BAC vector pBeloBAC II (A) or kafirin gene-coding region (B). The arrows in (A) indicate two bands detected from event TG14 by BAC sequences; the arrows in (B) indicate the two kafirin-specific bands detected from event TG14.

Analysis of the size of the BAC DNA inserted into the maize genome

To determine the approximate size of the DNA fragment that got integrated into the genome of event TG14, genomic DNA was digested by BamHI and SpeI. The first enzyme cuts at position 50 485 and 85 724, the second one cuts at positions 51 045, 81 604 and 84 771. The positions for the first copy of the kafirin gene-coding sequence are from 52 584 to 53 390 and the last copy from 81 707 to 82 513. Therefore, BamHI produces a fragment of ~35.2 kb comprising the entire cluster with all 10 gene copies while the last 3' copy of the kafirin gene has a polymorphism that creates a SpeI site. When the kafirin-coding sequence is used as probe, then one should observe one band with BamHI and one large and a small band with SpeI (Figure 1).

Uncut and digested DNA from two individual T2 seeds derived from the transmission of event TG14 from T1 plants, digested DNA from non-transformed seed and purified BAC DNA were separated on an agarose gel by pulsed-field gel electrophoresis to resolve large-sized DNA fragments (Materials and Methods). The fractionated DNA was blotted onto a membrane and hybridized with a kafirin probe. The results are presented in Figure 3. In the SpeI-digested lanes, we detected bands of the correct size in both samples of event TG14, which migrated to similar positions as the bands detected in the positive control BAC SB40L16 digestion and had the expected sizes of 30.5 and 3.2 kb, respectively. No bands were detected in the non-transformed control lanes. If we consider the BamHI-digested lanes, hybridization signals on both transgenic plant DNAs formed a compressed band ~35 kb, causing the deformation ('V' shaped) of the signals due to the differential DNA concentrations in the samples. The center of these appeared to be in a position slightly forward to the band detected in the BAC DNA included as a positive control. But judging at the edges of the genomic DNA lanes, the size was still in agreement with that of the purified BAC DNA sample. Therefore, with

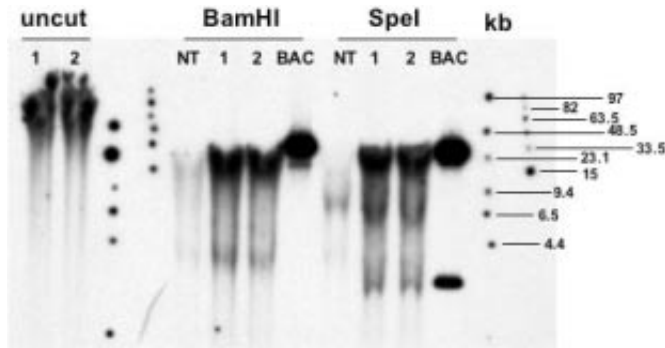


Figure 3. Pulsed-field gel electrophoresis analysis of genomic DNA from event TG14. Genomic DNAs extracted from two regenerated plants (1 and 2) of transgenic event TG14 were either uncut or digested with BamHI or SpeI. DNA of non-transgenic plant (NT lanes) or BAC SB40L16 (BAC lanes) were included as negative and positive controls, respectively. The DNAs were separated on a 1% agarose gel using a CHEF apparatus as described under Materials and Methods. DNA was blotted onto a nylon membrane, and hybridized with a ^{32}P -labeled kafirin probe. The PFGE marker band sizes are indicated on the right.

both the diagnostic restriction enzymes, we were able to detect the expected DNA fragments comprising the kafirin gene cluster in SB40L16. We conclude that TG14 contains the intact kafirin gene cluster and because it also contains sequences including the vector backbone, the integrated DNA fragment was more than 45 kb in length.

Expression and segregation of the kafirin gene cluster in maize

Since one integration event has a single, non-rearranged DNA segment containing the entire kafirin gene cluster integrated into the maize genome, it should have the potential to express this set of genes during the development of transgenic seeds. Because the 22 kDa kafirins and zeins have nearly the same length but differ slightly in their amino acid composition (Table 1), an analysis of the prolamins fraction of seeds derived from the TG14 event, which should contain both zeins and kafirins, would not be distinguishable by SDS-PAGE. However, because of their different amino acid composition, one would expect that zeins and kafirins would be separated by IEF gel electrophoresis. This technique separates proteins based on their pI value instead of size. Indeed, when the prolamins fraction was extracted from single T2 and T3 maize seeds and subjected to SDS-PAGE analysis, we could only detect enhanced 22 kDa bands in TG14 samples relative to non-transformed seed, but not differently migrating bands (data not shown).

Therefore, IEF protein gel electrophoresis was used instead. Prolamins were extracted from 10 T2 seeds derived from a selfed T1 plant (TG14) and 10 T3 seeds that originated from the cross of two T2 siblings (TG14). Parental lines tested positive for the kafirin cluster as assessed by Southern hybridization (data not shown). Prolamins extracted from two sorghum seeds and from three non-transgenic maize seeds were used as controls. The 22 kDa kafirins (SB lanes, Figure 4) have a unique pI value compared to the zeins in non-transgenic maize seeds (NT lanes, Figure 4). Indeed in the test samples, the kafirins migrate accordingly as a prominent band to a position discernible from that of the zeins (indicated by an

Table 1. Comparison of amino acid composition of 22 kDa zeins and kafirins

Amino acids	22 kDa zein	22 kDa kafirin
Ala	40	46
Arg	3	2
Asn	14	16
Asp	0	0
Cys	1	1
Gln	50	56
Glu	1	1
Gly	3	2
His	4	2
Ile	13	12
Leu	48	46
Lys	1	1
Met	5	3
Phe	10	10
Pro	22	21
Ser	19	14
Thr	10	9
Trp	0	1
Tyr	7	8
Val	15	17
Total	266	268

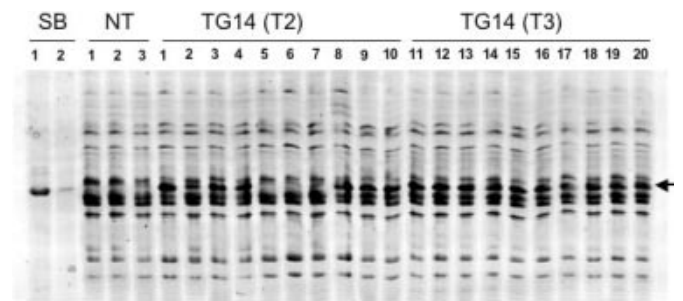


Figure 4. IEF gel analysis of alcohol-soluble proteins from event TG14. Seed proteins containing the zein/kafirin prolamins fraction were extracted with 70% ethanol/1% 2-mercaptoethanol. About 5 μg of 10 T2 (lanes 1–10) and 10 T3 (lanes 11–20) individual seed protein samples of transgenic event TG14 were loaded onto the gel and subjected to isoelectric focusing as described under Materials and Methods. Two sorghum seed samples (SB lanes) and three non-transgenic seed samples (NT lanes) were included as positive or negative controls for kafirins. An arrow on the right-hand side indicates the position of the 22 kDa kafirin band.

arrow in Figure 4). Therefore, the kafirin cluster integrated as event TG14 is expressed after the first and second seed transmission in maize seeds.

However, in the lanes corresponding to T2 seeds, a band co-migrating with the kafirin band of the sorghum sample was only observed in 7 out of 10 times. This is not unexpected because T1 plants are heterozygous for the transgene, which when selfed should produce a quarter of its seeds without the transgene. Although this analysis is just a small sampling of T2 progeny, the genetic segregation is consistent with the Southern-blot analysis showing that the TG14 event is a single locus. Moreover, T2 seeds homozygous for the absence of the transgene, exhibit the same profile as non-transformed plants, confirming the unique pI value of the kafirin band shown in the sorghum lanes of the IEF gel. Segregation of the transgene in T2 seeds produces either heterozygous or homozygous progeny for the transgenic locus. Therefore, it appears that the T3 seeds were derived from two T2 parents (sib-crossed) that were both homozygous for the transgenic locus, as all lanes

exhibit nearly uniform, high level accumulation of the kafirin band in their prolamin fraction. The kafirin relative to the zein levels in the T3 seeds even appeared to be consistently higher than in the positive T2 seed samples. However, this slight variation could also be due to samples that are still heterozygous compared to the homozygous T3 seeds. The accumulation of transgene-derived kafirin protein did not seem to interfere with zein accumulation, as judged by the identical concentrations of IEF zein bands, which served as loading controls. Because kafirins like zeins have a 21 amino acid signal peptide, which in an unprocessed form would significantly differ in its mobility and charge (20), it also appeared that the kafirins were properly processed in maize endosperm. Therefore, the kafirin genes in maize were genetically stable and their expression strength was not changed through two rounds of meiosis.

DISCUSSION

Inserting large intact cloned DNA fragments into plant genomes

While previous efforts had shown that it was possible to insert large cloned fragments into plant genomes, we have shown that 10 linked functional genes can be introduced at the same time in their original chromosomal context. There were five key steps for the insertion of large DNA fragments including those containing tandem gene families. The first step was the DNA preparation itself. To circumvent conjugation of cloned-DNA with *Agrobacterium*, DNA was isolated from *E.coli* directly. The coating of gold particles with BAC-sized DNA had to occur with very careful mixing of the sample to avoid DNA shearing, which presented the major limit in transforming large sized DNA fragments. The second step was direct transformation of plant cells from pre-embryogenic callus (sponge like callus) with the DNA by particle bombardment. To enable selection of transformed cells a DNA mixture of the BAC DNA and DNA containing the expressible marker gene had to be used. We had tested these co-transformation methods previously to optimize the proper molar ratio between target DNA and selectable marker DNA (15). The third step was the screening of individual transgenic events by Southern-blot analysis. The direct DNA uptake transformation method was known to produce complex integration events, which appeared to be the result of illegitimate recombination of different sized DNA fragments (21). In addition, complex integration events could also be unstable during meiosis. Although in most events that have been investigated in this study, the integrated DNA copy and pattern was identical among different individuals, in one event (TG7, Figure 2B) we noticed significant variation in copy number and also in the DNA restriction pattern. The simplest way to circumvent these types of complex structures and DNA rearrangements during meiosis was to reduce the frequency of DNA fragmentation through pure DNA samples, careful coating of particles and the screening of multiple events. Only one out of eight events met our criteria; still, the frequency that such simple integration events occurred was manageable from an experimental point of view, particularly since transformation efficiency was now >10% (15). Fourth, selected events had to be characterized with restriction digests producing one or two large DNA

fragments that were resolved by pulsed-field gel electrophoresis so that the integrity and size of DNA fragments could be directly compared with the donor DNA. The fifth step was the genetic test. Independent of the restriction enzyme analysis, integrity and stability of the DNA could be confirmed by the programmed expression of the genes that were contained in the donor DNA and the segregation of the presence and absence of the products of the transgenes in a Mendelian fashion.

Potential role of chromatin configuration in large DNA segments

Nevertheless, a particular challenge in this example was that the 10 linked genes were highly homologous. Here, the first barrier was to propagate such repetitive DNA in bacteria. Although *E.coli* strains were known to tolerate such DNA structures, such an experience was still lacking for *Agrobacterium* strains. The second barrier was repeat-induced gene silencing in plants (22). Because the 3.5 kb kafirin repeat was highly conserved and its genes almost produced identical proteins (97–100%), these sequence repeats could have induced the above mentioned gene silencing effect. In this respect, it might have been critical to add several kilo base pairs of contiguous sequences that naturally flanked the kafirin cluster in plants. The high-level expression in transgenic T3 seeds also indicated that *cis*-acting sequences like enhancers or insulators, recognized by transacting and chromatin-remodeling factors, were contained within the inserted BAC fragment. Therefore, the insertion of large DNA fragments could also be critical in providing important DNA sequences for mimicking the chromatin structure that ensured a proper configuration of transgenes. From these results it appeared that these factors were sufficiently conserved between sorghum and maize, so that it was feasible to exchange valuable multigenic traits within the grass family in a way that was usually only achieved by breeding. Moreover, the accumulation of kafirins, without repression of zeins, holds the promise that protein levels in seeds might be raised without loss of yield due to biotechnological applications (23).

The role of whole-genome-duplication in the regulation of gene expression

While it might not be surprising that tissue-specific gene expression was preserved between sorghum and maize, we recently found evidence that its *trans*-acting factors indeed had diverged sufficiently to adapt to different specificities. Previously, it was assumed that the expression of 22 kDa zeins in maize required the transcriptional activator Opaque-2 (O2) (24). However, when the expression of individual members of the 22 kDa zein gene subfamily were examined in the absence of O2 *in vivo*, it was found that some members, *zp22/6* and *zp22/D87*, used a different transcriptional activator than O2 (3). Although this regulatory gene had not been cloned, yet because of the large size of the family of *b-zip* transcription factors in all plants, a possible explanation could come from understanding the origin of the maize genome. We have recently shown that maize resulted from the hybridization of two progenitors that separated from the progenitor of sorghum almost instantaneously, ~11.9 mya (14). Although after hybridization maize might have lost up to half of its duplicated genes, it also retained many of them, with

Table 2. Sequence motifs in promoter regions of prolamin genes^a

Gene	O2-box 1 (Z1)	P-box	O2-box 2 (Z2)	O2-box 3 (Z3)	Size	In frame
<i>azs22;4</i>	TCACATGTGT (328)	GTGTAAAG (324)	CATGCATGTCAT (302)	TCCACGTAGA (292)	267	0
<i>azs22;8</i>	TCACATGTGT (325)	GTGTAAAG (321)	CATGCATGTCAT (299)	TCCACGTAGA (289)	267	0
<i>azs22;10</i>	TCACATGTGT (328)	GTGTAAAG (324)	Ct TGCATGTCAT (302)	TCCACGTAGA (292)	267	0
<i>zp22/6</i>	TCACATGTGT (328)	GTGTAAAG (324)	Ct TGCATGTCAT (302)	TCCACGTAGA (292)	267	0
<i>zp22/D87</i>	TCc CATGTGT (326)	GTGTAAAG (322)	CATGCATGTCAT (300)	TCCACGTAGA (290)	238	0
<i>azs22;12</i>	TCg CATGTGT (325)	GTGTAAAG (321)	g ATGCATGTCAT (299)	TCCAa GTAt A (290)	267	0
<i>azs22;16</i>	TCACATGTGT (326)	GTGTAAAG (322)	CATGCATGTCAT (300)	TCCACa TAA A (290)	263	0
<i>22-kafirin</i>	Ta ACATGTGT (325)	GTGTAAAG (321)	CATG t t a t TTc T (295)	TCCAa a TAGA (364)	263	0
<i>22-coixin</i>	aCg Ct a t TGT (327)	t TGTAAG (323)	CATGCATGTCAT (301)	TCCA a TAGA (272)	267	0

^aSequence motifs are as described by Muth *et al.* (31). Genes in boldface are expressed; letters in boldface indicate changes from the consensus sequences.

some of them changing not their function but their specificity like the *r1* and the *b1* genes (25). The same redundancy might also exist for the *O2* gene in maize, while sorghum should have only a single copy of the *O2* gene. Indeed, an *O2* homolog controlling the expression of 22 kDa kafirins had been characterized previously (26). This protein exhibited a high level of amino acid similarity to its counterpart in maize; particularly at the DNA-binding domain, which was 90% identical and classified these transcription factors as the basic leucine zipper family.

The *cis*-acting sequences in the regulation of 22 kDa prolamin genes

Since the 22 kDa zein genes were regulated by two different transcriptional regulators, one might wonder whether this was due to the evolution of the promoter target sites. The prolamin box (27,28) and *O2* boxes 1 and 2 (24,29–31) were found ~300 bp upstream of the coding regions of 22 kDa zein genes in maize, 22-kDa kafirin genes in sorghum (3,13) and 22-kDa coixin genes in *Coix lachryma-jobi* (30). *Coix*, like sorghum, is a very close relative of maize, where amplification of the 22 kDa prolamin gene had also occurred after its progenitor separated from the two progenitors of maize and the progenitor of sorghum (32). Although previous studies had shown that *O2* bound specifically to the TCCACGTAGA sequence and sequences containing this motif were capable of conferring transcriptional activation in hybrid promoters tested in transient expression systems (29), we could observe that *O2* could work without this sequence motif *in vivo* (Table 2). While both *zp22/6* and *zp22/D87* had this *O2* box and were expressed in the absence of *O2*, *azs22;16* had two mutations in the *O2* box, but required *O2* for its expression. The 22 kDa prolamin promoters from sorghum and *Coix* also had mutations altering the TCCACGTAGA sequence motif. Although these results did not contradict the findings that the TCCACGTAGA sequence could be used as a binding site for transcriptional activation by *O2*, they clearly did not reflect the physiological conditions under which transcriptional activation occurred. Subsequent observations suggested that secondary binding sites could compensate for the lack of the conserved TCCACGTAGA motif (31). Substituted *O2* boxes were bound cooperatively and concerted changes might restore a general high-binding affinity and activity (33). For instance, the *azs22;12* gene in maize inbred BSSS73, was the only gene with an intact coding region, but with mutations in the canonical and auxiliary *O2* boxes

at the same time (Table 2). Therefore, one could argue that this gene prevented *O2* function because of cumulative defective binding sites. However, the 22 kDa kafirin promoter had also mutations in the same sequence motifs as the *azs22;12* gene, but apparently was functional in the TG14 transgenic maize line. It also had been shown that the maize *O2* protein activated transcription of a 22 kDa kafirin promoter (26), the 22 kDa coixin upstream sequences (30), a wheat glutenin promoter (34) and a rice glutelin promoter (35). However, if all *cis*-acting sequences were compared, the only motif that was conserved in sequence and position among maize, sorghum and *Coix* was TGTAAG, which was also called the endosperm or prolamin box (Table 2). Interestingly, a DNA-binding protein from maize had been described as that which is bound to the prolamin box and interacted with the purified *O2* protein (36). Furthermore, it had been shown that the prolamin-box binding protein specifically recognized the TGTAAG motif (37) and acted as an inhibitor of *O2* binding to the *O2* box, indicating a hierarchical system of the multiple interactions among trans-acting factors (38). Therefore, it will be interesting to see whether the transgenic-kafirins cluster is expressed when the TG14 is introgressed into an *O2* background.

In conclusion, in this study, we successfully transformed a 22 kDa kafirin gene cluster, containing the entire 22 kDa kafirin gene family, into maize by the particle bombardment method. The transformed kafirin genes were highly expressed in maize endosperm even without a sorghum regulator. This appeared to be the first report of direct incorporation of an entire heterologous gene cluster (gene family) into plants. This approach could provide a great opportunity for crop improvement and functional genomics of gene families.

ACKNOWLEDGEMENTS

We wish to thank Kathy Ward for technical assistance and Mike Peterczak for greenhouse and field management. This work has been supported by the Department of Energy (DOE) grant no. DE-FG05-95ER20194 to J.M.

REFERENCES

- Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A. and Felsenfeld, G. (2002) The insulation of genes from external enhancers and silencing chromatin. *Proc. Natl Acad. Sci. USA*, **99**, 16433–16437.

2. Leister, D. (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends Genet.*, **20**, 116–122.
3. Song, R., Llaca, V., Linton, E. and Messing, J. (2001) Sequence, regulation, and evolution of the maize 22-kDa α zein gene family. *Genome Res.*, **11**, 1817–1825.
4. Hamilton, C.M., Frary, A., Lewis, C. and Tanksley, S.D. (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc. Natl Acad. Sci. USA*, **93**, 9975–9979.
5. Liu, Y.G., Shirano, Y., Fukaki, H., Yanai, Y., Tasaka, M., Tabata, S. and Shibata, D. (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. *Proc. Natl Acad. Sci. USA*, **96**, 6535–6540.
6. Song, J., Bradeen, J.M., Naess, S.K., Helgeson, J.P. and Jiang, J. (2003) BIBAC and TAC clones containing potato genomic DNA fragments larger than 100 kb are not stable in *Agrobacterium*. *Theor. Appl. Genet.*, **107**, 958–964.
7. Choi, S., Begum, D., Koshinsky, H., Ow, D.W. and Wing, R.A. (2000) A new approach for the identification and cloning of genes: the pBACwch system using Cre/lox site-specific recombination. *Nucleic Acids Res.*, **28**, e19.
8. Ercolano, M.R., Ballvora, A., Paal, J., Steinbiss, H.-H., Salamini, F. and Gebhardt, C. (2004) Functional complementation analysis in potato via biolistic transformation with BAC large DNA fragments. *Mol. Breed.*, **13**, 15–22.
9. Ueda, T. and Messing, J. (1993) Manipulation of amino acid balance in maize seeds. *Genet. Eng.*, **15**, 109–130.
10. Song, R. and Messing, J. (2002) Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. *Plant Physiol.*, **130**, 1626–1635.
11. Song, R. and Messing, J. (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. *Proc. Natl Acad. Sci. USA*, **100**, 9055–9060.
12. Song, R., Llaca, V. and Messing, J. (2002) Mosaic organization of orthologous sequences in grass genomes. *Genome Res.*, **12**, 1549–1555.
13. DeRose, R.T., Ma, D.P., Kwon, I.S., Hasnain, S.E., Klassy, R.C. and Hall, T.C. (1989) Characterization of the kafirin gene family from sorghum reveals extensive homology with zein from maize. *Plant Mol. Biol.*, **12**, 245–256.
14. Swigonova, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J.L. and Messing, J. (2004) Close split of maize and sorghum genome progenitors. *Genome Res.*, **14**, 1916–1923.
15. Segal, G., Song, R. and Messing, J. (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. *Genetics*, **165**, 387–397.
16. Frame, B.R., Zhang, H., Cociolone, S.M., Sidorenko, L.V., Dietrich, C.R., Pegg, S.E., Zhen, S., Schnable, P.S. and Wang, K. (2000) Production of transgenic maize from bombarded Type II callus: effect of gold particle size and callus morphology on transformation efficiency. *In Vitro Cell Dev. Biol. Plant*, **36**, 21–29.
17. Woo, S.S., Jiang, J., Gill, B.S., Paterson, A.H. and Wing, R.A. (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acids Res.*, **22**, 4922–4931.
18. Christensen, A.H. and Quail, P.H. (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Res.*, **5**, 213–218.
19. Das, O.P., Cruz-Alvarez, M., Chaudhuri, S. and Messing, J. (1990) Molecular methods for genetic analysis of maize. *Methods Mol. Cell Biol.*, **1**, 213–222.
20. Coleman, C.E., Lopes, M.A., Gillikin, J.W., Boston, R.S. and Larkins, B.A. (1995) A defective signal peptide in the maize high-lysine mutant floury 2. *Proc. Natl Acad. Sci. USA*, **92**, 6828–6831.
21. Makarevitch, I., Svitashv, S.K. and Somers, D.A. (2003) Complete sequence analysis of transgene loci from plants transformed via microprojectile bombardment. *Plant Mol. Biol.*, **52**, 421–432.
22. Ye, F. and Signer, E.R. (1996) RIGS (repeat-induced gene silencing) in *Arabidopsis* is transcriptional and alters chromatin configuration. *Proc. Natl Acad. Sci. USA*, **93**, 10881–10886.
23. Stoger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard, J., Williams, S., Keen, D., Perrin, Y., Christou, P. and Fischer, R. (2000) Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. *Plant Mol. Biol.*, **42**, 583–590.
24. Schmidt, R.J., Ketudat, M., Aukerman, M.J. and Hoschek, G. (1992) Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. *Plant Cell*, **4**, 689–700.
25. Lai, J., Ma, J., Swigonova, Z., Ramakrishna, W., Linton, E., Llaca, V., Tanyolac, B., Park, Y.-J., Jeong, O.-Y., Bennetzen, J.L. and Messing, J. (2004) Gene loss and movement in the maize genome. *Genome Res.*, **14**, 1924–1931.
26. Pirovano, L., Lanzini, S., Hartings, H., Lazzaroni, N., Rossi, V., Joshi, R., Thompson, R.D., Salamini, F. and Motto, M. (1994) Structural and functional analysis of an Opaque-2-related gene from sorghum. *Plant Mol. Biol.*, **24**, 515–523.
27. Ueda, T., Wang, Z., Pham, N. and Messing, J. (1994) Identification of a transcriptional activator-binding element in the 27-kDa zein promoter, the –300 element. *Mol. Cell. Biol.*, **14**, 4350–4359.
28. Marzábal, P., Busk, P.K., Ludevid, M.D. and Torrent, M. (1998) The bifactorial endosperm box of γ -zein gene: characterisation and function of the Pb3 and GZM cis-acting elements. *Plant J.*, **16**, 41–52.
29. Ueda, T., Wawerczak, W., Ward, K., Sher, N., Ketudat, M., Schmidt, R.J. and Messing, J. (1992) Mutations of the 22- and 27-kD zein promoters affect transactivation by the Opaque-2 protein. *Plant Cell*, **4**, 701–709.
30. Yunes, J.A., Cord Neto, G., da Silva, M.J., Leite, A., Ottoboni, L.M.M. and Arruda, P. (1994) The transcriptional activator Opaque-2 recognizes two different target sequences in the 22 kD-like α -prolamin genes. *Plant Cell*, **6**, 237–249.
31. Muth, J.R., Müller, M., Lohmer, S., Salamini, F. and Thompson, R.D. (1996) The role of multiple binding sites in the activation of zein gene expression by Opaque-2. *Mol. Gen. Genet.*, **252**, 723–732.
32. Ottoboni, L.M.M., Leite, A., Yunes, J.A., Targon, M.L.P.N., de Souza Filho, G.A. and Arruda, P. (1993) Sequence analysis of 22 kDa-like α -coixin genes and their comparison with homologous zein and kafirin genes reveals highly conserved protein structure and regulatory elements. *Plant Mol. Biol.*, **21**, 765–778.
33. Yunes, J.A., Vettore, A.L., da Silva, M.J., Leite, A. and Arruda, P. (1998) Cooperative DNA binding and sequence discrimination by the Opaque2 bZIP factor. *Plant Cell*, **10**, 1941–1955.
34. Holdsworth, M.J., Munoz-Blanco, J., Hammond-Kosack, M., Colot, V., Schuch, W. and Bevan, M.W. (1995) The maize transcription factor Opaque-2 activates a wheat glutenin promoter in plant and yeast cells. *Plant Mol. Biol.*, **29**, 711–720.
35. Wu, C.-Y., Suzuki, A., Washida, H. and Takaiwa, F. (1998) The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by Opaque-2 in transgenic rice plants. *Plant J.*, **14**, 673–683.
36. Vicente-Carbajosa, J., Moose, S.P., Parsons, R.L. and Schmidt, R.J. (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.
37. Wang, Z., Ueda, T. and Messing, J. (1998) Characterization of the maize prolamin box-binding factor-1 (PBF-1) and its role in developmental regulation of the zein multigene family. *Gene*, **223**, 321–332.
38. Wang, Z. and Messing, J. (1998) Modulation of gene expression by DNA-protein and protein-protein interactions in the promoter region of the zein multigene family. *Gene*, **223**, 333–345.