Molecular Characterization of a Novel Methionine-Rich δ-Kafirin Seed Storage Protein Gene in Sorghum (*Sorghum bicolor* L.)

L. Izquierdo[1,2] and I. D. Godwin[1,3]

ABSTRACT

We have isolated, cloned, and characterized a 660 bp full length cDNA encoding a putative seed storage protein gene, δ-kafirin (AY834250), from developing seeds of sorghum (*Sorghum bicolor*) inbred line SPV-475. Translation of the DNA sequence predicts a 16-kDa polypeptide (precursor) of 147 amino acids, rich in methionine residues (17%). Reverse Transcriptase-PCR (RT-PCR) and Real Time-PCR revealed δ-kafirin is only expressed in developing seeds. δ-Kafirin exhibited 96% identity with another methionine-rich sorghum seed storage protein (AY043223). Interestingly, this δ-kafirin gene contained two insertions rich in ATG codons encoding five more methionine residues than AY043223. Comparison of the sorghum δ-kafirin proteins with other seed storage proteins from maize, rice, and Brazil nut revealed conserved domains, mainly at the N-terminus. This similarity, in particular to the zeins, suggests these proteins shared a common ancestor and that the variation observed occurred after the separation of the species.

Cereal Chem. 82(6):706–710

Cereals are the most important crops in the world. In addition to their primary importance as starch staples, they provide protein for the nutrition of humans and mammalian and avian livestock. The livestock are unable to synthesize 10 of the 20 amino acids essential for protein synthesis. Cereal seed storage proteins also play an important role in food processing technology and baking quality (Shewry and Halford 2002).

Cereal seed storage proteins are synthesized at high levels in the endosperm, with temporal and spatial control of synthesis during endosperm development. Their main biological function is to supply nitrogen and amino acid reserves for the germinating seedling (Shewry et al 1995). At the time of formation, storage proteins are protected against uncontrolled premature degradation by several mechanisms. The major mechanism is through the sequestration from the cytoplasm into specialized membrane-bound organelles called protein bodies (PB) (Taylor et al 1985; Coleman et al 1997; Oria et al 2000). Seed storage proteins are expressed as precursors that undergo posttranslational modification before deposition in the PB (Higgins 1984; DeRose et al 1989).

Historically, storage proteins have been classified into four groups (albumins, globulins, prolamins, and glutelins) based on their extraction and solubility in a series of solvents such as water, dilute saline, alcohol-water mixtures, and dilute alkali or acid (Müntz 1998; Shewry and Halford 2002).

In maize, sorghum, and coix, the related grass of Asian origin (*Coix lachryma-jobi*), the most abundant seed storage proteins are the prolamins, which account for ≥50% of the total protein. They are soluble in water-alcohol mixtures and present a high proportion of proline and glutamine (Shewry et al 1995). They have been reclassified into four groups (α, β, γ, and δ) based on solubility properties, electrophoretic mobility, and nucleotide sequences (Shull et al 1991). The major sorghum prolamins are called kafirins. Kafirins share many properties with zeins, the extensively studied maize prolamins. It is believed that the prolamins in the Triticeae (wheat, barley, rye) and the Panicoideae (maize, sorghum, millet) have a different evolutionary origin (Shewry and Halford 2002).

Sulfur-rich seed storage proteins have been reported throughout the angiosperms, including dicot and monocotyledonous plants (Swarup et al 1995; Fischer et al 1996; Lai and Messing 2002). Evolutionary analysis of some of these proteins suggests they are the result of gene duplications that took place before the split into monocot and dicotyledonous plants (Fischer et al 1996). The discovery of these proteins has opened up possibilities of modifying crops to provide a higher sulfur content in human and animal diets. Nutritional improvement by genetic transformation has been performed in potato (Chakraborty et al 2000), lupin (Molvig et al 1997), and alfalfa (Tabe et al 1995). Another approach is to use genes that encode for methionine-rich (MR) proteins to transform plants that could act as bioreactors where the amino acid can be "harvested" and used in food technology or animal feeding (Jiang and Sun 2002).

Sorghum (*Sorghum bicolor* L.) is the fifth most important cereal crop in the world. It is widely grown throughout the semiarid tropics and subtropics for human and animal nutrition, with over 300 million people depending on it as a staple food. Sorghum seed storage proteins are not well characterized. Here we report the cloning and characterization of a sorghum cDNA that encodes a seed-specific δ-kafirin protein, rich in methionine residues. This is the first published report of the characterization of a δ-kafirin. We analyzed the structural relationship of this kafirin (AY834250) with similar proteins, in particular the zeins, and will discuss its potential use in improvement of animal and human diets.

MATERIALS AND METHODS

Nucleic Acids

The Indian inbred line SPV475 was used for all experiments. This line, developed by the National Research Centre for Sorghum, in Hyderabad, India, is widely used as a cultivar and parent line in the development of food-quality sorghum cultivars for the postrainy season in India. Total RNA from roots, stems, leaves, and three stages (early, mid-, and late-grain fill) of developing grains was prepared using the Total RNA isolation system kit (Promega, Madison, WI) and the protocol of Apel and Kloppstech (1978). For Northern analysis, 10 μg of total RNA of each sample was size-fractionated on 1% denaturing formaldehyde gel and blotted onto Hybond N+ membranes (Roche Diagnostics, Sydney, Australia) following the procedure of Sambrook et al (2001). Strand cDNA synthesis was performed using an oligodT primer and SuperScript-II Reverse Transcriptase system (Invitrogen, Carlsbad, CA) and samples were treated with DNAse to eliminate any residual DNA.

RNA from developing seeds was used as a template to amplify cDNA ends with a 5'/3' RACE kit (Roche Diagnostics, Sydney, Australia). We designed primers (sequences available on request) by first flanking conserved regions of the kafirin gene sequences

[1] School of Land and Food Sciences, The University of Queensland, St Lucia, QLD 4072, Australia.
[2] Present address: Centre for Plant Conservation Genetics, Southern Cross University, Lismore NSW 2480 Australia.
[3] Corresponding author. Phone: +61-7-3365-2141. Fax: +61-7-3365-1177. E-mail: i.godwin@uq.edu.au

DOI: 10.1094/CC-82-0706

```
1                                                    TGCCTACGACTCACTATAGGGCGAATTGGGCCCTCTAGATG      41

42   CATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTACACCT CATCTACCACCACCGCC      113
                        CACAATCGCATATCTACT-----ATTCTC----TTGGAAGCGAGGAA---ACCACCGCC

114  ATGGCAGCCAAGATGTTTGCATTGTTTGCACTCCTAGCTCTTTGTGCAAGCGCCACAAGTGCGACCCATATT      185
     ATGGCAGCCAAGATGTTTGCATTGTTTGCACTCCTAGCTCTTTGTGCAAGCGCCACAAGTGCGACCCATATT

186  CCAGGGCACTTGCCACTAGTGATGCCATTGGGTACCATGAACCCATGCACGCAGTACTGCATGATGCAACAG      257
     CCAGGGCACTTGCCACTAGTGATGCCATTGGGTACCATGAACCCATGCACGCAGTACTGCATGATGCAACAA

258  AGGTTTGCCAGATTGTTAGCGTGGCCAATCCCGATGCTACAGCAACTGTTGGCCTTACCGCTTCAGCCGGCG      329
     AGGTTTGCCAGATTATTGGCGTGGCCAATCCCGATGCTACAGCAATTGT----------CGCTTCAGCCGGCG

330  TATCAGACGCCAATGATGATGCCGAACATGATGCCACCAATGATGATGAGGCCGACGATGATGCCACCAATG      401
     TATCAGACGCCAATGACGATGCCGAACATGATGCCACCAATGA----------

402  ACG ATGATGCCGAGCGTGATGTCATTGCCACAATGTCACTGTGATGCCATCTCCCAGATTATGCAGCAACAG      473
     --CG ATGATGCCCAGCATGATGTCGTTGCCACAATGTCACTGTGATGCCATCTCCCAGATTATGCAGCAACAG

474  CAGTTACCACTCATGTTCAACCCAA CAGCCATGGCGA TCCCACCCATGTTCTTACAGCAAC CCTTCGTTAGT      545
     CAGTTACCATTCATGTTCAACCCAA CAGCCATGGCGA TCCCACCCATGTTCTTACAGCAAC CCTTTGTTAGT

546  TCTGCATTCTAGATAGAAATATTTGTGTTGTGTACTGAATAATAAAGTTGACATGCCATCGCATGTGACTCA      617
     TCTGCATTCTAGATAGAAATATTTTTGTTGT ----ACTGAATAATAAAGTTGACATGCCATCGC--- GTGACTCA

618  TTATTAGGAATAAAACAAGCCAATAAATTGCTTTTTCCTTATT
     TTATTA------ATCTAG----------
```

Fig. 1. Nucleotide sequence of the two δ-kafirins reported from sorghum: AY834250 (this study) and AY043223 (GenBank). Shaded letters represent mismatched nucleotides between the two sequences. For AY834250 sequence, the start codon (bold letters), the stop codon (boxed), and the putative polyadenylation signals (underlined letters) are indicated.

TABLE I
Sequences Used in Comparative Analysis[a]

Species	Accession Number (GenBank)	Type	% Methionine Residues (mature protein)
Zea mays (maize)	T03381	zein	
	AY047317	zein	
	U31541	18 kDa δ-zein	27.0
	AF371265	zein	
	M23537	10 kDa	22.5
	AY104139	10 kDa δ-zein	23.0
	U25674.1	zein	
	M72700	zein	
	AF461049	zein	
Oryza sativa (rice)	S08219	prolamin	20.0
Bertholletia excelsa (Brazil nut)	P04403	2S-albumin	19.0
Sorghum bicolor (sorghum)	AY043223	δ-kafirin	20.0

[a] Values taken from Swarup et al (1995).

from the GenBank to amplify 3' ends. A PCR fragment at ≈700 bp was consistently amplified, ligated into TOPO-TA vector (Invitrogen), and transformed into *E. coli* cells. Cells were plated onto media containing IPTG (isopropyl-β-D-thiogalactopyranoside [Progen, Brisbane, Australia]) and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside [Progen]) to allow for blue-white selection or recombinants. White colonies were isolated and sequenced with M13 universal primers and GenBank was BLAST-searched for nucleotide and amino acid similarities. Multiple sequence alignment and phylogenetic relationships with other plant sequences were performed using ClustalW, Seqboot, Protdist, Neighbor, and Consensus programs from BioManager server (http://biomanager.angis.org.au) The nucleotide and protein sequences used for comparative analysis were taken from the GenBank library (Table I).

Quantitative RT-PCR

cDNA samples (≈100 ng) from different tissues were used in a quantitative PCR analysis on a 7700 Sequence detector (Applied Biosystems, Melbourne, Australia). A 25-μL sample was prepared containing 2× SYBER green PCR master mix (Applied Biosystems), 100 n*M* of forward (Met-Dig5': 5'-AGACGCCAATGATG ATGCC-3') and reverse (Met-Dig3': 5'-CAGCAACCCTTCGTT AGTTC-3') primers. The amplification program included one denaturing step at 95°C for 10 min, followed by an amplification step of 45 cycles for 10 sec at 95°C, 1 min at 60°C, and 1 min at 72°C. The PCR product was detected using the fluorescent dye Sybergene (Applied Biosystems). To test the efficiency of the primers, a standard curve was generated using a cDNA sample dilution series. For standardization we included samples for amplification of the 26S ribosomal RNA gene. The primers used in this case were 5'-TTAACAGCCTGCCCACCCTGG-3' and 5'-ATCC ATTTTGCCGACTTCCC-3' as forward and reverse, respectively. All samples were analyzed in duplicate and a negative, template-free control for each set of primers was added. The expression levels of the δ-kafirin gene were expressed relative to the 26S ribosomal RNA gene. An additional RT-PCR experiment using the same cDNA samples and Met primers in a 50-μL reaction mixture was performed following the Elongase system protocol (Invitrogen, Carlsbad, CA) with a lower 58°C annealing temperature.

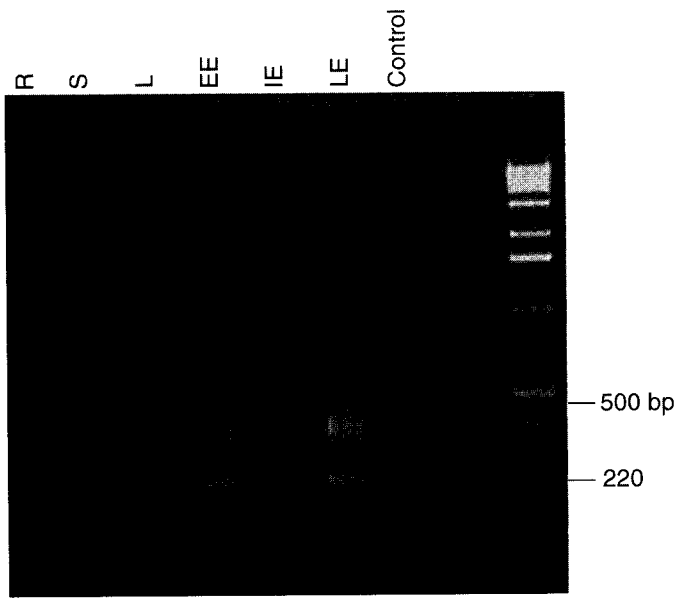

Fig. 2. RT-PCR analysis for the δ-kafirin gene in roots (R), stem (S), leaves (L), and three stages of developing seeds: ≈10–12 (early endosperm, EE); ≈18–20 (intermediate endosperm, IE); and ≈25 (late endosperm, LE) days after flowering (DAF).

Hybridization Experiments

A nonradioactive Northern blot was performed to obtain cDNA clone for a template in a PCR reaction to generate a Digoxigenin labelled fragment of 209 bp by incorporation of DIG-dUTP (PCR-DIG probe synthesis kit, Roche Diagnostics, Sydney, Australia). The primers used for the amplification were Met-Dig5' (forward) and Met-Dig3' (reverse). This PCR product was used as a probe for RNA detection. Hybridization was conducted overnight at 42°C following the DIG-system protocol (Roche Diagnostics).

RESULTS AND DISCUSSION

cDNA and Amino Acid Sequence

Using the 3'RACE technique, we isolated a cDNA fragment from developing sorghum seeds. The complete length of this cDNA clone was 660 bp (Fig. 1) and contained one polypeptide-coding region of 147 amino acids with a molecular mass of 16.4 kDa. The polypeptide begins at the second ATG codon at position 114 and ends at position 554. The second ATG was considered the start of the polypeptide because it was very closely located to an identical motif (CCGCCATGGC) to the consensus sequence reported for zeins (Heidecker et al 1991). The sequence had two noncoding regions of 44 and 103 bp in the 5' and 3' ends, respectively. Three polyadenylation signal sequences (AATAAAT) were determined at 16, 29, and 69 nt upstream of the poly (A) tail. Moreover, a potential hairpin loop of 8 bp was found between the first two adenylation signals; this was similar to the one reported for a 16 kDa rice prolamin (Shyur et al 1992). The clone is rich in hydrophobic amino acids including methionine (17%), proline (10.2%), leucine (12.2 %), and alanine (10.9%). The methionine residues, although clustered throughout the polypeptide, were mainly concentrated in the center of the molecule.

δ-Kafirin Expression

Northern blots were unable to detect δ-kafirin transcript from any tissue. Because of the high homology with the GenBank sorghum sequence AY043223, we had a very narrow selection for a sequence to use as a probe or for designing primers that were specific to our transcript. The expresion of AY043223 sequence was detected by Northern (data not shown). This might indicate that our δ-kafirin is much lower expressed than AY043223. RT-PCR analysis detected expression in early (≈12 DAF) and late stages (≈20–25 DAF) of endosperm development (Fig. 2). The 215 nt RT-PCR products were sequenced to confirm their identity. However, RT-PCR experiment also revealed expression (although low) of δ-kafirin transcript in intermediate stages (IE) of development. Both techniques confirmed endosperm-specific expression

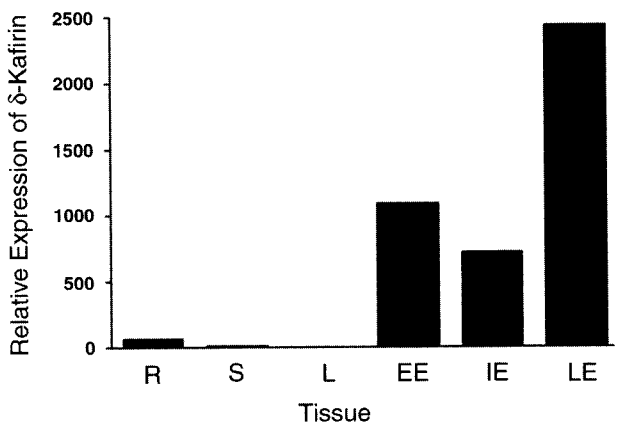

Fig. 3. Relative expression of δ-kafirin gene in roots (R), stem (S), leaves (L), early endosperm (EE, 10–12 days), intermediate endosperm (IE, 15–20 days), and late endosperm (LE, >21 days). Data were normalized using values from expression of the 26S ribosomal gene.

from the GenBank to amplify 3' ends. A PCR fragment at ≈700 bp was consistently amplified, ligated into TOPO-TA vector (Invitrogen), and transformed into *E. coli* cells. Cells were plated onto media containing IPTG (isopropyl-β-D-thiogalactopyranoside [Progen, Brisbane, Australia]) and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside [Progen]) to allow for blue-white selection or recombinants. White colonies were isolated and sequenced with M13 universal primers and GenBank was BLAST-searched for nucleotide and amino acid similarities. Multiple sequence alignment and phylogenetic relationships with other plant sequences were performed using ClustalW, Seqboot, Protdist, Neighbor, and Consensus programs from BioManager server (http://biomanager.angis.org.au) The nucleotide and protein sequences used for comparative analysis were taken from the GenBank library (Table I).

Quantitative RT-PCR

cDNA samples (≈100 ng) from different tissues were used in a quantitative PCR analysis on a 7700 Sequence detector (Applied Biosystems, Melbourne, Australia). A 25-µL sample was prepared containing 2× SYBER green PCR master mix (Applied Biosystems), 100 n*M* of forward (Met-Dig5': 5'-AGACGCCAATGATG ATGCC-3') and reverse (Met-Dig3': 5'-CAGCAACCCTTCGTT AGTTC-3') primers. The amplification program included one denaturing step at 95°C for 10 min, followed by an amplification step of 45 cycles for 10 sec at 95°C, 1 min at 60°C, and 1 min at 72°C. The PCR product was detected using the fluorescent dye Sybergene (Applied Biosystems). To test the efficiency of the primers, a standard curve was generated using a cDNA sample dilution series. For standardization we included samples for amplification of the 26S ribosomal RNA gene. The primers used in this case were 5'-TTAACAGCCTGCCCACCCTGG-3' and 5'-ATCC ATTTTGCCGACTTCCC-3' as forward and reverse, respectively. All samples were analyzed in duplicate and a negative, template-free control for each set of primers was added. The expression levels of the δ-kafirin gene were expressed relative to the 26S ribosomal RNA gene. An additional RT-PCR experiment using the same cDNA samples and Met primers in a 50-µL reaction mixture was performed following the Elongase system protocol (Invitrogen, Carlsbad, CA) with a lower 58°C annealing temperature.

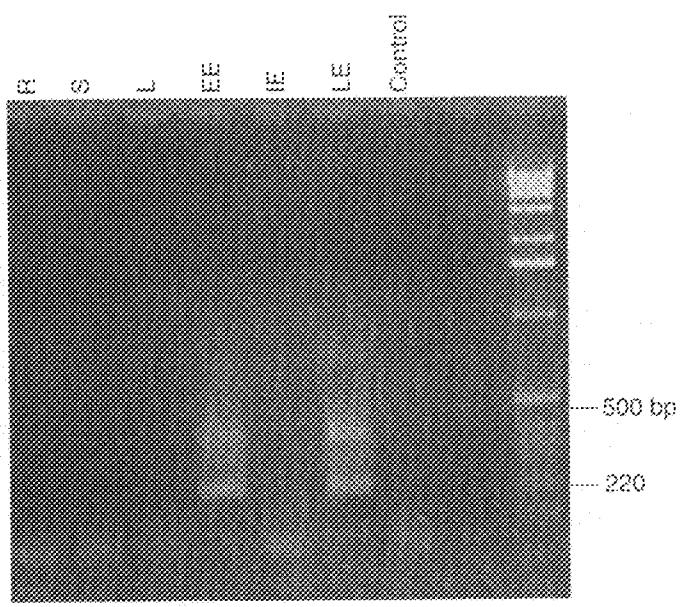

Fig. 2. RT-PCR analysis for the δ-kafirin gene in roots (R), stem (S), leaves (L), and three stages of developing seeds: ≈10–12 (early endosperm, EE); ≈18–20 (intermediate endosperm, IE); and ≈25 (late endosperm, LE) days after flowering (DAF).

Hybridization Experiments

A nonradioactive Northern blot was performed to obtain cDNA clone for a template in a PCR reaction to generate a Digoxigenin labelled fragment of 209 bp by incorporation of DIG-dUTP (PCR-DIG probe synthesis kit, Roche Diagnostics, Sydney, Australia). The primers used for the amplification were Met-Dig5' (forward) and Met-Dig3' (reverse). This PCR product was used as a probe for RNA detection. Hybridization was conducted overnight at 42°C following the DIG-system protocol (Roche Diagnostics).

RESULTS AND DISCUSSION

cDNA and Amino Acid Sequence

Using the 3'RACE technique, we isolated a cDNA fragment from developing sorghum seeds. The complete length of this cDNA clone was 660 bp (Fig. 1) and contained one polypeptide-coding region of 147 amino acids with a molecular mass of 16.4 kDa. The polypeptide begins at the second ATG codon at position 114 and ends at position 554. The second ATG was considered the start of the polypeptide because it was very closely located to an identical motif (CCGCCATGGC) to the consensus sequence reported for zeins (Heidecker et al 1991). The sequence had two noncoding regions of 44 and 103 bp in the 5' and 3' ends, respectively. Three polyadenylation signal sequences (AATAAAT) were determined at 16, 29, and 69 nt upstream of the poly (A) tail. Moreover, a potential hairpin loop of 8 bp was found between the first two adenylation signals; this was similar to the one reported for a 16 kDa rice prolamin (Shyur et al 1992). The clone is rich in hydrophobic amino acids including methionine (17%), proline (10.2%), leucine (12.2 %), and alanine (10.9%). The methionine residues, although clustered throughout the polypeptide, were mainly concentrated in the center of the molecule.

δ-Kafirin Expression

Northern blots were unable to detect δ-kafirin transcript from any tissue. Because of the high homology with the GenBank sorghum sequence AY043223, we had a very narrow selection for a sequence to use as a probe or for designing primers that were specific to our transcript. The expresion of AY043223 sequence was detected by Northern (data not shown). This might indicate that our δ-kafirin is much lower expressed than AY043223. RT-PCR analysis detected expression in early (≈12 DAF) and late stages (≈20–25 DAF) of endosperm development (Fig. 2). The 215 nt RT-PCR products were sequenced to confirm their identity. However, RT-PCR experiment also revealed expression (although low) of δ-kafirin transcript in intermediate stages (IE) of development. Both techniques confirmed endosperm-specific expression.

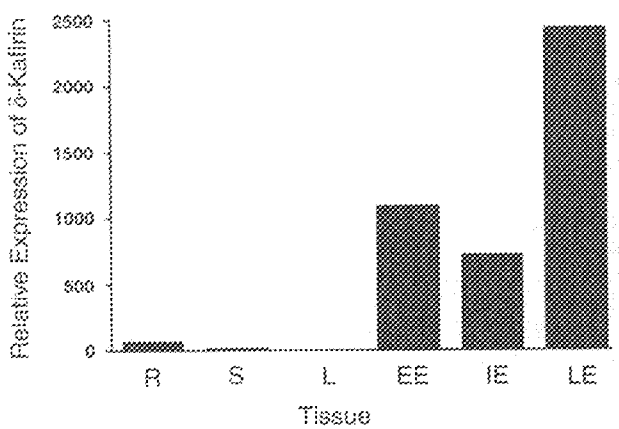

Fig. 3. Relative expression of δ-kafirin gene in roots (R), stem (S), leaves (L), early endosperm (EE, 10–12 days), intermediate endosperm (IE, 15–20 days), and late endosperm (LE, >21 days). Data were normalized using values from expression of the 26S ribosomal gene.

of the gene, and that level of transcript is the highest at later stages (LE) of developing seeds (Fig. 3). The lack of expression for the IE tissue in the RT-PCR experiment suggests it is a more sensitive method than standard PCR.

Comparison with Other MR Seed Storage Proteins

Our sequence, AY834250, showed 85% identity with another methionine-rich (MR) seed storage protein reported for sorghum in GenBank (AY043223) with significant alignment between nucleotides 59–632 (Fig. 1). δ-Kafirin AY834250 gene is 139 nt longer. There are seven short insertions and 17 substitutions from nucleotides 66–104 and then after nucleotide 578. In the central region of δ-kafirin between nucleotides 105–577, the homology is almost complete except for three substitutions and two insertions, one of nine nucleotides located at position 307 and the other of 30 nt starting at position 373. Both insertions are rich in ATG codons, which when translated encode a total of 25 methionine residues to our δ-kafirin. This reiteration of ATG codons has also been previously reported for an 18 kDa zein (U31541) and other zeins (Heidecker et al 1991). These are believed to have arisen as a result of an unequal crossing-over in repetitive motifs in the central part of the coding region (Swarup et al 1995).

```
                    1         11        21        31        41        51
T03381      MAA---KMFALFALLALCATATSATHIPGHLSPLLMPLATMNPWMQYCMKQQGVANLLAW
AY047317    MAA---KMFALFALLALCATATSATHIPGHLSPLLMPLATMNPWMQYCMKQQGVANLLAW
U31541      MAA---KMFALFALLALCATATSATHIQALVT-TTDAIGYHEPWMQYCMKQQGVANLLAW
AF371265    MAA---KMFALFALLALCATATSATHIPGHLSPLLMPLATMNPWMQYCMKQQGVANLLAW
AY043223    MAA---KMFALFALLALCASATSATHIPGHLP-LVMPLGTMNPCTQYCMMQQRFARLLAW
δ-kafirin   MAA---KMFALFALLALCASATSATHIPGHLP-LVMPLGTMNPCTQYCMMQQRFARLLAW
M23537      MAA---KMLALFALLALCASATSATHIPGHLP-PVMPLGTMNPCMQYCMMQQGLASLMAC
AY104139    MAA---KMLALFALLALCASATSATHIPGHLP-PVMPLGTMNPCMQYCMMQQGLASLMAC
AF461049    MAA---KMLALFALLALCASATSATHIPGHLP-PVMPLGTMNPCMQYCMMQQGLASLMAC
U25674.1    MAA---KMLALYALLALCGSGTSATHIPAHLP-PAMPLGTMNRCMQYCMMQQGLASLMAC
S0821       MAAYTSKIFALFALIALSASATTAITTMQYFP-PTLAMGTMDPCRQYMMQTLGMGSSTAM
P04403      MAK----ISVAAAALLVLMALGHATAFRATVTTTVVEEENQEECREQMQRQQMLSHCRMY
M72708      MKM----VIVLVVCLALSAASASAMQMPCPCAGLQGLYGAGAGLTTMMGAGGLYPYAEYL

                   61        71        81        91       101       111
T03381      PTLMLQQLLASPLQ-CQMPMMMPGMMPPMTMMPMPSMMPSMMVPTMMSPMTMASMMPPMM
AY047317    PTLMLQQLLASPLQQCQMPMMMPGMMPPMTMMPMPSMMPSMMVPTMMSPMTMASMMPPMM
U31541      PTLMLQQLLASPLQQCQMPMMMPGMMPPMTMMPMPSMMPS-MVPTMMSPMTMASMMPPMM
AF371265    PTLMLQQLLASPLQPCQMPMMMPGMMPPMTMMPMPNMMPSMMVPTMMSPITIASMMPPMM
AY043223    PIPMLQQLS----------------------------------------------------L
δ-kafirin   PIPMLQQLA----------------------------------LP------------L
M23537      PSLMLQQLLALP--------------------------------LQTMPVMMPQMM
AY104139    PSLMLQQLLALP--------------------------------LQTMPVMMPQMM
AF461049    PSLMLQQLLALP--------------------------------LQTMPVMMPNMM
U25674.1    PSLMLQQLLALP--------------------------------PQTSPVTIPKMM
S0821       FMSQPMALLQ----------------------------------------------------Q
P04403      MRQQMEESPYQT--------------------------------MP-RRGMEPHMS
M72708      RQPQCSPLAAAPYYAG----------------------------CGQTSAMYQPLRQ

                  121       131       141       151       161       171
T03381      MPSMISPMTMPSMMPSMIMPTMMSPMIMPSMMPPMMMPSMVSPMMMPNMMTVPQCYSGSI
AY047317    MPSMISPMTMPSMMPSMIMPTMMSPMIMPSMMPPMMMPSMVSPMMMPNMMTVPQCYSGSI
U31541      MPSMISPMTMPSMMPSMIMPTMMSPMIMPSMMPPMMMPSMVSPMMMPNMMTVPQCYSGSI
AF371265    MPNMVSPMMMPSMMPSTMTPSMMPPIMMPSMIPPMMMPSMVSSMIMPNMMTVPQCYSDSI
AY043223    QPAYQTPMMPNMMP----------------------------PMTMMPSMMSLPQCHCDAI
δ-kafirin   QPAYQTPMMPNMMPPMMP----RPTMMP------------PMTMMPSVMSLPQCHCDAI
M23537      TPNMMSPLMMPSMMSPMVLP----------------SMMSQIMMP------QCHCDAV
AY104139    TPNMMSPLMMPSMMSPMVLP----------------SMMSQMMMP------QCHCDAV
AF461049    TPNMMSPLMMPSMMSPMVLP----------------SMMSQMMMPCRMLLPQCHCDAV
U25674.1    KPNMMYPLMMPSMMSPMGLP----------------SMMSQMMMP------QCLCDAA
S0821       QCCMQLQGMMPQCHCG----------------------TSCQMMQSMQQV-ICAGLGQ
P04403      ECCEQLEGMDESCRCEG----------------------LRMMMMRMQQEEMQPRGEQM
M72708      QCCQQQMRMMDVQSVAQQLQMMMQLERAAAASS-----SLYEPALMQQQQQLLAAQGLNP

                  181       191       201       211
T03381      SHIIQQ-QQLPFMFSPTAMAIPPMFLQQPFVGAAF------
AY047317    SHIIQQ-QQLPFMFSPTAMAIPPMFLQQPFVGAAF------
U31541      SHIIQQ-QQLPFMFSPTAMAIPPMFLQQPFVGAAF------
AF371265    SHIIQQ-QQLPFMFSPTAVAISPMFLQQPFVGAAF------
AY043223    SQIMQQ-QQLPFMFNPTAMAIPPMFLQQPFVSSAF------
δ-kafirin   SQIMQQ-QQLPLMFNPTAMAIPPMFLQQPFVSSAF------
M23537      SQIMLQ-QQLPFMFNPMAMTIPPMFLQQPFVGAAF------
AY104139    SQIMLQ-QQLPFMFNPMAMTIPPMFLQQPFVGAAF------
AF461049    SQIMLQ-QQLPFMFNPMAMTIPPMFLQQPFVGAAF------
U25674.1    SQIMLQ-QQLPFMFNQMAMTIPPMFLQQPFVGAEF------
S0821       QQMMKMAMQMPYMCN-----MAFVNFQLSSCGCC-------
P04403      RRMMRLAENIPSRCN-----LSPMRCPMGGSIAGF------
M72708      MAMMMA-QNMPAMGGLYQYQYQLPSYRTNPCGVSAAIPPYY
```

Fig. 4. Alignment of deduced amino acid sequence of δ-kafirin with methionine-rich (MR) proteins from sorghum (AY043223), maize (T03381, AY047317, U31541, AF371265, AF461049, M23537, AY104139, U25674, and M72708), rice (S0821), and Brazil nut (P04403). Shaded letters represent differences between the two sorghum sequences. Gaps in sequences were to improve alignment.

Protein alignment of the two sorghum clones with 11 other MR proteins (from maize, rice, and Brazil nut) revealed the presence of conserved motifs (Fig. 4). The longest conserved block of residues (19–22) is located at the N-terminal ($MAAKM^{F}/_{L}AL^{F}/_{Y}$ $ALLALC^{A}/_{G}{}^{S}/_{T}ATSAT$) which has been reported to represent a signal peptide that targets the proteins to the PB organelles within the ER. This peptide is cleaved during maturation to yield the form found in mature seed endosperm in maize and rice (Kirihara et al 1988; DeRose et al 1989; Masumara et al 1990). Other short conserved regions are also found through these sequences. There is a C-terminal domain ($^{P}/_{S}PMFLQQPFV^{G}/_{S}{}^{A}/_{S}{}^{A}/_{E}F$) shared between sorghum and maize prolamins. The phylogenetic tree (Fig. 5) showed zeins comprise a very heterogeneous group of proteins. One of these groups (AF371265, T03381, U31541, and AY047317) had a higher homology (78% of identical residues, on average) with the sorghum MR kafirins than with other zeins.

DNA alignment of the two sorghum sequences with three of the zeins (DNA sequence for T03381 was not available) that grouped closely in the phylogenetic tree revealed (data not shown) an extensive homology (>70% overall), suggesting a common ancestor between zeins and kafirins. DeRose (1989) used a conserved motif on the C-terminal of kafirins as a probe to hybridize to the sorghum genome, detecting 20 kafirin-like sequences (140 zeins reported for maize). Some kafirins presumably did not cross-hybridize with the conserved sequence and, hence, numbers may be signi-

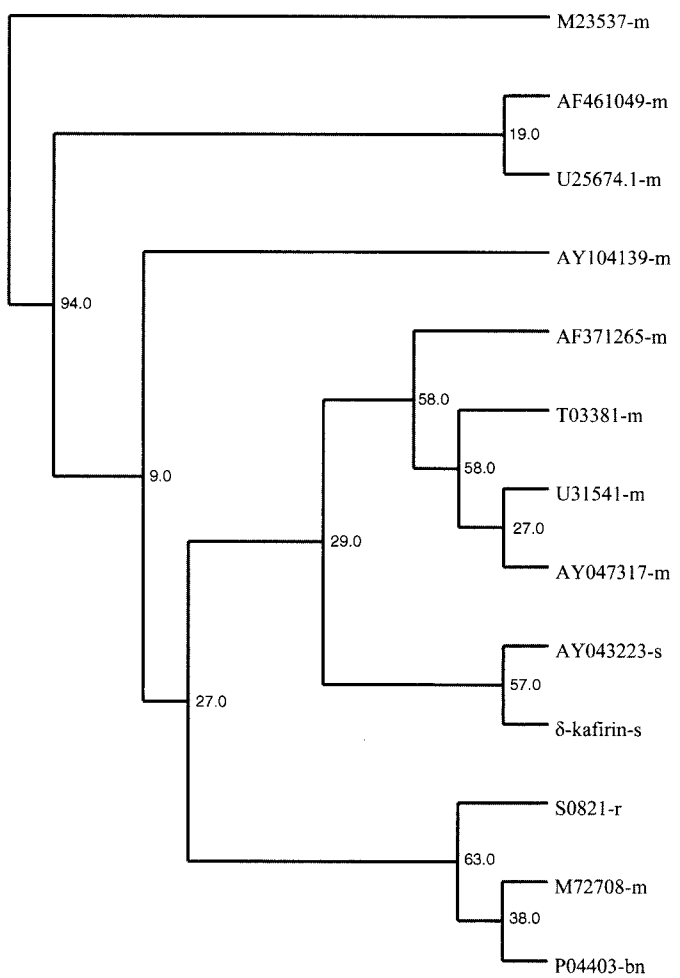

Fig. 5. Phylogenetic relationships among 13 methionine-rich (MR) proteins using the Neighbor joining algorithm. Tree was built using Clustal W, Seqboot, Protdist, Neighbor, and Consensus software from BioManager server. Numbers on branches represent bootstrap values. Letters following each accession number indicate species (m: maize; r: rice; s: sorghum, bn: brazil nut).

ficantly higher. Further experimentation could be performed to develop a series of probes representing all the kafirin and zein classes to screen the sorghum genome more extensively. In maize, δ-zeins represent a small number of proteins compared with the other groups (Woo et al 2001). Woo et al (2001) used in situ hybridization to monitor expression of a δ-zein (U32541) and found that the level of transcript increments with the time of development. However, expression was always confined to two small regions within the endosperm.

CONCLUSIONS

Our results suggest this novel gene encodes a precursor polypeptide that is cleaved to render a mature form of the protein with ≈20% methionine content. This represents a high methionine content, comparable to other similar proteins reported in other species (Table I). This attribute makes δ-kafirin a potential candidate to use for genetic engineering of cereal and pulse crops for animal and human nutrition. Hoffman et al (1987) introduced a chimeric zein gene between the 5' and 3' flanking sequences of the bean β-phaseolin gene that directs the tissue-specific expression of the gene into tobacco seeds. There are several successful examples of the genetic engineering crops for improved sulfur amino acid content for nutritional improvement (Molvig et al 1997; Lee et al 2003). We are currently developing transgenic sorghums with this storage protein overexpressed in the endosperm by using a stronger endosperm-specific promoter.

Preliminary RFLP (Random Fragment Length Polymorphism) analysis in sorghum cultivars (data not shown) demonstrated there is some genetic variation for the trait in sorghum germplasm, with six different alleles and evidence of duplication detected among a diverse set of 25 inbred lines. Hence, there is potential for improvement within sorghum using conventional or marker-assisted breeding. However, it is unlikely that the protein is expressed at sufficient levels in wild-type sorghum to significantly improve the level of methionine in the grain. Research to detect and analyze more sulfur-rich kafirins would lead to a greater understanding of seed storage protein structure and expression in sorghum. This information would be useful for sorghum crop improvement for human and animal diets. The high methionine level of this protein may also be extremely useful for the improvement of other cereals and pulse crops.

ACKNOWLEDGMENTS

We thank David Jordan, sorghum breeder of the Qld DPI&F for discussions regarding this manuscript. We acknowledge the provision of a UQ Research Development Grant in allowing this work to be completed.

LITERATURE CITED

Apel, K., and Kloppstech, K. 1978. The plastid membrane of barley (*Hordeum vulgare*). Light-induced appearance of mRNA coding for the apoprotein of the light of harvesting chlorophyll a/b protein. Eur. J. Biochem. 85:581-588.

Chakraborty, S., Chakraborty, N., and Datta, A. 2000. Increased nutritive value of transgenic potato by expressing a nonallergenic seed albumin gene from *Amaranthus hypochondriacus*. PNAS USA 97:3724-3729.

Coleman, C. E., Dannenhoffer, J. M., and Larkins, B. A. 1997. The prolamins of maize, sorghum and coix. Pages 257-288 in: Cellular and Molecular Biology of Plant Seed Development. B. A. Larkins and I. K. Vasil, eds. Kluwer Academic: Boston.

DeRose, R. T., Ma, D., Kwon, I., 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 Molec. Biol. 12:245-256.

Fisher, H., Chen, L., and Wallish, S. 1996. The evolution of angiosperm seed proteins: A methionine-rich legume subfamiliy present in lower angiosperm clades. J. Molec. Biol. 43:399-404.

Heidecker, G., Chaudhuri, S., and Messing, J. 1991. Highly clustered zein gene sequences reveal evolutionary history of the multigene family. Genomics 10:719-732.

Higgins, T. J. 1984. Synthesis and regulation of major proteins in seeds. Ann. Rev. Plant Physiol. 35:191-221.

Hoffman, L. M., Donaldson, D. D., Bookland, R., Rashka, K., and Herman, E. M. 1987. Synthesis and protein body deposition of maize 15-kd zein in transgenic tobacco seeds. EMBO J. 6:3213-3221.

Jiang, L. W., and Sun, S. M. 2002. Membrane anchors for vacuolar targeting: Applications in plant bioreactors. Trends Biotechnol. 20:99-102.

Kirihara, J. A., Petri, J. B., and Messing, J. 1988. Isolation and sequence of a gene encoding a methionine-rich 10 kDa protein from maize. Gene 71:359-370.

Lai, J., and Messing, J. 2002. Increasing maize seed methionine by mRNA stability. Plant J. 30:395-402.

Lee, T. T. T., Wang, M. C., Hou, R. C., Chen, L., Su, R., Wang, C., and Tzen, J. C. 2003. Enhanced methionine and cysteine levels in transgenic rice seeds by the accumulation of sesame 2S albumin. Biosci. Biotechnol. Biochem. 67:1699-1705.

Masumara, T., Ibino, T., Kidzu, K., Mitsukawa, N., Tanaka, K., and Fujii, S. 1990. Cloning and characterization of a cDNA encoding a rice 13 kDa prolamin. Molec. Gen. Genet. 221:1-7.

Molvig, L., Tabe, L. M., Eggum, B. O, Moore, A., Craig, S., Spencer, D., and Higgins, T. 1997. Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. PNAS USA 94:8393-8398.

Müntz, K. 1998. Deposition of storage proteins. Plant Molec. Biol. 38:77-99.

Oria P. M., Hamaker, B. R., Axtell, J. D., and Huang, C. 2000. A highly digestible mutant cultivar exhibits a unique folded structure of endosperm protein bodies. PNAS USA 97:5065-5070.

Sambrook, J., and Russell, D. 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Shewry, P. R., and Halford, N. 2002. Cereal storage proteins: Structures, properties and role in grain utilization. J. Exp. Bot. 53:947-958.

Shewry, P. R., Napier, J. A., and Tatham, A. S. 1995. Seed storage proteins: Structures and biosynthesis. Plant Cell 7:945-956.

Shull, J. M., Waterson, J. J., and Kirleis, A. W. 1991. Proposed nomenclature for the alcohol-soluble proteins (kafirins) of *Sorghum bicolor* (L.) Moench based on molecular weight, solubility and structure. J. Agric. Food Chem. 39:83-87.

Shyur, L., Wen, T., and Chen, C. 1992. cDNA cloning and gene expression of the major prolamins of rice. Plant Molec. Biol. 20:323-326.

Swarup, S., Timmermans, M., Chaudhuri, S., and Messing, J. 1995. Determinants of the high methionine trait in wild and exotic germplasm may have escaped selection during early cultivation of maize. Plant J. 8:359-368.

Tabe, L. M., Wardley-Richardson, T., Ceriotti, A., Aryan, A., McNabb, W., Moore, A., and Higgins, T. A. B. 1995. A biotechnological approach to improving the nutritive values of alfalfa. J. Anim. Sci. 73:2752-2759.

Taylor, J. R., Schussler, L., and Linderberg, v.d.W., N. 1985. Protein body formation in starchy endosperm of developing *Sorghum bicolor* (L.) Moench, seeds. South Africa J. Bot. 51:35-40.

Woo, Y., Hu, D., Larkins, B., and Jung, R. 2001. Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. Plant Cell 13:2297-2317.

[Received April 14, 2005. Accepted August 24, 2005.]