Theor Appl Genet (2010) 121:1227–1237
DOI 10.1007/s00122-010-1383-9

ORIGINAL PAPER

Allelic variation of the β-, γ- and δ-kafirin genes in diverse *Sorghum* genotypes

H. K. C. Laidlaw · E. S. Mace · S. B. Williams · K. Sakrewski · A. M. Mudge · P. J. Prentis · D. R. Jordan · I. D. Godwin

Received: 16 December 2009 / Accepted: 3 June 2010 / Published online: 19 June 2010
© Springer-Verlag 2010

Abstract The β-, γ- and δ-kafirin genes were sequenced from 35 *Sorghum* genotypes to investigate the allelic diversity of seed storage proteins. A range of grain sorghums, including inbred parents from internationally diverse breeding programs and landraces, and three wild *Sorghum* relatives were selected to encompass an extensive array of improved and unimproved germplasm in the Eusorghum. A single locus exists for each of the expressed kafirin-encoding genes, unlike the multigenic α-kafirins. Significant diversity was found for each locus, with the cysteine-rich β-kafirin having four alleles, including the first natural null mutant reported for this prolamin subfamily. This allele contains a frame shift insertion at +206 resulting in a premature stop codon. SDS-PAGE revealed that lines with this allele do not produce β-kafirin. An analysis of flour viscosity reveals that these β-kafirin null lines have a difference in grain quality, with significantly lower viscosity observed over the entire Rapid ViscoAnalyser time course. There was less diversity at the protein level within the cysteine-rich γ-kafirin, with only two alleles in the cultivated sorghums. There were only two alleles for the δ-kafirin locus among the *S. bicolor* germplasm, with one allele encoding ten extra amino acids, of which five were methionine residues, with an additional methionine resulting from a nucleotide substitution. This longer allele encodes a protein with 19.1% methionine. The Asian species, *S. propinquum*, had distinct alleles for all three kafirin genes. We found no evidence for selection on the three kafirin genes during sorghum domestication even though the δ-kafirin locus displayed comparatively low genetic variation. This study has identified genetic diversity in all single copy seed storage protein genes, including a null mutant for β-kafirin in *Sorghum*.

Communicated by E. Guiderdoni.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1383-9) contains supplementary material, which is available to authorized users.

H. K. C. Laidlaw · S. B. Williams · A. M. Mudge · P. J. Prentis · I. D. Godwin (✉)
School of Land, Crop and Food Sciences, The University of Queensland, Brisbane, QLD 4072, Australia
e-mail: i.godwin@uq.edu.au

E. S. Mace · K. Sakrewski · D. R. Jordan
Department of Employment, Economic Development and Innovation, Hermitage Research Station, Agri-Science Queensland, Warwick, QLD 4370, Australia

Present Address:
H. K. C. Laidlaw
CSIRO Plant Industry, PO Box 1600, Canberra, ACT 2601, Australia

Introduction

The prolamin storage proteins found in sorghum endosperm are called kafirins. Four subclasses, α-, β-, γ- and δ-kafirin, have been described based on their protein characteristics and more recently on DNA sequence. Each kafirin shares a high degree of homology with the equivalent class of protein from maize, the zeins (Garratt et al. 1993; Watterson et al. 1993), the coixins from the related C4 grass, *Coix lachryma-jobii*, commonly known as coix or Job's tears, and the caneins from sugarcane. The kafirins are stored in discrete protein bodies, which form a tight matrix with the starch granules in mature endosperm and contribute to grain hardness, digestibility and processing quality of sorghum (Hamaker and Bugusu 2003; Ezeogu et al. 2005; Shewry and Halford 2002; Wong et al. 2009).

The α-kafirins are the most abundant class, overall comprising 70–80% of total kafirins (Hamaker et al. 1995). The distribution of α-kafirins is more predominant in the outer layers of the grain, comprising 80–84% of the total kafirin fraction in vitreous endosperm and 66–71% in opaque endosperm (Watterson et al. 1993). The α-kafirins are encoded by a multigenic family of 23 genes, although only 19 are known to be expressed (Xu and Messing 2008). Approximately 10 α-kafirin genes are found at a single locus, in a large tandem repeat (Song et al. 2004). In contrast, there are only one or two genes encoding the β-, γ- and δ-kafirins, such as the single β-kafirin gene reported by Chamba et al. (2005). The β-kafirin protein accounts for 7–8% of the total kafirin fraction in vitreous and 10–13% in opaque endosperm (Watterson et al. 1993). The cysteine-rich γ-kafirin, first cloned and characterized by Freitas et al. (1994), is more abundant than β-kafirin, comprising 9–12% of the total kafirin fraction in the vitreous and 19–21% in the opaque endosperm (Watterson et al. 1993). Izquierdo and Godwin (2005) characterized the single copy δ-kafirin gene, which is expressed at very low levels and encodes a methionine-rich protein. δ-Kafirin is thought to comprise less than 1% total seed storage protein in mature grain.

The β- and γ-kafirins are highly cysteine-rich proteins and appear to form both intra- and inter-molecular disulfide bonds. This in turn leads to a higher degree of protease resistance in these seed storage proteins. As these kafirins are on the periphery of the protein body in the mature grain (Oria et al. 2000), they contribute to sorghum's lower protein digestibility relative to other cereals (Belton et al. 2006). Cooked sorghum also has a lower protein digestibility compared to maize, which is not desirable for its use as a staple food, as monogastric animal feed or in industrial processes. This is likely to be caused by polymerization of the kafirins through disulfide bonding (and other factors as reviewed by Duodu et al. 2003) and conformational changes in secondary structure resulting in lower susceptibility to proteolysis (Emmambux and Taylor 2009).

Among other factors, the success of selection for genetic gain is dependent on the presence of sufficient genetic variation in the breeding population. The recent availability of the sorghum whole genome sequence (from the inbred genotype BTx623; Paterson et al. 2009) allows for the precise *in silico* identification of the kafirin loci and the design of primers to characterize these loci. While the kafirins have been characterized at a protein level and more recently at the DNA sequence level, the allelic diversity of these important cereal storage proteins has not been investigated. To assess the allelic diversity of β-, γ- and δ-kafirin genes and to identify novel variants at these loci, the expressed kafirin genes were sequenced in sorghum lines from a wide geographical and genetic background. We also developed tightly linked PCR-based markers and mapped these genes onto a consensus genetic linkage map of sorghum, linked to the whole genome sequence. Implications for molecular breeding for altered grain quality characteristics are discussed. Finally, we tested for evidence that any of these loci had experienced a history of selection during domestication.

Materials and methods

Plant material

The β-, γ- and δ-kafirin genes were sequenced from a diverse set of 32 inbred grain sorghum lines and three wild relatives in the genus *Sorghum* (Table 1), consisting of a subset of the 94 diverse lines characterized in Mace et al. (2008). Sorghum lines were included from diverse geographical locations and breeding programs, including Australia/QPIF, Africa, the USA/TAMU and India/ICRISAT. The wild sorghum relatives, all from the Eusorghums, are interfertile with the cultivated sorghums and include the Asian *S. propinquum* and the African weedy types *S. bicolor* spp. *verticilliflorum* and *S. bicolor* spp. *drummondii*.

DNA extraction, PCR amplification and sequencing

Total genomic DNA was extracted from each sorghum line using a modified CTAB-based extraction protocol (Saghai-Maroof et al. 1984; Doyle and Doyle 1987).

Primers were designed from available sequences for each kafirin gene, including in-house sequences and those available at NCBI for β- and γ-kafirin. The sorghum whole genome sequence (WGS) trace archive at NCBI was used to identify flanking sequences for the design of primers to amplify single, full-length amplicons for δ-kafirin. Combinations of primers bKafF/bKafR, gKafF/gKafR and dKafF/dKafR (Supplementary Table 1) were used to amplify the full-length ORF of β-, γ- and δ-kafirin, respectively. For amplification of the β- and δ-kafirin genes, the 25 μL PCR consisted of 4 mM $MgCl_2$, 0.4 mM dNTPs, 400 nM primers, 1.05 units of Expand High Fidelity Enzyme (Roche Diagnostics GmbH, Germany), and 10 ng of genomic DNA. All reactions were performed on a MyCycler Thermal Cycler (Bio-Rad Laboratories, Inc., USA). To amplify the β-kafirin fragment, the following conditions were used: 94°C for 3 min, followed by 30 cycles of 94°C for 45 s, 58°C for 45 s and 72°C for 60 s, and a 7 min final extension at 72°C. For the amplification of the δ-kafirin fragment, the following conditions were used: 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 45 s, and a 7 min final extension at 72°C. For the amplification of the γ-kafirin

Table 1 Origin and race of the diverse set of 32 inbred sorghum lines and three wild relatives

Code	Aliases	Race	Origin
B923296	B23296	Complex	Australian breeding program
QL12		Complex	Australian breeding program
QL39		Complex	Australian breeding program
QL41		Complex	Australian breeding program
R931945-2-2	R31945-2-2	Complex	Australian breeding program
R890562	R90562	Complex	Australian breeding program
S. bicolor spp. *verticilliflorum*	*S. arundinaceum*	Wild relative	Endemic Australia
S. propinquum		Wild relative	S.E. Asia
296B		Complex	China
Ai4		Complex	China breeding program
LR9198		Complex	China breeding program
M35-1		Durra	India
ICSV400		Caudatum derivative	Indian breeding program
ICSV745		Complex	Indian breeding program
B35	BTx642	Durra	Ethiopia
IS 12611C	SC 111, PI 534146	Caudatum	Ethiopia
IS 12648C	SC 157-6, PI 276818	Kafir-caudatum	Ethiopia
IS 12661		Caudatum	Ethiopia
IS 8525	PI 563092	Kafir	Ethiopia
S. bicolor spp. *drummondii*	PI 330272, AusTRCF 300263	Wild relative	Ethiopia
SC 170-6-8	Partially converted IS 12661	Caudatum	Ethiopia
IS 12572C	SC 62C, PI 152730	Complex	Kenya
IS 17214	SC 1075, NSL 365700	Unknown	Nigeria
Kaura (black glumes)	Kaura (black glume) Nth Nig	Caudatum derivative	Nigeria
IS 3151		Caudatum	South Africa
IS 22457C	SC 1552, PI 569903	Caudatum	Sudan
BOK11		Kafir	US breeding program
BTx3197	Combine Kafir 60	Kafir	US breeding program
BTx623		Kafir-caudatum	US breeding program
KARPER 669		Complex	US breeding program
RTx2737		Complex	US breeding program
RTx7000		Kafir-caudatum	US breeding program
Hegari		Caudatum	Unknown
KS 115		Durra	Unknown
R9733		Unknown	Unknown

gene, a concentration of 2 mM $MgCl_2$ was used at an annealing temperature of 60°C in the β-kafirin thermocycling program.

All PCR products were purified using the Wizard® SV PCR Clean-Up System (Promega Corporation, USA). DNA sequencing was performed on an ABI 3130*xl* Genetic Analyzer following the manufacturer's instructions (Applied Biosystems, USA). Direct sequencing of the PCR products was performed using the forward primers (bKafF, gKafF, dKafF) and with the reverse primers (bKafR, gKafR, dKafR) (Supplementary Table 1) to confirm rare or low frequency SNPs.

Sequence analysis

Signal peptide predictions were carried out using SignalP (http://www.cbs.dtu.dk/services/SignalP).

To test for a history of selection on kafirin genes during sorghum domestication, we used the HKA test (Hudson et al. 1987) using DNASP (Librado and Rozas 2009). This method uses divergence and polymorphism data at two or more loci to determine whether sequence variation at individual loci is statistically different from one another (i.e., higher or lower) under the null hypothesis of neutral evolution. The HKA test generates an expected number of

segregating sites and sequence differences between species assuming all loci share the same population history. A significant difference between the observed and expected test statistics indicates that the assumption (that all loci share the same neutral population history) was violated, possibly as a result of selection during domestication. We conducted the test with the Asian species, *S. propinquum*, as an outgroup in analyses with all sites and silent sites only.

Mapping to chromosomes

The β- and γ-kafirin loci were mapped in two previously described mapping populations; R931945-2-2/IS 8525 RIL population ($n = 146$; Parh 2005; Mace et al. 2008, 2009) and the BTx623/IS3620C RIL population ($n = 137$; Menz et al. 2002; Mace et al. 2009). Both mapping populations have been used previously as component maps to construct a high density consensus map for sorghum (Mace et al. 2009). The δ-kafirin locus was mapped only in the R931945-2-2/IS 8525 RIL population using the sequencing primers, as these revealed a size difference between the R931945-2-2 and IS 8525 amplicons. For the β- and γ-kafirin loci, primers were designed for SSRs in flanking sequences identified using Phytozome (http://www.phytozome.net/cgi-bin/gbrowse/sorghum/) and FastPCR (Kalendar 2007). Genotyping data for the three kafirin loci were integrated with the previously collected marker data for each mapping populations and mapped as previously described (Mace et al. 2008, 2009) using the MultiPoint software (http://www.multiQTL.com). The position of the three new kafirin loci were projected onto the sorghum consensus map using the "neighbors" approach described by Cone et al. (2002). The graphical representation of the map was drawn using MapChart software (Voorrips 2002). The physical bp positions of the three loci were identified using the BLASTn function (E threshold <0.001) against the sorghum whole genome sequence via Phytozome (http://www.phytozome.net).

Protein extraction and gene electrophoresis

Sorghum grains were ground in a coffee grinder to a fine powder. A 100 mg sample was treated twice with 50 mM Tris–HCl at pH 7.8, 100 mM KCl and 5 mM EDTA at room temperature to separate the water/salt-soluble albumin/globulin fraction by centrifugation. The prolamin fraction (including the kafirins) was solvent extracted twice in a buffer consisting of 60% tertiary butanol, 0.5% Na acetate and 2% β-mercaptoethanol.

SDS-PAGE to separate the kafirins was performed on pre-cast 15% Tris–HCl polyacrylamide gels (Bio-Rad Laboratories Inc, USA) and electrophoresed in Tris/glycine/SDS buffer (Bio-Rad Laboratories Inc, USA), according to the manufacturer's instructions. Proteins were visualized after staining with Biosafe Coomassie blue stain (Bio-Rad), according to the manufacturer's instructions.

Starch viscosity analysis

Viscosity analyses were performed on a selection of 9 of the 35 sorghums used in this study. An additional β-kafirin null line, SC382 from Nigeria, was included in this analysis. Starch viscosity analyses were determined using the rapid viscosity analyzer (RVA-4) (Newport Scientific, Warriewood NSW 2102) essentially as described by Sopade et al. (2006) with data analysis software (Thermocline). Each sample was held at 50°C and stirred at 960 revolutions min^{-1} for 10 s followed by constant stirring at 160 revolutions min^{-1} for 50 s. With constant stirring at 160 revolutions min^{-1}, the samples were heated from 50 to 95°C in 3 min and 42 s (equivalent to 12.2°C min^{-1}), held at 95°C for 2 min and 30 s, cooled to 50°C in 3 min and 48 s (equivalent to 11.8°C min^{-1}) and finally held at 50°C for 5 min. The total time for each test was 15 min, and duplicate analyses were conducted to obtain the reported values.

Results

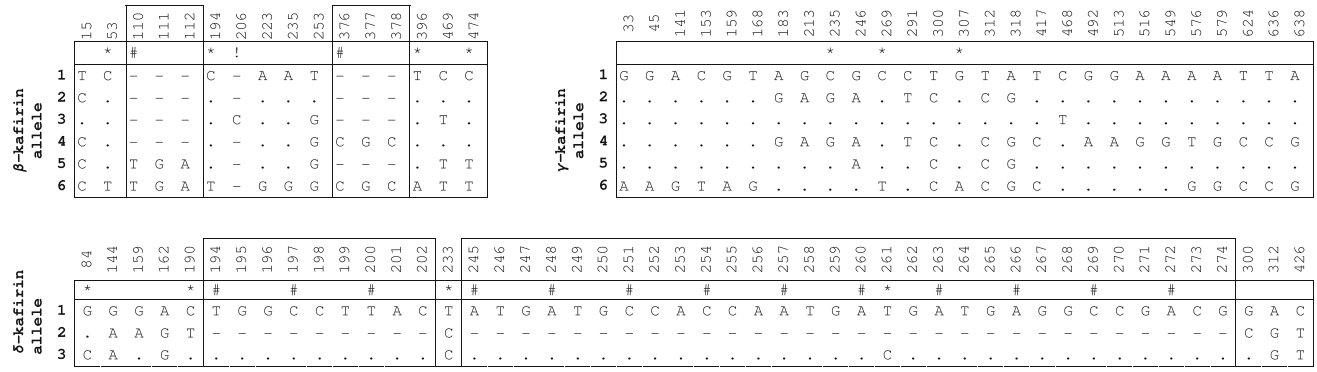
Kafirin sequencing and allelic diversity

The β-, γ- and δ-kafirin genes were sequenced from 32 inbred sorghum lines and three wild relatives from the genus *Sorghum*. Allelic variants were identified and nucleotide polymorphisms for all genes are summarized in Table 2. All of the allelic variants previously identified and included in the NCBI database were found in this study (β-kafirin AJ717660 with incomplete 3′ end is allele type 2; γ-kafirin M73688 is allele type 1 and X62480 is allele type 2; δ-kafirin AY043223 is allele type 2 and AY834250 is allele type 1 with two possible SNPs). Nucleotide sequences for new allele types have been deposited as Genbank accessions GU732401–GU732412 (Table 3). A summary of allelic variants of the three kafirins in the 35 *Sorghum* genotypes is shown in Table 3. This demonstrates the wide range of allelic combinations that exist in these lines and clearly shows the unique nature of the alleles seen in the wild relative, *S. propinquum*.

β-kafirin

The β-kafirin coding sequence varies between 579 and 585 nucleotides in length and encodes a putative amino acid

Table 2 Nucleotide polymorphisms in the alleles of the β-, γ- and δ-kafirin coding sequences



Numbers on the top correspond to the nucleotide position within each kafirin sequence. Nucleotide changes are indicated by standard abbreviations, while dots represent identity with the first allele and a dash represents its absence. Changes that alter the predicted amino acid residue are designated by an asterisk (*) and indels of a single nucleotide are indicated by an exclamation mark (!). Indels in the predicted amino acid sequence are indicated by a hash symbol (#) at the first position of the codon and are boxed

sequence of 193 and 195 residues, which includes a predicted signal peptide of 20 residues. Four allelic types were identified in our panel of grain sorghum (>50% carried allele 1), and *S. bicolor* spp. *verticilliflorum* and *S. propinquum* both carried a unique allele. Four lines carried an allele with a single cytosine insertion at the +206 position (allele 3) that caused a frame shift and led to an early termination codon. Subsequent protein analysis using SDS-PAGE has shown that these lines, as represented by QL12, do not produce β-kafirin (Fig. 1). The full transcript of the *S. propinquum* allele includes three alanine to valine and one leucine to glutamine substitution resulting from single nucleotide substitutions, and an additional methionine and alanine insertion. The allele found in line ICSV400 (allele 4) also contained the same methionine insertion. It should also be noted that the previously published β-kafirin sequence (allele 2; Chamba et al. 2005) has three nucleotides missing immediately preceding the predicted stop codon. The 3′ primer used by Chamba et al. (2005) was designed from the homologous β-zein and β-coixin sequences, which lack this extra codon.

γ-kafirin

The γ-kafirin coding sequence is 639 nucleotides in length and encodes a putative amino acid sequence of 213 residues, including a predicted signal peptide of 19 residues. Five different alleles were identified in grain sorghum and a unique allele in *S. propinquum*. Line IS 12661 has ambiguous sequence at four nucleotide positions, suggesting that it is heterozygous at this locus. Almost all the nucleotide changes were silent in the grain sorghum alleles resulting in only two protein variants with a single amino acid difference. This difference was caused by a cytosine to guanine substitution at nucleotide 235 that resulted in a proline to alanine substitution in the protein sequence (alleles 2 and 4). The *S. propinquum* allele had a proline to leucine substitution as a result of a nucleotide substitution at nucleotide +269 and a valine to isoleucine substitution as a result of a nucleotide substitution at +307.

δ-Kafirin

The δ-kafirin coding sequence varies between 405 and 444 nucleotides in length. This encodes a putative amino acid sequence of between 135 and 148 residues, including a predicted signal peptide of 21 residues. Two alleles were identified in cultivated grain sorghum and both were equally represented across the diverse sample of genotypes. A unique allele was identified in *S. propinquum*, although only one nucleotide difference was at a new position not variable between the other two allelic types. The two major alleles differed by the presence (allele 1 represented by QL39) or absence (allele 2 represented by QL41) of two insertions. These were a 9-bp insertion from +194 to +202 (encoding ALP residues) and a 30-bp insertion from +266 to +295. The second insertion is methionine rich and in a region of repetitive motifs, encoding PMMMRPTMMP. Allele 1 had one other nucleotide substitution encoding a methionine residue; hence overall, this allele encoded a protein with six more methionine residues compared to the protein encoded by allele 2. The allele with the insertions was more methionine rich than the allele without them (19.1 and 15.9% of the mature protein, respectively).

Table 3 Allelic composition of kafirin genes in each sorghum line

Sorghum line	β	γ	δ
S. propinquum	6	6	3
S. bicolor spp. *drummondii*	1	4	1
Kaura (black glumes)	3	4	2
IS 8525	1	4	2
KS115	1	4	2
BTx3197	1	2	2
QL41	1	2	2
BOK11	1	2	2
RTx7000	1	2	1
R931945-2-2	1	2	1
B35	1	2	1
BTx623	2	2	1
R9733	4	2	1
S. bicolor spp. *verticilliflorum*	5	3	2
QL12	3	1	2
IS 17214	3	1	2
LR9198	2	3	2
IS12648C	1	3	2
Ai4	1	3	2
M35-1	1	3	2
IS 12661	1	1, 5	2
B923296	1	1	2
IS12572C	1	1	2
IS 3151	1	5	1
IS 22457C	2	1	1
IS 12611C	2	1	1
QL39	2	1	1
SC170-6-8	2	1	1
RTx2737	3	1	1
ICSV400	4	1	1
Hegari	4	1	1
296B	1	3	1
ISCV745	1	1	1
R890562	1	1	1
KARPER669	1	1	1

Each allele has been numbered sequentially and colored similarly, based on abundance in this set of germplasm. The allele numbering corresponds to the allele types listed in Table 2. Allelic variants correspond to the following Genbank accessions: β-kafirin alleles 1–6, GU732401–GU732406, respectively; γ-kafirin allele 1, M73688; allele 2, X62480; alleles 3–6, GU732407–GU732410, respectively; δ-kafirin allele 1, GU732411; allele 2, AY043223; allele 3, GU732412. Alleles that encode identical protein sequences are β-kafirin alleles 1, 2; γ-kafirin alleles 1, 3 and 5, and alleles 2 and 4

Tests for selection

The HKA test with both total sites and silent sites only detected no departure from the null hypothesis of neutral evolution This suggests that the kafirin genes shared similar population histories during *Sorghum* domestication (all χ^2 tests $P > 0.1$).

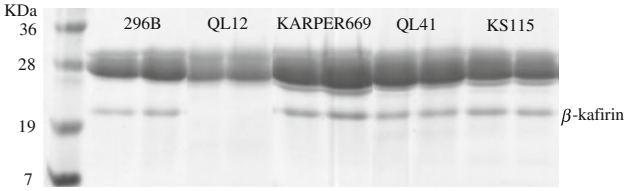


Fig. 1 Natural variation of the β-kafirin protein in five inbred sorghum lines, as shown on an SDS-PAGE gel. Duplicate kafirin extracts were loaded for each inbred line. The β-kafirin protein is absent in QL12, which has a single cytosine insertion at the +206 position that causes a frame shift and leads to an early termination codon

Mapping

Genetic linkage maps of chromosomes 2, 9 and 10 are presented in Fig. 2. The γ-kafirin locus maps to SBI-02, at position 123 cM on the consensus map and sequence maps to 60,425,497 bp and corresponds to the gene Sb02g025510. The β-kafirin locus maps to SBI-09, at position 35 cM on the consensus map and sequence maps to 173,690 bp and corresponds to the gene Sb09g000360. The δ-kafirin locus maps to the centromeric region of SBI-10 at position 58 cM, closely linked to the SvPEPCAA marker, an (AT)n SSR identified in the ninth intron of the phosphoenolpyruvate carboxylase gene (PEPC; Taramino et al. 1997), and sequence maps to 20,416,917 bp and corresponds to the gene Sb10g013050.

Starch viscosity analysis

Viscosity properties (Fig. 3) were determined for nine of the lines used in this study, plus one additional β-kafirin null line identified in our laboratory as part of a related project (a line of Nigerian origin, SC382, A. Mudge, pers. comm.). Significant variation among the lines was observed during the entire time course of analysis. The two β-kafirin null lines showed the lowest viscosity levels during all stages, including peak and final viscosity. The first peak is the water absorption stage and is predominantly dependent on protein content, with higher protein content generally associated with higher gelatinization temperature (P. Sopade, pers. comm.).

Discussion

A survey of the β-, γ- and δ-kafirin genes of sorghum has revealed allelic diversity at each locus. Within a diverse set of 35 *Sorghum* genotypes, six alleles were identified for both β- and γ-kafirin, compared to only three alleles for δ-kafirin. New allelic variants have been found for both β- and γ-kafirin genes.

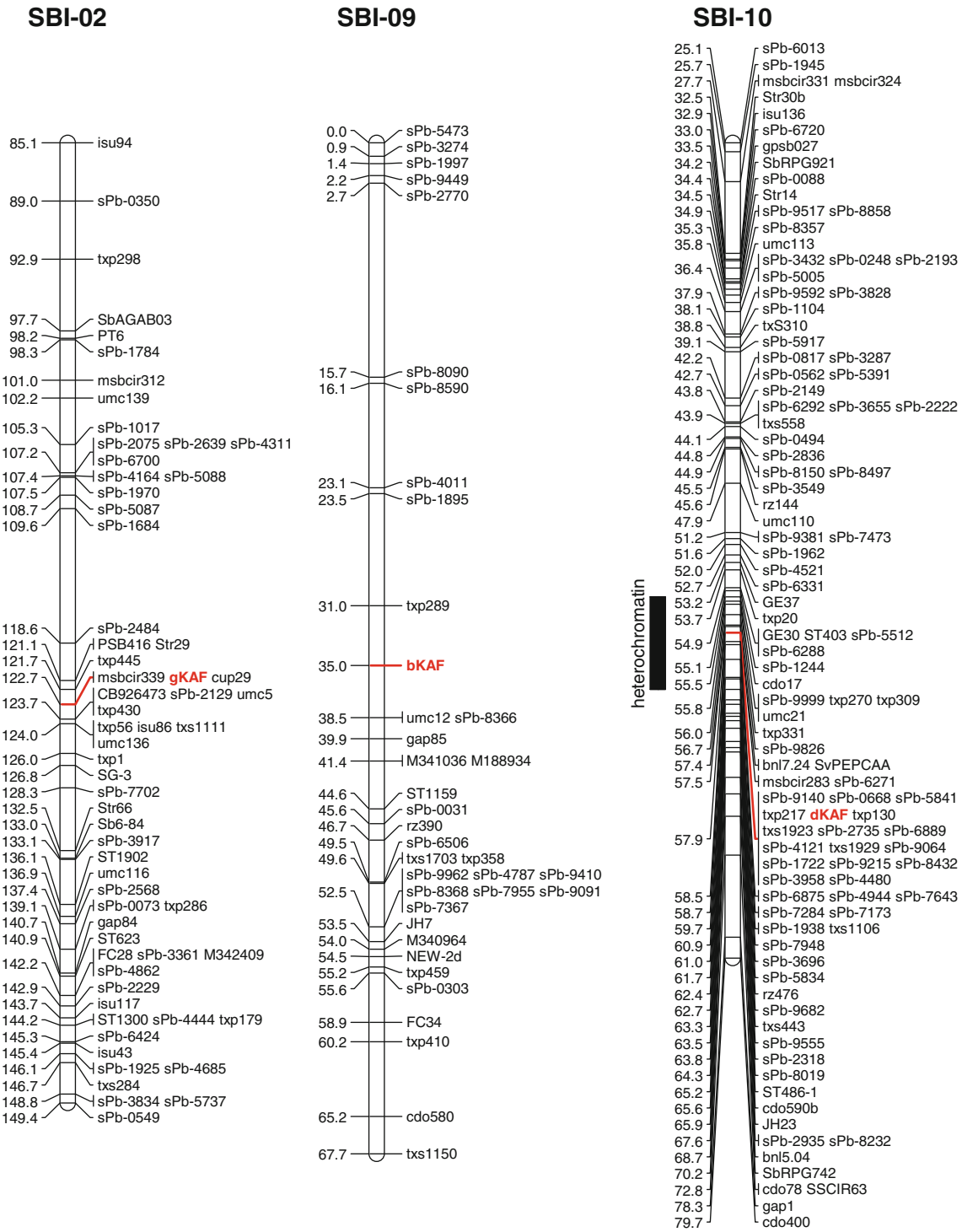


Fig. 2 Location of the three kafirin loci, projected onto the sorghum consensus map (Mace et al. 2009), based on genetic linkage mapping in two independent mapping populations

Screening the β-kafirin locus has revealed the presence of an allele in four grain sorghum lines (QL12, RTx2737, IS17214 and Kaura) with a single nucleotide insertion that causes a frame shift and premature termination codon. The absence of the β-kafirin protein in one of these lines (QL12) was demonstrated on an SDS-PAGE gel (Fig. 1). This represents the first published report of a β-kafirin null mutant. Two β-kafirin null lines were shown to have different viscosity properties compared to a selection of other sorghum lines (Fig. 3). As the β-kafirin protein is thought

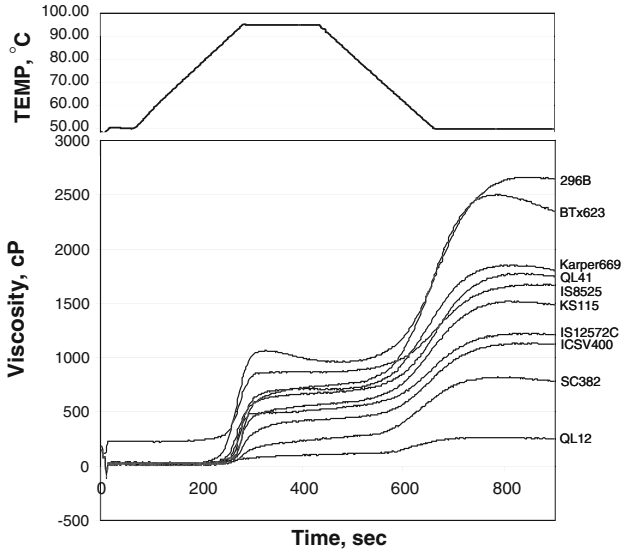


Fig. 3 Analysis of starch viscosity of 10 diverse sorghum lines using the Rapid ViscoAnalyser. The β-, γ- and δ-kafirin allelic composition was known for nine of the lines (this study). An additional line from Nigeria (SC382) was also included. Both SC382 and QL12 are β-kafirin null lines

to be present around the outside of the protein body, it is possible that its absence in the null mutants may affect the structure of the protein body and therefore influence grain quality characteristics as demonstrated by these differences in viscosity. The identification of these lines should be useful for further investigations into the role of β-prolamins in the formation and structure of the protein body, which is yet to be clearly defined. Further examination using scanning electron microscopy will also be useful for determining whether the null mutation causes any major disruption to the starch–protein matrix. The four lines with the null allele are all yellow endosperm types, but are otherwise normal in appearance. It is unlikely that this is a causal relationship. The endosperm color is under multigenic control (Salas Fernandez et al. 2008) and two lines, Karper 669 and KS115, also have yellow endosperm but do not have the null allele. The four lines observed to carry the β-kafirin null allele are likely to be identical by descent for this gene. The lines Kaura (black glumes) and IS17214 both originate from Nigeria. Lines QL12 and RTx2737 share a common ancestor known as Short Kaura, a derivative of Kaura. Further research is required to identify whether this null allele may have been selected as a result of a specific end-use quality parameter or other phenotype.

In maize, developing protein bodies initially form as small accretions of β- and γ-zeins, as demonstrated in immunolocalization experiments (Lending and Larkins 1989). The α- and δ-zeins subsequently accumulate within this network as discrete locules that expand and aggregate to fill the interior. This results in the β- and γ-zeins being concentrated around the peripheral layer of the more mature protein body. Subsequent work by Kim et al. (2002) suggests that zein protein interactions are key determining factors in protein body assembly. Using a yeast two-hybrid system, strong interactions were found between the 16-kDa γ- and 15-kDa β-zein, and between the 10-kDa δ- and the α-zeins. These interactions are generally consistent with the model of protein body formation that has been proposed from the immunolocalization experiments. Transgenic studies have also shown that co-expression of both β- and δ-zein genes results in more stable formation of protein bodies when compared to δ-zein being expressed alone (Bagga et al. 1997). As more genomic sequences have become available, it has also been possible to study the ancestry and gene amplification of the kafirins within the grasses (Xu and Messing 2009). This recent study discusses the diversification of the prolamins as a function of their compartmentalization within the seed as storage proteins. The proposed evolutionary pathway, with β- and γ-kafirin as the older prolamins, aligns well with their proposed roles in protein body formation, i.e., with their initial deposition playing a central role in the structural and functional properties of the protein body (Xu and Messing 2009). The β-kafirin protein is typically a minor component of the protein body and it appears that some sorghum lines can function without it, or have another protein filling its functional role. Further work is required to determine whether the absence of a β-kafirin protein affects the structure of the protein body and whether this subsequently alters the digestibility and grain quality in these lines.

Until recently, the δ-kafirin protein had not been identified at the protein level *in planta* and it was presumed to be present at low abundance (Izquierdo and Godwin 2005), unlike the δ-zein protein that can be detected by SDS-PAGE. A 15-kDa protein from a concentrated protein extract has now been shown to be the δ-kafirin protein using N-terminal sequencing (Mokrane et al. 2009). It is interesting to note that the δ-kafirin gene has the lowest level of allelic diversity, perhaps suggesting that this locus is a recent target of positive selection pressure. This contrasts with the 18-kDa δ-zeins that have significant variability in their amino acid sequence (Woo et al. 2001). These authors propose that the δ-zein genes are under low evolutionary selection pressure to maintain their primary protein structure and have had little or no selection on the alleles during domestication. Although the δ-kafirins and δ-zeins share significant sequence homology, it appears that the selective pressures on these loci may have been quite different. As the δ-prolamins appeared before the *Panicoideae* (including sorghum and maize) split from the other cereals (Xu and Messing 2008), further investigation is needed to explain the apparent lack of diversity among the δ-kafirin alleles.

Our results provide no evidence for directional selection on the β-, γ- and δ-kafirin genes associated with the domestication of sorghum. Hamblin et al. (2007) recently reported a similar finding of no evidence for directional selection for 15 candidate loci in the starch metabolism pathway of sorghum. An inability to detect selection at these kafirin and starch metabolism loci may be explained partly by the fact that it is difficult to detect selection when all loci used in an HKA test are under fairly high levels of selective constraint. Domesticated sorghums also have had a history of disruptive selection for different end uses during domestication (flat or fermented bread, ground or fermented porridge, brewing, animal feed, etc.; Doggett 1988) and this may have obscured any signal of directional selection on kafirin genes examined in this study. Future research examining directional selection at kafirin or starch metabolism loci during sorghum domestication should focus within germplasm developed for different end uses to avoid the complicating issue of disruptive selection.

The two δ-kafirin alleles from *S. bicolor* are different due to the 9 and 30 base pair insertions. This difference is similar to an allelic variant of the 10-kDa δ-zein that has an 18-nucleotide insertion in a nearby region of the protein (Kim and Krishnan 2003). However, it is interesting to note that the 30-nucleotide insertion in the δ-kafirins is in the same region of the peptide where an insertion of more than 50 amino acids creates the major difference between the 10- and 18-kDa δ-zeins. This region in the 18-kD δ-zein contains multiple Met-Met-X-Pro repeat motifs, which are similar to the smaller δ-kafirin insertion containing two of these motifs. The 30-nucleotide insertion in the δ-kafirin gene contains a 22-bp tandem repeat of the sequence immediately adjacent to the insertion site. A 12-bp sequence from within this repeat is also present just upstream of this site, highlighting the repetitive nature of the nucleotide sequence within this methionine-rich region of the δ-kafirin. In contrast to the δ-zeins with a methionine content of 22.5 and 25.3%, respectively, the two δ-kafirin alleles have a lower content of 15.9 and 19.1%. This difference between the δ-kafirin alleles was more comparable to that seen between the 10- and 18-kDa δ-zeins, rather than between the 10 k-Da δ-zein alleles of 22.2 and 22.5% (with and without the insertion, respectively). There is significant variation in maize with respect to δ-zein expression that is due to both the combination of the δ-zein alleles present and also the alleles of a post-transcriptional regulator (Wu et al. 2009). Unlike the δ-kafirin alleles reported here, several null alleles for both of the δ-zein loci were found, and these were caused by frame shift mutations or transposon insertion (Wu et al. 2009). In maize lines where both δ-zein null mutations were combined, the grain had a normal phenotype.

It is apparent from the sorghum genome sequence that chromosome 2 contains a tandem duplication of the γ-kafirin gene with less than 7 Kb between them (Sb02g025510 and Sb02g025490), which are homologous to the 27- and 50-kDa γ-zein genes, respectively. The predicted amino acid sequences of these two γ-kafirins share only 28% identity, while the 50-kDa sequence has 86% amino acid identity with a partial "50-kDa γ-canein" from *Saccharum officinarum* (ABP64792). The 50-kDa γ-kafirin is also predicted to have a single intron, unlike any of the other kafirin genes investigated here. The 50-kDa γ-zein has a similar expression pattern to the other γ-zeins, but with significantly fewer transcripts (Woo et al. 2001). As no ESTs for the 50-kDa γ-kafirin have been identified, allelic diversity at this locus was not investigated in the current study.

The γ-kafirin protein sequence is also highly conserved, with nearly all allelic variants having no effect on the amino acid sequence, suggesting that it may also be under some level of functional selection. A functional role for γ-zeins in the assembly of the protein body in maize has been previously suggested based on immunolocalization studies (Lending and Larkins 1989), gene expression analysis (Woo et al. 2001) and on the self-assembly properties of the PPPVHL repeat motif that is common to all the γ-prolamins (as discussed by Belton et al. 2006). Transgenic studies have also shown that the accumulation of α-zein is enhanced when co-expressed with γ-zein (Coleman et al. 1996), demonstrating the interactions between the prolamin subclasses that are involved in protein body development and the importance of the γ-prolamins in this process. Previous studies by Oria et al. (2000) have related the enzyme-resistant periphery formed by disulfide bonds between β- and γ-kafirins with the poor digestibility of α-kafirin in the interior of the protein body, suggesting that these may be key targets for improving the digestibility of sorghum grain.

Any influence that the different kafirin alleles have on sorghum grain quality and digestibility remains to be investigated. Sorghum lines with similar allele types can readily be identified from Table 3, which also allows the identification of lines carrying a desired allelic combination. This information can also be used to select suitable parental lines for the generation of germplasm with any other desired combination of alleles to further investigate their influence on grain quality. Additional allelic variants may also be identified with continued screening of other sorghum germplasm. One study that demonstrated the complexity of the relationship between the kafirin proteins and feed quality characteristics was reported by Hicks et al. (2001). One of their key findings was the significant variation in the proportions of kafirins found among their parent lines and hybrids. It is interesting to note that one of their

male parent lines (Tx2737) is expected to contain the null β-kafirin allele, which may explain some of the combining ability effects that were observed. Studies on the homologous 27-kDa γ-zein demonstrated that mutation of residue 155 from cysteine to alanine (C155A) resulted in a substantial increase in protease digestibility of the altered protein (Lee and Hamaker 2006). They noted that this was a highly conserved Cys residue across all cereal γ-prolamins. We observed no variability for the corresponding residue (C138) in the γ-kafirin alleles reported in the current study.

Assuming the various alleles of the kafirin loci identified in this paper contribute to important grain quality traits, they may become the subject of marker-assisted selection. It is critical therefore to have an understanding of the other genes that might be influenced by selection for specific kafirin alleles. For example, the γ-kafirin locus falls within the QTL region of stg3, a major QTL for the stay-green drought resistance trait (Harris et al. 2007). If drought resistance and grain quality were both of interest to a breeding program, then any marker-assisted selection for γ-kafirin would need to take into account the linkage phase of the desirable alleles for both traits. In addition, the position of two loci within a chromosome does not always provide a good indication of the degree of recombination between the loci. In one of the populations used in this study for example, the δ-kafirin locus and SvPEPCAA marker were very closely linked, yet they were more than 26 Mb apart on chromosome 10.

Wild relatives and landrace accessions should be important sources of genetic variation for crop improvement. A screen of zeins in a diverse set of maize inbred lines, landraces and teosinte accessions has recently revealed the accumulation of several novel zeins in teosinte and substantial variability among the γ- and δ-zein proteins (Flint-Garcia et al. 2009). Differences in the accumulation of methionine in the grain and in the expression levels of δ-zeins within another diverse set of maize genotypes have also been reported (Swarup et al. 1995). Without an obvious phenotype, these high methionine alleles were proposed to have been lost early during domestication and selection, resulting in little genetic variability remaining in the current germplasm (Swarup et al. 1995). While only one accession of the wild sorghum relative *S. propinquum* was included in this study, it was shown to have unique alleles for all three kafirin genes. Further investigation into the genetic diversity of the kafirin genes in this, and other related species, may lead to the identification of substantial new diversity. As the species can be readily hybridized with cultivated sorghums (Chittenden et al. 1994), *S. propinquum* represents a significant genetic resource for the introgression of diverse grain properties into *S. bicolor*. Significant opportunities also exist for using the variation described here for further studies to investigate the importance of different kafirin alleles in sorghum quality and, in particular, for using the identified β-kafirin null mutants to further dissect the role and importance of β-kafirin proteins in the mechanism of protein body formation and seed development in general.

References

Bagga S, Adams HP, Rodriguez FD, Kemp JD, Sengupta-Gopalan C (1997) Coexpression of the maize δ-zein and β-zein genes results in stable accumulation of δ-zein in endoplasmic reticulum-derived protein bodies formed by β-zein. Plant Cell 9:1683–1696

Belton PS, Delgadillo I, Halford NG, Shewry PR (2006) Kafirin structure and functionality. J Cereal Sci 44:272–286

Chamba EB, Halford NG, Forsyth J, Wilkinson M, Shewry PR (2005) Molecular cloning of β-kafirin, a methionine-rich protein of sorghum grain. J Cereal Sci 41:381–383

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* × *S. propinquum*, suitable for high-density mapping, suggests ancestral duplication of sorghum chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Coleman CE, Herman EM, Takasaki K, Larkins BA (1996) The maize γ-zein sequesters α-zein and stabilizes its accumulation in protein bodies in transgenic tobacco endosperm. Plant Cell 8:2335–2345

Cone KC, McMullen MD, Bi IV, Davis GL, Yim Y-S, Gardiner JM, Polacco ML, Sanchez-Villeda H, Fang Z, Schroeder SG, Havermann SA, Bowers JE, Paterson AH, Soderlund CA, Engler FW, Wing RA, Coe EH (2002) Genetic, physical and informatics resources for maize. On the road to an integrated map. Plant Physiol 130:1598–1605

Doggett H (1988) Sorghum, 2nd edn. Longman, London

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf material. Phytochem Bull 19:11–15

Duodu KG, Taylor JRN, Belton PS, Hamaker BR (2003) Factors affecting sorghum protein digestibility. J Cereal Sci 38:117–131

Emmambux MN, Taylor JRN (2009) Properties of heat-treated sorghum and maize meal and their prolamin proteins. J Agric Food Chem 57:1045–1050

Ezeogu LI, Duodu KG, Taylor JRN (2005) Effects of endosperm texture and cooking conditions on the in vitro starch digestibility of sorghum and maize flours. J Cereal Sci 42:33–44

Flint-Garcia SA, Bodnar AL, Scott MP (2009) Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. Theor Appl Genet 119:1129–1142

Freitas FA, Yunes JA, Silva MJ, Arruda P, Leite A (1994) Structural characterization and promoter activity analysis of the *gamma* kafirin gene from sorghum. Mol Gen Genet 245:177–186

Garratt R, Oliva G, Caracelli I, Leite A, Arruda P (1993) Studies of the zein-like α-prolamins based on an analysis of amino acid sequences: implications for their evolution and three-dimensional structure. Proteins 15:88–99

Hamaker BR, Bugusu BA (2003) Overview: Sorghum proteins and food quality. In: Belton PS, Taylor JRN (eds) Workshop on the proteins of sorghum and millets: enhancing nutritional and functional properties for Africa (AFRIPRO), pp 2–4. Available via Afripro conference proceedings. http://www.afripro.org.uk

Hamaker BR, Mohamed AA, Habben JE, Huang CP, Larkins BA (1995) Efficient procedure for extracting maize and sorghum

kernel proteins reveals higher prolamin contents than the conventional method. Cereal Chem 72:583–588
Hamblin MT, Salas Fernandez MG, Tuinstra MR, Rooney WL, Kresovich S (2007) Sequence variation at candidate loci in the starch metabolism pathway in sorghum: prospects for linkage disequilibrium mapping. Crop Sci 47:S125–S134
Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Klein R, Mullet J (2007) Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58:327–338
Hicks C, Bean SR, Lookhart GL, Pedersen JF, Kofoid KD, Tuinstra MR (2001) Genetic analysis of kafirins and their phenotypic correlations with feed quality traits, in vitro digestibility, and seed weight in grain sorghum. Cereal Chem 78:412–416
Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159
Izquierdo L, Godwin ID (2005) Molecular characterization of a novel methionine-rich delta-kafirin seed storage protein gene in sorghum (*Sorghum bicolor* L.). Cereal Chem 82:706–710
Kalendar R (2007) FastPCR: a PCR primer design and repeat sequence searching software with additional tools for the manipulation and analysis of DNA and protein. http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm
Kim WS, Krishnan HB (2003) Allelic variation and differential expression of methionine-rich δ-zeins in maize inbred lines B73 and W23a1. Planta 217:66–74
Kim CS, Woo Y-M, Clore AM, Burnett RJ, Carneiro NP, Larkins BA (2002) Zein protein interactions, rather than the asymmetric distribution of zein mRNAs on endoplasmic reticulum membranes, influence protein body formation in maize endosperm. Plant Cell 14:655–672
Lee S-H, Hamaker BR (2006) Cys 155 of 27 kDa maize g-zein is a key amino acid to improve its in vitro digestibility. FEBS Lett 580:5803–5806
Lending CR, Larkins BA (1989) Changes in the zein composition of protein bodies during maize endosperm development. Plant Cell 1:1011–1023
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452
Mace ES, Xia L, Jordan DR, Halloran K, Parh DK, Huttner E, Wenzl P, Kilian A (2008) DArT markers: diversity analyses and mapping in *Sorghum bicolor*. BMC Genomics 9:26
Mace ES, Rami J-F, Bouchet S, Klein PP, Klein RE, Kilian A, Wenzl P, Xia L, Sakrewski K, Jordan DR (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biol 9:13
Menz MA, Klein RR, Mullet J, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP(R), RFLP and SSR markers. Plant Mol Biol 48:483–499
Mokrane H, Lagrain B, Gebruers K, Courtin CM, Brijs K, Proost P, Delcour JA (2009) Characterization of kafirins in Algerian sorghum cultivars. Cereal Chem 86:487–491
Oria MP, Hamaker BR, Axtell JD, Huang CP (2000) A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proc Natl Acad Sci USA 97:5065–5070
Parh DK (2005) DNA-based markers for ergot resistance in sorghum. PhD thesis. University of Queensland
Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann M, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556
Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018
Salas Fernandez MG, Hamblin M, Li L, Rooney WL, Tuinstra MR, Kresovich S (2008) Quantitative trait loci analysis of endosperm color and carotenoid content in sorghum grain. Crop Sci 48:1732–1743
Shewry PR, Halford NG (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot 53:947–958
Song RT, Segal G, Messing J (2004) Expression of the sorghum 10-member kafirin gene cluster in maize endosperm. Nucleic Acids Res 32:8
Sopade PA, Hardin M, Fitzpatrick P, Desmee H, Halley P (2006) Macromolecular interactions during gelatinisation and retrogradation in starch-whey systems as studied by rapid visco-analysis. Int J Food Eng 2:7
Swarup S, Timmermans MCP, Chaudhuri S, 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
Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME (1997) Characterisation and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72
Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78
Watterson JJ, Shull JM, Kirleis AW (1993) Quantitation of α-, β-, and γ-kafirins in vitreous and opaque endosperm of *Sorghum bicolor*. Cereal Chem 70:452–457
Wong JH, Lau T, Cai N, Singh J, Pedersen JF, Vensel WH, Hurkman WJ, Wilson JD, Lemaux PG, Buchanan BB (2009) Digestibility of protein and starch from sorghum (*Sorghum bicolor*) is linked to biochemical and structural features of grain endosperm. J Cereal Sci 49:73–82
Woo Y-M, Hu DWN, Larkins BA, Jung R (2001) Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. Plant Cell 13:2297–2317
Wu Y, Goettel W, Messing J (2009) Non-Mendelian regulation and allelic variation of methionine-rich delta-zein genes in maize. Theor Appl Genet 119:721–731
Xu JH, 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 USA 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