JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

# Grain Sorghum Proteomics: Integrated Approach toward Characterization of Endosperm Storage Proteins in Kafirin Allelic Variants

Julia E. Cremer,*[,†] Scott R. Bean,[‡] Michael M. Tilley,[‡] Brian P. Ioerger,[‡] Jae B. Ohm,[§] Rhett C. Kaufman,[‡] Jeff D. Wilson,[‡] David J. Innes,[∥] Edward K. Gilding,[⊥] and Ian D. Godwin[†]

[†]School of Agriculture and Food Sciences and [⊥]Institute for Molecular Bioscience, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia

[‡]Center for Grain and Animal Health Research, Agricultural Research Service, United States Department of Agriculture, Manhattan, Kansas 66502, United States

[§]Cereal Crops Research Unit, Agricultural Research Service, United States Department of Agriculture, Fargo, North Dakota 58108, United States

[∥]Department of Agriculture, Fisheries and Forestry, St Lucia, Brisbane, QLD 4072, Australia

Supporting Information

**ABSTRACT:** Grain protein composition determines quality traits, such as value for food, feedstock, and biomaterials uses. The major storage proteins in sorghum are the prolamins, known as kafirins. Located primarily on the periphery of the protein bodies surrounding starch, cysteine-rich $\beta$- and $\gamma$-kafirins may limit enzymatic access to internally positioned $\alpha$-kafirins and starch. An integrated approach was used to characterize sorghum with allelic variation at the kafirin loci to determine the effects of this genetic diversity on protein expression. Reversed-phase high performance liquid chromatography and lab-on-a-chip analysis showed reductions in alcohol-soluble protein in $\beta$-kafirin null lines. Gel-based separation and liquid chromatography−tandem mass spectrometry identified a range of redox active proteins affecting storage protein biochemistry. Thioredoxin, involved in the processing of proteins at germination, has reported impacts on grain digestibility and was differentially expressed across genotypes. Thus, redox states of endosperm proteins, of which kafirins are a subset, could affect quality traits in addition to the expression of proteins.

**KEYWORDS:** *seed storage proteins, sorghum, kafirin, mass spectrometry, HPLC, digestibility*

## INTRODUCTION

Improving agricultural productivity is imperative to feeding an expanding global population set to peak at 9 billion by the year 2050.[1] Diminishing soil quality and increasing demand for limited water reserves, coupled with unpredictable weather patterns associated with climate change, are placing mounting pressure on farming systems. Growers are being prompted to adopt more nutrient efficient crop alternatives[2] and to invest in improving drought resistance through conventional and genomics assisted breeding.[3] Sorghum produces grain with greater water and nutrient use efficiency, yielding up to 33% more biomass per unit water than maize, a close crop relative.[4,5] However, grain digestibility and nutritional value is less optimal in sorghum compared to maize, due in part to the highly cross-linked nature of cysteine-rich storage proteins in the endosperm.[6]

The commercial value of cereals is largely determined by the physiochemistry of the endosperm and the composition and interactions of protein and starch present there. Soft endosperm sorghum varieties are more digestible, but may exhibit increased susceptibility to environmental stresses, including molds and drought.[7] Considerable variation occurs in endosperm protein content across the cereals. Maize, sorghum, and millet, of the subfamily Panicoideae, contain around 80% prolamin, whereas rice and oats, of the subfamily Festucoideae, contain greater proportions of albumin and globulin. Barley, wheat and rye, classified in the same subfamily as rice and oats, but grouped in the tribe Triticeae, exhibit greater similarity in protein composition to the Panicoideae, but with a higher albumin/globulin (A/G) content.[8]

The kafirins are classified into groups according to various properties, including molecular weight, solubility, structure, and amino acid composition. The $\beta$-kafirins (18 kDa), $\delta$-kafirins (13 kDa), and $\gamma$-kafirins (28 kDa) are rich in cysteine and methionine. Cysteine enrichment contributes to the intra- and interconnectivity of the protein matrix surrounding starch.[9] The $\alpha$-kafirins (22−26 kDa)[10] are rich in nonpolar amino acids and do not cross-link as extensively, forming mainly intramolecular disulfide bonds. There are 23 members of the $\alpha$-kafirin family. In maize, 42 members of the homologous 19−22 kDa $\alpha$-zein family have been identified, including the wild-type allele for the *floury2* mutation.[11,12] Although extensive sequence homology and functional similarity exists between sorghum and maize storage

Received: May 19, 2014
Revised: August 26, 2014
Accepted: September 1, 2014
Published: September 1, 2014





dx.doi.org/10.1021/jf5022847 | J. Agric. Food Chem. 2014, 62, 9819−9831

proteins, sorghum protein bodies are more highly cross-linked, leading to increased levels of covalent bonding and protein polymerization in the matrix.[8,9] This increased level of cross-linking reduces overall protein digestibility, which compounds nutritional deficiencies of sorghum, such as low inherent levels of lysine and threonine, and often needs to be supplemented in food and feed.[13]

Identification and characterization of endosperm proteins and the elements regulating their synthesis, localization, and degradation facilitates enhancement of amino acid content and starch accessibility for improved grain nutritional value.[14] This study involves the evaluation of grain protein composition in sorghum lines with allelic variation in kafirin storage proteins using a range of proteomic tools, including reversed-phase HPLC (RP-HPLC), lab on a chip (LOC), SDS-PAGE, and liquid chromatography−tandem mass spectrometry (LC−MS/MS). The focus of the work is to identify candidate proteins, which may impact on quality traits, particularly those which are differentially expressed across the genotypes.

## MATERIALS AND METHODS

**Plant Material.** Twenty-eight inbred sorghum lines, with previously determined allelic variation in the $\beta$-, $\gamma$-, and $\delta$-kafirin storage proteins, were included in the study.[15] The primary focus of the analysis was on comparisons between the $\beta$-kafirin null line QL12 and 296B, an important food line of Indian origin. Plants were cultivated under field conditions at the University of Queensland, Gatton campus, QLD, Australia, over the 2008/2009 and 2011/2012 summer growing seasons.

**Protein Extractions.** *Albumin/Globulin (A/G).* Wholegrain flour (100 mg) prepared from kernels ground in liquid nitrogen using a mortar and pestle was mixed in 1 mL of extraction solvent (50 mM Tris·HCl pH 7.8, 100 mM KCl, and 5 mM EDTA). The sample was vortexed and centrifuged, and a 500 $\mu$L aliquot of supernatant was removed to a new tube.[16] The process was repeated by adding an additional 1 mL of extraction solvent to the pellet, repeating the extraction process and removing an additional 500 $\mu$L of supernatant to the same tube.

*Prolamin.* The A/G pellet from the above extraction procedure was retained and washed with 1 mL of distilled water ($dH_2O$), and 1 mL of solvent was added (60% *tert*-butyl alcohol, 0.5% sodium acetate w/v, and 2% $\beta$-mercaptoethanol ($\beta$-ME) v/v). The pellet was then vortexed and centrifuged, and 500 $\mu$L of supernatant was transferred to a new tube. The procedure was repeated once, and a further 500 $\mu$L of supernatant was removed to the same tube.

*Glutelin/Residual Proteins.* The pellet from the prolamin extraction was washed with 1 mL of $dH_2O$ and 1 mL of sodium borate buffer (125 mM pH 10.0 containing 1% SDS w/v and 1% $\beta$-ME v/v) was added to the pellet. The suspension was vortexed and centrifuged, and 500 $\mu$L of supernatant was transferred to a new tube. The procedure was repeated once, and a further 500 $\mu$L of supernatant was removed to the same tube.

**Reversed-Phase HPLC (RP-HPLC).** Protein samples were analyzed using an Agilent 1100 HPLC system (Agilent, Foster City, CA), fitted with a Poroshell column of varying stationary phases (as indicated below for the various protein fractions). The system employed a binary gradient with a constant flow rate of 0.7 mL/min, and a column temperature was maintained at 55 °C. Solvents for RP-HPLC included water containing 0.1% trifluoroacetic acid (TFA) w/v (A), and acetonitrile (ACN) containing 0.07% TFA w/v (B), with gradient flow specifications as follows: 0−18 min, 45%−60% B; 18−19 min, decreased to 45% B; followed by a 7 min post run. Absorbance was measured with a UV detector at 214 nm.

*Albumin/Globulins.* After extraction, 250 $\mu$L A/G aliquots were freeze-dried in a SpeedVac overnight. Prior to RP-HPLC analysis, the aliquots were resuspended in 100 $\mu$L of 50% ethylene glycol to stabilize proteins.[17] Samples (5 $\mu$L) were analyzed using a 2.1 × 75 mm Poroshell 300 SB C3 column.

*Prolamins.* After extraction, protein samples (1 mL) were alkylated by adding 33 $\mu$L of 4-vinylpyridine and vortexed for 10 min. Samples (5 $\mu$L) were then injected directly for RP-HPLC analysis using 2.1 × 75 mm Poroshell 300 SB C18 column.[16] Proteins were detected by UV at 214 nm.

**Lab on a Chip (LOC).** The lab-on-a-chip procedure was carried out on an Agilent 2100 Bioanalyser using the Protein 80 assay kit (Agilent Technologies, Palo Alto, CA). A 1:1 mixture of protein to denaturing buffer was used.

**Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE).** *Protein Sample Cleanup.* Samples were cleaned and concentrated with a 2-D Clean-Up Kit (GE Healthcare Ltd.). Concentrations were measured using a 2-D Quant Kit (GE Healthcare Ltd.) according to the manufacturer's instructions.

*One-Dimensional (1D) SDS-PAGE.* A 10 $\mu$L sample was resuspended in 5 $\mu$L of 3X loading buffer (100 mM Tris, 1% SDS (w/v), 0.01% bromophenol blue, 15% glycerol, 0.05% $\beta$-ME), and heated to 95 °C, and then chilled on ice and loaded onto a small format precast anyKd (10−250 kDa range) gradient gel (Bio-Rad, Hercules, CA), with 15 $\mu$L of ColorPlus prestained marker (broad range 7−175 kDa) added to the first and last wells. Gels were run in SDS buffer (25 mM Tris, 200 mM glycine, 0.1% SDS w/v) at 200 V for 40 min until the loading dye front reached the bottom of the gel. Gels were fixed in 50% MeOH, 7% acetic acid for 1 h with shaking, and then stained in ∼50 mL of Coomassie stain overnight. Gels were destained in 1% acetic acid for 30 min, and then rinsed with $dH_2O$ and scanned.

*Two-Dimensional (2D) Isoelectric focusing.* A rehydration solution containing 100 $\mu$g of protein was prepared to a total volume of 125 $\mu$L (20 $\mu$L of protein [concentration 5 $\mu$g/$\mu$L], 100 $\mu$L of rehydration buffer [7 M urea, 2 M thiourea, 2% CHAPS, 0.5% (v/v) IPG buffer, 0.0002% bromophenol blue], 4 $\mu$L of fresh 1 M dithiothreitol (DTT), 1.25 $\mu$L of carrier ampholytes [pH 3−11]). The sample mixture was loaded onto 7 cm IPG strip (3−11 nonlinear gradient (NL) small strips), placed in a carrier coffin, and 0.8 mL of dry strip cover fluid was applied to minimize evaporation and urea crystallization. IEF was run for a total of ∼16 000 V h at 20 °C on the IPGphor with the following program settings: 30 V for 14 h (rehydration), 100 V for 2 h (Step-n-hold), 500 V for 1.5 h (Step-n-hold), 1000 V for 1.5 h (Step-n-hold), 5000 V for 1 h (gradient), 5000 V for 2 h (Step-n-hold), and 50 V (hold). It was ensured that the voltage reached at least 5000 V to achieve a standard protein gradient. Strips were removed from the coffin and either used immediately for 2D SDS-PAGE gel analysis or stored at −20 °C until required.

*Two-Dimensional SDS-PAGE.* IPG strips were saturated with MES running buffer [50 mM 2-morpholinoethanesulfonate (MES), 50 mM Tris base, 0.01% SDS w/v, 1 mM EDTA], 6 M urea, 30% glycerol, 2% SDS w/v, and 0.001% bromophenol blue by first equilibrating in MES buffer plus 50 mg/mL (DTT) for 15 min with shaking to reduce disulfide bonds, and then soaking strips in MES buffer plus 125 mg/mL iodoacetamide (IAA) for 15 min with shaking to alkylate the free cysteine residues. IPG strips were fitted into Bis/Tris small format precast gels (15%) (Invitrogen, Carlsbad, CA). The loading well was then sealed with ∼0.8 mL agarose sealing solution (0.5% agarose, 0.001% bromophenol blue). The gel was run in 1X MES running buffer at low voltage (25−50 V) for 30 min, and then the voltage was increased to 125 V and run until the loading dye front disappeared off the bottom of the gel (approximately 1.5−2 h). Gels were fixed in 50% MeOH, 7% acetic acid for 1 h with shaking, and then transferred to a Sypro Ruby silver stain solution overnight with shaking and destained for 30 min (10% MeOH, 7% acetic acid) prior to scanning using a Typhoon 9400 (GE Healthcare). The gel was then stained in Coomassie overnight with shaking to visualize spots for excision from gel and subsequent mass spectrometric analysis. Coomassie stained gels were destained in 1% acetic acid and scanned with Odyssey at 700 nm.

**Mass Fingerprinting (LC−MS/MS).** *Reduction/Alkylation of Gel Pieces.* Protein spots were manually excised from gels and destained in 50 mM ammonium bicarbonate (ABC)/50% acetonitrile (ACN), 2 × 200 $\mu$L, with shaking for 3−4 h or overnight. For 1D gel pieces, buffer/destain was removed from gel pieces with pipetting and the sample was soaked in 40 $\mu$L of 10 mM DTT to reduce cysteine residues. Gel pieces were incubated at 60 °C for 30 min, DTT solution was removed, and 40 $\mu$L of 55 mM fresh IAA was added. Samples were then incubated at room temperature for 30 min in the dark. IAA was discarded and (for 1D

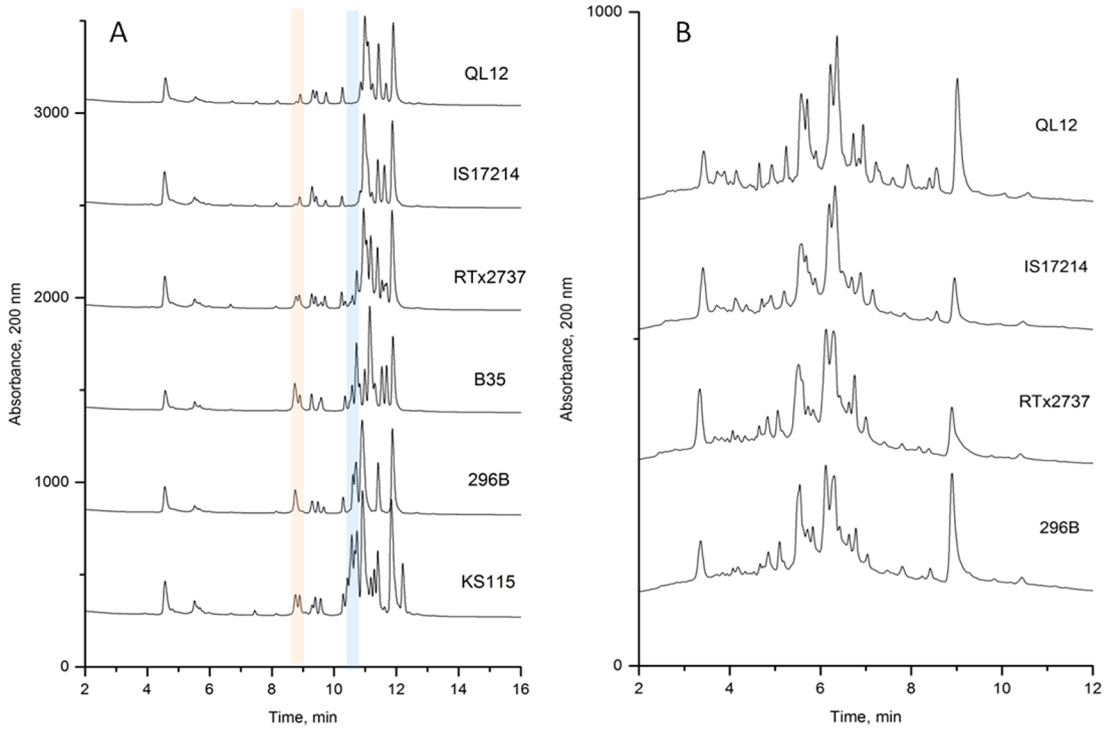


**Figure 1.** RP-HPLC profiles for the (A) alcohol-soluble (prolamin) fraction and (B) water/salt-soluble (albumin/globulin) fraction across $\beta$-kafirin null allelic variants QL12, 1517214, and RTx2737, and lines with normal $\beta$-kafirin content, including B35, 296B, and KS115. In the prolamin fraction $\beta$-kafirin null lines showed a reduction in the size of a peak eluted at 9 min (outlined in orange). Peak distribution profiles at 10−11 min elution times were similar across lines with normal $\beta$-kafirin content, while a significant proportion of protein in this area of the chromatogram was missing in the $\beta$-kafirin null mutants (outlined in blue).

and 2D gel pieces), 100 $\mu$L of 50 mM ammonium bicarbonate (ABC) was added, with vortexing/shaking for 1−2 min, removal of solution, and repeated washing. Gel pieces were then dehydrated in 50 $\mu$L of 100% ACN for 5 min, with vortexing.

*Enzymatic Digestion and Peptide Extraction.* Gel pieces were rehydrated in a 50 mM ABC solution containing 8 $\mu$L of trypsin (10 ng/$\mu$L) (sequencing-grade modified trypsin, Promega, Madison, WI) for 10−20 min at 4 °C. An additional volume (6−16 $\mu$L) of 50 mM ABC buffer was then added, depending on the size of the gel piece, and samples were incubated overnight at 37 °C. Then, 50 $\mu$L of 50% ACN/0.1% TFA was added to of gel pieces and samples were sonicated in a water bath for 10 min, and centrifuged briefly, and supernatant was transferred to a new tube. An additional 50 $\mu$L aliquot of 50% ACN/0.1% TFA was added, sonication was repeated, and supernatant was combined with the first extract. Supernatant was lyophilized in a SpeedVac at 45 °C, and peptides were resuspended in 10 $\mu$L of 5% ACN/0.1% TFA. A Ziptip cleanup was carried out according to the manufacturer's instructions (Millipore) for removal of acrylamide contamination prior to MS analysis.

*LC−MS/MS.* Digested peptides were run on LC−MS/MS using an ESI-QTOF instrument. MS parameters were similar to those in Kappler and Nouwens,[18] with the following modifications: Samples were desalted for 5 min on an Agilent C18 trap (0.3 × 5 mm, 5 $\mu$m), followed by separation on a Vydac Everest C18 column (300 A, 5 $\mu$m, 150 mm × 150 $\mu$m) at a flow rate of 1 $\mu$L/min, using a gradient of 10−60% buffer B over 30 min, where buffer A = 1% ACN/0.1% formic acid and buffer B = 80% ACN/0.1% formic acid. Eluted peptides were directly analyzed on a TripleTof 5600 mass spectrometer (ABSciex) using a Nanospray III interface. Gas and voltage were set as required. MS TOF scan across $m/z$ 350−1800 was performed for 0.5 s, followed by data-dependent acquisition of 20 peptides with intensity above 100 counts across $m/z$ 40−1800 (0.05 s per spectrum) with rolling collision energy. MS data were converted to Mascot generic format and submitted to Mascot.

*MS Data Analysis.* The automated Mascot search engine (Matrix Science, London, U.K.) was used to identify best matched protein sequences to the peptides detected by MS. Trypsin was specified as the proteolytic enzyme, and carbamidomethyl (C) of cysteine and oxidation (M) of methionine residues were taken into account. Charged states of 2+, 3+, and 4+ were considered for parent ions. Using a Ludwig-based search of plant species, a profile of best matched protein(s) was generated employing an algorithm to rank the proteins identified based on their peptide mass fingerprints. Individual ion scores were calculated as $-10 \times 10g(P)$, where $P$ is the probability that the observed match is a random event. Peptides identified with an ion score of >40 were considered to indicate identity or extensive similarity. Each of the peptide identifications was manually inspected and verified to ensure that the spectra were composed of a wide series of intense fragments, which could be designated as major fragments (b or y) of the proposed peptide. Each protein match also generated a protein score in the MS/MS search report, which is the sum of the highest ions scores for each distinct sequence. The exponentially Modified Protein Abundance Index (emPAI) provided an approximate, relative quantification of the proteins present in the sample. Functionally characterized homologues to putative uncharacterized sorghum proteins identified with MS were identified through BLAST searches based on FASTA sequences derived through queries to the Uniprot database.

**Statistical Analysis.** Absorbance values from RP-HPLC of alcohol-soluble proteins were interpolated by a spline method, and absorbance area values were calculated to 0.01 min intervals using MATLAB (release 2012a, The MathWorks, Natick, MA). Linear correlation coefficients were calculated between an individual absorbance area and digestibility and shown as a continuous spectrum over RP-HPLC retention time.

## RESULTS AND DISCUSSION

**RP-HPLC Profiling of Seed Proteins.** Profiling of alcohol-soluble prolamin (Figure 1A and Supporting Information, Figure S1) and water/salt-soluble albumin/globulin (A/G) (Figure 1B

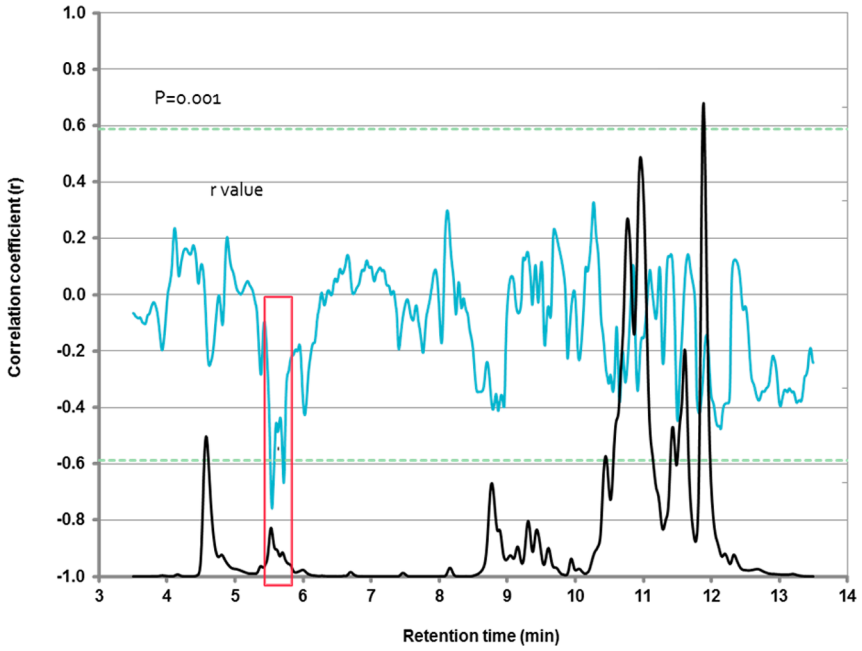


**Figure 2.** RP-HPLC analysis showing a significant negative correlation between protein digestibility (%) and an alcohol-soluble kafirin protein peak eluting at 5.5−6 min on the chromatogram (outlined in red). Analysis was carried out across 26 diverse sorghum lines (listed in Table 1), including the genotypes evaluated for ethanol conversion by Cremer et al.[20]

**Table 1. Total Protein, % Protein Digestibility, and Total Starch Measurements for Sorghum Grain Lines with Diverse Kafirin Allelic Backgrounds**

| | | | | uncooked | | cooked | | | |
|---|---|---|---|---|---|---|---|---|---|
| genotype/line | $\beta$ | $\gamma$ | $\delta$ | total protein (%) | digestibility (%) | total protein (%) | digestibility (%) | starch (%) | moisture |
| RTx7000/FF-RTx7000 | 1 | 2 | 1 | 12.5 | 51.5 | 14.3 | 21.2 | 70.1 | 8.1 |
| IS22457C | 2 | 1 | 1 | 11.9 | 60.2 | 13.4 | 24.0 | 54.9 | 7.3 |
| IS8525 | 1 | 4 | 2 | 13.9 | 18.1 | 15.1 | 7.6 | 62.2 | 6.6 |
| QL12 | 3 | 1 | 2 | 14.6 | 58.4 | 15.5 | 25.3 | 56.3 | 7.9 |
| IS12572C/FF-6214E | 1 | 1 | 2 | 10.6 | 30.8 | 14.0 | 26.0 | 62.9 | 10.4 |
| Ai4 | 1 | 3 | 2 | 12.7 | 45.1 | 13.8 | 21.8 | 62.5 | 10.5 |
| BTx3197 | 1 | 2 | 2 | 12.8 | 40.5 | 15.4 | 25.4 | 66.5 | 10.3 |
| Hegari/Early Hegari | 4 | 1 | 1 | 11.8 | 54.4 | 15.2 | 29.0 | 71.0 | 7.5 |
| ISCV745 | 1 | 1 | 1 | 10.3 | 60.2 | 12.1 | 32.0 | 70.4 | 10.7 |
| B923296 | 1 | 1 | 2 | 11.9 | 61.2 | 13.3 | 22.3 | 64.0 | 7.6 |
| ISCV400 | 4 | 1 | 1 | 11.9 | 56.9 | 13.6 | 29.6 | 68.1 | 7.6 |
| QL39 | 2 | 1 | 1 | 11.0 | 40.1 | 13.3 | 17.0 | 67.3 | 9.5 |
| Karper 669 | 1 | 1 | 1 | 12.6 | 55.3 | 14.0 | 25.6 | 67.3 | 9.7 |
| QL41 | 1 | 2 | 2 | 10.7 | 51.6 | 12.5 | 22.7 | 68.0 | 9.7 |
| R931945-2-2 | 1 | 2 | 1 | 11.6 | 62.3 | 12.9 | 17.3 | 66.8 | 6.1 |
| R89562/R890562-1-1 | 1 | 1 | 1 | 13.2 | 46.9 | 14.9 | 18.7 | 67.3 | 6.6 |
| M35-1/FF-BM351 | 1 | 3 | 2 | 12.9 | 48.7 | 14.3 | 15.8 | 67.1 | 9.2 |
| KS115 | 1 | 4 | 2 | 16.5 | 40.8 | 18.1 | 16.8 | 64.1 | 9.6 |
| IS17214 | 3 | 1 | 2 | 11.4 | 59.2 | 13.5 | 33.8 | 61.9 | 10.8 |
| SC1270-6-8/FF-SC1270-6-8 | 2 | 1 | 1 | 12.0 | 59.0 | 14.0 | 17.3 | 62.5 | 9.6 |
| LR91918 | 2 | 3 | 2 | 10.3 | 56.0 | 12.6 | 26.4 | 63.1 | 9.4 |
| IS12611C-F-F_ SC11114E | 2 | 1 | 1 | 11.3 | 54.0 | 13.8 | 22.5 | 64.7 | 8.1 |
| R9733 | 4 | 2 | 1 | 14.6 | 59.6 | 16.8 | 14.7 | 66.8 | 5.8 |
| BOK11/FF-BPK11 | 1 | 2 | 2 | 11.7 | 65.0 | 13.5 | 15.9 | 60.5 | 7.2 |
| BTx623/FF-BTx623 | 2 | 2 | 1 | 12.9 | 50.8 | 15.3 | 15.3 | 64.9 | 9.9 |
| B35/FF-B35 | 1 | 2 | 1 | 11.7 | 51.4 | 13.4 | 19.2 | 63.0 | 8.8 |
| 296B/FF-296B | 1 | 3 | 1 | 10.0 | 68.2 | 12.8 | 28.6 | 58.2 | 10.8 |
| RTx2737/FF-BTx2737 | 3 | 1 | 1 | 13.3 | 58.1 | 15.2 | 19.0 | 59.3 | 7.9 |

and Supporting Information, Figure S2) protein fractions across 28 sorghum lines using RP-HPLC revealed a range of diversity, particularly in the alcohol-soluble kafirin-containing fraction. Variation observed across wild-type and mutant lines could


dx.doi.org/10.1021/jf5022847 | J. Agric. Food Chem. 2014, 62, 9819−9831

contribute to phenotypic differences, such as increased flour viscosity and fermentation efficiency.[19,20] Among the β-kafirin null lines QL12, IS17214, and RTx2737, a high degree of similarity was observed in the alcohol-soluble peak distributions compared to lines with functional β-kafirin alleles. An approximately 50% reduction in the height of a peak eluting at 9 min was observed in β-kafirin null lines compared to wild type (Figure 1A, highlighted in orange), indicating a likely elution profile for β-kafirin (under the analysis conditions described here) and presenting a potential marker for the protein. Alternately, the β-kafirin peak could have been masked by a series of larger peaks eluting from the column at 10−11 min, which were absent in null lines (Figure 1A, highlighted in blue). Further work to positively identify the β-kafirin in RP-HPLC separations under these exact conditions is needed. RP-HPLC analysis of the water/salt-soluble fraction showed less similarity among peak profiles for the β-kafirin null mutants relative to other lines. However, substantial quantitative differences in peak heights were observed among the genotypes and there was a high level of variability across the sample set in a peak eluting at 9 min, which was unrelated to fluctuations in total sample protein concentration.

Interestingly, a statistically significant negative correlation was identified between a set of protein peaks eluted in the kafirin fraction and protein digestibility across sorghum grain lines, providing evidence for links between seed protein composition and end-use traits (Figure 2). Further research is needed to identify the protein(s) present in this peak; however, given the sequential extraction procedure utilized in this work, it is likely that this is a member of the kafirin family.

**Lab-on-a-Chip (LOC) Analysis of Seed Proteins.** LOC analysis, employing microfluidic chip−capillary electrophoresis, generated size-separated protein profiles for alcohol-soluble (prolamin) (Figure 3) and water/salt-soluble (A/G) fractions (Supporting Information, Figure S3). The prolamin fraction showed a large set of peaks present in the 22−26 kDa size range, representing the previously characterized α- and γ-kafirins. Small peaks were also visible at 11 and 19 kDa, with the 19 kDa peak representing the β-kafirins and the 11 kDa peak potentially representing the δ-kafirins. The β-kafirin peak was diminished in size in β-kafirin nulls, but, interestingly, a peak or set of peaks was visible at 19−20 kDa across all lines. This indicates that an additional protein of a similar size to β-kafirin may be present in the peak. This protein could represent a different protein, such as peptidyl-prolyl cis/trans isomerase (PPIase), identified in the same band/spot as β-kafirin through SDS-PAGE and LC−MS/MS, as outlined in the section SDS-PAGE and LC−MS/MS Protein Profiling. Chip profiles indicate that several cultivars in addition to the β-kafirin null mutants contain relatively low levels of β-kafirin, including 296B, a highly digestible line.[20] LOC analysis also revealed the presence of a small peak visible at 46 kDa in alcohol-soluble samples, which could represent a high molecular weight kafirin or kafirin dimer. In subsequent analysis, a protein spot extracted from the corresponding size range on 2D gels was identified through LC−MS/MS as a ∼36 kDa γ-prolamin homologue (Supporting Information, Figure S4 and Supplementary Table 1), thus providing sequence information for this kafirin entity at the protein level. As observed with RP-HPLC, the LOC profiling of the water/salt-soluble fraction shows that QL12 contains a relatively high content and diversity of A/G's compared to other genotypes (Supporting Information, Figure S3). High-lysine sorghum varieties, such as P721, exhibit an elevated A/G content, with increased nutritional value.[21] This indicates that QL12, among other lines, may contain increased levels of lysine and other essential amino acids due to a high A/G content.

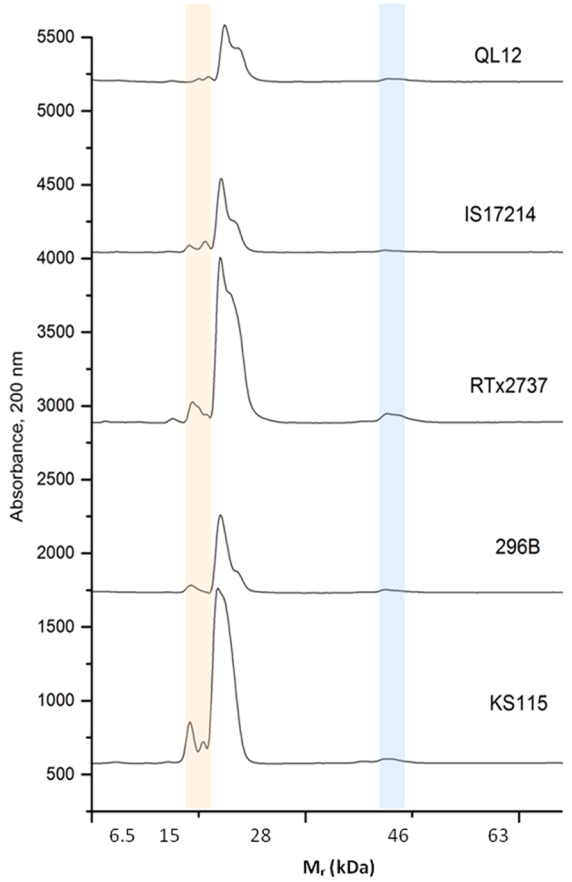


**Figure 3.** Alcohol-soluble proteins visualized on lab on a chip (LOC) across β-kafirin null lines QL12, 1517214, and RTx2737, and normal β-kafirin lines 296B and M35. A 19 kDa β-kafirin peak (highlighted in orange) precedes a set of large 22−25 kDa α- and γ-kafirin peaks, and a HMW prolamin is visible across the genotypes as a small peak at ∼50 kDa (highlighted in blue).

**SDS-PAGE and LC−MS/MS Protein Profiling.** Profiling of water/salt-soluble (A/G), alcohol-soluble (prolamin), and alkali-soluble (glutelin) protein fractions across the β-kafirin null mutant QL12 and wild-type line 296B using SDS-PAGE gel separation methods, coupled with mass spectrometry (LC−MS/MS) (Figure 4 and Supporting Information, Figure S4 and Supplementary Table 1), identified a range of proteins involved in storage protein biochemistry and starch metabolism. One-dimensional gel separation of each of the protein fractions by size (Figure 4) showed that the A/G fraction contains a highly concentrated, complex array of proteins present across a broad size range, similar to profiles generated in previous studies.[22] The prolamin fraction was less complex and more concentrated within the 23−26 kDa size range, which contains the more abundant α-kafirin storage proteins. The alkali-soluble glutelin fraction exhibited some similarities to the prolamin fraction, likely due to incomplete extraction of the prolamin fraction, but contained a wider diversity of protein bands visible across a greater size range, which was also verified on 2D gels, and has been reported in previous studies.[23,24] It is hypothesized that the glutelin proteins are more diverse because they both play a role in connecting the protein:starch matrix and provide a source of


dx.doi.org/10.1021/jf5022847 | J. Agric. Food Chem. 2014, 62, 9819−9831

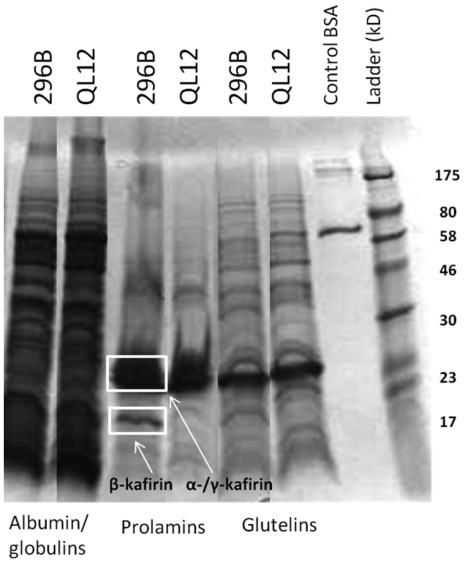


**Figure 4.** One-dimensional SDS-PAGE separation of water/salt-soluble (albumin/globulins), alcohol-soluble (prolamins), and alkali-soluble (glutelins) from normal $\beta$-kafirin allelic variant 296B, and $\beta$-kafirin null QL12. Excised protein bands (outlined in white) were digested with trypsin and analyzed using LC−MS/MS (results listed in Supporting Information, Supplementary Table 1). Ladder: P7709V ColorPlus Prestained Protein Marker, Broad Range (7−175 kDa) Mw ladder.

hydrolytic enzymes for the breakdown of starch and protein during germination.[25]

**Characterization of Kafirin Subclasses and the $\beta$-Kafirin Null Mutation.** Each kafirin subclass in the alcohol-soluble fraction was isolated using SDS-PAGE and identified through LC−MS/MS. $\alpha$- and $\gamma$-kafirins were visualized within the 22−25 kDa size range on the 1D SDS-PAGE gel (Figure 4, outlined in white). Using 2D SDS-PAGE separation of proteins by both size and charge, the $\delta$-kafirin protein was isolated in the prolamin fraction for genotype 296B (spot 11) (Supporting Information, Figure S4 and Supplementary Table 1). Matches to $\alpha$- and $\gamma$-kafirins were also generated on 2D gels in prolamin gel spots 8 and 11 for genotype QL12, and with spot 9 returning $\alpha$-kafirin as a top hit. Differential expression of the $\beta$-kafirin protein was detected in the alcohol-soluble fraction across mutant and wild-type germplasm. This provided a positive control for the MS technique and allowed for further characterization of the mutant at the protein level. One-dimensional separation of the prolamin fraction revealed altered expression of $\beta$-kafirin in a 19 kDa protein band, present in 296B and absent in QL12 (Figure 4, outlined in white), as predicted based on previous molecular characterization.[15] On 2D gels differential expression of $\beta$-kafirin was identified in 296B prolamin spot 11 and glutelin spot 4 (Supporting Information, Figure S4 and Supplementary Table 1). Identification of $\beta$-kafirin in the glutelin fraction (296B spot 4) may reflect the incomplete sequential extraction of alcohol-soluble proteins in the previous step.[26] $\beta$-Kafirin was completely absent in QL12, with the exception of a weak hit in the 2D prolamin spot 1, which had a low Mascot score achieved across two peptide matches and was discarded due to poor-quality spectral data. This match may represent a truncated form of the protein, as it was identified in a smaller size range on the gel. Although protein expression was generally similar across the genotypes within the range of spots analyzed on 2D gels, differential expression of a low molecular weight thioredoxin (Trx) enzyme was identified in the water/salt-soluble fraction, where spot 4 was present in 296B and absent in QL12. This was the only Trx identified across the protein fractions analyzed in the study, with the finding replicated in triplicate (Figure 5 and Supporting Information , Supplementary Table 1). Trx catalyzed the conversion of seed proteins from the oxidized to the reduced state during germination, with significant impacts on protein digestibility and grain nutrition.

**Sorghum Homologue to 50 kDa $\gamma$-Prolamin.** High molecular weight (HMW) $\gamma$-prolamins (~50 kDa) have been identified and characterized across a range of plant species,

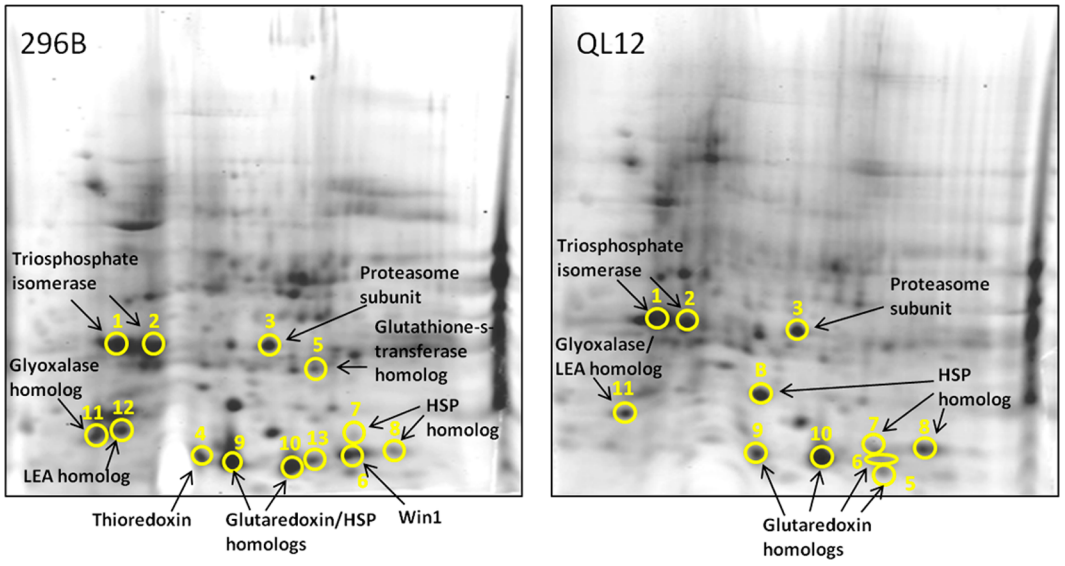


**Figure 5.** Two-dimensional SDS-PAGE analysis of the water/salt-soluble protein fraction in $\beta$-kafirin null allelic variant compared with wild-type 296B. Protein samples were loaded onto 7 cm IPG strips (3−11 NL) and run on IPGphor machine for isoelectric focusing. SDS-PAGE gels (4−12% Bis/Tris small format precast) were utilized for separation of proteins by size (Mw). Protein spots (circled in yellow) were excised across a range of size and pI, digested with trypsin, and identified with LC−MS/MS. Spot 4 was identified as a differentially expressed thioredoxin, present in 296B and absent in QL12.


dx.doi.org/10.1021/jf5022847 | J. Agric. Food Chem. 2014, 62, 9819−9831

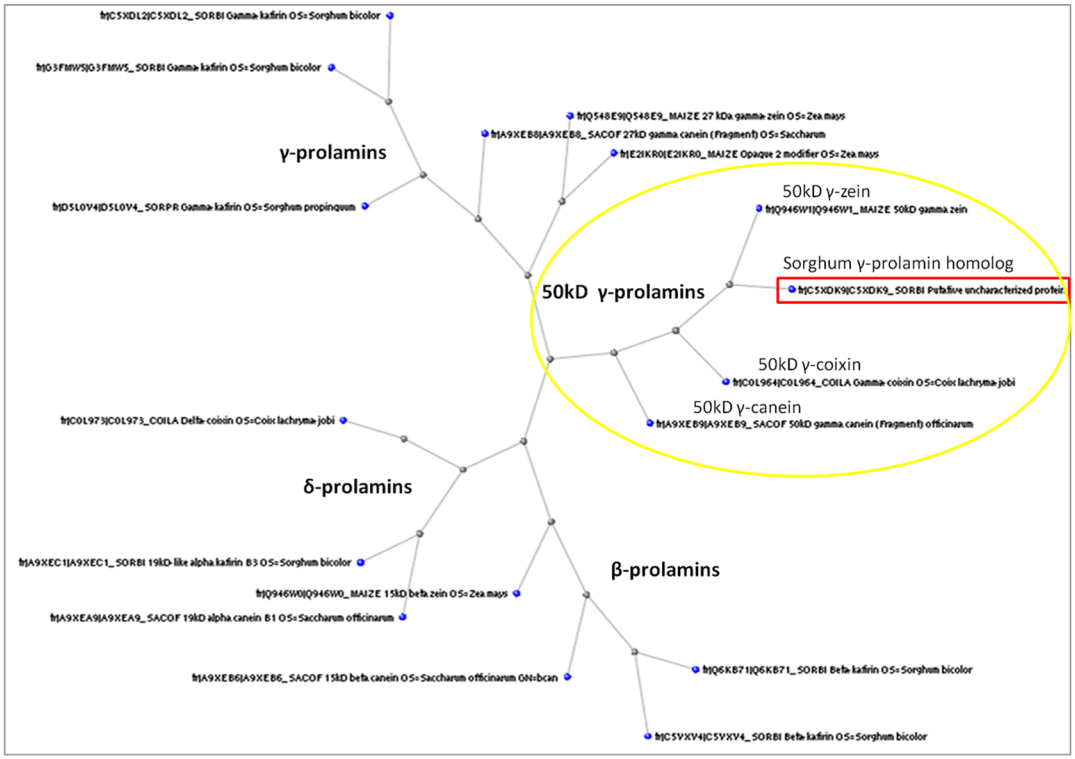


**Figure 6.** Phylogenetic neighbor joining tree view of the alignment of protein sequences representing prolamin subclasses across grass species using the Grishin protein method. Uncharacterized sorghum γ-prolamin homologue groups with 50 kDa γ-zein and γ-canein of maize and sugarcane, and groups away from the LMW γ-prolamins.

including maize and wheat.[27,28] Previous studies on sorghum have identified a protein band at ∼45 kDa in the alcohol-soluble fraction using 1D SDS-PAGE. However, this protein has not yet been characterized at the genetic level.[15] The HMW protein band was previously reported as a kafirin dimer because it diminishes in intensity upon treatment with increasing concentrations of reducing agent, indicating that it can be broken down into smaller peptides.[8,29] However, because the original 45 kDa band is still intact on reduced gels, it has also been suggested that the protein may be linked by bonds that are not broken by reducing agents.[30] In the present study, this HMW protein was visible in reduced alcohol-soluble protein samples from both genotypes QL12 and 296B, on 1D and 2D gels. LC−MS/MS data indicate that this protein represents a homologue to HMW γ-prolamin identified in closely related species (Supporting Information, Figure S4 and Supplementary Table 1). Across both genotypes, 2D prolamin spot 4 returned a match for a putative uncharacterized sorghum protein (accession number C5XDK9), exhibiting homology to 50 kDa γ-zein, γ-canein, and γ-coixin (Figure 6). The sorghum γ-prolamin homologue is larger than any previously characterized kafirin, with a calculated Mw 36 615 Da. LOC analysis of the prolamin fraction revealed the presence of two small peaks at ∼44 and 46 kDa across the majority of the sample collection (Figure 3, highlighted in blue), which may correspond to two genetic variants of the HMW γ-prolamin.[8,27] In a previous study using LOC, similar peak profiles were also observed for the HMW prolamin peak in reduced alcohol-soluble protein samples extracted from three grain genotypes (QL12, QL41, and 296B) from the kafirin allelic variant sample set, which had been harvested in a previous growing season.[31] Across our data set, quantitative peak area values were attained through LOC for the peaks in the 44−46 kDa size range and these data were aligned with readings for grain quality parameters. The general structure and distribution of prolamins in the protein bodies has been shown to be uniform across maize and sorghum. In maize, 50 kDa γ-zeins localize to the periphery of the protein bodies, similar to other γ-zeins.[27] Immunocytochemistry shows that γ-kafirins also localize to the periphery of protein bodies, and may prevent proteolysis of internally positioned α-kafirins.[32] Sorghum lines with increased digestibility exhibit a change in the protein body structure from spherical to invaginated, with the γ-kafirins located from the periphery to the folds of the structure, resulting in an increased exposure of the α-kafirins to proteolytic breakdown.[33] In maize, a clear physical distribution of zein classes is observed within the seed. Protein bodies in the subaleurone layer are smaller and contain mainly β- and γ-zeins, while those encapsulating starch in the inner endosperm are larger and contain continuous central regions of α-zeins with β- and γ-zeins located on the periphery.[34] Construction of mutant γ-zein proteins has pinpointed the proline-rich N-terminal domain as being critical in wild-type protein body development.[35] Furthermore, deletion of a cysteine-rich γ-zein domain results in abnormally structured protein bodies. The different kafirin classes exhibit variable solubilities according to their degree of polymerization and cross-linking.[29] Interactions between β- and γ-kafirins on the periphery of protein bodies may limit enzyme accessibility to α-kafirin, impeding digestion of protein and starch encased in the matrix. This study presents evidence for the seed-specific expression of a high molecular weight γ-prolamin homologue in sorghum. Characterization of this protein in sorghum could enhance efforts to further distinguish the roles of the different classes of kafirins in maintaining endosperm connectivity and access to starch.

**Identification of Proteins with Potential Impacts on Grain Quality.** Profiling of water/salt-, alcohol-, and alkali-soluble proteins across the sample population of kafirin allelic

variants using mass spectrometry identified proteins with potential various impacts on the structure of the protein−starch matrix, such as enzymes functioning in protein cross-linking and the mobilization of starch during germination[36,37] (Table 2, Figure 5−7, and Supporting Information, Figure S4 and Supplementary Table 1). Identification and localization of these proteins in the mature grain indicates they may have an impact on endosperm development and/or germination. Deciphering the precise activities of these proteins in sorghum will involve additional analysis of the transcriptome and proteome throughout grain development.

*Thioredoxins (Trx) and Glutaredoxins (Grx).* Thioredoxins are small proteins containing a site with a redox-active disulfide, which functions in the reversible oxidation of protein SH groups to a disulfide bridge[38] (Table 2). Differential expression of a sorghum Trx protein, C5XB72, homologous to maize Trx B6SX54, was observed in the A/G fraction, with 2D protein spot 4 present in 296B and absent in QL12 (Figure 5). Matches to sorghum Trx had a high emPAI protein quantification score, pI 5.79, and Mw 13 061 Da. Trx labeling studies have shown in vitro and in vivo that the enzyme catalyzes the reduction of seed proteins during germination.[39] Therefore, Trx has been linked to enhanced grain digestibility in wheat and sorghum.[40,41] Alternate expression of Trx in this study may be associated with the *β*-kafirin null mutation and could have potential downstream impacts on digestibility, in addition to effects of the mutation.

The sorghum homologue to a maize glutaredoxin Grx_C2.2−glutaredoxin subgroup 1 was identified in the A/G fraction in 296B spots 9 and 10, and in QL12 spots 5 and 6, which localized to the same size range (Mw ∼13 kDa) (Figure 5). Grx, including the maize Grx_C2.2 identified in this study, catalyze the enzymatic reaction where proteins with reduced sulfide groups are converted to those with oxidized disulfide bonds (Table 2). Grx have been implicated in the oxidative stress response through regeneration of enzymes involved in peroxide and methionine sulfoxide reduction.[42] These proteins therefore have dual roles in protein aggregation and stress responses, creating opportunities for the development of streamlined grain improvement strategies affecting multiple traits.

*Protein Disulfide Isomerases (PDIs).* PDIs function as molecular chaperones in disulfide-mediated protein folding. They contain two Trx domains with a redox site.[43] Matches to HMW PDI from maize (C0PLF0 and A5A5E7) were generated in glutelin spot 2 across both genotypes (Supporting Information, Figure S5 and Supplementary Table 2). These proteins appear to migrate to the same location on A/G and prolamin gels, but the spots were not analyzed in this fraction using LC−MS/MS. However, it is possible that they share the same identity as the PDIs identified in the glutelin fraction because their distribution on the gel is similar. Again, the presence of these spots across multiple fractions may be a result of incomplete sequential extraction of the proteins.

Previous studies indicate that PDI mutations can result in irregular starch granule formation and chalky grain phenotypes.[44] The rice mutant *esp2* lacks protein disulfide isomerase-like (PDIL) 1:1, but shows enhanced expression of the thiol disulfide oxidoreductase OsEro1.[45] The grain exhibits altered seed storage protein compartmentalization through inhibition of disulfide bond formation. It has been proposed that the formation of native disulfide bonds in proglutelins also depends on an electron transfer pathway involving the OsEro1 and the PDI-like OsPDIL.[46] In the *floury2* mutant, PDIL is up-regulated, where a mutation in a signaling peptide results in the abnormal

**Table 2. Protein Candidates Identified in Sorghum Grain with Reported Effects on Protein−Starch Matrix Structure, Grain Quality, and Stress Responses**

| accession no. | protein | 296B spot | QL12 spot | protein fraction | homologues | function |
|---|---|---|---|---|---|---|
| C5XB72/C5YGL1 (sorghum) | thioredoxin | 4 | absent | A/G | B6SX54_maize | converts seed proteins from the oxidized to the reduced state during germination |
| C5YBX1 (sorghum) | glutaredoxin Grx_C2.2−glutaredoxin subgroup 1 | 9 and 10 | 5 and 6 | A/G | B6THA1_maize | converts proteins with reduced sulfide groups to those with oxidized disulfide bonds |
| F2DDK1 (barley) | peroxiredoxin | 5 | 5 | prolamin | HB6T2Y1_maize and PR2E1_ORYSJ (rice) | interacts with glutaredoxins, thioredoxins, and cyclophilin as both reductants and non-dithio−disulfide exchange proteins |
| C5XT06, B3GQV9, and C5Z9C6 (sorghum) | peptidyl-prolyl cis/trans isomerase (PPIase) | 9, 10, and 11 | 9 and 10 | prolamin/glutelin | maize B4FZZ2 and B4FY3T, wheat AZLM55, sugarcane C7E3V7 | facilitates protein folding through slow isomerization of peptide bonds in oligopeptides and through the amino acid proline in cellular proteins |
| A5A5E7 (maize) | protein disulfide isomerase (PDI) | 2 | 2 | glutelin | maize C0PLF0 and A5A5E7 | promotes the correct disulfide pairing in proteins |
| P81368 (sorghum) | α-amylase inhibitor | 1, 2, 8, and 11 | 1 and 2 | prolamin | IAA5_SORI (sorghum) | slows the conversion of starch to sugars (and ethanol) |


dx.doi.org/10.1021/jf5022847 | J. Agric. Food Chem. 2014, 62, 9819−9831

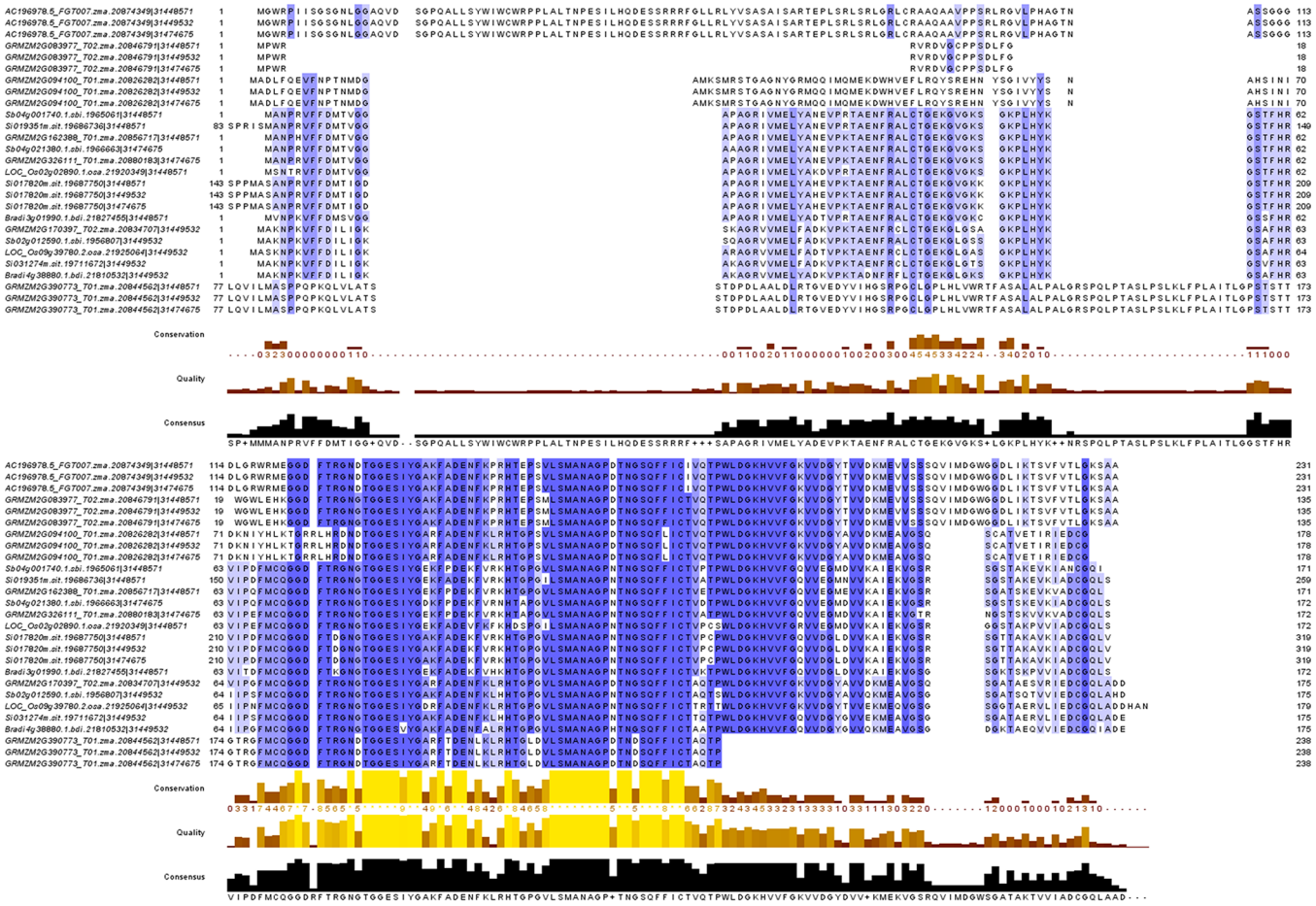

**Figure 7.** ClustalW alignment of peptidyl-prolyl cis/trans isomerase (PPIase) peptide sequences across major grass species, illustrating a high level of sequence homology. PPIases identified in sorghum grain prolamin and glutelin protein fractions (Supporting Information, Figure S4 and Supplementary Table 1) are included. Dark purple indicates exact homology; light purple indicates highly conserved homology.

processing and accumulation of small precursor $\alpha$-zeins and high levels of luminal binding protein (BiP) in irregularly shaped protein bodies.[47] Mutations in sorghum protein isomerases, such as PDI, may have a similar effect on protein body formation.

*Peptidyl-prolyl cis/trans Isomerases (PPIases).* PPIases catalyze protein folding through isomerization of peptide bonds in oligopeptides and on proline residues in cellular proteins.[43,48,49] PPIases were identified in the 19 kDa band colocalized with $\beta$-kafirin on the 296B 1D prolamin gel, as well as in the 26 kDa band, colocalized with $\alpha$-kafirin (Figure 4). On 2D gels, PPIases were detected in the 296B prolamin fraction gel spots 9, 10 and 11, and in QL12 gel spots 9 and 10, as well as in the QL12 glutelin fraction spot 4 (Supporting Information, Figure S4 and Supplementary Table 1). Sorghum PPIases identified in the study (accession numbers C5XT06, C5Z9C6, and B3GQV9), consistently localized to the same size region (~18 kDa) on 1D and 2D PAGE gels, and within the same pI range pH 8–9. PPIases colocalized with $\beta$- and $\delta$-kafirins in the 296B 2D prolamin gel spot 11 and with $\alpha$-kafirin in QL12 spot, as well as in the prolamin fraction on the 1D SDS-PAGE gel, indicating that they are hydrophobic enzymes, which may interact with the kafirins.

The PPIase family is encoded by multiple genes, exhibiting some redundancy in plants.[50] An alignment of the sorghum PPIases identified in this study, with orthologues from closely related species, reveals a high level of sequence homology at the protein level (Figure 7). The sorghum PPIases exhibited close homology to their counterparts in *Zea mays* (B4FZZ2 and B4FY3T), *Triticum aestivum* (AZLM55), *Saccharum officinarum* (C7E3V7), *Citrus sinensis* (D0ELH5) and *Gerbera* (A8CYN7). PIN1-type PPIases are encoded by multiple genes and are present in a wide variety of plant species. A four amino acid insertion site is situated next to the phosphor-specific recognition site of the active site, regulating the activity of the enzyme.[51] The PPIase activity of cyclophilins is regulated by thioredoxin,[52] also identified in this study and discussed above. PIN1At in *Arabidopsis* encodes a PPIase, which regulates flowering time through phosphorylation-dependent prolyl cis/trans isomerization of key regulatory pathway specific transcription factors.[53] Alterations to protein folding status catalyzed by PPIases with the cis/trans isomerization of proline imidic peptide bonds could also affect protein body composition in sorghum and other grain crops.[54]

*Heat Shock Protein (HSP)/Luminal Binding Protein (BiP).* HSP/BiP chaperones are key regulators of protein body biogenesis, with significant impacts on grain quality traits and composition[55] (Table 2). HSPs/BiPs were isolated predominantly from the water/salt-soluble (A/G) and alkali-soluble (glutelin) protein fractions across both genotypes (Figure 5 and Supporting Information, Figure S5 and Supplementary Table 2). A/G spot 2 produced a strong match to sorghum protein (accession number C5XY25), which exhibits homology to the *Oryza sativa* 19 kDa class II HSP. The protein has a calculated Mw 19 982 Da and pI 5.71, appropriate to its position on the gel.


dx.doi.org/10.1021/jf5022847 | J. Agric. Food Chem. 2014, 62, 9819–9831

A sorghum homologue to a maize 16.9 kDa class I heat shock protein was also identified (C5XQR9) in a lower molecular weight region of the gel within a more basic pI range, relative to spot 2 discussed above. The HSP homologue C5XQR9 localized to an appropriate position on the gel, corresponding to Mw 17 121 Da and pI 6.18. Matches to C5XQR9 were observed in 296B spots B, 7, 8, 9, and 10, and in QL12 spots B and 7. QL12 spot 8 contained an additional sorghum homologue (accession number C5XML7) to a maize 17.5 kDa class II HSP.

In addition, several HMW (50−75 kDa) putative uncharacterized proteins with homology to HSP chaperones or luminal binding proteins (BiPs) were identified in the alkali-soluble glutelin fraction across both 296B and QL12 genotypes (accession numbers C5YU58, C5WNX8, and C5XPN2). Sorghum HSP C5YU58 has also been identified in grain proteomic analyses carried out by other researchers.[56] In maize, elevated HSP/BiP levels during development are linked to mutations in $\alpha$- and $\gamma$-zeins resulting in an opaque or floury phenotype.[55,57,58] BiP has been found to associate with the surface of rice protein bodies, assisting in prolamin deposition through disulfide bonding at specific cysteine residues.[59] In wheat, these chaperones are localized to the interior, rather than on the surface of protein bodies, indicating diverse roles for HSP/BiP in protein body aggregation across different plant species.[60]

*α-Amylase Inhibitors.* Sorghum $\alpha$-amylase inhibitors were isolated from the prolamin fraction in both genotypes (Figure 5, Table 2). The enzyme was isolated from a low molecular weight (10−15 kDa) size range, within pI 7−8. The inhibitor was identified in 296B gel spots 1, 2, 8, and 11, and in QL12 spots 1 and 2, which was observed across duplicate gels. The activity of $\alpha$-amylase inhibitors in the grain affects starch digestion into glucose, slowing the conversion of starch to ethanol during the fermentation process.[61] Lines expressing low levels of $\alpha$-amylase inhibitor may be appropriate candidates for the biofuels industry. $\alpha$-Amylase inhibitors also play a role in plant defense, where they deter crop destruction by snails and birds in tannin-containing sorghum and have been shown to enhance insect resistance in wheat.[62−64]

**Non-Prolamin Proteins in the Protein−Starch Matrix.** Non-prolamin proteins, such as albumins and globulins, are rich in essential amino acids, including lysine and tryptophan, and provide the embryo with additional readily accessible nitrogen reserves during germination. Sorghum homologues to the major nonkafirin storage proteins were identified across all protein solubilities (Supporting Information, Figure S4 and Supplementary Table 1). In the water/salt-soluble fractions sorghum homologues to globulin and cupin-like proteins in *Zea mays* were abundant, including accession numbers C5WY16 and C5WQD2.

Globulins are saline-soluble secondary storage proteins, with many belonging to the cupin superfamily. For example, 7S globulins function exclusively as storage proteins, but are not required for normal seed function.[65] Globulin S-1 (63 kDa) and globulin S-2 (45 kDa) collectively represent ∼20% of seed protein content in maize and share amino acid sequence similarity with the 7S seed proteins of wheat and legumes.[66] These proteins have significant impacts on the nutritional quality of the grain.

Proteins identified in the alkali-soluble glutelin fraction included a range of vicilin- and legumin-like storage proteins. The sorghum homologue to uncleaved maize legumin, accession number C5YY38, and the globulin/vicilin-like sorghum homologue C5WUN6, were identified in this fraction. Immuno-localization studies in *Medicago trunculata* using antivicilin antibodies show preferential targeting of vicilins to the periphery of the protein bodies.[67] Sorghum glutelins have not yet been extensively characterized, but it is hypothesized that they may form complex highly linked protein networks encasing protein bodies and providing additional structure to the protein−starch matrix.[68,69] Protein members of the legumin superfamily are most abundant in legumes, oats, and rice. Wheat legumin-like protein, or triticin, accumulates in globulin inclusion bodies at the periphery of prolamin bodies. Overexpression of pea legumin in wheat forms crystalline patterns contributing to the altered structure of the protein−starch matrix.[70,71]

A range of proteins were identified in the glutelin fraction in addition to legumin and vicillin, which included HSPs/BiPs, cell division cycle proteins, homologues to maize caleosin, glutathione *S*-transferases, RuBisCo large subunit binding protein, and wheat MOTHER OF FT AND TFL1 (MFT), which regulates seed dormancy and the onset of germination.[72] Significant amounts of $\alpha$-, $\beta$-, and $\delta$-kafirins were also identified in this fraction, although quantification scores indicate that their concentrations were considerably less than in the prolamin fraction, and their presence may be due to incomplete sequential extraction of the alcohol-soluble fraction as previously noted.

Across the study, protein spots excised from HMW areas of 2D gels generally produced matches specifically to HMW proteins, while spots excised from LMW areas tended to produce matches to both high and low molecular weight proteins. This indicates that spots in LMW areas of the gel may contain a mix of intact proteins and protein subunits, or cleaved products of larger proteins. For example, sorghum homologues to HMW globulin S-1 and S-2 (C5WY16 and C5WQD2), cupin family proteins (C5WUN and C5X0T3), and uncleaved legumin (C5YY38) were isolated from every protein fraction, from both high and low molecular weight areas of the gel, whereas LMW proteins (∼13 kDa), such as thioredoxin, glutaredoxin, and $\alpha$-amylase inhibitors, were isolated exclusively from the lower molecular weight area of the gel. Highly abundant sorghum homologues to glyceraldehyde-3-phosphate dehydrogenase as well as various HSPs were detected in the glutelin fraction. These proteins were generally found in the area of their calculated pI values, but localized across a broader size range. Because the presence of these proteins was so widespread, the matches were not listed in results tables (Supporting Information, Supplementary Table 1), unless they represented the only quality match returned for a protein spot.

**Sorghum Proteomics.** Our proteomic analysis compiles sequence and biochemical information describing a range of proteins affecting sorghum endosperm structure and composition. The derived data set will further augment annotation of the sorghum proteome and facilitate identification of potential targets for improved grain quality. Furthermore, the data provide a basis for comparative analysis with other major grain crops. Possible avenues for utilizing protein sequence data include the development of high-lysine varieties with increased A/G content for enhanced nutritional quality, as well as the modification of enzyme-regulated protein aggregation in the endosperm for increased starch availability.

Differential regulation of thioredoxin in the $\beta$-kafirin null mutant QL12 indicates that expression of the enzyme may be either directly or indirectly linked to the $\beta$-kafirin mutation. Evaluation of Trx expression at the RNA/protein level across additional $\beta$-kafirin mutants and at varying stages of development

could provide further insights into this relationship. Future research on previously uncharacterized proteins identified in this study may reveal elements related to grain quality parameters, such as digestibility and flour pasting properties. In particular, the HMW $\gamma$-prolamin homologue identified provides an interesting candidate for further functional analysis as it has been shown to localize to the periphery of the protein bodies in maize, possibly limiting enzymatic access to internally located $\alpha$-zeins and starch. Identification of proteins with desirable amino acid content and enhanced digestibility will also benefit future research initiatives to improve sorghum grain nutritional value and palatability.

The commercial success of grain crops with altered protein composition, such as high lysine and increased digestibility lines, can be limited by negative pleiotropic effects accompanying changes in the amino acid profile. For example, maize endosperm protein mutants, *o2* (*opaque2*) and *fl2* (*floury2*) have an improved amino acid content, but exhibit a number of undesirable traits such as reduced grain yield and increased susceptibility to diseases and pests. This work supports the development of high throughput screening methods for biomarkers associated with grain quality and endeavors for the genetic improvement and biofortification of sorghum through molecular breeding and transformation. The identification and characterization of proteins impacting on various grain quality parameters in sorghum, such as nutritional quality, grain hardness, and stress resistance, will assist breeders in their introgression of genetic factors associated with these traits into established breeding lines. The rapid introgression and selection of desirable grain quality traits will require the linking of the proteome with the genome. Next generation genome resequencing[73] and genotyping by selection[74] tools are now publically available and are being incorporated into sorghum prebreeding for other grain quality traits such as digestibility.[75] This will ultimately lead to more cost-effective and efficient plant breeding for improved sorghum varieties tailored to specific end uses.

## ■ ASSOCIATED CONTENT

### Supporting Information

Supplementary Table 1 contains a list of protein matches to LC−MS/MS generated peptide profiles. Figure S1 contains RP-HPLC peak profiles of the alcohol-soluble (prolamin) proteins, and Figure S2 contains RP-HPLC profiles for the water/salt-soluble protein fraction across sorghum grain samples. Figure S3 contains LOC profiling of the A/G fraction in $\beta$-kafirin null variants compared to wild type. Figure S4 presents 2D SDS-PAGE gels for A/G, prolamin, and glutelin fractions for wild-type 296B and $\beta$-kafirin null QL12. Figure S5 and Supplementary Table 2 show 2D SDS PAGE/LC−MS/MS identification HMW alkali-soluble (glutelin) proteins. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*Tel.: +61-3365-2141. Fax: +61-3365-2141. E-mail: j.cremer@uq.edu.au.

### Funding

The work is supported by funding from an ARC linkage grant with Pacific Seeds, by the Andersons Research Program, and by the USDA—ARS, Manhattan, KS, USA.

### Notes

Names are necessary to report factually on available data; however, the U.S. Department of Agriculture neither guarantees nor warrants the standard of the product, and use of the name by the U.S. Department of Agriculture implies no approval of the product to the exclusion of others that may also be suitable. USDA is an equal opportunity provider and employer.

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Extensive support in proteomic techniques was kindly provided by Dr. Amanda Nouwens at the SCMB protein facility (The University of Queensland). Identification of proteins with LC−MS/MS was facilitated through open access queries to the Mascot server.

## ■ ABREVIATIONS USED

A/G, albumin/globulins; BiP, luminal binding proteins; Grx, glutaredoxins; HPLC, high performance liquid chromatography; IEF, isoelectric focusing; HSP, heat shock proteins; HMW, high molecular weight; kDa, kilodalton; LMW, low molecular weight; LC−MS/MS, liquid chromatography−tandem mass spectrometry; LOC, lab on a chip; MS, mass spectrometry; Mw, molecular weight; PDI, protein disulfide isomerase; PPIase, peptidyl-prolyl cis/trans isomerase; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; Trx, thioredoxin

## ■ REFERENCES

(1) United Nations. *World Population to 2300*; United Nations: New York, 2004.

(2) Rijsberman, F. R. Water scarcity: Fact or fiction? *Agric. Water Manage.* **2006**, *80*, 5−22.

(3) Ortiz, R.; Iwanaga, M.; Reynolds, M. P.; Wu, H.; Crouch, J. H. Overview on Crop Genetic Engineering for Drought-prone Environments. *J. SAT Agric. Res.* [Online] **2007**, *4* (1), 30 pp. http://www.icrisat.org/journal/SpecialProject/sp3.pdf.

(4) Steduto, P.; Katerji, N.; Puertos-Molina, H.; Unlu, M.; Mastrorilli, M.; Rana, G. Water-use efficiency of sweet sorghum under water stress conditions Gas-exchange investigations at leaf and canopy scales. *Field Crops Res.* **1997**, *54*, 221−234.

(5) Kim, M.; Day, D. F. Composition of Sugar Cane, Energy Cane, and Sweet Sorghum Suitable for Ethanol Production at Louisiana Sugar Mills. *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 803−807.

(6) Wong, J. H.; Lau, T.; Cai, N.; Singh, J.; Pedersen, J. F.; Vensel, W. H.; Hurkman, W. J.; Wilson, J. D.; Lemaux, P. G.; Buchanan, B. B. Digestibility of protein and starch from sorghum (Sorghum bicolor) is linked to biochemical and structural features of grain endosperm. *J. Cereal Sci.* **2009**, *49*, 73−82.

(7) Rooney, L. W. Sorghum and Millets. In *Cereal Grain Quality*, Henry, R. J., Kettlewell, P. S., Eds.; Chapman and Hall: Dordrecht, Netherlands, 1996; pp 153−177.

(8) Belton, P. S.; Delgadillo, I.; Halford, N. G.; Shewry, P. R. Kafirin structure and functionality. *J. Cereal Sci.* **2006**, *44*, 272−286.

(9) De Mesa-Stonestreet, N. J.; Alavi, S.; Bean, S. R. Sorghum Proteins: The Concentration, Isolation, Modification, and Food Applications of Kafirins. *J. Food Sci.* **2010**, *75*, R90−R104.

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

(11) Xu, J. H.; Messing, J. Amplification of prolamin storage protein genes in different subfamilies of the Poaceae. *Theor. Appl. Genet.* **2009**, *119*, 1397−412.

(12) Song, R.; Llaca, V.; Linton, E.; Messing, J. Sequence, Regulation, and Evolution of the Maize 22-kD $\alpha$ Zein Gene Family. *Genome Res.* **2001**, *11*, 1817−1825.

(13) Luis, E. S.; Sullivan, T. W.; Nelson, L. A. Nutrient Composition and Feeding Value of Proso Millets, Sorghum Grains, and Corn in Broiler Diets. *Poultry Sci.* **1982**, *61*, 311−320.

(14) Matta, N.; Singh, A.; Kumar, Y. Manipulating seed storage proteins for enhanced grain quality in cereals. *Afr. J. Food Sci.* **2009**, *3*, 439−446.
(15) Laidlaw, H.; Mace, E.; Williams, S.; Sakrewski, K.; Mudge, A.; Prentis, P.; Jordan, D.; Godwin, I. Allelic variation of the beta-, gamma- and delta-kafirin genes in diverse *Sorghum* genotypes. *Theor. Appl. Genet.* **2010**, *121*, 1227−1237.
(16) Bean, S. R.; Ioerger, B. P.; Blackwell, D. L. Separation of Kafirins on Surface Porous Reversed-Phase High-Performance Liquid Chromatography Columns. *J. Agric. Food Chem.* **2011**, *59*, 85−91.
(17) Bean, S. R.; Tilley, M. Separation of Water-Soluble Proteins from Cereals by High-Performance Capillary Electrophoresis (HPCE). *Cereal Chem.* **2003**, *80*, 505−510.
(18) Kappler, U.; Nouwens, A. S. The molybdoproteome of Starkeya novella—insights into the diversity and functions of molybdenum containing proteins in response to changing growth conditions. *Metallomics* **2013**, *5*, 325−334.
(19) Shewayrga, H.; Sopade, P. A.; Jordan, D. R.; Godwin, I. D. Characterisation of grain quality in diverse sorghum germplasm using a Rapid Visco-Analyzer and near infrared reflectance spectroscopy. *J. Sci. Food Agric.* **2012**, *92*, 1402−1410.
(20) Cremer, J. E.; Liu, L.; Bean, S. R.; Ohm, J. B.; Tilley, M.; Wilson, J. D.; Kaufman, R. C.; Vu, T. H.; Gilding, E. K.; Godwin, I. D.; Wang, D. Impacts of kafirin allelic diversity, starch content and protein digestibility on ethanol conversion efficiency in grain sorghum. *Cereal Chem.* **2014**, *91*, 218−227.
(21) Guiragossian, V.; Chibber, B. A. K.; Van Scoyoc, S.; Jambunathan, R.; Mertz, E. T.; Axtell, J. D. Characteristics of proteins from normal, high lysine, and high tannin sorghums. *J. Agric. Food Chem.* **1978**, *26*, 219−223.
(22) Reddy, N. P. E.; Vauterin, M.; Frankard, V.; Jacobs, M. Biochemical Characterisation and Cloning of α-Kafirin Gene from Sorghum (Sorghum bicolor L Moench). *J. Plant Biochem. Biotechnol.* **2001**, *10*, 101−106.
(23) Vendemiatti, A.; Rodrigues Ferreira, R.; Humberto Gomes, L.; Oliveira Medici, L.; Antunes Azevedo, R. Nutritional Quality of Sorghum Seeds: Storage Proteins and Amino Acids. *Food Biotechnol.* **2008**, *22*, 377−397.
(24) Wilson, C. M.; Shewry, P. R.; Miflin, B. J. Maize endosperm proteins compared by sodium dodecyl sulphate gel electrophoresis and isoelectric focussing. *J. Cereal Chem.* **1981**, *58*, 275−281.
(25) Wu, Y. V.; Wall, J. S. Lysine content of protein increased by germinating of normal and high lysine sorghum. *J. Agric. Food Chem.* **1980**, *28*, 455−458.
(26) Taylor, J. R. N.; Novellie, L.; Liebenberg, N. v. d. W. Sorghum protein body composition and ultrastructure. *J. Cereal Chem.* **1984**, *61*, 69−73.
(27) Woo, Y.-M.; Hu, D. W.-N.; Larkins, B. A.; Jung, R. Genomics Analysis of Genes Expressed in Maize Endosperm Identifies Novel Seed Proteins and Clarifies Patterns of Zein Gene Expression. *Plant Cell* **2001**, *13*, 2297−2317.
(28) Ferrante, P.; Masci, S.; D'Ovidio, R.; Lafiandra, D.; Volpi, C.; Mattei, B. A proteomic approach to verify in vivo expression of a novel γ-gliadin containing an extra cysteine residue. *Proteomics* **2006**, *6*, 1908−1914.
(29) Wang, Y.; Tilley, M.; Bean, S.; Sun, X. S.; Wang, D. Comparison of Methods for Extracting Kafirin Proteins from Sorghum Distillers Dried Grains with Solubles. *J. Agric. Food Chem.* **2009**, *57*, 8366−8372.
(30) Mokhawa, G.; Kerapeletswe-Kruger, C. K.; Ezeogu, L. I. Electrophoretic analysis of malting degradability of major sorghum reserve proteins. *J. Cereal Sci.* **2013**, *58*, 191−199.
(31) Pandit, P. *Molecular characterisation of the alpha-kafirin multigene family for the genetic improvement of sorghum grain quality*. Dissertation, The University of Queensland, Brisbane, 2008.
(32) Oria, M. P.; Hamaker, B. R.; Shull, J. M. Resistance of Sorghum α-, β-, and γ-Kafirins to Pepsin Digestion. *J. Agric. Food Chem.* **1995**, *43*, 2148−2153.
(33) Oria, M. P.; Hamaker, B. R.; Axtell, J. D.; Huang, C.-P. A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5065−5070.
(34) Lending, C. R.; Larkins, B. A. Changes in the zein composition of protein bodies during maize endosperm development. *Plant Cell* **1989**, *1*, 1011−1023.
(35) Geli, M. I.; Torrent, M.; Ludevid, D. Two structural domains mediate two sequential events in [gamma]-zein targeting: Protein endoplasmic reticulum retention and protein body formation. *Plant Cell* **1994**, *6*, 1911−1922.
(36) Cho, M.-J.; Wong, J. H.; Marx, C.; Jiang, W.; Lemaux, P. G.; Buchanan, B. B. Overexpression of thioredoxin h leads to enhanced activity of starch debranching enzyme (pullulanase) in barley grain. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14641−14646.
(37) Alkhalfioui, F.; Renard, M.; Vensel, W. H.; Wong, J.; Tanaka, C. K.; Hurkman, W. J.; Buchanan, B. B.; Montrichard, F. Thioredoxin-Linked Proteins Are Reduced during Germination of Medicago truncatula Seeds. *Plant Physiol.* **2007**, *144*, 1559−1579.
(38) Holmgren, A. Thioredoxin catalyzes the reduction of insulin disulfides by dithiothreitol and dihydrolipoamide. *J. Biol. Chem.* **1979**, *254*, 9627−9632.
(39) Alkhalfioui, F.; Renard, M.; Montrichard, F. Unique properties of NADP-thioredoxin reductase C in legumes. *J. Exp. Bot.* **2007**, *58*, 969−978.
(40) Wong, J. H.; Cai, N.; Tanaka, C. K.; Vensel, W. H.; Hurkman, W. J.; Buchanan, B. B. Thioredoxin Reduction Alters the Solubility of Proteins of Wheat Starchy Endosperm: An Early Event in Cereal Germination. *Plant Cell Physiol.* **2004**, *45*, 407−415.
(41) Sahoo, L.; Lindquist, D. J. L.; Pedersen, J. F.; Kaur, R.; Wong, J. H.; Buchanan, B. B.; Lemaux, P. G. Effect Of Transgenes From Sorghum On The Fitness Of Shattercane x Sorghum Hybrids. In *62nd Annual Meeting of the North Central Weed Science Society 2007*; Curran Associates: Red Hook, NY, 2007; p 45.
(42) Rouhier, N.; Lemaire, S. D.; Jacquot, J. P. The role of glutathione in photosynthetic organisms: Emerging functions for glutaredoxins and glutathionylation. *Annu. Rev. Plant Biol.* **2008**, *59*, 143−166.
(43) Freedman, R. B. A protein with many functions? *Nature* **1989**, *337*, 407−408.
(44) Kim, Y.; Yeu, S.; Park, B.; Koh, H.-J.; Song, J.; Seo, H. Protein Disulfide Isomerase-Like Protein 1−1 Controls Endosperm Development through Regulation of the Amount and Composition of Seed Proteins in Rice. *PLoS One* **2012**, *7* (9), e44493.
(45) Takemoto, Y.; Coughlan, S. J.; Okita, T. W.; Satoh, H.; Ogawa, M.; Kumamaru, T. The Rice Mutant esp2 Greatly Accumulates the Glutelin Precursor and Deletes the Protein Disulfide Isomerase. *Plant Physiol.* **2002**, *128*, 1212−1222.
(46) Onda, Y.; Kumamaru, T.; Kawagoe, Y. ER membrane-localized oxidoreductase Ero1 is required for disulfide bond formation in the rice endosperm. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 14156−14161.
(47) Li, C.; Larkins, B. Expression of protein disulfide isomerase is elevated in the endosperm of the maize floury-2 mutant. *Plant Mol. Biol.* **1996**, *30*, 873−882.
(48) Gething, M.-J.; Sambrook, J. Protein folding in the cell. *Nature* **1992**, *355*, 35−45.
(49) Brandts, J.; Halvorsen, H.; Brennan, M. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **1975**, *14*, 4953−4963.
(50) He, Z.; Li, L.; Luan, S. Immunophilins and Parvulins. Superfamily of Peptidyl Prolyl Isomerases in Arabidopsis. *Plant Physiol.* **2004**, *134*, 1248−1267.
(51) Yao, J.-L.; Kops, O.; Lu, P.-J.; Lu, K. P. Functional Conservation of Phosphorylation-specific Prolyl Isomerases in Plants. *J. Biol. Chem.* **2001**, *276*, 13517−13523.
(52) Motohashi, K.; Koyama, F.; Nakanishi, Y.; Ueoka-Nakanishi, H.; Hisabori, T. Chloroplast Cyclophilin Is a Target Protein of Thioredoxin: Thiol Modulation of the Peptidyl-prolyl cis-trans Isomerase Activity. *J. Biol. Chem.* **2003**, *278*, 31848−31852.

(53) Wang, Y.; Liu, C.; Yang, D.; Yu, H.; Liou, Y.-C. Pin1At Encoding a Peptidyl-Prolyl cis/trans Isomerase Regulates Flowering Time in Arabidopsis. *Mol. Cell* **2010**, *37*, 112−122.
(54) Duodu, K. G.; Tang, H.; Grant, A.; Wellner, N.; Belton, P. S.; Taylor, J. R. N. FTIR and solid state 13C spectroscopy of proteins of wet cooked and popped sorghum and maize. *J. Cereal Sci.* **2001**, *33*, 261−269.
(55) Zhang, F.; Boston, R. Increases in binding protein (BiP) accompany changes in protein body morphology in three high-lysine mutants of maize. *Protoplasma* **1992**, *171*, 142−152.
(56) Buckner, R. J. *Developmental and crosslinking factors related to low protein digestibility of sorghum grain*. Dissertation, Purdue University, West Lafayette, IN, 1997.
(57) Kim, C. S.; Gibbon, B. C.; Gillikin, J. W.; Larkins, B. A.; Boston, R. S.; Jung, R. The maize Mucronate mutation is a deletion in the 16-kDa γ-zein gene that induces the unfolded protein response1. *Plant J.* **2006**, *48*, 440−451.
(58) Kim, C. S.; Hunter, B. G.; Kraft, J.; Boston, R. S.; Yans, S.; Jung, R.; Larkins, B. A. A Defective Signal Peptide in a 19-kD α-Zein Protein Causes the Unfolded Protein Response and an Opaque Endosperm Phenotype in the Maize De*-B30 Mutant. *Plant Physiol.* **2004**, *134*, 380−387.
(59) Muench, D. G.; Wu, Y.; Zhang, Y.; Li, X.; Boston, R. S.; Okita, T. W. Molecular Cloning, Expression and Subcellular Localization of a BiP Homolog from Rice Endosperm Tissue. *Plant Cell Physiol.* **1997**, *38*, 404−412.
(60) Levanony, H.; Rubin, R.; Altschuler, Y.; Galili, G. Evidence for a novel route of wheat storage proteins to vacuoles. *J. Cell Biol.* **1992**, *119*, 1117−1128.
(61) Dreher, M. L.; Dreher, C. J.; Berry, J. W.; Fleming, S. E. Starch digestibility of foods: A nutritional perspective. *Crit. Rev. Food Sci. Nutr.* **1984**, *20*, 47−71.
(62) Farmer, E.; Ryan, C. Inter plant communication: air-borne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7713−7716.
(63) Gatehouse, A.; Fenton, K.; Jepson, I.; Pavey, D. The effects of a-amylase inhibitors on insect storage pests: inhibition of a-amylase in vitro and effects on development in vivo. *J. Sci. Food Agric.* **1986**, *37*, 727−734.
(64) Yetter, M.; Saunders, R.; Boles, H. Alpha-amylase inhibitors from wheat kernels as factors in resistance to postharvest insects. *Cereal Chem.* **1979**, *56* (4), 243.
(65) Kriz, A. L.; Wallace, N. H. Characterisation of the maize globulin-2 gene and analysis of 2 null alleles. *Biochem. Genet.* **1991**, *29*, 241−254.
(66) Belanger, F. C.; Kriz, A. L. Molecular Characterization of the Major Maize Embryo Globulin Encoded by the *Glb1* Gene. *Plant Physiol.* **1989**, *91*, 636−643.
(67) Abiraches-Darmency, M.; Dessaint, F.; Benlicha, E.; Schneider, C. Biogenesis of protein bodies during vicilin accumulation in Medicago truncatula immature seeds. *BMC Res. Notes* **2012**, *5*, 409.
(68) Shull, J. M.; Watterson, J. J.; Kirleis, A. W. Purification and immunocytochemical localization of kafirins in Sorghum bicolor (L. Moench) endosperm. *Protoplasma* **1992**, *171*, 64−74.
(69) Taylor, J.; Taylor, J. R. N. Protein Biofortified Sorghum: Effect of Processing into Traditional African Foods on Their Protein Quality. *J. Agric. Food Chem.* **2011**, *59*, 2386−2392.
(70) Arcalis, E.; Marcel, S.; Altmann, F.; Kolarich, D.; Drakakaki, G.; Fischer, R.; Christou, P.; Stoger, E. Unexpected Deposition Patterns of Recombinant Proteins in Post-Endoplasmic Reticulum Compartments of Wheat Endosperm. *Plant Physiol.* **2004**, *136*, 3457−3466.
(71) Stöger, E.; Parker, M.; Christou, P.; Casey, R. Pea Legumin Overexpressed in Wheat Endosperm Assembles into an Ordered Paracrystalline Matrix. *Plant Physiol.* **2001**, *125*, 1732−1742.
(72) Nakamura, S.; Abe, F.; Kawahigashi, H.; Nakazono, K.; Tagiri, A.; Matsumoto, T.; Utsugi, S.; Ogawa, T.; Handa, H.; Ishida, H.; Mori, M.; Kawaura, K.; Ogihara, Y.; Miura, H. A Wheat Homolog of MOTHER OF FT AND TFL1 Acts in the Regulation of Germination. *Plant Cell* **2011**, *23*, 3215−3229.
(73) Mace, E. S.; Tai, S.; Gilding, E. K.; Li, Y.; Prentis, P. J.; Bian, L.; Campbell, B. C.; Hu, W.; Innes, D. J.; Han, X.; Cruickshank, A.; Dai, C.; Frère, C.; Zhang, H.; Hunt, C. H.; Wang, X.; Shatte, T.; Wang, M.; Su, Z.; Li, J.; Lin, X.; Godwin, I. D.; Jordan, D. R.; Wang, J. Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **2013**, *4*, No. 2320.
(74) Morris, G. P.; Ramu, P.; Deshpande, S. P.; Hash, C. T.; Shah, T.; Upadhyaya, H. D.; Riera-Lizarazu, O.; Brown, P. J.; Acharya, C. B.; Mitchell, S. E.; Harriman, J.; Glaubitz, J. C.; Buckler, E. S.; Kresovich, S. Population genomic and genome-wide association studies of agro-climatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 453−458.
(75) Gilding, E. K.; Frère, C. H.; Cruickshank, A.; Rada, A. K.; Prentis, P. J.; Mudge, A. M.; Mace, E. S.; Jordan, D. R.; Godwin, I. D. Allelic variation at a single gene increases food value in a drought-tolerant staple cereal. *Nat. Commun.* **2013**, *4*, 1483.