Theor Appl Genet (2008) 117:555–563
DOI 10.1007/s00122-008-0799-y

ORIGINAL PAPER

Globulins are the main seed storage proteins in *Brachypodium distachyon*

Debbie L. Laudencia-Chingcuanco · William H. Vensel

Received: 25 July 2007 / Accepted: 12 May 2008 / Published online: 5 June 2008

Abstract *Brachypodium distachyon* is being developed as a model system to study temperate cereals and forage grasses. We have begun to investigate its utility to understand seed development and grain filling by identifying the major seed storage proteins in a diploid accession Bd21. With the use of ID SDS–PAGE and mass spectrometry we detected seven major storage protein bands, six of which were identified as globulins. A subset of the major seed proteins isolated from three hexaploid accessions, Bd4, Bd14 and Bd17 were also identified as globulins. Several *Brachypodium* cDNAs clones encoding globulin were completely sequenced. Two types of globulin genes were identified, *Bd.glo1* and *Bd.glo2*, which are similar to maize 7S and oat 12S globulins, respectively. The derived polypeptide sequences of the globulins contain a typical signal peptide sequence in their polypeptide N-termini and two cupin domains. *Bd.glo1* is encoded by a single copy gene, whereas, *Bd.glo2* belongs to a gene family.

Introduction

Brachypodium distachyon is being developed as a model system to study temperate cereals and forage grasses (Draper et al. 2001; Hasterok et al. 2004; Vogel et al. 2006a, b). *Brachypodium* belongs to the subfamily Pooideae, which includes economically important crops like wheat, barley and oat. *B. distachyon*'s physical and genomic attributes make it suitable for functional genomics studies. Its small diploid genome (~355 Mbp) is just about twice that of the dicot model plant *Arabidopsis thaliana* (Bennett and Leitch 2005; Vogel et al. 2006a) and has a relatively low repetitive DNA content (Catalan et al. 1995; Hasterok et al. 2006). Its small physical stature, self-fertility, short lifecycle, simple growth requirements and competence to be efficiently transformed (Draper et al. 2001; Christiansen et al. 2005; Vogel et al. 2006a) makes it amenable for large-scale mutagenesis studies.

Rice (*Oryza sativa*) has been promoted as a model organism for monocots because it is a major crop with a relatively small genome (~451 Mbp). It already has fairly well-developed genomics resources e.g. its genome has been sequenced, thousands of ESTs and full-length cDNAs are available, and mutant lines have been generated. However, it is not an ideal system to address biological questions specific to temperate grasses such as freezing tolerance, vernalization requirements and cold-season crop resistance to certain pathogens. Furthermore, its relatively longer life cycle, large physical stature and special semi-aquatic growth requirement makes working with rice expensive and time-consuming. The development of *Brachypodium* as an experimental model organism for grasses will strongly complement rice.

In January 2006, an International Brachypodium Initiative was formed to promote collaboration among researchers to develop genetic and genomic resources to establish *Brachypodium* as an experimental system for biological investigations in grasses (http://www.brachypodium.org/IBI). A diverse collection of inbred diploid and polyploid accessions has been shared (http://www.brachypodium.org/Bd_Accessions.html; http://www.aber.ac.uk/plantpathol/germplasm.htm). The sequencing of ESTs and the *Brachypodium* genome is under way (http://www.jgi.doe.gov/sequencing/why/CSP2007/brachypodium.html). BAC

Communicated by J. W. Snape.

D. L. Laudencia-Chingcuanco (✉) · W. H. Vensel
Western Regional Research Center, USDA-ARS,
Albany, CA 94710, USA
e-mail: dlc@pw.usda.gov

libraries have been constructed, molecular markers are being characterized, mutagenized and mapping populations are being generated.

With a large seed relative to its plant size, *B. distachyon* will be suitable for studies in seed development and grain filling. Unlike rice caryopsis, *B. distachyon* caryopsis morphology includes a crease—a feature that has been of interest to manipulate to increase the recovery of flour from milled crops like wheat and barley. With its small plant size, self-fertility, good seed set, transformability and ease of growth, mutational and molecular approaches could be more conveniently utilized in *B. distachyon* to dissect and manipulate grain morphology and biochemical pathways with the aim to improve major cereals' end-use quality.

To understand the process of grain filling, we have begun to characterize the major seed storage compounds, starch and storage proteins, and the enzymes involved in their synthesis. In this paper, we report that globulins are the major seed storage proteins in *B. distachyon*.

Materials and methods

Plant materials and protein gel electrophoresis

B. distachyon caryopses for accessions Bd4, Bd14, Bd17 and Bd21 were generously provided by Dr. John Vogel, USDA-ARS, Albany, CA (Vogel et al. 2006a). *Brachypodium* total seed protein was prepared by crushing a caryopsis between a glassine paper with a hammer and the remains suspended in 200 µl PAGE loading buffer [2% SDS, 50 mM DTT, 62.5 mM Tris (pH 8.5), 10% glycerol, 0.1 mg/ml pyronin] (Kasarda et al. 1998). In total, 2–3 µl of protein extracts were loaded onto 4–12% acrylamide Novex NuPAGE Bis–Tris gel (Invitrogen, Carlsbad, CA) and separated using the Novex MES SDS running buffer. Protein band molecular weights were estimated using Multimark 12 (Invitrogen) proteins as molecular weight markers. Gels were stained overnight using Brilliant Blue G (Sigma–Aldrich, St. Louis, MO) and destained with water. Gels were photographed and analyzed using the Versadoc Quantity one software (Bio-Rad Laboratories, Hercules, CA).

Mass spectrometry-based protein identification

Protein-containing bands from SDS–PAGE gels were identified by electron spray ionization mass spectrometry as previously described (Vensel et al. 2005). In preparation for identification by mass spectrometry the protein-containing bands, excised from 1D electrophoretic (1DE) gels, were deposited in a 96-well reaction plate from a DigestPro robot (Itavis, Koeln, DE) that automatically performs destaining, reduction, alkylation and tryptic digestion of the protein. The sample receiving plate from the DigestPro was placed in the autosampler that was interfaced to the QSTAR PULSAR *i* quadrupole time-of-flight (TOF) mass spectrometer (Applied Biosystems/MDS Sciex, Toronto, CA). Gradient RP-HPLC and QSTAR instrument parameters for operation and data collection were as previously described (Vensel et al. 2005).

Tandem mass spectra were extracted from the QSTAR AnalystQS wiff files using Mascot Daemon (http://www.matrixscience.com/). The software used to analyse the MS/MS data files was obtained from the Global Proteome Machine (GPM) organization. The local installation of the GPM open-source software (http://www.thegpm.org/) was used for visualization and analysis of the data. The spectrum modeler X! TANDEM (Craig and Beavis 2004; Fenyo and Beavis 2003) that is a part of the GPM software was used to match MS/MS fragmentation data to peptide sequences using X! Tandem (version 2005.06.01.2) that had been configured to search a file containing 185,450 protein sequence entries from the NCBI nonredundant database (050601). The file contained sequences from the NCBI nr-Other-Viridiplantae.fasta, nr-Arabidopsis-thaliana.fasta, nr-Oryza-sativa.fasta sequences as well as a copy of the cRAP.fasta (common repository of adventitious proteins) available at www.thegpm.org. X! Tandem was configured to operate assuming that the digestion enzyme trypsin was used. The iodoacetamide derivative of cysteine was specified in X! Tandem as a fixed modification. Oxidation of methionine and iodoacetamide derivative of cysteine were specified in X! Tandem as variable modifications. Leucine and isoleucine are isobaric (similar mass) and cannot be distinguished by the analysis.

Globulin cDNA clone identification and characterization

DNA sequences of globulin genes for proteins that matched the Bd21 protein band were obtained from NCBI database. The gene sequence was used to query the *Brachypodium* EST database at http://wheat.pw.usda.gov/bEST. Identity of the putative *Brachypodium* globulin clones were verified by DNA sequencing.

DNA sequencing and analysis

Plasmids from *Brachypodium* cDNA clones were isolated and purified using the QIAGEN Spin Miniprep Kit (Valencia, CA). The nucleotide sequence of the cDNA inserts were determined using the ABI Prism BigDye Teminator v3.1 Cycle Sequencing kit and the ABI 3730 Genetic Analyzer (Applied Biosytems, Foster City, CA). The sequencing reaction was carried out in a 10 µl volume containing 300–500 ng DNA template, 3.2 pmol primer, 0.5 µl

BigDye Terminator and 2 µl 5× sequencing buffer. The cycle sequencing was carried out in an MJ Tetrad Thermocycler with the following conditions: 96°C for 5 min, followed by 40 cycles of three-step PCR (rapid ramp to 96°C for 10 s, rapid ramp to 50°C for 5 s and rapid ramp to 60°C for 4 min.). The initial gene-specific primers were designed from the original EST sequence or from the sequence derived using an M13F universal primer. Subsequent gene-specific primers were designed from the resulting sequences until the whole cDNA clone was fully sequenced at 3× or more coverage.

Sequence data were analyzed using the SeqMan and MegAlign software from DNAStar (Madison, WI) and Jalview. SignalP 3.0 server was used to predict the putative signal peptide (Bendtsen et al. 2004) from the deduced protein sequence.

Southern hybridization

Approximately 10 ug of *B. distachyon* Bd21 genomic DNA was digested with restriction enzymes *Hind*III, *Eco*R1, *Bam*H1, *Sac*I, *Sac*II and *Sst*I; resolved in a 1% agarose gel and transferred by capillary action onto a GeneScreen Plus membrane (NEW Research Products, Boston, MA). DNA was cross-linked to the nylon membrane by UV-light treatment at 200 mJ using a Stratalinker (Stratagene). The blot was hybridized with ^{32}P-labeled DNA probe in 20–40 ml hybridization buffer (15% formamide, 0.2 M sodium phosphate pH 7.2, 1 mM EDTA, 1% bovine serum albumin, 7% SDS) at 58°C overnight. After hybridization, the blot was washed twice in 2 × SSC–1.0% SDS at 58°C, and twice at 0.2 × SSC–0.1% SDS at 65°C for 30 min each under constant agitation. The hybridization signals were detected by exposing the blot on a KODAK X-OMAT film (Eastman Kodak Co., Rochester, NY). The probes were generated by amplifying fragment of the cDNA clones using gene-specific primers. Amplified products were purified using a QIAquick PCR Purification kit (QIAGEN Inc., Valencia, CA).

Results

Identification of the major seed storage proteins

To identify the major seed storage proteins in *B. distachyon* accession 21 Bd21), total protein from two caryopses were independently extracted and resolved in a denaturing 1D polyacrylamide gel (SDS–PAGE). Seven dominant bands ranging from 20 to 57 kD in size were visualized after staining (Fig. 1). When total protein was extracted from caryopses with excised embryos, the same seven prominent bands were observed suggesting that these proteins are primarily from the endosperm.

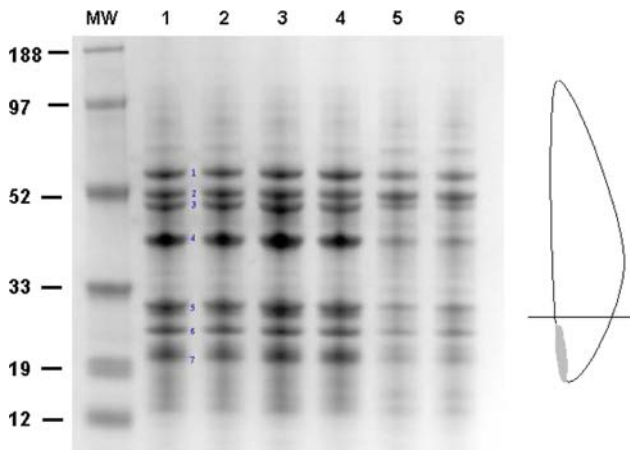

Fig. 1 SDS–PAGE of reduced *Brachypodium* total seed proteins, (1–2) whole caryopsis (3–4) endosperm section (5–6) embryo section. The number adjacent to the major protein bands is the designated band ID for each protein. A diagram of a *Brachypodium* caryopsis on the *right* indicates where the cut was made to separate the endosperm from the embryo section (embryo designated by *dark* area). The estimated size of each molecular weight marker band (in kiloDaltons, kD) is indicated on the *left*

Each of the seven major bands was separately excised from the gel for protein analysis by mass spectrometry. Two bands that migrated close to other proteins were carefully separated: the upper protein band at 29 kD and the lower protein band at 20 kD were designated protein bands 5 and 7, respectively. Table 1 shows the sequences of the peptides that gave the closest match to the spectra of the trypsin-digested protein fragments derived from each band.

Protein bands 1 and 2 gave the same peptide sequence of VAVLEANPR. A database search showed that this peptide sequence is present in the partial sequence of *Globulin-1* (*Glb1*) encoded protein derived from *Zea mays* subsp. *parviglumis* (GenBank accession AAC31462) (Hilton and Gaut 1998). *Glb1* encodes a 7S globulin, which is the major storage protein found in maize embryo and aleurone tissue (Belanger and Kriz 1989; Kriz and Schwartz 1986), and maps in the long arm of maize chromosome 1 (Schwartz 1979).

Two to three protein fragments—identified from band 3 indicate that band 3 protein is similar to oat 12S globulin protein (GenBank accession 1515394A). The 12S globulins are the major storage proteins in oat endosperm (Brinegar and Peterson 1982; Chesnut et al. 1989).

Bd21 protein band 4 gave two different peptide sequences both of which correspond to unknown proteins sequenced from rice.

Peptides derived from band 5 are similar to a subset of those obtained for band 3. Protein database search shows that band 5 is also most similar to the oat 12S globulin that was identified for band 3 protein. Some of the fragment sequences obtained for bands 6 and 7 also overlapped with

Table 1 Peptides from Bd21 identified by mass spectrometry

Protein band	Size (kD)	Caryopsis 1 peptides	Caryopsis 2 peptides	Best hit GB ID	*E* value	Protein ID
1	57	VAVLEANPR		AAC31462	−1.4	Globulin
2	52	VAVLEANPR	VAVLEANPR	AAC31462	−1.3	Globulin
				AAC31462	−1.2	
3	50	SQAGITEYFDEQNEQFR	SQAGITEYFDEQNEQFR	1515394A	−20.8	Globulin
			VFDVNNNANQLEPR	1515394A	−24.8	
			ALPVDVLANAYR			
4	42	QTSGINNHVREGEDPQK	CQAQPSRAQPPPPSTPM	XP_450420	−1.3	Unknown
				XP_449986	−1.9	
5	29	SQAGITEYFDEQNEQFR	SQAGITEYFDEQNEQFR	1515394A	−1.9	Globulin
			VFDVNNNANQLEPR	1515394A	−6.5	
6	25		PILNLVQLSATR	BAD28627	−27.5	Globulin
			TNANSMVSHLAGK			
			ANSMVSHLAGK			
			PVDVIANAYR			
7	20		SSIFRALPTDVLANAYR	1311273A	−9.4	Globulin

Blank space indicates that no peptide fragment was successfully identified. Best hit Genbank accession ID in column 5 refers to candidate protein identified by BlastX that best matched the peptides in columns 3 and 4

the sequences derived from band 3. Bands 6 and 7 show similarity to rice globulin proteins known as glutelins, which are the major endosperm storage proteins in rice (Higuchi and Fukazawa 1987; Okita et al. 1989; Yamagata et al. 1982).

Globulins are the major seed storage proteins in other *Brachypodium* lines

To further verify whether globulins are the major storage proteins in *B. distachyon*, the identities of major seed storage proteins in three other accessions were also determined by SDS–PAGE and mass spectrometry (Table 2). The identities of a subset of the major bands resolved in an SDS–PAGE gel as shown in Fig. 2 indicates that globulins are also the major seed storage proteins in Bd4, Bd14 and Bd17, three accessions with hexaploid genomes (Vogel et al. 2006a).

Bd4 57 kD protein (band1) contains the peptide sequence VAVLEANPR that matched the sequence of AAC31462, maize *Glb1* globulin protein, the same protein that was identified for band 1 of Bd21. Bd14 band 5 detected two peptides SQAGITEYFDEQNEQFR and FDVNNNANQLEPR, which are similar to those obtained for Bd21 band 3.

Three protein bands in Bd17 gave detectable fragments. Two peptides were obtained for band 4: GRITHLNSKNFPTLNLVQMSATR and ALPVDVLANAYR. A protein database search identifies Bd17 band 4 to be similar (*E* value $10^{-10.3}$) to oat 12S globulin protein 1515394A as identified by Bd21 band 3. Bd17 band 7 peptide (CTGVFIR) matched CAA54153, another 12S globulin from oat. Band 8 peptide (SQAGITEYFDEQNEQFR) best matched oat 12S globulin 1515394A, similar to the protein identified as Bd21 band3.

Identification and sequencing of the *Brachypodium* globulin cDNA clones

More than 20,000 *Brachypodium* ESTs from five different cDNA libraries have been sequenced and are available in the GenBank database (Vogel et al. 2006a). To identify

Table 2 Peptides from Bd4, Bd14 and Bd17 identified by mass spectrometry

Accession	Protein band ID	Peptide(s)	Best hit GB ID	*E* value	Protein ID
Bd4	1	VAVLEANPR	AAC31462	−1.3	Globulin
Bd14	5	SQAGITEYFDEQNEQFR FDVNNNANQLEPR	1515394A	−5.8	Globulin
Bd17	4	GRITHLNSKNFPTLNLVQMSATR ALPVDVLANAYR	1515394A	−10.3	Globulin
Bd17	7	CTGVFIR	CAA54153	−1.8	Globulin
Bd17	8	SQAGITEYFDEQNEQFR	1515394A	−1.9	Globulin

Protein band ID refers to the position of the major bands in a lane in Fig. 2. Best match Genbank accession ID in column 4 refers to candidate protein identified by BlastX that best matched the peptides in column 3

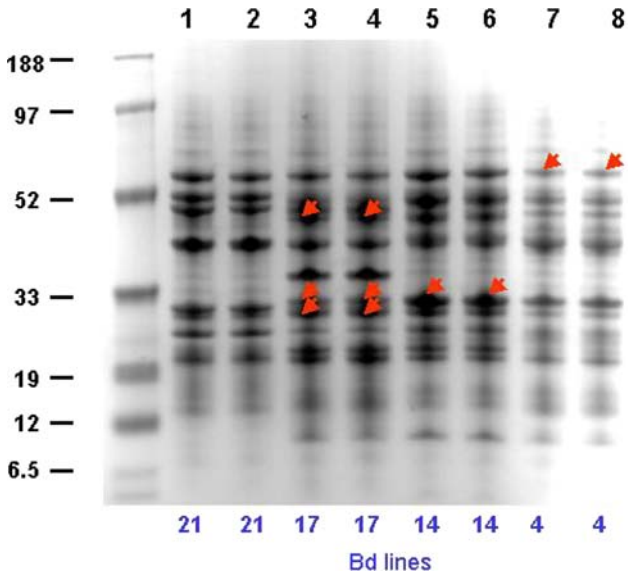

Fig. 2 SDS–PAGE polypeptide patterns of *Brachypodium distachyon* caryopsis proteins (1–2) Bd21, diploid; (3–4) Bd17, hexaploid; (5–6) Bd14, hexaploid; (7–8) Bd4, hexaploid. *Arrows* indicate bands that gave detectable peptides and whose identities were determined by mass spectrometry: Bd17 bands 4, 7 and 8; Bd 14 band 5 and Bd4 band 1

Brachypodium clones that potentially code for the above major seed proteins, the DNA sequences for the candidate protein that best matched each of the Bd21 protein bands were obtained from GenBank using TBLASTN (Altschul et al. 1990). The DNA sequences were then used to query a *Brachypodium* EST database (http://wheat.pw.usda.gov/bEST). The EST clones identified were resequenced to verify their identity.

To obtain cDNA clones coding for Bd21 band 1 or band 2 polypeptides, the partial sequence for maize globulin gene AF064219 (from which AAC31462 protein sequence was deduced) was used. Only one EST clone, generated from the callus cDNA library, was identified (at *E* value 7×10^{-35}). The cDNA clone contains a 1.778 kb insert and analysis of its sequence indicated that it encodes a globulin gene, herein designated *Bd.glo1*, with the deduced polypeptide sequence containing the peptide VAVLEANPR (Fig. 3). It is not clear, however, whether this clone codes for the globulin protein found in the seed. The sequence of the full insert of *Bd.glo1* cDNA did not end in a poly-A tail suggesting that it may be incomplete. To verify whether the cDNA has an incomplete clone, the full sequence of the derived protein and DNA sequence of *Bd.glo1* were used to query the NCBI protein and DNA database. A vicilin-like 7S globulin protein from maize gave the best match (*E* value 6×10^{-120}) to BD.GLO1 and a barley 7S globulin gene best match (*E* value 2×10^{-83}) to *Bd.glo1* DNA sequence. The sequence of BD.GLO1 is missing the last 57 amino acids of the polypeptide compared to maize 7S globulin protein and 22 residues compared to the protein derived from the barley 7S globulin gene (data not shown), indicating that *Bd.glo1* encodes a partial globulin gene.

Oat 12S globulin 151539A, which best matches Bd21 protein band 3, has 99.6% identical sequence to the deduced

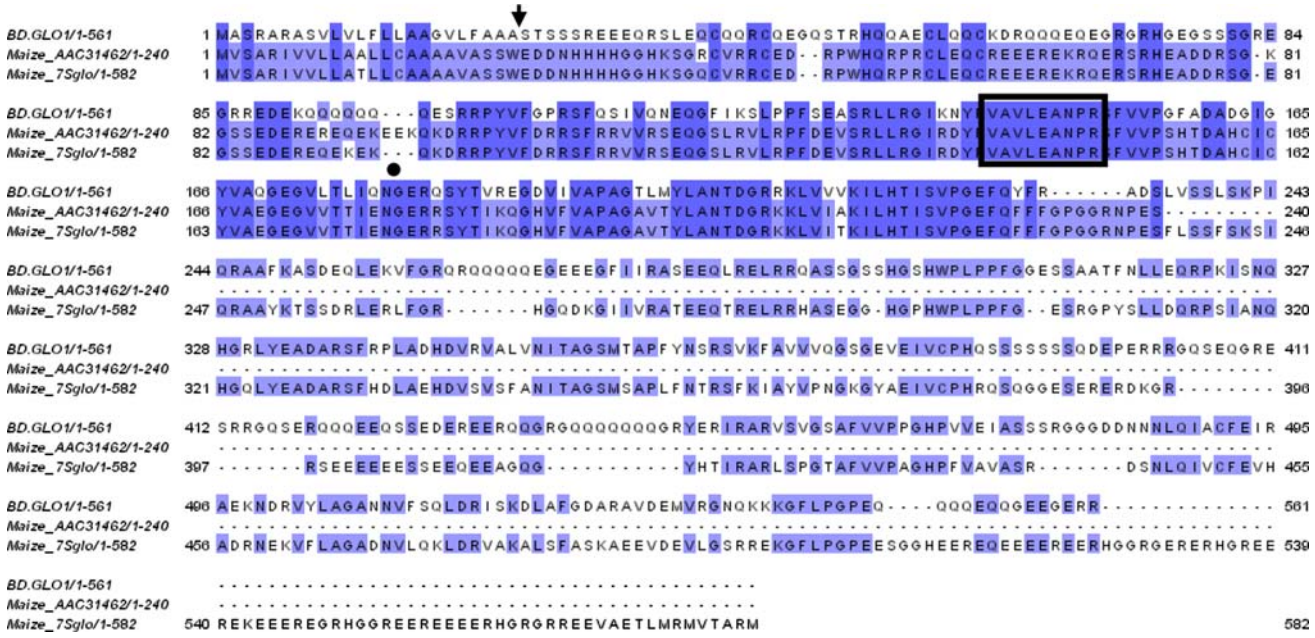

Fig. 3 Protein sequence alignment of BD.GLO1 and maize 7S globulins. The deduced polypeptide sequence of BD.GLO1 was aligned with maize 7S globulins accession AAC31462 and CAA41809. Identical sequences are indicated by *shaded* residues. The *arrow* at the top of the strand indicates the putative cleavage site of the presumed signal peptide. The *boxed* sequence corresponds to the peptide sequence detected by mass spectroscopy for Bd21 protein band 1 and 2. The *solid circle* indicates the potential cleavage site for asparaginyl endopeptidase

polypeptide from M21405. M12405 as a query sequence identified six similar *Brachypodium* ESTs with an *E* value of 10^{-10} or lower, all of which were from the *Brachypodium* seed cDNA library. Four of the clones were successfully resequenced (two failed to grow). Full-length sequences of the cDNA clones indicate that three were similar to oat 12S globulin and one was a prolamin-like gene. One of the 12S globulin-like clones was a pseudogene (data not shown). The other two, herein designated *Bd.glo2a* and *Bd.glo2b* cDNAs, are 99.2% identical to each other. The 1.640 kb *Bd.glo2a* and 1.657 kb *Bd.glo2b* sequences differed only in three nucleotides, one at the 5′ UTR and 2 at the 3′ UTR. It is possible that these two cDNAs are allelic transcripts of the same gene. The identical deduced polypeptides from *Bd.glo2a* and *Bd.glo2b* are 69.8% similar to oat 12S globulin 151539A polypeptide sequence (Fig. 4).

The rice DNA sequence for XM_450420, which encodes an unknown protein similar to Bd21 band 4, did not identify any *Brachypodium* ESTs from the database.

The candidate genes that best matched Bd21 protein bands 5–7 identified the same set of ESTs identified for protein band 3.

Brachypodium globulin protein structure

The primary sequences of the derived proteins from *Bd.glo1* and *Bd.glo2* show the typical domains found in 7S and 12S globulins, respectively (Figs. 3, 4). These globulins are synthesized as preproteins, which undergo proteolytic processing during maturation and deposition into storage organelles. Both the amino termini of the derived polypeptide of BD.GLO1 and BD.GLO2 displayed the canonical features of a signal peptide: a basic residue near the N-terminus and a leucine-rich hydrophobic core. Although the exact cleavage site has not been determined by N-terminal sequencing, the cleavage sites most likely occur between residue A24 and S25 in BD.GLO1 and between residue A24 and Q25 in BD.GLO2 based on von Heijne's rule for specifying signal peptides (Bendtsen et al. 2004; von Heijne 1986). BD.GLO2 without the 24 amino acid signal peptide results in a 50.7 kD protein which is close to the size determined for Bd21 band 3 protein.

Earlier reports have shown that as globulin protein is transported and deposited in the storage vacuoles, a specific post-translational cleavage occurs resulting in a mature subunit consisting of basic and acidic polypeptides (Jung et al. 1998; Walburg and Larkins 1983). The cleavage, catalyzed by an asparaginyl endopeptidase, occurs after an asparagine that is usually followed by a glycine residue. This conserved cleavage site is usually located seven amino acids before the cysteine residue involved in the disulfide linkage between the acidic and basic polypeptides (Jung et al. 1998). BD.GLO1 only has one potential cleavage site between N79 and G80. However, the downstream cysteine residue was absent. BD.GLO2 has six potential cleavage sites, but only two are located in a hydrophilic region of the protein (data not shown) that are likely to be available for enzyme action. Of the two it is more likely that N291–G292 is the cleavage site since it is the only pair followed by a cysteine residue at the right position.

Cleavage of BD.GLO2 between N291 and G292 will result in an acidic 29.7 kD polypeptide (with deduced

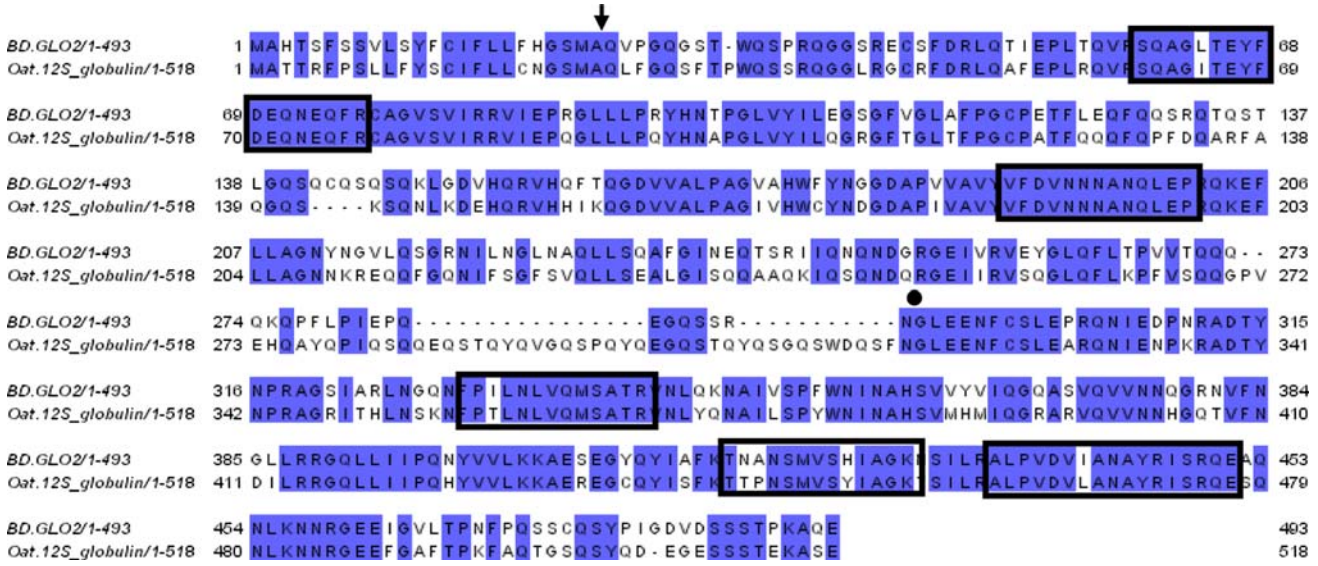

Fig. 4 Protein sequence alignment of BD.GLO2 and oat 12S globulin. The deduced polypeptide sequence of BD.GLO2 was aligned with oat 12S globulin accession AAA32720. *Shaded* residues indicate identical sequences. The *arrow* at the top of the strand indicates the cleavage site of the putative signal peptide. The *boxed* sequences correspond to the peptide sequence detected by mass spectroscopy for Bd21 protein bands 3, 5–7. The *solid circle* indicates the potential cleavage site for asparaginyl endopeptidase

isoelectric point of 6.0) and a basic 22 kD polypeptide (deduced isoelectric point of 9.1). These two polypeptides closely match the size of Bd21 protein bands 5 and 7, respectively. The derived sequence for the acidic 29.kD polypeptide contains the sequence for the peptides determined for band 5, whereas the sequence of the basic 22 kd polypeptide closely matches the sequence for the peptide determined for band 7. Thus, it is possible that Bd21 bands 5 and 7 are bi-products of the cleavage of proteins in band 3.

Both BD.GLO1 and BD.GLO2 are members of the cupin superfamily of proteins (Dunwell 1998; Dunwell et al. 2001; Khuri et al. 2001; Marchler-Bauer and Bryant 2004). Each protein contains two of the conserved cupin domains, thus are categorized as bicupins. The cupin domain contains a conserved beta-barrel structure, which was first described in wheat germins, thermostable glycosylated proteins produced during germination (Lane et al. 1991). Extensive sequence analysis has now identified representative members of this protein family in both prokaryotes and eukaryotes (Dunwell et al. 2000, 2004). The major seed storage globulins in plants (i.e. vicilin, legumin, glycinin, glutenins) are members of this family (Dunwell 1998).

Southern hybridization analysis

Southern hybridization analysis of Bd21 genomic DNA indicates that *Bd.glo1* hybridized strongly to a single band suggesting that it is a single copy gene. *Bd.glo2*, on the other hand, hybridized strongly to multiple bands implying that it belongs to a gene family (Fig. 5). Cross hybridization was not observed between the two gene probes at the stringency used. This was expected since the *Bd.glo1* and *Bd.glo2* are only 11.3% similar at the nucleotide level.

Discussion

The phylogenetic placement of *Brachypodium* makes it a more appropriate system than rice, the traditional monocot model organism, to explore agronomic traits of interest in temperate grasses, which includes important crops like wheat, barley and oat.

To begin to investigate its utility as an experimental model plant and to dissect grain development and grain filling we identified its major seed storage proteins. Storage proteins account for about half the seed proteins in cereals, and are a major source of protein in human diet. Understanding the mechanism of seed storage protein synthesis, trafficking and deposition could provide insights into development of effective approaches to improve cereal grain nutritional quality and functional properties in end-use products.

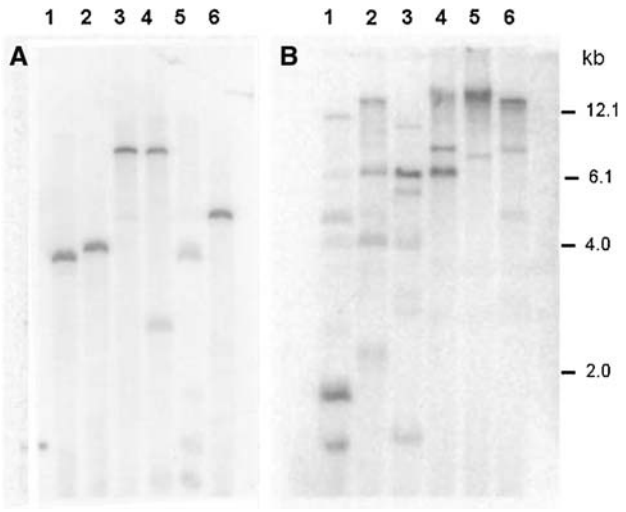

Fig. 5 Southern blot analysis of Bd21 genomic DNA using *Bd.glo1* and *Bd.glo2* as probes. Brachypodium genomic blots hybridized with A. *Bd.glo1* and B. *Bd.glo2*. The restriction enzymes used to digest the genomic DNA areas follows: lanes 1: *Hind*III 2: *Eco*R1 3: *Bam*H1 4: *Sac*I 5: *Sac*II and 6: *Sst*I. Sizes of molecular markers (in kilobase pairs nucleotide, kb) are indicated on the right side of the panels

Plant proteins are traditionally classified based on their solubility in a series of solvents. Water-soluble proteins are classified as albumins, saline-soluble proteins as globulins and alcohol-soluble proteins as prolamins. Proteins in these solubility classes are present in seeds but one type usually predominates in different plant families. For example, globulins are the predominant storage proteins in rice and oat, whereas, prolamins predominate in maize and in Triticeae (wheat, barley and rye). Progress in genomic techniques and the development of analytical algorithms designed to compare sequences, however, has led to the explosion of deduced protein sequences that has uncovered similarities in molecular structure among proteins in different protein solubility classes. Our understanding of the evolutionary relationship between seed storage proteins will greatly improve with the identification of members of each protein family based on molecular domain similarities. Whole genome sequences from model organisms like *Brachypodium* will propel progress towards this goal.

Brachypodium seed storage globulins

Six of the seven major protein bands of *Brachypodium* seed extracts resolved on 1DE gels are globulins ranging in size from 20 to 57 kD. These globulins fall into two classes, the 7S (vicillin-like) and 12S (legumin-like) globulins. It has been previously reported that both the types of globulins undergo post-translational proteolysis and modification as they are synthesized into the endoplasmic reticulum and deposited in the storage vacuoles. The legumin-like globulins are cleaved into basic and acidic polypeptides upon

maturation. Bd21 protein band 3, 5–7 belong to the legumin-like globulins.

We hypothesize that the three smaller bands (5–7) may be derived from the cleavage of a precursor protein similar to band 3. Consistent with this hypothesis, *Bd.glo2*, the seed globulin cDNA clone identified encodes a 493 amino acid preprotein (BD.GLO2) of 54.7 kD in size. Cleavage of the putative 24 amino acid signal peptide at the N-terminus of its polypeptide will yield a 50.7 kD protein, which is similar to the size of Bd21 band 1 protein. BD.GLO2 also contains a conserved cleavage site for asparaginyl endopeptidase, which cleaves globulins to basic and acidic polypeptides. The cleavage of the 50.7 kD BD.GLO2 between N291 and G292 residues would yield a 202 amino acid basic polypeptide and a 267 amino acid acidic polypeptide. The 22.3 kD basic polypeptide is well within the size of the 20 kD and 25 kD Bd21 protein bands 6 and 7. The 29.7 kD acidic polypeptide, on the other hand, correlates well with the size of Bd21 protein band 5. Furthermore, the 22.3 kD basic polypeptide contains the sequences for peptides derived from Bd21 protein bands 6 and 7, whereas, the 29.7 kD acidic polypeptide contains the sequences of the peptides derived for Bd21 protein band 5 by mass spectrometry. The acidic and basic polypeptides correspond to the N-terminal and the C-terminal halves of BD.GLO2. Our Southern blot analysis indicates that *Bd.glo2* belongs to a gene family. Bd21 protein bands 6 and 7 may be C-terminal halves of proteins encoded by two distinct members of *Bd.glo2* family.

Cupin family

The cupin family is comprised of proteins with a common beta-barrel structure. The family includes proteins from both prokaryote and eukaryote in origin. The cupin motif can be present in one or more copies in a protein. In plants, it is found as a single domain in fern sporulins and wheat germins/oxalate oxidase proteins, whereas, two domains are present in several seed storage globulins e.g. maize 7S and oat 12S globulins. Analysis of cupins identified at least 18 different functional classes including cell wall synthesis, desiccation-tolerance, defense response and transcription regulation of genes (Dunwell et al. 2001, 2004).

The two classes of globulin genes in *Brachypodium*, *Bd.glo1* and *Bd.glo2*, identified here are only 11% similar at the nucleotide level but their deduced proteins share the same molecular motifs, identifying them as members of the cupin superfamily. BD.GLO1 protein sequence is more similar to other plant bicupins like a sucrose binding protein from *Vicia faba* (AJ292221, 31%) and a globulin from wheat (WHTGLB1A, 54%) than to BD.GLO2. Conversely, BD.GLO2 protein is more similar to triticin (AAB27108, 54%), another wheat globulin, and to oat 12S globulin (AAA32720, 69.8%) than to BD.GLO1.

Brachypodium is evolutionarily closer to *Avena* than *Triticum*

The phylogeny of the grass family based on the combined data from chloroplast restriction sites, sequences of six nuclear encoded genes and plant morphology shows that *Brachypodium* is ancestral but equally closely related to *Avena* and *Triticum* (Kellogg 2001). Although both globulin and prolamin-like proteins are present in *Brachypodium*, like in *Avena*, globulins are its major seed storage proteins. Data from this work would imply that globulins are the ancestral storage protein (since they are found in *Avena* and *Brachypodium*), whereas prolamins are derived in the Triticeae. It would be interesting to determine whether the switch to prolamins as the main seed storage proteins in *Triticum* after it diverged from *Avena* occurred before or after it separated from *Bromus*. Analysis of the promoters and expression of globulin genes from these sister pooids might provide fundamental insights into the regulation and evolution of seed storage proteins in grasses.

Concluding remarks

The *Brachypodium* globulin genes identified in this study could serve as a biomarker to study the processes behind grain filling and a vehicle to dissect seed storage biology in cereals. The new globulin gene sequences from *Brachypodium* could be used in comparative studies of seed storage protein evolution in plants. With the ease of transforming *Brachypodium* and the availability of its sequenced genome (www.Brachypodium.org) fluorescent tagged globulin proteins could be used as *in planta* biosensors to dissect the regulation of storage protein synthesis, accumulation, transport and deposition into protein bodies and degradation during seed germination. Because *Brachypodium* is amenable to large-scale mutagenesis projects, *Brachypodium* overexpressing a globulin transgene fused to a selectable marker can be used to identify genes that regulate seed storage protein gene expression. As in *Brachypodium*, most cereal seeds contain both saline-soluble globulins and alcohol-soluble prolamins. Globulins have been shown to be easier to digest and contain a more balanced set of the amino acids essential to the human diet compared to prolamins. Understanding the regulation of seed storage protein in *Brachypodium* could provide new knowledge that could be used as a basis for the development of cereals with a wider spectrum of seed protein content and quality.

Acknowledgments We would like to thank Tiffany Chui and Xiaoyang Liu and Dr. Naxin Huo, for excellent technical assistance. The authors would also like to thank Drs. Grace Chen and Michael Gitt, and Gerard Lazo for the critical reading of the manuscript. Specific product name mentioned in this article does not constitute an endorsement and does not imply a recommendation over other suitable products. The USDA-ARS CRIS Project 5325-21000-011 funded this work.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Belanger FC, Kriz AL (1989) Molecular characterization of the major maize embryo globulin encoded by the Glb1 gene. Plant Physiol 91:746–750

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: Progress, problems and prospects. Ann Bot 95:45

Brinegar AC, Peterson DM (1982) Separation and characterization of oat globulin polypeptides. Arch Biochem Biophys 219:71–79

Catalan P, Shi Y, Armstrong L, Draper J, Stace CA (1995) Molecular phylogeny of the grass genus Brachypodium P.Beuav. based on RFLP and RAPD analysis. Bot J Linn Soc 117:263

Chesnut RS, Shotwell MA, Boyer SK, Larkins BA (1989) Analysis of avenin proteins and the expression of their mRNAs in developing oat seeds. Plant Cell 1:913–924

Christiansen P, Andersen CH, Didion T, Folling M, Nielsen KK (2005) A rapid and efficient transformation protocol for the grass Brachypodium distachyon. Plant Cell Rep 23:751–758

Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20:1466–1467

Draper J, Mur LA, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge AP (2001) *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol 127:1539–1555

Dunwell JM (1998) Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. Biotechnol Genet Eng Rev 15:1–32

Dunwell JM, Khuri S, Gane PJ (2000) Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. Microbiol Mol Biol Rev 64:153–179

Dunwell JM, Culham A, Carter CE, Sosa-Aguirre CR, Goodenough PW (2001) Evolution of functional diversity in the cupin superfamily. Trends Biochem Sci 26:740–746

Dunwell JM, Purvis A, Khuri S (2004) Cupins: the most functionally diverse protein superfamily? Phytochemistry 65:7–17

Fenyo D, Beavis RC (2003) A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal Chem 75:768–774

Hasterok R, Draper J, Jenkins G (2004) Laying the cytotaxonomic foundations of a new model grass, *Brachypodium distachyon* (L.) Beauv. Chromosome Res 12:397–403

Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G (2006) Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using BAC landing with fluorescent in situ hybridization. Genetics 173(1):349–362

von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14:4863–4690

Higuchi W, Fukazawa C (1987) A rice glutelin and a soybean glycinin have evolved from a common ancestral gene. Gene 55:245–253

Hilton H, Gaut BS (1998) Speciation and domestication in maize and its wild relatives: evidence from the globulin-1 gene. Genetics 150:863–872

Jung R, Scott MP, Nam YW, Beaman TW, Bassuner R, Saalbach I, Muntz K, Nielsen NC (1998) The role of proteolysis in the processing and assembly of 11S seed globulins. Plant Cell 10:343–357

Kasarda DD, Woodard KM, Adalsteins AE (1998) Resolution of high molecular weight glutenin subunits by a new SDS–PAGE system incorporating a neutral pH buffer. Cereal Chem 75:70–71

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

Khuri S, Bakker FT, Dunwell JM (2001) Phylogeny, function, and evolution of the cupins, a structurally conserved, functionally diverse superfamily of proteins. Mol Biol Evol 18:593–605

Kriz AL, Schwartz D (1986) Synthesis of globulins in maize embryos. Plant Physiol 82:1065–1075

Lane BG, Bernier F, Dratewka-Kos E, Shafai R, Kennedy TD, Pyne C, Munro JR, Vaughan T, Walters D, Altomare F (1991) Homologies between members of the germin family inhexaploid wheat and similarities between these wheat germins and certain *Physarum spherulins*. J Biol Chem 266:10461–10469

Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32:327–331

Okita TW, Hwang YS, Hnilo J, Kim WT, Aryan AP, Larson R, Krishnan HB (1989) Structure and expression of the rice glutelin multigene family. J Biol Chem 264:12573–12581

Schwartz D (1979) Analysis of the size alleles of the Pro gene in maize: evidence for a mutant protein processor. Mol Gen Genet 174:233–240

Vensel WH, Tanaka CK, Cai N, Wong JH, Buchanan BB, Hurkman WJ (2005) Developmental changes in the metabolic protein profiles of wheat endosperm. Proteomics 5:1594–1611

Vogel JP, Garvin DF, Leong OM, Hayden DM (2006a) Agrobacterium-mediated transformation and inbred line development in the model grass *Brachypodium distachyon*. Plant Cell Tissue Organ Cult 84:199

Vogel JP, Gu YQ, Twigg P, Lazo GR, Laudencia-Chingcuanco D, Hayden DM, Donze TJ, Vivian LA, Stamova B, Coleman-Derr D (2006b) EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. Theor Appl Genet 113:186–195

Walburg G, Larkins BA (1983) Oat seed globulin: subunit characterization and demonstration of its synthesis as a precursor. Plant Physiol 72:161–165

Yamagata H, Sugimoto T, Tanaka K, Kasai Z (1982) Biosynthesis of storage proteins in developing rice seeds. Plant Physiol 70:1094–1100