

# Molecular Basis of the Waxy Endosperm Starch Phenotype in Broomcorn Millet (*Panicum miliaceum* L.)

Harriet V. Hunt,<sup>\*1</sup> Kay Denyer,<sup>2</sup> Len C. Packman,<sup>3</sup> Martin K. Jones,<sup>4</sup> and Christopher J. Howe<sup>5</sup>

<sup>1</sup>McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, United Kingdom

<sup>2</sup>John Innes Centre, Norwich Research Park, Norwich, United Kingdom

<sup>3</sup>Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

<sup>4</sup>Department of Archaeology, University of Cambridge, Cambridge, United Kingdom

<sup>5</sup>Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

Research for this work was carried out at the McDonald Institute for Archaeological Research and the Department of Biochemistry, University of Cambridge, and at the John Innes Centre, Norwich.

\*Corresponding author: E-mail: hvh22@cam.ac.uk.

Associate editor: Peter Lockhart

## Abstract

Waxy varieties of the tetraploid cereal broomcorn millet (*Panicum miliaceum* L.) have endosperm starch granules lacking detectable amylose. This study investigated the basis of this phenotype using molecular and biochemical methods. Iodine staining of starch granules in 72 plants from 38 landrace accessions found 58 nonwaxy and 14 waxy phenotype plants. All waxy types were in plants from Chinese and Korean accessions, a distribution similar to that of the waxy phenotype in other cereals. Granule-bound starch synthase I (GBSSI) protein was present in the endosperm of both nonwaxy and waxy individuals, but waxy types had little or no granule-bound starch synthase activity compared with the wild types. Sequencing of the *GBSSI* (*Waxy*) gene showed that this gene is present in two different forms (L and S) in *P. miliaceum*, which probably represent homeologues derived from two distinct diploid ancestors. Protein products of both these forms are present in starch granules. We identified three polymorphisms in the exon sequence coding for mature GBSSI peptides. A 15-bp deletion has occurred in the S type *GBSSI*, resulting in the loss of five amino acids from glucosyl transferase domain 1 (GTD1). The second *GBSSI* type (L) shows two sequence polymorphisms. One is the insertion of an adenine residue that causes a reading frameshift, and the second causes a cysteine–tyrosine amino acid polymorphism. These mutations appear to have occurred in parallel from the ancestral allele, resulting in three *GBSSI*-L alleles in total. Five of the six possible genotype combinations of the S and L alleles were observed. The deletion in the *GBSSI*-S gene causes loss of protein activity, and there was 100% correspondence between this deletion and the waxy phenotype. The frameshift mutation in the L gene results in the loss of L-type protein from starch granules. The L isoform with the tyrosine residue is present in starch granules but is nonfunctional. This loss of function may result from the substitution of tyrosine for cysteine, although it could not be determined whether the cysteine isoform of L represents the functional type. This is the first characterization of mutations that occur in combination in a functionally polyploid species to give a fully waxy phenotype.

**Key words:** granule-bound starch synthase, broomcorn millet, *Panicum miliaceum*, waxy starch, cereal.

## Introduction

The evolution and diversification of crop plants has been shaped over thousands of years by conscious and unconscious human selection on a wide range of phenotypic traits. In plants cultivated primarily as a carbohydrate source, including cereals, loci that determine starch quality have been among the key targets for human modification (Whitt et al. 2002; Olsen et al. 2006). Plant starch comprises two polymers: amylose, a predominantly linear chain of  $\alpha$ -1,4-linked glucosyl units, and amylopectin, a highly branched molecule with short  $\alpha$ -1,4-linked glucosyl chains linked by  $\alpha$ -1,6-linkages. Cereal endosperm starch typically comprises 15–30% amylose and 70–85% amylopectin. Amylose synthesis is catalyzed by granule-bound starch synthase I (GBSSI, EC 2.4.1.242), which catalyzes the exten-

sion of the amylose molecule by the transfer of  $\alpha$ -D-glucose from adenosine diphosphate (ADP)-glucose.

Mutations at the *GBSSI* (also called *Waxy* or *Wx*) locus that result in amylose being present in low levels or absent in endosperm starch (Fukunaga et al. 2002)—the waxy phenotype—have been characterized in several species. In barley, a 413-bp deletion in the promoter and 5' untranslated region of the *GBSSI* gene disrupts transcription, reducing the production of GBSSI protein to a fraction of its levels in the nonwaxy type (Domon et al. 2002; Patron et al. 2002). In maize and in foxtail millet (*Setaria italica* [L.] P. Beauv.), gene expression is disrupted by diverse transposable element insertions (Wessler and Varagona 1985; Wessler et al. 1990; Marillonnet and Wessler 1997; Kawase et al. 2005). In rice, a G  $\rightarrow$  T mutation at the splice site of intron 1 results in incomplete processing of pre-mRNA

(Hirano and Sano 1991; Wang et al. 1995; Hirano et al. 1998; Isshiki et al. 1998). In sorghum, a single-nucleotide polymorphism (SNP) has been identified, which causes a Glu to His mutation that may inactivate the GBSSI protein (McIntyre et al. 2008). In all the above diploid crops, the waxy phenotype has arisen by spontaneous mutations. In hexaploid bread wheat, a waxy line has been produced by crossing partially waxy cultivars with spontaneous mutations in the three GBSSI homeologues in the A, B, and D genomes (Nakamura et al. 1995). These have been characterized, respectively, as a 23-bp deletion in the *Wx-A1b* allele that affects transcript splicing or a 173-bp transposon insertion in an exon; a deletion of the entire GBSSI transcription unit in the *Wx-B1b* allele; and a 588-bp deletion in the *Wx-D1b* allele that leads to the deletion of 30 amino acids from the C-terminus (Vrinten et al. 1999; Saito et al. 2004). Distinct single-nucleotide insertions and deletions causing frameshift mutations have been identified as responsible for the independent origin of null *Wx-A1* alleles in emmer wheat (Saito and Nakamura 2005).

With the exception of waxy strains of emmer and bread wheats, which have been generated by modern breeding programs, waxy varieties of cereals originated in east and southeast Asia, where they have been selected in response to a cultural preference for glutinous-type (i.e., waxy) starchy foods (Sakamoto 1996; Olsen and Purugganan 2002). *Panicum miliaceum* L. (broomcorn, proso, or common millet), a tetraploid species ( $2n = 4x = 36$ ), is one of the world's oldest and historically most important domesticated cereals; recent data attest to its presence as early as 10,000 cal BP in northern China (Crawford 2009; Lu et al. 2009). Its probable center of domestication was in this region (although with possible independent domestications further west in Eurasia; Zohary and Hopf 2000; Jones 2004; Hunt et al. 2008), and there is currently considerable interest in its early cultivation and domestication, as revealed through diverse analytical techniques including stable isotopic analysis, phytoliths, and lipid biomarkers (Jacob et al. 2008; Barton et al. 2009; Lu et al. 2009). These factors make this crop an interesting species in which to investigate the selection of waxy mutants. Distinct terms indicating glutinous broomcorn millet are recorded in Chinese classical texts dating back some 2,000 years (Sakamoto 1996). Waxy millet therefore provides an ideal system to explore the biochemistry and molecular biology that lie behind ancient food-preference choices.

Previous work (Graybosch and Baltensperger 2009) has shown that the waxy trait in this tetraploid species is determined by recessive alleles at two loci. In the current study, we have provided the first characterization of these loci, obtaining DNA sequence for the region corresponding to the full length of the mature GBSSI protein for two GBSSI homeologues. We found evidence for the presence of these two forms of GBSSI in starch granules. At one locus, we have established a clear link between a genetic mutation and its biochemical effect of loss of GBSSI activity, leading to loss of amylose synthesis and waxy-type endosperm. At the other locus, we identified three alleles, of which two result in the loss of active GBSSI protein. It was not possible to determine whether the third allele represents the func-

tional version of this homeologue, but consideration of the allele distributions and genotype combinations enables us to establish a hypothesis for the evolution of the waxy phenotype in this species.

## Materials and Methods

### Plant Material

Accessions of *P. miliaceum* germplasm were provided by the USDA-ARS North Central Regional Plant Introduction Station, Ames, IA, and by the Vavilov Research Institute, St Petersburg, Russia. The majority of the accessions were described as having landrace status. Grain from 38 accessions was grown in greenhouses at the University of Cambridge Botanic Garden. Between one and three plants per accession (72 plants in total) were analyzed for endosperm starch phenotype and subsequently for GBSSI genotype (table 1). Eight of these 72 plant samples (highlighted in table 1)—accession 3o plant #1, 3y #1, 04 #1, 47 #1, 70 #1, 71 #1, 76 #1, and 82 #1—were analyzed for starch protein, starch synthase (SS) activity, and sequenced for DNA corresponding to the full length of the mature GBSSI protein.

### Starch Phenotyping

The development of endosperm starch is determined by the triploid ( $3n$ ) endosperm genome (Sano 1984; McIntyre et al. 2008). *Panicum miliaceum* is largely self-pollinated, but crosspollination may exceed 10% (Baltensperger 1996); we excluded the possibility of this by bagging plants at the first sign of panicle development in perforated cellophane bags (Focus Packaging & Design Ltd, Louth, Lincolnshire, United Kingdom). Grains were harvested at maturity, crushed individually between glass slides, and stained with Lugol's solution (10% [w/v] KI [Sigma–Aldrich Ltd., Gillingham, Dorset, United Kingdom], 5% [w/v] I<sub>2</sub> [Sigma–Aldrich Ltd.]), diluted 100-fold with water immediately prior to use. Three grains were analyzed per plant. Starch-granule color was observed under 20× objective magnification on a microscope (Nikon, Tokyo, Japan). Slides were photographed using a DN100 camera (Nikon).

### Extraction and Analysis of Starch Protein

Starch protein from the eight plant samples highlighted in table 1 was extracted and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). A single mature grain was dehusked and crushed in an Eppendorf tube. The powdered grain was resuspended in 80- $\mu$ l extraction buffer (50 mM 2-[4-(2-hydroxyethyl) piperazin-1-yl]ethanesulfonic acid (HEPES) pH 8.0, 1 mM dithiothreitol (DTT), 10 mM ethylenediaminetetraacetic acid (EDTA) and centrifuged in a benchtop centrifuge at 14,000 rpm for 5 min at 4 °C. The supernatant was transferred to a second Eppendorf tube, and the pellet was resuspended in 80- $\mu$ l extraction buffer to give equivalent volumes in the pellet and supernatant fractions. Twenty-microliter sample buffer (55 mM Tris–HCl pH 6.8, 2% SDS, 10% glycerol, 11 mg ml<sup>-1</sup> DTT, bromophenol blue to color) was added to each fraction, and they were

**Table 1.** Accessions of *Panicum miliaceum* Used in This Study.

Accession Number	Origin	Variety	Accession Source and Code	Individual Plant Number	MSR11 Genotype	M12R12 Genotype	Int5LFR3 Genotype	Overall Genotype	Starch Granule Phenotype
1	China Beijing		USDA-ARS, PI408805	1		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				3		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
2	Kazakhstan		USDA-ARS, PI346938	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
3o	Korea south		USDA-ARS, Ames 12701	1	15-bp deletion	—	A	S <sub>-15</sub> /L <sub>Y</sub>	Waxy
				2	15-bp deletion	—	A	S <sub>-15</sub> /L <sub>Y</sub>	Waxy
				3	15-bp deletion	—	A	S <sub>-15</sub> /L <sub>Y</sub>	Waxy
3y	Korea south		USDA-ARS, Ames 12701	1	15-bp deletion	—	A	S <sub>-15</sub> /L <sub>Y</sub>	Waxy
				2	15-bp deletion	—	A	S <sub>-15</sub> /L <sub>Y</sub>	Waxy
4	Kyrgyzstan		USDA-ARS, PI346936	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
12	China Shanxi	sanguineum	VIR, #8966	1		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
16	Mongolia	mongolicum, tephrum	VIR, #464	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				3		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
24	Kyrgyzstan	flavum	VIR, #8525	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
29	Kazakhstan Aqtobe	dacicum	VIR, #3773	1		Heterozygote	G	S <sub>0</sub> /L <sub>C</sub> /S <sub>0</sub> /L <sub>f</sub> heterozygote	Nonwaxy
				2		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
31	Ukraine Poltava	flavum	VIR, #3009	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
				2		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
38	Azerbaijan	coccineum	VIR, #1546	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
				2		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
47	Russia Samara		VIR, #9052	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
				2		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
48	Russia Tatarstan	flavum	VIR, #2804	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
53	Russia Orenburg	sanguineum	VIR, #9438	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
54	Russia Altai	flavum	VIR, #2392	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
55	Russia Omsk	vitellinum	VIR, #2825	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
56	Russia Buryatia	tephrum	VIR, #8222	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				3		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
57	Russia Irkutsk	mongolicum	VIR, #316	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				3		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
58	Russia Amur	tephrum	VIR, #8508	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				3		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
59	Russia Amur	tephrum	VIR, #8545	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
60	Russia Primorskiy Kray	badium	VIR, #8571	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
				2		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
61	Russia Primorskiy Kray	tephrum	VIR, #50	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
66	Mongolia		VIR, #509	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
67	China NE	tephrum	VIR, #1175	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy

Table 1. Continued

Accession Number	Origin	Variety	Accession Source and Code	Individual Plant Number	M5R11 Genotype	M12R12 Genotype	Int5Lfr3 Genotype	Overall Genotype	Starch Granule Phenotype
68	China Manchuria	sibiricum	VIR, #1371	1	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy
				2	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy
69	China NE	tephrum	VIR, #1999	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
70	China NE	subcooccineum	VIR, #2012	1	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy
71	China NW		VIR, #2282	1		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				3		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
72	China NW	flavum/aureum	VIR, #2301	1		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
				2		—	A	S <sub>0</sub> /L <sub>Y</sub>	Nonwaxy
75	China	aureum	VIR, #3750	1	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy
76	China NW	ochroleucum	VIR, #3790	1		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
				2		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
77	China NE	album	VIR, #8803	1		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
				2	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy
78	China N	sanguineum	VIR, #8966	1		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
79	China NW		VIR, #9079	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
80	China NW		VIR, #9095	1		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
				2		—	G	S <sub>0</sub> /L <sub>C</sub>	Nonwaxy
82	China Heilongjiang	atrocastaneum	VIR, #9205	1	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy
				2	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy
84	China		VIR, #10155	1		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
				2		A	G	S <sub>0</sub> /L <sub>f</sub>	Nonwaxy
85	China		VIR, #10238	1	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy
				2	15-bp deletion	A	G	S <sub>-15</sub> /L <sub>f</sub>	Waxy

A = frameshift mutation  
G gives cysteine codon,  
A gives tyrosine codon

NOTE.—Origin and variety information as given by the germplasm source institution. Key: USDA-ARS = US Department of Agriculture-Agricultural Research Service; VIR = Vavilov Institute of Research. Results of genotyping at the S and L loci, and starch phenotyping with iodine, are shown. Where genotype is not specifically indicated, plants did not have the 15-bp deletion at the S locus and did not have the additional (frameshift) A at the L locus. Entries in bold typeface indicate plants analyzed by SDS-PAGE and Western blotting, assayed for SS activity, and sequenced for the full-length exon 2–exon 14 region.

heated at 90 °C for 3 min. The samples were cooled and centrifuged, and the supernatant was analyzed by SDS-PAGE on 7.5% polyacrylamide gels that were electroblotted onto polyvinylidene difluoride (PVDF) membrane. Antiserum was raised to a 60-kDa protein excised from an SDS-PAGE gel of granule-bound proteins from developing barley (*Hordeum vulgare* L.) endosperm (Smith AM, personal communication). Rabbit sera containing the resulting antibarley endosperm GBSSI antibodies in a 1:2,000 (v/v) dilution, and the secondary antibody IgA antirabbit phosphatase conjugate, were used to develop blots following the method of Denyer, Barber, et al. (1997). In a second experiment, powdered grain was resuspended and washed three times in 2% SDS (instead of extraction buffer) then once in H<sub>2</sub>O. The resulting suspension was spun through 80% CsCl and washed once more in H<sub>2</sub>O prior to the addition of sample buffer and subsequent analysis as above.

### GBSS Activity Assays

SS activity assays were carried out on the same eight plant samples analyzed with SDS-PAGE. Three replicate starch

preparations were made for each plant sample. Starch preparation followed a method modified from Denyer et al. (1995). Twenty grains were dehusked and ground in a chilled mortar, then resuspended in 1-ml extraction buffer (0.1 M Tris–acetate pH 7.0, 0.5 M NaCl, 1 mM DTT, and 1 mM EDTA). The homogenate was centrifuged in a benchtop centrifuge at 13,000 rpm for 3 min at room temperature and the supernatant discarded. The pellet was washed in the same way once more with 1-ml extraction buffer, three times in 1-ml wash buffer (50 mM Tris–acetate pH 8.0, 1 mM DTT, and 1 mM EDTA), and once in ice-cold 100% acetone. The pellet was air dried and stored at –20 °C prior to assaying.

SS assays were carried out in 100 µl volumes containing 100 mM Bicine (pH 8.5), 25 mM potassium acetate, 10 mM DTT, 5 mM EDTA, 20 µl of 50 mg ml<sup>-1</sup> potato amylopectin in water (freshly boiled and then cooled to room temperature), 2 mM ADP[U<sup>14</sup>C]glucose (at 4.6 GBq mol<sup>-1</sup>), and 20 µl starch suspension (50 mg ml<sup>-1</sup> starch in 100 mM 3-morpholinopropane-1-sulfonic acid (MOPS) [pH 7.2], 5 mM MgCl<sub>2</sub>, 50 ml l<sup>-1</sup> glycerol, and 2 mM DTT, 1 g l<sup>-1</sup>

bovine serum albumin). Assay mixtures were incubated for 30 min at 25 °C, and then 3 ml ice-cold 75% (v/v) aqueous methanol containing 1% (w/v) KCl was added to stop the reaction. Controls were stopped by the addition of methanol/KCl immediately after the addition of the starch-granule suspension to the rest of the assay. After at least 5-min precipitation in methanol/KCl, starch was collected by centrifugation in a benchtop centrifuge at 3,000 rpm for 5 min at room temperature. The supernatant was discarded, and the starch pellet was resuspended in 300 µl water. The methanol–KCl precipitation and water resuspension steps were repeated a further two times. After the final resuspension in water, 3 ml Hisafe II scintillant was added, and radioactivity in starch was determined by scintillation counting. All assays and controls were performed in triplicate for each replicate starch sample.

### DNA Extraction, Amplification, and Analysis

Leaf tissue was freeze dried and ground using a Qiagen Tissue Lyser. DNA was extracted using either a Qiagen Plant DNeasy kit (Qiagen Ltd, Crawley, West Sussex, United Kingdom) or a hexadecyltrimethylammonium bromide (CTAB) protocol modified from Rogers and Bendich (1994). In the latter, powder was resuspended and incubated in a mix of 500 µl CTAB buffer (2% [w/v] CTAB, 0.1 M Tris–HCl pH 8.0, 1.4 M NaCl, 0.02 M EDTA), 50 µl Sarkosyl buffer (10% [w/v] *N*-lauryl sarcosine, 0.1 M Tris–HCl pH 8.0, 0.02 M EDTA), and 5 µl β-mercaptoethanol) at 60 °C for 1 h. One volume of chloroform/isoamyl alcohol (24: 1) was then added and the mixture vortexed to form an emulsion. Following 3-min centrifugation at 13,000 rpm in a benchtop centrifuge, the supernatant was transferred into a clean tube, and the chloroform/isoamyl alcohol clean-up stage was repeated. Two-third volume of ice-cold isopropanol was added to the supernatant. Precipitated DNA was pelleted by 3-min centrifugation at 13,000 rpm in a benchtop centrifuge, and pellets were washed in 500 µl 70% (v/v) ethanol. Following a final centrifugation as above, the ethanol supernatant was discarded and pellets air dried prior to resuspension in 100 µl water.

Initial experimentation with primers from Fukunaga et al. (2002) to the *S. italica* GBSSI sequence consistently demonstrated the presence of two GBSSI loci in *P. miliaceum*, which we designated the short (“S”) and long (“L”) genes. Following further extensive experimentation with primers designed against the *S. italica* sequence, we used the following protocols to amplify the region corresponding to the entire mature peptide for the S and L loci in the eight plants previously analyzed for starch protein and enzyme activity. The L locus was amplified in a single polymerase chain reaction (PCR) using the primers FPSLVVC3 and Rstop3 (table 2), in 50 µl volumes using 1× Finnzymes HF buffer (New England Biolabs, Hitchin, Hertfordshire, United Kingdom), 200 µM deoxynucleoside triphosphates (dNTPs), 0.3 µM of each primer, 3% dimethyl sulfoxide, 1 U Finnzymes Phusion High-Fidelity DNA Polymerase (New England Biolabs). Cycling conditions were 30 s

at 98 °C; 40 cycles of 10 s at 98 °C, and 2 min 30 s at 72 °C; final extension step of 10 min at 72 °C. The S locus could not be reliably amplified as a single product, apparently due to the presence of a poly-C motif in intron 7 that often caused the reaction to fail. We therefore amplified the S locus in three overlapping fragments. The 5′ and 3′ regions were amplified using the primer combinations FPSLVVC3 and ex7S<sub>ext</sub>, and int7S<sub>f</sub> and Rstop3, respectively. Reactions were performed as for the amplification of the L locus above, except using 40 cycles of a three-step PCR cycle of 10 s at 98 °C, 30 s at 71 °C/64 °C, respectively, 1 min at 72 °C. We amplified the central fragment of the S gene using a nested PCR strategy. The primary PCR used the primers int5S<sub>f</sub> and R11 and was carried out according to the same protocol as the above reactions, with the differences that MgCl<sub>2</sub> was added to give a final concentration of 2 mM and the three-step cycling program was as follows: 40 cycles of 20 s at 98 °C, 30 s at 64 °C, and 30 s at 72 °C. The cleaned product from this reaction was used as template in a secondary PCR using the primers M17 and Ex8r, which followed the same protocol as the primary PCR.

The primers FPSLVVC3 and Rstop3 overlapped with the N- and C-termini of the mature GBSSI peptide, as expected by comparison with GBSSIs from other species (supplementary fig S1, Supplementary Material online), by 3 and 21 nt, respectively. Peptide mass fragment analysis (see below) was therefore used to confirm that the N- and C-termini matched the amino sequence anticipated by comparison with *S. italica*.

Following identification of exon polymorphisms, one shorter fragment within the S sequence and two within the L sequence were amplified for all 72 plant samples analyzed for starch phenotype, using the primers M5 and R11, M12 and R12, and int5L<sub>f</sub> and R3, respectively (fig. 1 and table 2 show positions and details of primers used). PCRs were carried out on an Eppendorf MasterCycler thermocycler, in 25-µl volumes containing 1× PCR buffer, 0.3 µM of each primer, 200 µM dNTPs, 1.3 U DNA polymerase (Expand High-Fidelity PCR System, Roche Diagnostics Ltd., Burgess Hill, United Kingdom) and 1-µl template DNA. Reactions for the M5-R11 and M12-R12 fragments contained 1.5 mM MgCl<sub>2</sub>; reactions for the int5L<sub>f</sub>-R3 fragment contained 2.5 mM MgCl<sub>2</sub>. Cycling conditions for the M5-R11 and M12-R12 fragments were as follows: 2 min at 94 °C; 30 cycles of 30 s at 94 °C, 30 s at 58 °C, and 1 min 30 s at 72 °C; final extension step of 7 min at 72 °C. Cycling conditions for the int5L<sub>f</sub>-R3 fragment were 2 min at 94 °C, 35 cycles of 45 s at 94 °C, 30 s at 54 °C, and 2 min at 72 °C, with a final extension step of 7 min at 72 °C.

PCR products were checked for size on Tris/acetate/EDTA buffer–agarose gels, excised when necessary, and cleaned using an illustra GFX PCR purification kit (GE Healthcare, Amersham, United Kingdom) or a Qiagen gel purification kit (Qiagen Ltd). Cycle sequencing was performed using primers for each fragment as shown in table 2. Contigs were assembled in ChromasPro version 1.41 (Technelysium Pty, Ltd., Tewantin, Australia). Sequences were aligned in MEGA

**Table 2.** Primers Used for PCR and DNA Sequencing.

Amplicon	PCR Primers	Sequencing Primers	Sequence (5' to 3')	Direction	Source
L gene, exon 2–exon 14	FPSLVVC3	FPSLVVC3	TCCCCTCCCTCGTCTGCGC	F	This study
	Rstop3	Rstop3	AGGGAGCGGCCACGTTCTCCTT	R	This study
		Ex5R	CTGGCAAAGCAGGCTGAAACG	R	This study
		R3	TGGTAGTTGCTCTTGAGGTA	R	
		M2ext	GAGGTTCTTCCACTGCTACAAGC	F	This study
		M16	CACCGATTGCTCTGCATT	F	This study
		R16	TCACGACTAACCCATCACGC	R	This study
		M4	TTTCTGCATCCACAACATCT	F	Umeda et al. (1991)
		R13	GTAATTGCTCTTGCTGGGGTCCC	R	This study
		M12	CGTGACCATCTTCTCTGTA	F	This study
		R7	ACGAGTCCACCGTGACGC	R	Umeda et al. (1991)
		M5	GGACGTCAGCGAGTGGGACC	F	Umeda et al. (1991)
		M7	CACCAGCCGCTTCGAGCCCT	F	Umeda et al. (1991)
	S gene, exon 2–exon 7/intron 7	FPSLVVC3	FPSLVVC3	TCCCCTCCCTCGTCTGCGC	F
ex7Srext		ex7Srext	GGGGCAAACCTTTGCATTCTGTAGATG	R	This study
		Ex5R	CTGGCAAAGCAGGCTGAAACG	R	This study
		M2ext	GAGGTTCTTCCACTGCTACAAGC	F	This study
		R3	TGGTAGTTGCTCTTGAGGTA	R	Umeda et al. (1991)
		M17	TACACACTGCCATCTGTTGTGC	F	This study
S gene, intron 5–exon 8	int5Sf		TGCATTTAAACAAGGGGCGAGTACTG	F	This study
	R11		CAGGCACACTGCTCCCAATG	R	This study
	secondary PCR				
	M17	M17	TACACACTGCCATCTGTTGTGC	F	This study
S gene, intron 7–exon 14	Ex8r	Ex8r	CGAAGGATGACCTGAACCTCTCA	R	This study
	int7Sf	int7Sf	TCGTTTGTATGTTGCTAGACGCC	F	This study
	Rstop3	Rstop3	AGGGAGCGGCCACGTTCTCCTT	R	This study
		R13	GTAATTGCTCTTGCTGGGGTCCC	R	This study
		R11	CAGGCACACTGCTCCCAATG	R	This study
		R7	ACGAGTCCACCGTGACGC	R	Umeda et al. (1991)
		M5	GGACGTCAGCGAGTGGGACC	F	Umeda et al. (1991)
L gene, intron 5–exon 7	int5Lf	int5Lf	ATGTTTGAATGAATGCTCC	F	This study
	R3		TGGTAGTTGCTCTTGAGGTA	R	Umeda et al. (1991)
	M12	M12	CGTGACCATCTTCTCTGTA	F	This study
L gene, intron 8–intron 9	R12		CGACGACGAACTCTCAACAC	R	This study
	M5	M5	GGACGTCAGCGAGTGGGACC	F	Umeda et al. (1991)
S gene, exon 9–intron 10	R11		CAGGCACACTGCTCCCAATG	R	This study

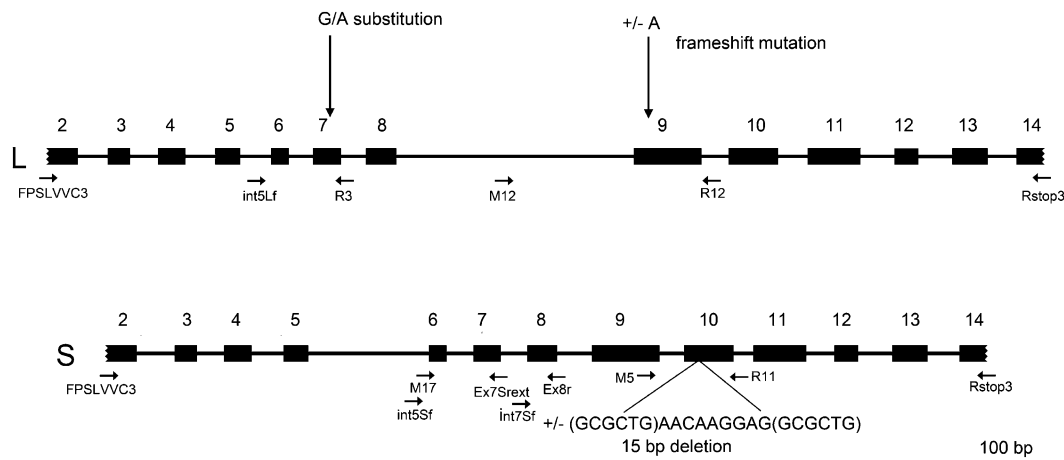
version 4.0 (Tamura et al. 2007) and predicted intron–exon boundaries and amino acid sequence established by reference to the published sequence for *S. italica* (Fukunaga et al. 2002). Sequences have been submitted to GenBank; accession numbers are given in table 3.

We aligned the predicted amino acid sequences in *P. miliaceum* with GBSS sequences downloaded from GenBank for a range of monocots and dicots using the ClustalW alignment tool in MEGA 4.0. A maximum likelihood tree of relationships between these taxa, with 1,000 bootstrap replicates, was estimated using the PhyML online web server (Guindon et al. 2005), using the Jones, Taylor, and Thornton (JTT) + I + G model of protein-sequence evolution, selected using the Akaike information criterion in ProtTest (Abascal et al. 2005). Tree files were edited using Dendroscope version 2.2 (Huson et al. 2007).

### Proteomics

In an initial experiment, starch-granule protein from the samples 4 #1/47 #1 (samples combined), 76 #1, 3y #1, and 82 #1 was analyzed to identify the proteins present. Gel plugs (1 mm × 1 mm) containing a band corresponding to a protein approximately 52 kDa in size were cut from

the SDS-PAGE gel, washed twice for 20 min in freshly prepared 400 mM ammonium bicarbonate: 100% acetonitrile (1:1), twice in 100% acetonitrile for 1 min and then for 15 min, and then air dried for 10 min. The plugs were incubated for 3 h at 37 °C with 5 μl 10 mM ammonium bicarbonate containing 50-ng modified porcine trypsin (Promega, Madison, WI). After the addition of 5 μl 5% formic acid, the plugs were incubated at room temperature for 10 min. A Dionex U3000 high performance liquid chromatography system was used to deliver the peptides at a flow rate of 150 nl min<sup>-1</sup> to the mass spectrometer (LTQ Orbitrap, Thermo Electron Corp., Runcorn, Cheshire, United Kingdom). Peptides were trapped and desalted using a precolumn (C18 pepmap100, LC Packings) and then separated on an analytical column (self-pulled to a length of 12-cm and 50-μm ID and self-packed with Waters BEH130 C18, 1.7 μm) with a gradient of 5–45% acetonitrile in water/0.1% formic acid at 0.66% increase per minute. The mass spectrometer was operated in positive ion mode with a nanospray source and a capillary temperature of 200 °C. The source voltage and focusing voltages were tuned for the transmission of peptide Met-Arg-Phe-Ala (m/z 524). Data-dependent analysis was carried out in



**Fig. 1.** Structure of the GBSSI sequence analyzed in *Panicum miliaceum* showing the L and S gene types (putative homeologues) and the relative sizes of the exons (wide bars, numbered analogously to *Setaria italica*) and introns. Approximate positions and 5' to 3' direction of primers used for PCR are marked with arrows (primers not to scale). The mutations identified in the L and S loci are identified.

Orbitrap-IT parallel mode (collision induced dissociation fragmentation) on the five most abundant ions in each cycle. The Orbitrap was run with a resolution of 30,000 over the MS range from *m/z* 400 to *m/z* 1,800 and an MS target of  $1e^6$  and 1-s maximum scan time. The MS2 was triggered by a minimal signal of 5,000 with a target of  $2e^4$  and 200-ms scan time. For selection of 2+ and 3+ charged precursors, charge state and monoisotopic precursor selection was used. Collision energy was 35, and an isolation width of 2 was used. Dynamic exclusion was set to 1 count and 60-s exclusion with an exclusion mass window of  $-0.5$  to  $+1.2$ . Raw files were processed in Bioworks to generate data files. The merged data files were used to search the SPTrEMBL (Viridiplantae) database with Mascot 2.2 (Matrixscience) (in-house), with a peptide tolerance of 5 ppm and a fragment tolerance of 0.6 Da, allowing up to three missed cleavages.

In a second experiment, selected samples were reanalyzed to identify the C-terminal sequence of the proteins present. Bands were cut from the gel as above and washed for 30 min with 200  $\mu$ l 50% acetonitrile in 0.1M ammonium bicarbonate containing 5 mM Tris-[2-carboxyethyl]-phosphine, followed by the addition of 20  $\mu$ l 250 mM iodoacetamide in water, for a further 30 min. They were then washed for 30 min in 400  $\mu$ l 50% acetonitrile in 0.1 M ammonium bicarbonate and dried under vacuum for 10 min. The plugs were incubated for 18 h

at 37 °C with 25  $\mu$ l 100 mM ammonium bicarbonate containing 10  $\mu$ g  $ml^{-1}$  proteinase (either AspN endoproteinase or trypsin). Peptides were recovered by binding to a conditioned  $\mu$ C18 ZipTip (Millipore, United Kingdom), washed with 5% acetic acid, and eluted with 2–5  $\mu$ l of 70% methanol/0.2% formic acid. Analysis was by matrix assisted laser desorption ionisation (MALDI) mass spectrometry (Waters Micromass MaldiMX Micro) using  $\alpha$ -cyano-4-hydroxycinnamic acid matrix (10 mg  $ml^{-1}$  in 50% aqueous acetonitrile/0.1% trifluoroacetic acid) and by nanoelectrospray ms/ms using a Thermo Finnigan LCQ Classic instrument. Desalted sample was delivered using a static nanospray source (Proxeon Biosystems, Denmark) at 0.5 kV, and peaks of interest were interrogated manually for mass and fragmentation using standard parameters recommended by the manufacturer. Results were analyzed using Qual Browser in Xcalibur 1.2 (Thermo), Mascot (Matrix Science, United Kingdom) and custom spreadsheets.

## Results

### Starch Phenotyping

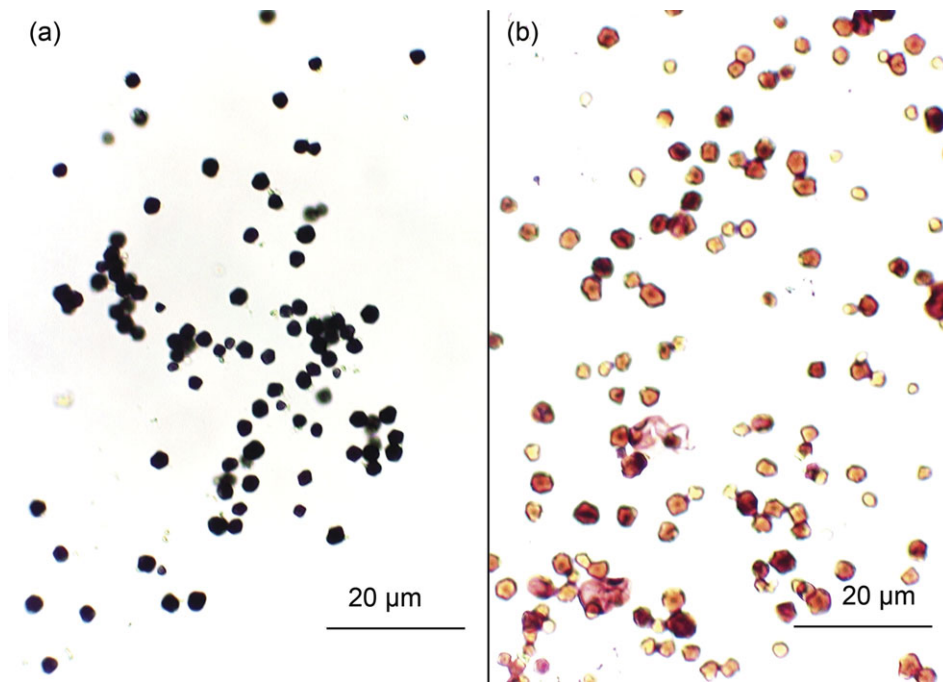
Fifty-eight plants, drawn from 31 different accessions, produced grain with endosperm starch that stained blue-black with iodine, indicating the presence of amylose (nonwaxy; fig. 2[a]). Fourteen plants, drawn from eight different accessions, produced grain with starch granules that stained red (waxy type; fig 2[b]).

### Starch Protein Analysis and GBSS Activity

We selected four nonwaxy and four waxy plant samples for analysis of starch protein. In all eight samples, two bands produced a crossreaction with antibarley GBSSI antiserum (four plant samples are shown in fig. 3[a]; the other four samples gave very similar results). One was approximately 52 kDa, slightly smaller than GBSSI in other species, which is typically close to 60 kDa (Denyer, Edwards, et al. 1997). Because each lane contains protein from the same amount of grain material, the intensity of the bands in the soluble and

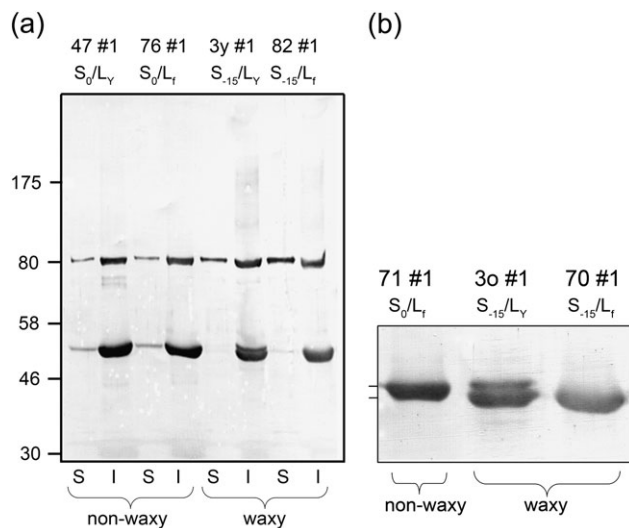
**Table 3.** Genbank Accession Numbers for the Eight Plants Sequenced for the Full-Length GBSSI Exon 2–Exon 14 Sequences.

Sample	Genotype	S Gene Sequence Accession Number	L Gene Sequence Accession Number
4 #1	S <sub>0</sub> /L <sub>C</sub>	GU199261	GU199253
47 #1	S <sub>0</sub> /L <sub>Y</sub>	GU199262	GU199254
71 #1	S <sub>0</sub> /L <sub>f</sub>	GU199263	GU199255
76 #1	S <sub>0</sub> /L <sub>f</sub>	GU199264	GU199256
3o #1	S <sub>-15</sub> /L <sub>Y</sub>	GU199265	GU199257
3y #1	S <sub>-15</sub> /L <sub>Y</sub>	GU199266	GU199258
70 #1	S <sub>-15</sub> /L <sub>f</sub>	GU199267	GU199259
82 #1	S <sub>-15</sub> /L <sub>f</sub>	GU199268	GU199260



**FIG. 2.** Starch granules of *Panicum miliaceum* stained with 0.01× Lugol's solution. (a) Nonwaxy type, stained blue–black (accession 58 plant 1). (b) Waxy type, stained red–brown (accession 85 plant 1).

insoluble fractions enables a comparison of the relative amount of the protein in each. Almost all of the 52-kDa protein was associated with the insoluble (granule-bound)



**FIG. 3.** (a) Western blot of soluble (S) and insoluble (I) protein fractions from endosperm starch, developed with anti-barley GBSSI antibody and IgA antirabbit phosphatase conjugate. The band at ~52 kDa represents the GBSSI protein. The larger ~80-kDa band most likely represents a second isoform of GBSS in broomcorn millet. (b) Detail of selected samples (insoluble protein fractions only) from a parallel blot (using different plants of the same genotypes) to that in figure 5(a). The  $S_{-15}/L_f$  genotype, lacking five amino acids, produces a slightly smaller GBSSI band than the  $S_0/L_f$  genotype. A double band can be seen in the  $S_{-15}/L_Y$  genotype, with a smaller band of the same size as in the  $S_{-15}/L_f$  sample and a larger band comparable in size with the  $S_0$  GBSSI. The larger band probably represents the (inactive)  $L_Y$  protein.

fraction in both nonwaxy and waxy samples. A second protein, approximately 80 kDa in size, was also observed in all samples and was distributed more evenly between the soluble and insoluble starch protein fractions. This protein most likely represents a second isoform of SS, probably SSII.

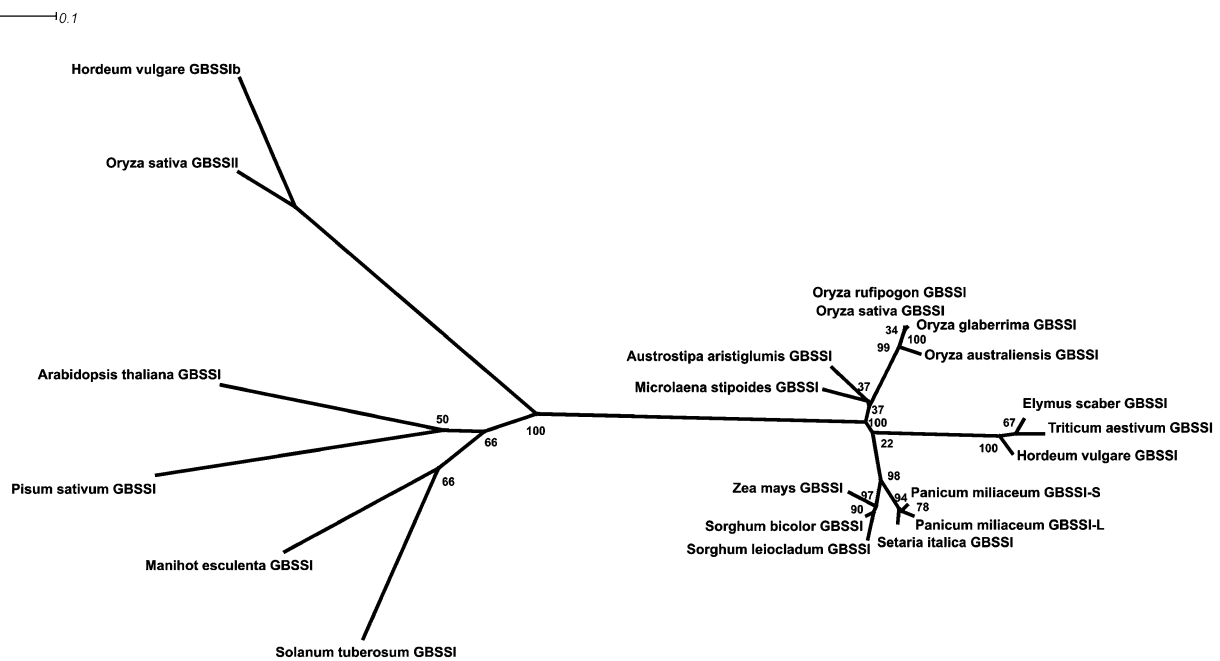
A close examination of the relative mobilities of the GBSSI proteins (fig. 3[b]) indicated that the predominant band in the waxy samples was of a slightly lower apparent molecular mass than that in the nonwaxy samples. In two of the waxy samples (3o #1 and 3y #1), a double band was seen, with the fainter upper band appearing equivalent in size to the band in the nonwaxy samples.

The four nonwaxy and four waxy plants analyzed for starch protein were assayed for GBSS activity. SS activity per mg starch was several-fold higher in the nonwaxy than in the waxy samples (table 6).

To check whether the waxy GBSSI in the buffer-insoluble pellet was entrapped in starch granules or insoluble because it was denatured, we carried out a second western blotting experiment in which pellets were washed with SDS and CsCl. This has previously been shown to remove proteins effectively from the outside of starch granules (Mu-Forster et al. 1996; Stoddard 1999) and should solubilize most insoluble denatured proteins. The results showed that some waxy GBSSI remained in the insoluble fractions, indicating that it was still at least partially granule bound (data not shown). Thus, loss of activity in these waxy mutant GBSSI proteins is not accompanied by loss of the ability to bind to granules.

### DNA Sequencing

Using a combination of primers designed against the *S. italica* GBSSI sequence and novel internal primers in



**Fig. 4.** Maximum likelihood tree of GBSS amino acid sequences, using the JTT + I + G model of protein evolution with 1,000 bootstrap replicates. Percentage bootstrap support values are shown. GenBank accession numbers for the GBSS proteins: *Arabidopsis thaliana* GBSSI NP\_174566; *Austrostipa aristiglumis* GBSSI ABU98330.1; *Elymus scaber* GBSSI ABU98331.1; *Hordeum vulgare* GBSSI CAA30755, GBSSIIb AAM74054; *Manihot esculenta* GBSSI CAA52273; *Microlaena stipoides* GBSSI ABU98332.1; *Oryza australiensis* GBSSI ABU98325.1; *Oryza glaberrima* GBSSI BAA01272; *Oryza rufipogon* GBSSI ABU98326.1; *Oryza sativa* GBSSI CAA46294, GBSSII AAL58572; *Panicum miliaceum* GBSSI-L type ADA61154, GBSSI S-type ADA61162 (this study); *Pisum sativum* GBSSI AAB26591; *Setaria italica* GBSSI BAC06486; *Solanum tuberosum* GBSSI CAA41359; *Sorghum bicolor* GBSSI Q43134; *Sorghum leiocladum* GBSSI ABU98327.1; *Triticum aestivum* GBSSI AAB26860; and *Zea mays* GBSSI P04713.

the *P. miliaceum* sequences, we were able to amplify and sequence 2 GBSSI products, around 3.6 and 3.2 kb in size, which we designated the large (henceforth L) and small (S) fragments, respectively (fig. 1). We identified exon–intron boundaries within the L and S sequences by alignment with the published sequence for *S. italica* GBSSI (AB089143) and numbered exons and introns correspondingly. Both the L and S sequences span the region from the start of the mature GBSSI peptide, midway through exon 2, to the end of exon 14. These two sequences were present consistently across accessions of *P. miliaceum*.

The alignment of *P. miliaceum* GBSS amino acid sequences with those of other species is available as [supplementary fig. S1](#), Supplementary Material online. All GBSSI sequences form a clade with 100% bootstrap support in the maximum likelihood tree, distinct from other isoforms of GBSS, demonstrating that both *Panicum* sequences are of GBSSI type (fig. 4) rather than of the GBSSIIb/GBSSII types typically expressed in leaves rather than endosperm. *Panicum miliaceum* GBSSI-S and -L types are closely related to *S. italica* GBSSI, forming a clade with 94% bootstrap support, consistent with the taxonomic position of these genera within Poaceae and their established relationship in grass phylogenies (e.g., Bouchenak-Khelladi et al. 2008). Within this clade, the two *Panicum* sequence types emerge as sister taxa with 78% bootstrap support.

High exon–sequence identity (95.3%) was seen between *P. miliaceum* L, *P. miliaceum* S, and *S. italica* GBSSI sequences, with 97.1% predicted amino acid identity. Eight

fixed amino acid differences were found between the *P. miliaceum* M4R9 L and S genes ([supplementary fig. S1](#), Supplementary Material online).

The GBSSI intron sequences aligned very weakly both between *P. miliaceum* L and *P. miliaceum* S and with *S. italica*. *Panicum miliaceum* L differed from S by a large deletion (ca. 300 bp) in intron 5 and a large insertion (ca. 700 bp) in intron 8. These indels accounted for most of the observed size difference between the two GBSSI sequences.

Three between-plant exon sequence polymorphisms were identified in *P. miliaceum* GBSSI. Polymorphism for a 15-bp indel was observed in the S gene, near the 5' end of exon 10, within the region containing the sequence GCGCTGAACAAGGAGGCGCTG. By comparison with other GBSSI sequences in the database, we inferred that this polymorphism arose from a deletion in some *P. miliaceum*. Because of the repetition of the motif GCGCTG in this region of the alignment, it was not possible to determine unambiguously the exact position of the deletion, but it results in a change of amino acid sequence from ...ALNKEAL... to ...AL... In the following discussion, we use  $S_0$  to refer to the allele or protein product lacking the deletion and  $S_{-15}$  for the allele or its product with the deletion. The repeated 6-bp motif suggests that this deletion may have arisen by recombination across this repeat.

The amino acid–sequence alignment showed that the ALNKEAL motif containing the site of the deletion in some S sequences is conserved among all Poaceae GBSSI



**Fig. 5.** Three-dimensional structure of the *Agrobacterium tumefaciens* glycogen synthase protein (Buschiazzo et al. 2004). The amino acids ALNKKAV that correspond to the ALNKEAL motif in the *Panicum miliaceum* GBSSI-S-type protein are highlighted and indicated with the arrow. A color version of this figure is available as [supplementary figure S2](#), Supplementary Material online.

sequences ([supplementary fig. S1](#), Supplementary Material online). The lysine residue in this motif is conserved across all SSs and in glycogen synthase of *Agrobacterium*. Threading of *Arabidopsis* SS amino acid sequence onto the 3D structure of glycogen synthase in *Agrobacterium* (Buschiazzo et al. 2004; Busi et al. 2007) shows that the five amino acid deletion in the S protein of *P. miliaceum* falls in a helix within the GTD1 ([fig. 5](#); a color version of this figure is available as [supplementary fig. S2](#), Supplementary Material online).

Two polymorphisms were observed in the L gene. One was an insertion/deletion of an adenine residue 19 nt from the 5' end of exon 9. By comparison with other GBSSI sequences, we inferred that the additional adenine represents an insertion in *P. miliaceum* that causes a shift in the reading frame, resulting in altered downstream inferred amino acid sequence and a novel stop codon 228 codons beyond the insertion. The second polymorphism was a substitution of an adenine for a guanine residue in exon 7, which causes a change from a cysteine codon to a tyrosine codon at amino acid position 249 (numbered according to the alignment in [supplementary fig. S1](#), Supplementary Material online). By comparison with other GBSSI sequences, we inferred that the cysteine is the wild-type amino acid, and the tyrosine represents a mutant form.

On the basis of these preliminary results, we designed fragment-specific primers, such that the primer pair M5/R11 amplifies the 391-bp region including the deletion site in the S gene, the primer pair M12/R12 amplifies the 632-bp region including the A insertion site in the L gene, and the primer pair int5Lf/R3 amplifies the 251-bp region including the SNP in the L gene. Genotyping results for the S and L loci are shown in [table 1](#).

At the L locus, all plants with the frameshift A insertion had the guanine residue at the SNP site, whereas plants without the frameshift mutation were polymorphic for the guanine-adenine substitution. Thus, we defined three

**Table 4.** Summary of Genotypes Observed in 72 Plants, with Associated Phenotypes.

	L <sub>C</sub>	L <sub>Y</sub>	L <sub>F</sub>
S <sub>0</sub>	31 Nonwaxy	15 Nonwaxy	11 Nonwaxy
S <sub>-15</sub>	0	5 Waxy	9 Waxy

NOTE.—One plant was heterozygous for S<sub>0</sub>/L<sub>C</sub> and S<sub>0</sub>/L<sub>F</sub>.

alleles of the L gene: L<sub>F</sub> (frameshift mutation), L<sub>C</sub> (no frameshift, cysteine codon at amino acid position 249), and L<sub>Y</sub> (no frameshift, tyrosine codon at amino acid position 249). We observed these L-gene alleles in combination with the S gene alleles S<sub>0</sub> and S<sub>-15</sub> such that five of the six possible combinations of genotypes were found. [Table 4](#) summarizes the number of plants of each genotype. No other polymorphic sites were seen in the exon sequence of either locus.

All of the 14 phenotypically waxy plants had the 15-bp deletion at the S locus. Nine of these plants had the L<sub>F</sub> allele with the additional A residue in exon 9, whereas five (all those from accessions 3o and 3y) had the L<sub>Y</sub> allele. Of the 58 nonwaxy plants, none had the 15-bp deletion at the S locus; at the L locus, 11 had the additional A (L<sub>F</sub> allele), 15 had no A insertion in combination with the tyrosine residue at aa position 249 (L<sub>Y</sub> allele), and 30 had no A insertion in combination with the cysteine codon at aa position 249 (L<sub>C</sub> allele). One plant was heterozygous for the L<sub>F</sub> and L<sub>C</sub> alleles ([table 1](#)). Because the plants (with this one exception for L) were homozygous at both the S and L loci, we can be confident that the endosperm genotypes were homozygous and identical to those of their parent plants, albeit with one extra copy. Thus, among 72 plants representing 38 different accessions, there was 100% correspondence between the waxy starch phenotype and the 15-bp deletion at the S locus. At the L locus, the waxy phenotype of plants with the genotypes S<sub>-15</sub>/L<sub>Y</sub> and S<sub>-15</sub>/L<sub>F</sub> is evidence that the L<sub>Y</sub> and L<sub>F</sub> alleles do not produce an active protein product. Because the L<sub>C</sub> allele does not occur in combination with the S<sub>-15</sub> allele, there is no direct evidence whether L<sub>C</sub> results in an active GBSSI protein or not.

### Proteomics

From the SDS-PAGE gels and the DNA sequence data, we inferred that the GBSSI protein present in each sample most likely represented the S<sub>0</sub> or S<sub>-15</sub> isoform, according to its genotype, and also the L<sub>C</sub> or L<sub>Y</sub> isoform in plants with those genotypes. We inferred that no L-type protein was present in plants with the L<sub>F</sub> allele, because we did not see a band of molecular mass ~45 kDa that would correspond to the truncated protein predicted from the frameshift. This suggests that the L<sub>F</sub> protein is not expressed or is unstable. The apparent lower molecular mass of the major band in the S<sub>-15</sub> compared with the S<sub>0</sub> samples is consistent with its missing five amino acids. It is likely that the larger band in the doublet in the samples with the genotype S<sub>-15</sub>/L<sub>Y</sub> represents the L<sub>Y</sub> protein. This band was absent from samples with the genotype S<sub>-15</sub>/L<sub>F</sub>.

To test these hypotheses regarding protein identity, we compared the peptide mass fingerprints of the GBSSI proteins excised from the lanes containing the insoluble protein fractions in the gel in [figure 3\(a\)](#) with those of in silico trypsin-digested plant proteins in the SPTrEMBL Viridiplantae database to which the *P. miliaceum* sequences  $S_0$ ,  $S_{-15}$ ,  $L_C$ , and  $L_Y$  had been manually added. We found that the top matches were all GBSSI proteins, confirming that the 52-kDa protein, present in both nonwaxy and waxy *P. miliaceum* starch, is indeed GBSSI. The top protein matches were  $S_0$  in the accessions with normal starch and  $S_{-15}$  in the accessions with waxy starch. Fragments enabling the discrimination of S from L protein were present in all samples ([table 5](#)). All genotypes had peptides specific to the S protein. These were numerous for all samples except the  $S_{-15}/L_f$  genotype (sample 82#1), which had low percentage coverage and overall score and showed only a few of the discriminatory fragments.  $L_C/L_Y$  genotypes also had numerous fragments specific to the L protein, which were almost or totally absent in  $L_f$  samples. This indicates that some of the GBSSI protein in these plants is contributed by the  $L_C$  or  $L_Y$  gene as appropriate. It was not possible to discriminate between the  $L_C$  and  $L_Y$  proteins, because the diagnostic fragment containing the cysteine–tyrosine residue was not found in any of the samples analyzed, probably because of its high molecular mass. To discriminate between  $S_0$  and  $S_{-15}$  isoforms, we looked for peptides characteristic of  $S_{-15}$ . These were present in  $S_{-15}$  genotype samples and almost or entirely absent from  $S_0$  genotype samples. Direct evidence for the presence of the  $S_0$  protein was limited because many of the fragments containing the additional five amino acids also occur in the  $L_C/L_Y$  proteins. This explains the presence of these fragments in the  $S_{-15}/L_Y$  genotype sample. The one fragment diagnostic of  $S_0$ , K.YDVSTAIAAKALNKE, is rare because it relies on a missed cleavage; it was found single-fold in the mixed  $S_0/L_C$  and  $S_0/L_Y$  genotype sample (which had the highest coverage and empAI score of any sample). We inferred the presence of the  $S_0$  protein in the  $S_0/L_{CY}$  and  $S_0/L_f$  genotypes indirectly from the absence of  $S_{-15}$  fragments in either, and in the latter sample, from the presence of  $S_0$  characteristic fragments in combination with the evidence suggesting the  $L_C$  or  $L_Y$  protein is lacking.

The low e-values of the peptides and manual checking of the ionic spectra confirmed the reliability of the fragment identifications. Very low levels of unexpected isoform-specific peptides are likely to be due to crosscontamination of the samples. The extensive coverage of the protein sequence and the comparative analysis of the four samples provide good corroborative evidence for the presence of waxy protein isoforms consistent with the genotype of each sample.

### N- and C-Terminal Identification

The 5' primer FPSLVVC3 and the 3' primer Rstop3 overlap with the N- and C-termini of the mature GBSSI protein, as predicted from alignment with other GBSSI sequences ([supplementary fig. S1](#), Supplementary

Material online), by 1 and 7 amino acids, respectively. This precludes identification of possible mutations in these sites from DNA sequencing, so to confirm the N-terminus of the mature peptide, we looked for the expected fragments in the peptide mass fingerprints. All four samples contained peptides matching the expected N-terminal sequence –.AAAGMNVVFGAEMAPWSK.T. Only one sample, genotype  $S_0/L_{CY}$ , with the highest percentage sequence coverage, showed the anticipated C-terminal fragment (K.ENVAAP–). We therefore carried out a second experiment using AspN proteinase to aim to recover a larger predicted C-terminal peptide (D489-P531; 4,457 Da average mass). This mass was successfully observed in  $S_0$  protein by MALDI analysis and its identity confirmed by electrospray ms/ms of the  $[M + 3H]^{3+}$  signal at 1,485.7m/z (monoisotopic). The Mascot score established its identity as the C-terminal peptide but with insufficient fragment coverage to confirm the C-terminal six residues (even though their accumulated mass was as expected). Solid identification was achieved by ms/ms of the  $[M + 4H]^{4+}$  signal at 1,114.5m/z. Manual interpretation of the spectrum identified a series of triply and quadruply charged b-ions which clearly derived from the C-terminus of the peptide. This allowed confident matching of the signals to the expected C-terminal sequence of IAPLAKENVAAP for  $S_0$  and  $S_{-15}$  protein samples ([fig. 6](#)).

### Discussion

Broomcorn millet represents a unique case among plant species with waxy mutants, in being a functional polyploid (Graybosch and Baltensperger 2009) in which waxy phenotypes have appeared and become established without deliberate modern breeding. This contrasts with the situation in wheat, in which partially waxy lines were discovered with mutations in one or two of the A, B, and D (in hexaploids) genome homeologues of GBSSI, and fully waxy tetraploid and hexaploid wheats have only very recently been synthesized through crossing appropriate partially waxy lines (Nakamura et al. 1993, 1995; Yamamori et al. 1994).

Although reticulate evolutionary relationships and the ancestry of polyploid taxa in the genus *Panicum* are not yet known, the presence of the S and L GBSSI genes in all samples is consistent with the inference from meiotic chromosome behavior by Hamoud et al. (1994) that *P. miliaceum* is an allotetraploid. We have observed comparable related pairs of other protein-coding nuclear sequences from *P. miliaceum* (Hunt HV, unpublished data). The L and S sequences would thus represent homeologues of the GBSSI gene derived from two (currently unknown) diploid ancestors of *P. miliaceum*. Further analysis of the evolutionary dynamics of Waxy sequences in the genus *Panicum* would be needed to confirm this interpretation; Fortune et al. (2007) demonstrated the presence of both paralogous and homeologous Waxy sequences in hexaploid *Spartina* species. However, the phylogenetic evidence clearly indicates that both the L and S types in

**Table 5.** MASCOT search output data from trypsin-digested samples from the gel in Fig. 3(a), showing identified fragments that enable discrimination of S and L proteins and S<sub>0</sub> and S<sub>-15</sub> proteins and the identification of N- and C-terminal sequences.

Sample	Genotype	Top protein match	Coverage (top match)	Score (top match)	Empai score (top match)	S-specific peptides	L-specific peptides
04 #1 /47 #1 (mixed sample)	S <sub>0</sub> , L <sub>C</sub> /L <sub>Y</sub>	S <sub>0</sub>	79%	10806	43.88	K.GGILEADK.V (31x) R.VMVISPR.Y (12x) K.TGFHMGR.L (2x) K.TGFHMGR.L (9x) K.YDVSTAIAAK.A (64x) K.VVGTPAYEEMVR.N (20x) K.VVGTPAYEEMVR.N (72x) R.AIKVVGTPAYEEMVR.N (1x)  *S <sub>0</sub> unique marker* K.YDVSTAIAAKALNK.E (2x)	R.VMVISPR.Y (6x) R.VMVISPR.Y (1x) K.YDVSTAISAK.A (20x) R.VVGTPVYEEMVR.N (5x) R.VVGTPVYEEMVR.N (32x) K.YIATKYDVSTAISAK.A (1x)
76 #1	S <sub>0</sub> , L <sub>f</sub>	S <sub>0</sub>	62%	14455	5.36	K.YDVSTAIAAK.A (21x) K.VVGTPAYEEMVR.N (3x) K.VVGTPAYEEMVR.N (32x)	K.YIATKYDVSTAISAK.A (1x)
3y #1	S <sub>-15</sub> , L <sub>Y</sub>	S <sub>-15</sub>	77%	5324	29.15	K.GGILEADK.V (26x) R.VMVISPR.Y (10x) K.SNYQSNGIYK.N (1x) K.VVGTPAYEEMVR.N (18x) K.VVGTPAYEEMVR.N (37x) K.TGFHMGR.L (3x) K.TGFHMGR.L (24x) K.YDVSTAIAAK.A (44x)	K.GGILEADR.V (7x) R.VMVISPR.Y (3x) R.VMVISPR.Y (8x) R.VVGTPVYEEMVR.N (6x) R.VVGTPVYEEMVR.N (15x) K.YDVSTAISAK.A (14x) K.YIATKYDVSTAISAK.A (1x)
82 #1	S <sub>-15</sub> , L <sub>f</sub>	S <sub>-15</sub>	56%	850	2.74	K.YDVSTAIAAK.A (1x)	none

Table 5. Continued

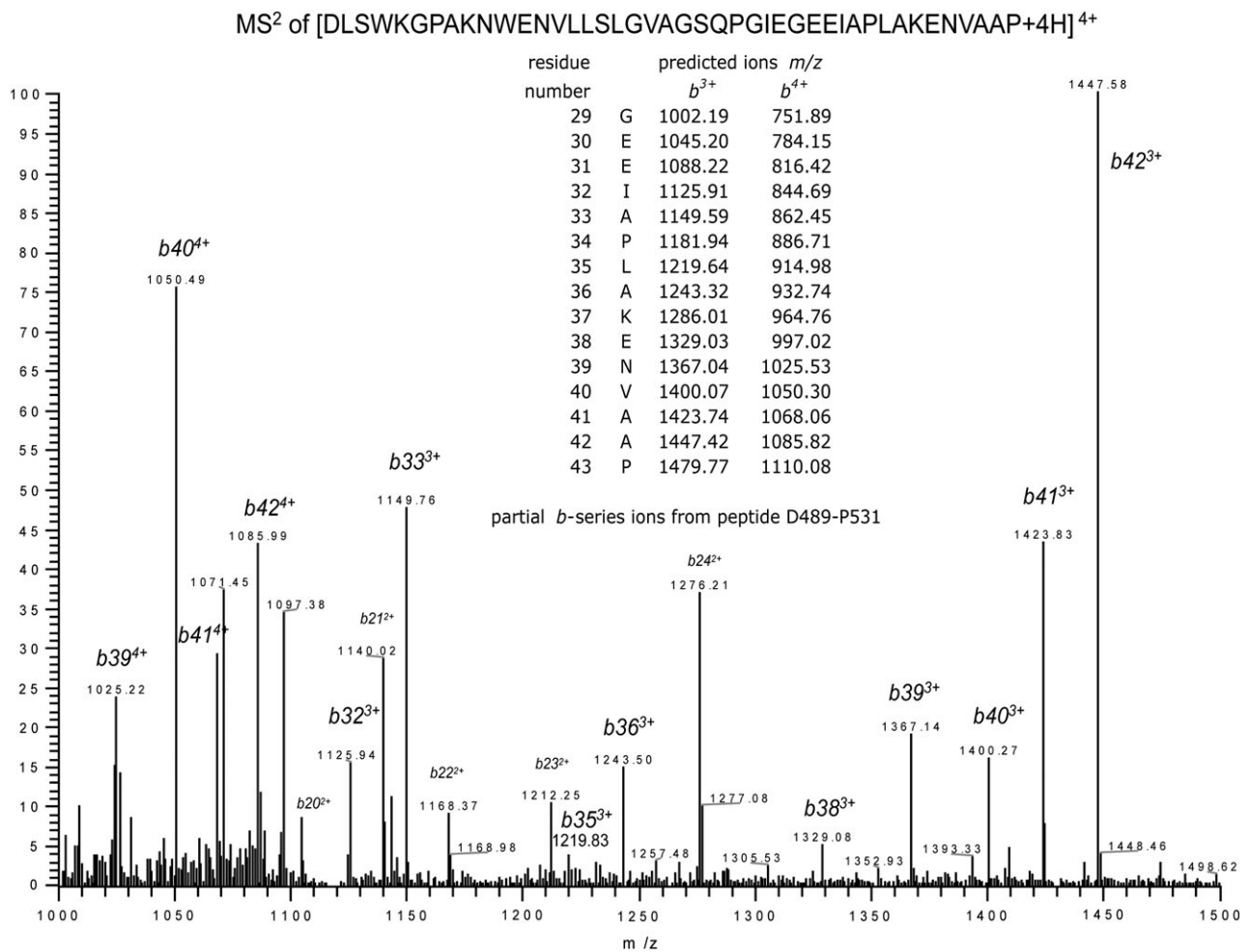
Sample	Peptides characteristic of $S_0$ (note: most are also attributable to $L_C/L_Y$ )	Peptides characteristic of $S_{-15}$	Peptides characteristic of $L_C$ (would also be attributable to $S_0/S_{-15}$ )	Peptides characteristic of $L_Y$	N-terminal peptides	C-terminal peptides
04 #1 /47 #1 (mixed sample)	K.EALQAAAAGLPVDR.K (90x) K.EALQAAAAGLPVDR.K.I (10x) K.ALNKEALQAAAAGLPVDR.K (2x)	K.ALQAAAAGLPVDR.K (4x)	none	none	-AAAGMNVVFGAEMAPWSK.T (2x) -AAAGMNVVFGAEMAPWSK.T (1x) -AAAGMNVVFGAEMAPWSK.T (2x) -AAAGMNVVFGAEMAPWSK.T (4x)	K.ENVAAP.- (1x)
76 #1	* $S_0$ unique marker* K.YDVSTAIAAKALK.E (2x) K.EALQAAAAGLPVDR.K (70x) K.EALQAAAAGLPVDR.K.I (1x)	none	none	none	-AAAGMNVVFGAEMAPWSK.T (1x) -AAAGMNVVFGAEMAPWSK.T (1x) -AAAGMNVVFGAEMAPWSK.T (7x)	none
3y #1	K.EALQAAAAGLPVDR.K (20x) K.EALQAAAAGLPVDR.K.I (3x)	K.ALQAAAAGLPVDR.K (38x) K.ALQAAAAGLPVDR.K.I (3x)	none	none	-AAAGMNVVFGAEMAPWSK.T (2x) -AAAGMNVVFGAEMAPWSK.T (2x)	none
82 #1	K.EALQAAAAGLPVDR.K (1x)	K.ALQAAAAGLPVDR.K (2x)	none	none	-AAAGMNVVFGAEMAPWSK.T (2x)	none

Only fragments identified as the top match for each query were included. The number of queries matching each peptide is shown in brackets after the sequence.

*P. miliaceum* represent *GBSSI* genes; neither represents the *GBSSIb/GBSSII* type isoforms that are typically found in nonendosperm tissues, such as the pericarp in wheat, and pods, roots, and nodules in pea and contribute to amylose synthesis in these organs (Hylton et al. 1996; Denyer, Barber, et al. 1997; Vrinten and Nakamura 2000; Edwards et al. 2002). It is likely that orthologues of the *GBSSIb/GBSSII* protein exist in *P. miliaceum* too but encoded by loci distinct from those coding for the L or S genes.

Waxy mutants in different species result variously from mutations that cause loss of *GBSSI* gene expression, loss of starch granule-bound protein, or loss of enzyme activity. The data accumulated from other species consistently indicate that *GBSSI* is the sole locus controlling endosperm starch waxiness. Polygenic control occurs only when this locus is duplicated in functional polyploids: The genetic data of Graybosch and Baltensperger (2009) indicated the existence of distinct waxy mutations at two *GBSSI* loci in *P. miliaceum*. Our SDS-PAGE data showed that the *GBSSI* protein was present in approximately equal amounts in both waxy and nonwaxy lines. In contrast, Graybosch and Baltensperger (2009) found only trace amounts of *GBSSI* in waxy lines; however, we have found that the recovery of *GBSSI* from starch can vary with genotype and extraction method (data not shown). The *SS* activity data demonstrated that the *GBSSI* protein in waxy types was nonfunctional. Consideration of the protein identities and genotype data indicates that, depending on the particular mutant alleles present in a given genotype, either one (S) or two (S and L) *GBSSI* loci contribute to this non-functional protein product. We have characterized the mutations at these two loci as follows.

Our experiments revealed two S-type *GBSSI* alleles ( $S_0$  and  $S_{-15}$ ) and three L-type *GBSSI* alleles ( $L_C$ ,  $L_Y$ , and  $L_f$ ). Of the six possible combinations of these S and L alleles, we found only five: None of the 72 plants analyzed had the genotype  $S_{-15}/L_C$ . The most common genotype was  $S_0/L_C$  (31 plants, from 17 accessions). We can postulate that this represents the ancestral genotype, for the following reasons: It is the most abundant genotype, and its distribution includes accessions from northern China, thought to be the center of diversity and origin of broomcorn millet (Crawford 2009; Hu et al. 2009). From the phenotype of the  $S_0/L_f$  plants, we know that the  $S_0$  allele encodes an active *GBSSI* protein. Given that no waxy mutants are known in any wild plant species (Sakamoto 1996; Shapter et al. 2009), it is highly likely that the diploid ancestors of *Panicum miliaceum*, and the newly formed tetraploid *P. miliaceum* or its wild progenitor, had S and L alleles encoding functional proteins. The data of Graybosch and Baltensperger (2009) indicate that a functional L-type allele still exists in the genepool. Thus, we hypothesize that  $L_C$  encodes this active *GBSSI*, although proof of this is lacking at present, in the absence of an  $S_{-15}/L_C$  genotype individual. Both the  $L_C$  and  $S_{-15}$  alleles are present at relatively high frequency in Chinese samples, but because *P. miliaceum* is strongly self-pollinating, it is possible that nonrandom mating has meant this genotype has not arisen or is



**Fig. 6.** Ms/ms fragmentation of [D489-P531 + 4H]<sup>4+</sup> ion, 1,114.5 *m/z*. A series of *b*<sup>3+</sup> and *b*<sup>4+</sup> ions can be seen which correspond to those expected from the peptide, confirming the sequence of the C-terminal section. Zoomscan of selected peaks confirmed 3+ charge state.

extremely rare. If our model is correct, the mutant S<sub>-15</sub> allele in combination with either the mutant L<sub>Y</sub> or mutant L<sub>F</sub> allele is required to produce a plant with the waxy phenotype.

Redundancy of homeologous gene copies in polyploid species means that it is not uncommon for one homeologue to lose function (Wendel 2000). It is possible that

following tetraploidization in *P. miliaceum*, a loss-of-function mutation in either the S or L *GBSSI* gene would have had little effect on starch phenotype. This is consistent with the results of Graybosch and Baltensperger (2009). We found two such genotypes with loss-of-function mutations in just the L gene: S<sub>0</sub>/L<sub>Y</sub> in 15 plants and S<sub>0</sub>/L<sub>F</sub> in 12 plants. We have shown that both the L<sub>Y</sub> and L<sub>F</sub> alleles confer loss of L-type *GBSSI* function. The mutant L<sub>Y</sub> and L<sub>F</sub> alleles arose independently, as demonstrated by the finding that no plant with the frameshift mutation in exon 9 also had the alanine substitution in exon 7, and they may have become widespread within the *P. miliaceum* distribution because, in isolation, they had little impact on the starch phenotype. This is demonstrated by the high frequency of the L<sub>Y</sub> allele in samples from western Russia, Ukraine and the Caucasus, far from the region where sticky-starch foods are preferred. It is also possible that a mutant L allele in an S<sub>0</sub> background may confer a partially waxy phenotype with a slightly lower than normal amylose content that is favored for some food uses. If so, these alleles may have spread more rapidly through the population due to active selection. There is evidence that partial waxy mutants of wheat were selected specifically for the production of udon

**Table 6.** Granule-Bound Starch Synthase Activity in Nonwaxy and Waxy Endosperm Starch Extracts of *Panicum miliaceum*.

Plant Sample	Phenotype	GBSS activity (nmol min <sup>-1</sup> mg <sup>-1</sup> starch)	
		Mean	Standard Error
4 #1	Nonwaxy	0.491	0.100
47 #1	Nonwaxy	0.250	0.085
71 #1	Nonwaxy	0.226	0.063
76 #1	Nonwaxy	0.433	0.160
3o #1	Waxy	0.021	0.010
3y #1	Waxy	0.026	0.010
70 #1	Waxy	0.017	0.006
82 #1	Waxy	0.017	0.002

The mean and standard error are of three extraction replicates.

noodles (Nakamura et al. 2002). The varieties of hexaploid wheat now grown for udon noodle production have one or two inactive waxy alleles. It is thought that the slightly lower amylose content of these genotypes confers desirable phenotypic properties and has thus been selected.

The mechanism of loss of activity in mutant L-type GBSSI clearly differs between alleles. For the  $L_f$  mutant allele, a frameshift due to the insertion of an additional nucleotide (A) is responsible for the loss of the L protein. The predicted protein product of the  $L_f$  gene has its GTD1 domain disrupted and truncated as a result of the frameshift within GTD5 and is therefore unlikely to be catalytically active. We did not see any evidence for this protein from immunoblotting, and it is probable that it is unstable or not synthesized at all (perhaps as a result of nonsense-mediated decay of the RNA). The  $L_\gamma$  allele is also evidently inactive, and we can hypothesize that it is the substitution of adenine for guanine at position 249 in exon 7 (which causes the substitution of tyrosine for cysteine in the amino acid sequence) that results in loss of L protein activity. The cysteine residue is conserved in all monocot GBSSI sequences and is located in GTD5, close to the active site.

It is likely that the loss-of-function mutation in the GBSSI-S gene arose in a line already carrying an L-gene mutation, because no  $S_{-15}/L_C$  genotype plants were found, and that this resulted in a plant with a fully waxy phenotype. A single mutation appears to account for loss of S protein activity in all samples analyzed: The 15-bp deletion was present in all plants with waxy-type grain, and conversely, all genotypes with this deletion had waxy grain. We found 14 waxy-phenotype plants: 9 plants with the  $S_{-15}/L_f$  genotype and 5 with the  $S_{-15}/L_\gamma$  genotype. It is unlikely that the  $S_{-15}$  mutation occurred more than once, and so, we suggest that the two genotypes that demonstrably result in waxy phenotypes,  $S_{-15}/L_\gamma$  and  $S_{-15}/L_f$  arose from hybridization between one of these genotypes and a phenotypically nonwaxy plant with the alternative mutant L allele in an  $S_0$  background.

The five amino acid deletion in the GBSSI-S protein, responsible for the loss of S protein activity in the *P. miliaceum*  $S_{-15}$  allele, lies in the GTD1 (fig. 5). The seven-amino-acid sequence spanning the deletion in the wild-type allele (ALNKEAL) is conserved in all monocot GBSSIs, suggesting it is functionally important. Our immunoblot, peptide fragment MS data and SS activity assay results show that the mutant GBSSI-S protein is present but nearly or completely inactive. This mutant GBSSI protein nevertheless persists in binding to starch granules, as has been found previously for some mutant GBSSIs in other species (e.g., one of the three mutant pea embryo GBSSI isoforms identified by Denyer et al. [1995] was found in normal amounts inside starch granules). We hypothesize that the five-amino-acid deletion in the *P. miliaceum* GBSSI-S protein results in a disruption of the GTD1, resulting in the observed loss of SS activity, but does not affect the binding of this protein to the starch granules.

It is possible that other mutant GBSSI alleles exist in *P. miliaceum*. However, our determination of the full-length amino acid sequences of the mature L and S GBSSI proteins,

inferred from DNA sequence and peptide mass fragment data in multiple waxy and nonwaxy samples, means that we can be confident that no other mutations exist that universally account for loss of functional protein at either locus.

All waxy-type individuals in this study came from accessions of East Asian (Chinese and Korean) origin. All samples originating elsewhere (Russia, Ukraine, the Caucasus, Central Asia, and Mongolia) were nonwaxy. Chinese samples showed a mixture of waxy and nonwaxy types both within and between accessions. This distribution of waxy phenotypes concurs with the data of Kimata and Negishi (2002), who found waxy and intermediate phenotypes in around 60% of *P. miliaceum* accessions from East Asia (China, Korea, and Mongolia) and at very low frequencies among accessions from elsewhere in Eurasia. These authors also found that nearly 100% of Japanese accessions were waxy or intermediate; we did not include any samples from Japan in this study, and we did not observe any of the intermediate phenotypes.

A similar geographic pattern is found in other cereals, comprising a mix of waxy and nonwaxy varieties in East and Southeast Asia, and near-exclusively nonwaxy varieties elsewhere in the range. This includes both species domesticated within the region (foxtail millet, rice) and those domesticated elsewhere and introduced into East Asia (barley, maize, sorghum, and Job's tears; Sakamoto 1996; Patron et al. 2002; Kawase et al. 2005; Olsen et al. 2006; Fan et al. 2008). Waxy types have not been recorded in wild cereals (Sakamoto 1996; Shapter et al. 2009). Selective sweeps in the Waxy genomic region have been demonstrated in rice and maize that reflect strong postdomestication selection, revealing the shaping of crop genome evolution through human choice (Olsen et al. 2006; Fan et al. 2008; Vaughan et al. 2008). A "glutinous-endosperm starch food culture," which arose early in the history of cereal farming in East and Southeast Asia and remained confined to the cultures of this region, is thought to have driven selection for waxy mutants arising in domesticated cereal populations.

The waxy broomcorn millet lines in this study all showed the same mutation in the GBSSI-S gene, implying a single origin for the loss of protein function in this homeologue. In the GBSSI-L gene, there are two nonfunctional alleles of independent origin. As in foxtail millet (*S. italica*), where waxy phenotypes have resulted from several different transposon insertion events (Kawase et al. 2005), waxy alleles in broomcorn millet may have distinct distribution geographies. Further investigation of the distribution of these alleles within and between populations, varieties, and landraces of cultivated broomcorn millet is of great relevance to understanding both the history of this crop's origins and spread and how the organization of crop plant diversity has been shaped by human food choices.

## Supplementary Material

Supplementary figures S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

This work was supported by a Leverhulme Trust research grant to M.K.J. and H.V.H. (ref. F/09707/B) and a Wellcome Trust Bioarchaeology Research Training Fellowship to H.V.H. (ref. 076815). We are grateful to the USDA-ARS and the Vavilov Research Institute for providing plant material. We thank Yo-Ichiro Sato (Research Institute for Humanity and Nature, Kyoto) for initial discussion, Pete Michna and colleagues at the University of Cambridge Botanic Garden for growing plants, Gerhard Saalbach (John Innes Centre) for proteomics analyses, Brendan Fahy (John Innes Centre) and Matt Lawes (University of Cambridge) for help in the laboratory, Mike Weldon (University of Cambridge) for protein-sequence analysis, Ellen Nisbet (University of South Australia) for phylogenetics advice, Kathryn Lilley (University of Cambridge) for proteomics advice, and Carla Lancelotti and Cameron Petrie (University of Cambridge) for assistance with graphics.

## References

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of bestfit models of protein evolution. *Bioinformatics* 21:2104–2105.
- Baltensperger DD. 1996. Foxtail and Proso millet. In: Janick J, editor. *Progress in new crops*. Alexandria (VA): ASHS Press. p. 182–190.
- Barton L, Newsome SD, Chen F-H, Wang H, Guilderson TP, Bettinger RL. 2009. Agricultural origins and the isotopic identity of domestication in northern China. *Proc Natl Acad Sci U S A*. 106:5523–5528.
- Bouchenak-Khelladi Y, Salamin N, Savolainen V, Forest F, van der Bank M, Chase MW, Hodkinson TR. 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Mol Phylogenet Evol*. 47:488–505.
- Buschiazzo A, Ugalde JE, Guerin ME, Shepard W, Ugalde RA, Alzari PM. 2004. Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. *Embo J*. 23:3196–3205.
- Busi MV, Palopoli N, Valdez HA, Fornasari MS, Wayllace NZ, Gomez-Casati DF, Parisi G, Ugalde RA. 2007. Functional and structural characterization of the catalytic domain of the starch synthase III from *Arabidopsis thaliana*. *Proteins* 70:31–40.
- Crawford GW. 2009. Agricultural origins in North China pushed back to the Pleistocene-Holocene boundary. *Proc Natl Acad Sci U S A*. 106:7271–7272.
- Denyer K, Barber LM, Burton R, et al. (12 co-authors). 1995. The isolation and characterization of novel low-amylose mutants of *Pisum sativum* L. *Plant Cell Env*. 18:1019–1026.
- Denyer K, Barber LM, Edwards EA, Smith AM, Wang TL. 1997. Two isoforms of the GBSSI class of granule-bound starch synthase are differentially expressed in the same plant. *Plant Cell Env*. 20:1566–1572.
- Denyer K, Edwards A, Martin C, Smith AM. 1997. The mechanism of amylose synthesis. In: Frazier PJ, Richmond P, Donald AM, editors. *Starch: structure and functionality*. Cambridge: Royal Society of Chemistry. p. 222–229.
- Domon E, Fujita M, Ishikawa N. 2002. The insertion/deletion polymorphisms in the waxy gene of barley genetic resources from East Asia. *Theor Appl Genet*. 104:132–138.
- Edwards A, Vincken J-P, Suurs LCJM, Visser RGF, Zeeman S, Smith AM, Martin C. 2002. Discrete forms of amylose are synthesized by isoforms of GBSSI in pea. *Plant Cell*. 14:1767–1785.
- Fan LJ, Quan LY, Leng XD, Guo XY, Hu WM, Ruan SL, Ma HS, Zeng MQ. 2008. Molecular evidence for post-domestication selection in the Waxy gene of Chinese waxy maize. *Mol Breeding*. 22:329–338.
- Fortune PM, Schierenbeck KA, Ainouche AK, Jacquemin J, Wendel JF, Ainouche ML. 2007. Evolutionary dynamics of Waxy and the origin of hexaploid *Spartina* species. *Mol Phylogenet Evol*. 43:1040–1055.
- Fukunaga K, Kawase M, Kato K. 2002. Structural variation in the Waxy gene and differentiation in foxtail millet [*Setaria italica* (L.) P. Beauv.]: implications for multiple origins of the waxy phenotype. *Mol Genet Genomics*. 268:214–222.
- Graybosch RA, Baltensperger DD. 2009. Evaluation of the waxy endosperm trait in proso millet (*Panicum miliaceum*). *Plant Breeding*. 128:70–73.
- Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res*. 33:W557–W579.
- Hamoud MA, Haroun SA, MacLeod RD, Richards AJ. 1994. Cytological relationships of selected species of *Panicum* L. *Biol Plantarum*. 36:37–45.
- Hirano HY, Eiguchi M, Sano Y. 1998. A single base change altered the regulation of the Waxy gene at the posttranscriptional level during the domestication of rice. *Mol Biol Evol*. 15: 978–987.
- Hirano HY, Sano Y. 1991. Molecular characterization of the Waxy locus of rice (*Oryza sativa*). *Plant Cell Physiol*. 32:989–997.
- Hu X, Wang J, Lu P, Zhang H. 2009. Assessment of genetic diversity in broomcorn millet (*Panicum miliaceum* L.) using SSR markers. *J Genet Genomics*. 36:491–500.
- Hunt HV, Vander Linden M, Liu X, Motuzaite-Matuzeviciute G, Jones MK. 2008. Millets across Eurasia: chronology and context of early records of the genera *Panicum* and *Setaria* from archaeological sites in the Old World. *Veget Hist Archaeobot*. 17:55–518.
- Huson DH, Richter DC, Rausch C, DeZulian T, Franz M, Rupp R. 2007. Dendroscope—an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460.
- Hylton CM, Denyer K, Keeling PL, Chang M-T, Smith AM. 1996. The effect of waxy mutations on the granule-bound starch synthases of barley and maize endosperms. *Planta* 198:230–237.
- Isshiki M, Morino K, Nakajima M, Okagaki RJ, Wessler SR, Izawa T, Shimamoto K. 1998. A naturally occurring functional allele of the rice waxy locus has a GT to TT mutation at the 5' splice site of the first intron. *Plant J*. 15:133–138.
- Jacob J, Disnar J-R, Arnaud F, Chapron E, Debret M, Lallier-Vergès E, Desmet M, Revel-Rolland M. 2008. Millet cultivation history in the French Alps as evidenced by a sedimentary molecule. *J Arch Sci*. 35:814–820.
- Jones MK. 2004. Between fertile crescents: minor grain crops and agricultural origins. In: Jones MK, editor. *Traces of ancestry: studies in honour of Colin Renfrew*. Cambridge: Oxbow Books. p. 127–134.
- Kawase M, Fukunaga K, Kato K. 2005. Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol Genet Genomics*. 274:131–140.
- Kimata M, Negishi M. 2002. Geographical distribution of proso millet, *Panicum miliaceum* L. on iodostarch and phenol reactions; with special reference to a northern propagation route into Japanese islands. *Environ Educ Stud, Tokyo Gakugei University*. 12:15–22.
- Lu H, Zhang J, Liu K, et al. (12 co-authors). 2009. Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *Proc Natl Acad Sci U S A*. 106:7367–7372.

- Marillonnet S, Wessler SR. 1997. Retrotransposon insertion into the maize waxy gene results in tissue-specific RNA processing. *Plant Cell*. 9:967–978.
- McIntyre CL, Drenth J, Gonzalez N, Henzell RG, Jordan DR. 2008. Molecular characterization of the waxy locus in sorghum. *Genome* 51:524–533.
- Mu-Forster C, Huang R, Powers JR, Harriman RW, Knight M, Singletary GW, Keeling PL, Wasserman BP. 1996. Physical association of starch biosynthetic enzymes with starch granules of maize endosperm. *Plant Physiol*. 111:821–829.
- Nakamura T, Vrinten P, Saito M, Konda M. 2002. Rapid classification of partial waxy wheats using PCR-based markers. *Genome* 45:1150–1156.
- Nakamura T, Yamamori M, Hirano H, Hidaka S. 1993. Decrease of Waxy (Wx) protein in 2 common wheat cultivars with low amylose content. *Plant Breeding*. 111:99–105.
- Nakamura T, Yamamori M, Hirano H, Hidaka S, Nagamine T. 1995. Production of Waxy (amylose-free) wheats. *Mol Gen Genet*. 248:253–259.
- Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD. 2006. Selection under domestication: evidence for a sweep in the rice Waxy genomic region. *Genetics* 173: 975–983.
- Olsen KM, Purugganan MD. 2002. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* 162:941–950.
- Patron NJ, Smith AM, Fahy BF, Hylton CM, Naldrett MJ, Rosnagel BG, Denyer K. 2002. The altered pattern of amylose accumulation in the endosperm of low-amylose barley cultivars is attributable to a single mutant allele of granule-bound starch synthase I with a deletion in the 5' non-coding region. *Plant Physiol*. 130:190–198.
- Rogers SO, Bendich AJ. 1994. Extraction of total DNA from plants, animals and fungi. In: Gelvin SB, Schilperoort RA, editors. *Plant molecular biology manual*, 2nd ed. Dordrecht (Netherlands): Kluwer. D1. p. 1–8.
- Saito M, Konda M, Vrinten P, Nakamura K, Nakamura T. 2004. Molecular comparison of waxy null alleles in common wheat and identification of a unique null allele. *Theor Appl Genet*. 108:1205–1211.
- Saito M, Nakamura T. 2005. Two point mutations identified in emmer wheat generate null Wx-A1 alleles. *Theor Appl Genet*. 110:276–282.
- Sakamoto S. 1996. Glutinous-endosperm starch food culture specific to Eastern and Southeastern Asia. In: Ellen R, Fukui K, editors. *Redefining nature: ecology, culture and domestication*. Oxford: Berg. p. 215–231.
- Sano Y. 1984. Differential regulation of waxy gene expression in rice endosperm. *Theor Appl Genet*. 68:467–473.
- Shapter FM, Eggler P, Lee LS, Henry RJ. 2009. Variation in granule-bound starch synthase I (GBSSI) loci amongst Australian wild cereal relatives (Poaceae). *J Cereal Sci*. 49:4–11.
- Stoddard FL. 1999. Survey of starch particle-size distribution in wheat and related species. *Cereal Chem*. 76:145–149.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596–1599.
- Umeda M, Ohtsubo H, Ohtsubo E. 1991. Diversification of the rice Waxy gene by insertion of mobile DNA elements into introns. *Jpn J Genet*. 66:569–586.
- Vaughan DA, Lu B-R, Tomooka N. 2008. The evolving story of rice evolution. *Plant Sci*. 174:394–408.
- Vrinten P, Nakamura T, Yamamori M. 1999. Molecular characterization of waxy mutations in wheat. *Mol Gen Genet*. 261:463–471.
- Vrinten PL, Nakamura T. 2000. Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues. *Plant Physiol*. 122:255–263.
- Wang ZY, Zheng FQ, Shen GZ, Gao JP, Snustad DP, Li MG, Zhang JL, Hong MM. 1995. The amylose content in rice endosperm is related to the posttranscriptional regulation of the Waxy gene. *Plant J*. 7:613–622.
- Wendel JF. 2000. Genome evolution in polyploids. *Plant Mol Biol*. 42:225–249.
- Wessler S, Tarpley A, Purugganan M, Spell M, Okagaki R. 1990. Filler DNA is associated with spontaneous deletions in maize. *Proc Natl Acad Sci U S A*. 87:8731–8735.
- Wessler SR, Varagona MJ. 1985. Molecular-basis of mutations at the Waxy locus of maize - correlation with the fine-structure genetic-map. *Proc Natl Acad Sci U S A*. 82:4177–4181.
- Whitt SR, Wilson LM, Tenaillon MI, Gaut BS, Buckler ES. 2002. Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci U S A*. 99:12959–12962.
- Yamamori M, Nakamura T, Endo TR, Nagamine T. 1994. Waxy protein-deficiency and chromosomal location of coding genes in common wheat. *Theor Appl Genet*. 89:179–184.
- Zohary D, Hopf M. 2000. *Domestication of plants in the Old World*, 3rd ed. Oxford: Oxford University Press.