A full-length cDNA clone encoding granule-bound starch synthase I (GBSSI = Waxy gene) from grain amaranth (Amaranthus cruentus L.) perisperm was isolated and characterized. Segregation of amylose content in F2 population suggested that the amylose content of A. cruentus is controlled by a single gene, Waxy (GBSSI). cDNA clone of this gene is 2076 bp in length and contains an open reading frame of 1821 bp corresponding to a polypeptide of 606 amino acids residues, including a transit peptide of 77 amino acids. Comparison of the cDNA and genomic sequences (3492 bp) suggested that the amaranth GBSSI gene has 12 introns, of which exons 1–13 contributed to the coding sequence. The mature protein shares 70.2–75.3% sequence identity with GBSSI of dicots and about 64.0–67.8% identity with those of monocots. This protein contains the conserved motif KTGGL found in other GBSSI proteins, which has been implicated as the active site in glycogen synthase. Sequence analysis predicted that GBSSI of amaranth has a transit peptide of 77 amino acids including FIR↓S, which is different cleavage site that of the other dicot species. These results will provide more useful information for understanding the structure/function relationship of this protein from amaranths perisperm.

Key Words: A. cruentus, grain amaranth, GBSSI, cDNA cloning, starch.

Introduction

Starch is one of the major storage components in the plant that is used in food or non-food industries (Ellis et al. 1998). Starch can be chemically fractionated into two types of glucan polymer: amylose and amylopectin. Amylose is a lightly branched linear molecule with low polymerization, whereas amylopectin is a much larger molecule with extensive branches resulting from α-1,6 linkages (Smith et al. 1997). Both types of polymers are elongated by starch synthase, which catalyzes the transfer of glucose (Glc) from Adenosine diphosphate (ADP)-Glc to growing glucan chains via a α-1,4-linkage (Martin and Smith 1995).

Granule bound starch synthase I (GBSSI), also known as the waxy protein, is responsible for the synthesis of amylose in the amyloplasts of plant storage organs (James et al. 2003, Tsai 1974). A role of GBSSI in amylose synthesis has been described for some plant species, such as barley (Rohde et al. 1988), rice (Hirano and Sano 1991), maize (Macdonald and Preiss 1985), potato (van der Leij et al. 1991) and amaranth (Okuno and Sakaguchi 1981). The GBSSI (= Waxy gene) plays a major part in determining functionality in food or non-food industries. Therefore, improving starch quality and functionality depends on understanding the action and effects of this gene.

The genus Amaranthus includes more than 60 species that grow in many areas of the world (Kauffman and Weber 1990). Amaranth grains still used to this day include the three species, A. caudatus, A. cruentus, and A. hypochondriacus, with high seed protein content and a balanced amino acid composition (Brenner et al. 2000, Kauffman and Weber 1990, Sauer 1993). Today, amaranth cultivation is developing in Central and South America, Africa, India, and China (Saunders and Becker 1984). In the case of Japan, this grain was first introduced regularly in 1980s from Rodale Research Institute, United States and a new grain amaranth variety “New Aztec” was developed in 2001 using “Mexico line”, which is a high-yielding and semi-dwarf line selected in Japan from the introduced genetic resources of grain amaranth (A. cruentus) originating in Mexico (Katsuta et al. 2001). Currently, this new grain, mostly A. cruentus, is cultivated gradually in some areas such as Iwate, Akita and Nagano prefecture.

The potential of both grain (seed) and vegetable (leaf) amaranth as a food resource has been reviewed extensively by many researchers (Becker et al. 1981, Morales et al. 1982).
Inheritance of waxyness in perisperm of *A. cruentus*

Pollinating the *F*<sub>1</sub> perisperm’s appearance, opaque or translucent, respectively. The waxy/nonwaxy grain can be easily distinguished by the perisperm of seeds, and controlled by a single major structural gene, with the *waxy* allele recessive to the *Waxy* allele (Okuno and Sakaguchi 1981, Sugimoto *et al.* 1985). Similar variation was also reported in *A. hypochondriacus* and *A. cruentus*, which has also nonwaxy and waxy types possessing 19.4–27.8% and 0% amylose contents (Inouchi *et al.* 1999).

To date, sequence information of the *GBSSI* structural genes from several species, such as maize (Klosgen *et al.* 1986), rice (Okagaki 1992), barley (Rohde *et al.* 1988), wheat (Clark *et al.* 1991, Murai *et al.* 1999), potato (van der Leij *et al.* 1991), sweet potato (Kimura *et al.* 2000) and foxtail millet (Fukunaga *et al.* 2002) have been published. To facilitate further studies of amaranth starch, it is essential to characterize the *GBSSI* gene and to establish its primary structures. However, the information of the complete sequence of *GBSSI* gene of amaranth has not been reported.

In this study, we report the inheritance of *waxy* in *A. cruentus* perisperm by using reciprocal crosses and present the complete coding sequence of the amaranth *GBSSI* (*A. cruentus*) and analyzed genomic structure.

Materials and Methods

*Plant materials*

The strains were classified into two types of starch granule in *Amaranthus cruentus*, the nonwaxy (USDA, Accession No. PI451826) and the waxy (USDA, Accession No. PI477914), by determining the amylose content from preliminary experiments (data not shown). These two strains were used for the study. Plants were grown in glasshouse conditions for 90 days. The grains harvested were used; the nonwaxy and the waxy accessions for protein analysis, and the nonwaxy accession for nucleic acid analysis, respectively.

Inheritance of waxyness in perisperm of *A. cruentus*

Coses were made between the nonwaxy (*WxWx*) and the waxy (*www*) accessions described above reciprocally in glasshouse conditions. *F*<sub>2</sub> population was generated by self-pollinating the *F*<sub>1</sub> plant. The genotype of each individual was determined based on the transparency of perisperm of the selfed seeds in *F*<sub>1</sub> and *F*<sub>2</sub> populations. In grain amaranth, the waxy/nonwaxy grain can be easily distinguished by perisperm’s appearance, opaque or translucent, respectively. A total of 100 individuals from each *F*<sub>2</sub> population were evaluated.

Detection of *GBSSI* activity by SDS-PAGE

Starch granules were purified from immature seeds, which were based on the methods of Nakamura *et al.* (1998) with some modifications. A 10 mg of purified starch was suspended with a sample buffer (0.5 M Tris-HCl, pH 6.8, 2.5% SDS, 10% glycerol, 2% 2-mercaptoethanol) and boiled for 3 min, then centrifuged at 15,000 rpm for 5 min, and the supernatant was subjected to gel. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was conducted using 10% separation gel and 4% stacking gel.

Isolation of total RNA and genomic DNA from the nonwaxy accession

Total RNA was isolated from the immature seeds using Plant RNA Reagent (Invitrogen) according to the manufacturer’s instruction. Genomic DNA was extracted from young leaves using the Cetyl trimethyl ammonium bromide (CTAB) method of Murray and Thompson (1980) with some modifications. DNA and RNA were quantified by UV absorption and by separation in 1.5% agarose formaldehyde gels, which were then visualized under UV light after staining with ethidium bromide.

Degenerate primer design

The alignment of the predicted amino acid sequences of peas (*Pisum sativum*, Accession No. X88789), tomato (*Lycopersicon esculentum*, Accession No. BT013430), and cassava (*Manihot esculenta*, Accession No. X74160) identified several conserved domains. Two conserved regions were used to design degenerate primers (Table 1) for *GBSSI* amplification from amaranth. Primer3 software (Rozen and Skaletsky 2000) ([http://primer3.sourceforge.net/](http://primer3.sourceforge.net/)) was used for designing primer sequences. All amino acid sequence data were obtained from NCBI/GeneBank database ([http://www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)).

Cloning of the full-length *GBSSI* cDNA

The first-strand *GBSSI* cDNA was constructed using the 3′- and 5′-Full RACE Core Set kit (TaKaRa) or Marathon cDNA amplification kit (Clontech). Reverse transcriptase-polymerase chain reaction (RT-PCR) and 3′- and 5′-rapid amplification of cDNA end (RACE) were performed using the primers listed in Table 1 (the AP1, AP2 primer supplied with the kit). Polymerase chain reaction was carried out using *ExTaq* polymerase (0.5 U/μl), 2.5 mM dNTP, 10 × PCR buffer (TaKaRa) and primers at 0.5 μM. Annealing temperatures used for each primer combination are listed in Table 2. The *GBSSI* gene was also amplified from genomic DNA. Amplified fragments were extracted from gel using QIAquick Gel Extraction Kit (QIAGEN) as specified by the manufacturers. The cDNA-amplified fragments were ligated into the pGEM vector using a TA cloning kit (Promega) and transformed into *Escherichia coli* cells (JM109).

Analysis of sequencing

The analysis of cDNA and genomic DNA sequencing
Molecular cloning of GBSSI cDNA from a grain amaranth

was performed using the automated ABI system with dye terminators (Applied Biosystems) as described by the manufacturer’s instruction. To identify multiple sequence alignment, analysis for the deduced amino acid sequences and nucleotide sequences was made using the ClustalX (Thompson et al. 1997) module within Geneious Pro 3.7.1 (Biomatters Ltd.). Protein molecular weight calculations were performed with a Geneious Pro 3.7.1. BLAST analysis searches were performed online with default settings (www.ncbi.nlm.gov/BLAST/). The transit peptide and cleavage site was predicted using the network programs of ChloroP (Emanuelsson et al. 1999) (http://www.cbs.dtu.dk/services/ChloroP) and TargetP (Emanuelsson et al. 2000) (http://www.cbs.dtu.dk/services/TargetP/). A phylogenetic tree of the GBSSI gene was obtained by analyzing the deduced amino acid sequence divergence of the GBSSI protein. Phylogenetic analysis was performed using the neighbor joining (NJ) method (Saitou and Nei 1987) in the PAUP*4.0 software (Swofford 1988).

Results

Detection of GBSS activity

The F1 plants in the two crosses from the reciprocal cross between the nonwaxy and the waxy amaranth showed nonwaxy genotype. The segregation pattern in the F2 population is shown in Table 3. The segregation for nonwaxy and waxy plants was clearly distinguished in a 3 : 1 ratio. The result indicated that the nonwaxy is controlled by a single gene.

To detect whether amaranth starch contains active GBSSI protein, starch granules were isolated and assayed for GBSSI activity and analyzed on SDS-PAGE gel (Fig. 1). The electrophoresis revealed that the nonwaxy type of A. cruentus observed about 67 kDa band, whereas there was no band in the waxy type. The existence or nonexistence of 67 kDa band is consistent with the seed phenotypes, nonwaxy or waxy.

Amplification of GBSSI cDNA sequences from amaranth

To obtain amaranth GBSSI cDNA sequences, reverse transcriptase-polymerase chain reaction (RT-PCR) was used and an 825-bp fragment that covers the N-terminal part of the protein was amplified. It was performed by selecting degenerate primers (see Table 1) derived from the conserved sequence found toward the N-terminal of all GBSSI proteins studied to date. The sequence of the PCR product showed 77% identity at the protein level with sweet potato GBSSI. This result indicates that the fragment was a partial sequence encompassing the conserved sequence of the GBSSI gene in amaranth. Then, the 825-bp cDNA fragment was used to design oligonucleotide primers for GBSSI amplification from amaranth.

Table 1. Primers used for amplifying and/or sequencing of GBSSI gene of amaranth

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5′→3′)</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSP1_dp*</td>
<td>AAAACTGGTGGWCTHGTAGTGTTC</td>
<td>F</td>
</tr>
<tr>
<td>GSP2_dp*</td>
<td>CAYTSCAACAGAAGGAGTTGATCG</td>
<td>F</td>
</tr>
<tr>
<td>GSP1-2</td>
<td>TTAGTAAAGTGAAGCTCTGAGCAA</td>
<td>F</td>
</tr>
<tr>
<td>GSP2-2</td>
<td>CAAAGTATTGCTGAATGTTAACA</td>
<td>F</td>
</tr>
<tr>
<td>GSP3</td>
<td>ATGGTGTCACCCGGTATGGAATA</td>
<td>F</td>
</tr>
<tr>
<td>GBSS1-F20</td>
<td>ATGGAAACAGTAACATCTCTCCTCCTAC</td>
<td>F</td>
</tr>
<tr>
<td>GBSS1-F12</td>
<td>CCACAAGATGTTAAGATGTTAAGTTAAGTTAAGT</td>
<td>F</td>
</tr>
<tr>
<td>Genomic F</td>
<td>TCATACAGAGATGAGTTGACACTA</td>
<td>F</td>
</tr>
<tr>
<td>Genomic F13</td>
<td>CCTCTAATCTCTACCTACCTGAAAT</td>
<td>F</td>
</tr>
<tr>
<td>Genomic F2</td>
<td>ATTACTCTTTTACGGTTATGACT</td>
<td>F</td>
</tr>
<tr>
<td>GBSS-F14</td>
<td>ACGGTTAAGTGAATTGTTAATCTCATA</td>
<td>F</td>
</tr>
<tr>
<td>GBSS-F16</td>
<td>ATGGTACCAACTGCCGTT</td>
<td>F</td>
</tr>
<tr>
<td>GBSS-R1*</td>
<td>WAGACGACTGTGCACCARTCTGGAAC</td>
<td>R</td>
</tr>
<tr>
<td>GBSS-R2-1*</td>
<td>ARAATACWCAAGACCTTCCTGCTCYTC</td>
<td>R</td>
</tr>
<tr>
<td>5GSP1</td>
<td>TGATCTACATCAGCAATTG</td>
<td>R</td>
</tr>
<tr>
<td>5GSP2</td>
<td>CATCTCTCCATAGAATGCGCATTACA</td>
<td>R</td>
</tr>
<tr>
<td>5GSP1-p</td>
<td>TTTAGCTGTGTCTCACAATTTTACT</td>
<td>R</td>
</tr>
<tr>
<td>5GSP2-p</td>
<td>AGAATGCGTTGCTCTCACAATTTTACT</td>
<td>R</td>
</tr>
<tr>
<td>Genomic R</td>
<td>CAGCAGCTTCCTCTGATAACAATA</td>
<td>R</td>
</tr>
<tr>
<td>GBSS-R12</td>
<td>AAGAATGACGAAATCTGCTTCT</td>
<td>R</td>
</tr>
<tr>
<td>GBSS-R13</td>
<td>GATATATACACAGGAGTCTACCATGTC</td>
<td>R</td>
</tr>
<tr>
<td>Genomic R2</td>
<td>ACACCAACTTCCTACCCACATCTGATA</td>
<td>R</td>
</tr>
<tr>
<td>GBSS-R15</td>
<td>TCTATGTGTCCTACATG</td>
<td>R</td>
</tr>
</tbody>
</table>

* degenerate primers

Table 2. Primer combinations and annealing temperatures used for amplification of fragments from amaranth GBSSI gene

<table>
<thead>
<tr>
<th>Template</th>
<th>Primer combination</th>
<th>Amplified region</th>
<th>Annealing temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA</td>
<td>GBSS-F16/GBSS-R15</td>
<td>Exon 1</td>
<td>55°C</td>
</tr>
<tr>
<td>cDNA</td>
<td>AP1/5GSP1-1</td>
<td>Exon 1</td>
<td>50°C</td>
</tr>
<tr>
<td>cDNA</td>
<td>GSP1_dp/5GSP2</td>
<td>Exon 2–Exon 4</td>
<td>50°C</td>
</tr>
<tr>
<td>cDNA</td>
<td>GSP2_dp/GBSS-R1</td>
<td>Exon 5</td>
<td>50°C</td>
</tr>
<tr>
<td>cDNA</td>
<td>GSP2-2/GBSS-R2-1</td>
<td>Exon 5–Exon 9</td>
<td>50°C</td>
</tr>
<tr>
<td>cDNA</td>
<td>GSP2-3/AP2</td>
<td>Exon 9–Exon 13</td>
<td>50°C</td>
</tr>
<tr>
<td>cDNA</td>
<td>GBSS1-F11/3’AP</td>
<td>3’ UTR</td>
<td>50°C</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>GBSS1-F20/5GSP2-1</td>
<td>Exon 1</td>
<td>54°C</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>GBSS1-F12/5GSP2</td>
<td>Exon 2–Exon 4</td>
<td>54°C</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>Genomic F/5GSP1</td>
<td>Exon 3–Exon 4</td>
<td>55°C</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>GSP1-2/Genomic R</td>
<td>Exon 5–Exon 7</td>
<td>55°C</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>GBSS1-F13/GBSS-R12</td>
<td>Exon 7–Exon 9</td>
<td>55°C</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>Genomic F2/GBSS-R13</td>
<td>Exon 9–Exon 11</td>
<td>54°C</td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>GBSS-F14/Genomic R2</td>
<td>Exon 11–Exon 13</td>
<td>54°C</td>
</tr>
</tbody>
</table>
The complete sequence of the full-length GBSSI cDNA was determined from seven positive cDNA clones by 3′- and 5′-RACE. It was determined in the 825-bp probe sequence between nucleotides at 635–1459 positions with 100% identity. The sequence of this clone is shown in Fig. 2, and the nucleotide sequence data reported in this paper was submitted in the DNA Data Bank of Japan (DDBJ)/European Molecular Biology Laboratory (EMBL)/GenBank nucleotide sequence databases with the accession number AB456685. The GBSSI cDNA is 2076 nucleotides including an open reading frame (ORF), 1821 nucleotides from the ATG at position 1 to the termination codon TAA at position 1819. In addition, the cDNA includes a 3′ untranslated region (UTR) of 231 nucleotides and a poly (A) tail of 24 nucleotides. However, 5′ UTR was not amplified in this study. A typical eukaryotic polyadenylation signal AATAAA (Joshi 1987) was found for the cDNA positions upstream from the poly (A) tail. These cDNA showed 69.7% identity with cassava GBSSI and 58.5–69.4% identity with the other plant species (Table 4).

Isolation and characterization of the GBSSI genomic clones

From the screening, seven positive genomic clones were isolated and the genomic structure of the amaranth GBSSI gene is shown in Fig. 3. The GBSSI genomic DNA is 3492 nucleotides including an intron of 1821 nucleotides. This sequence has no significant high homology with sequences of other genes encoding GBSSI in the National Center for Biotechnology Information (NCBI) databases using BLASTN. However, an amino acid sequence search using BLASTP showed high homology with other gene encoding GBSSI (http://www.ncbi.nlm.gov/BLAST/). Comparison of the cDNA sequences and genomic sequences suggested that the amaranth GBSSI gene has 12 introns, of which exons 1–13 contribute to the coding sequence. All introns in the amaranth GBSSI gene follow the universal GT-AG rule (Breathnach and Chambon 1981) and coincide with the AT-rich regions similar to those found in the gene of other dicots (Camirand et al. 1990, van der Leij et al. 1991). The exons and introns have a total guanine-cytosine (GC) content of 44.1% and 31.6%, respectively.

Table 3. Segregation in F2 of crosses between nonwaxy and waxy types

<table>
<thead>
<tr>
<th>Cross combination</th>
<th>F2 nonwaxy</th>
<th>F2 waxy</th>
<th>Total</th>
<th>Expected ratio</th>
<th>χ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>waxy × nonwaxy</td>
<td>73</td>
<td>27</td>
<td>100</td>
<td>3 : 1</td>
<td>0.21</td>
</tr>
<tr>
<td>nonwaxy × waxy</td>
<td>76</td>
<td>24</td>
<td>100</td>
<td>3 : 1</td>
<td>0.05</td>
</tr>
</tbody>
</table>

The amaranth GBSSI protein

The deduced amino acid sequence and the structure of amaranth GBSSI protein are shown in Fig. 2 and Fig. 4, respectively. The 1821 nucleotide ORF of the amaranth GBSSI cDNA gives a protein containing 606 amino acids with 67.3 kDa, which is reasonably close to the apparent molecular weight of the perisperm GBSSI as estimated on SDS-PAGE gels in this study (see Fig. 1). An amino acid sequence search of each genus in the reduced data set using BLAST showed high homology with GBSSI amino acid sequences from cassava (Manihot esculenta, Accession No. Q43784), snapdragon (Antirrhinum majus, Accession No. O82627), sweet potato (Ipomoea batatas, Accession No. ABW83791), potato (Solanum tuberosum, Accession No. ABI98288), perilla (Perilla frutescens, Accession No. AAG43519), soybean (Glycine max, Accession No. ABL96300), and numerous rice (Oryza sativa) (www.ncbi.nlm.gov/BLAST/).

The degree of sequence identity between the predicted mature GBSSI protein from amaranth and its homolog from various plant species (see Table 4) ranged from 64.0% to 75.3%. The amaranth GBSSI protein revealed three conserved regions, box 1, 2, and 3, which characterized the
Molecular cloning of GBSSI cDNA from a grain amaranth

The conserved regions showed similarity to all other known starch synthases and *Escherichia coli* glycogen synthase (Kumar et al. 1986, van der Leij et al. 1991). Sequence analysis predicted the presence of the transit peptide of 77 amino acids including cleavage site, IRS G, at the N-terminal. Therefore, the deduced mature protein and transit protein contain 529 amino acids with 58.6 kDa and 77 amino acids with 8.7 kDa, respectively. Comparison of the transit peptide sequences of amaranth and other plant species shows identity with dicots such as sweet potato, potato, cassava and pea by 43.7%, 32.9%, 30.6% and 22.4%, respectively. For monocots, it shows identity with that of rice, foxtail millet, wheat, barley and maize by 17.4%, 16.5%, 15.3%, 15.1%, 14.1% and 16.9% respectively (Table 4).

A phylogenetic tree was constructed on the basis of the sequence similarity of the mature proteins (Fig. 6). It was suggested that the GBSSI proteins be separated into two classes (Salehuzzaman et al. 1993). One class consists entirely of GBSSI from monocots. The second class consists entirely of GBSSI from dicots. The amaranth GBSSI was found in the group of dicots. In this group, amaranth was sister to the subgroup consisting of the rest of the dicots.
Detection of GBSS activity

We reported the identification of a gene encoding active GBSSI protein from grain amaranth (*A. cruentus*). The species had two types of starch, i.e., waxy and nonwaxy, and the perisperm of nonwaxy amaranth contained 67 kDa of active GBSSI protein (Fig. 1). *A. hypochondriacus*, the related species of *A. cruentus*, was also known to have two types of starch and 68 kDa of GBSSI protein in the nonwaxy type (Konishi *et al.* 1985). Moreover, we suggested the segregation pattern in the F$_2$ population of the cross between the nonwaxy and waxy perisperm is a good fit to the expected 3 : 1 ratio for monogenic inheritance (Table 3). This result was also accorded with the previous report by *A. hypochondriacus* which has been reported as completely segregated nonwaxy and waxy type in a 3 : 1 ratio in the next generation (Okuno and Sakaguchi 1982). Therefore, our results suggest that the amylose content in the perisperm starch of *A. cruentus* is determined by a single major structural gene, *Waxy* (*GBSSI*).

Isolation and characterization of the GBSSI gene in amaranth

This is the first report in which the GBSSI gene was sequenced and characterized in amaranth. The isolation of a cDNA clone of amaranth GBSSI was 2076 bp in length, containing ORF of 1821 bp and 3' UTR of sequence. The ORF encodes a 67.3 kDa protein of 606 amino acids. Then, we performed a genomic PCR-based approach, and seven positive different GBSSI genomic fragments were cloned and isolated. The combined sequence had 3492 nucleotides including an intron of 1821 nucleotides. Comparing the genomic sequence with the cDNA sequence, we identified intron positions. Total intron numbers in the GBSSI gene from amaranth were determined to be 12 divided by 13 exons (Fig. 3). This structure is similar to that of the nonwaxy type characterized in other species such as maize (Klosgen *et al.* 1980).

**Table 4.** Sequence comparisons of amaranth GBSSI cDNA and protein with those of other plant species

<table>
<thead>
<tr>
<th>Species* (NCBI/GenBank accession numbers of the nucleotide sequences)</th>
<th>Nucleotide sequence identity (%)</th>
<th>Full length protein identity (%)</th>
<th>Transit peptide identity (%)</th>
<th>Mature protein identity (%)</th>
<th>GC content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBSSI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dicotyledons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amaranth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cassava (X74160)</td>
<td>69.7</td>
<td>68.6</td>
<td>30.6</td>
<td>75.3</td>
<td>45.7</td>
</tr>
<tr>
<td>Potato (X58453)</td>
<td>69.3</td>
<td>67.8</td>
<td>32.9</td>
<td>73.8</td>
<td>44.8</td>
</tr>
<tr>
<td>Sweet potato (Ab071604)</td>
<td>69.4</td>
<td>68.7</td>
<td>43.7</td>
<td>73.1</td>
<td>47.1</td>
</tr>
<tr>
<td>Pea (X88789)</td>
<td>68.4</td>
<td>64.3</td>
<td>22.4</td>
<td>70.2</td>
<td>43.2</td>
</tr>
<tr>
<td>Perilla (AF210699)</td>
<td>67.5</td>
<td>67.1</td>
<td>–</td>
<td>72.7</td>
<td>47.9</td>
</tr>
<tr>
<td>Arabidopsis (NM103023)</td>
<td>66.8</td>
<td>65.5</td>
<td>–</td>
<td>72.7</td>
<td>45.2</td>
</tr>
<tr>
<td>Monocotyledons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rice (X65183)</td>
<td>62.1</td>
<td>60.6</td>
<td>17.4</td>
<td>67.8</td>
<td>61.6</td>
</tr>
<tr>
<td>Sorghum (SBU23945)</td>
<td>62.4</td>
<td>60.2</td>
<td>15.3</td>
<td>67.0</td>
<td>63.9</td>
</tr>
<tr>
<td>Wheat (X57233)</td>
<td>58.5</td>
<td>56.9</td>
<td>15.1</td>
<td>64.0</td>
<td>64.3</td>
</tr>
<tr>
<td>Foxtail millet (AB089141)</td>
<td>61.6</td>
<td>60.7</td>
<td>16.5</td>
<td>67.6</td>
<td>65.1</td>
</tr>
<tr>
<td>Barley (X07931)</td>
<td>61.5</td>
<td>59.5</td>
<td>14.1</td>
<td>65.9</td>
<td>64.1</td>
</tr>
<tr>
<td>Maize (X03935)</td>
<td>59.4</td>
<td>60.7</td>
<td>16.9</td>
<td>67.7</td>
<td>66.2</td>
</tr>
<tr>
<td><em>Escherichia coli</em> (AAA23870)</td>
<td>38.4</td>
<td>30.4</td>
<td>–</td>
<td>–</td>
<td>55.3</td>
</tr>
</tbody>
</table>

* All sequence data were obtained from NCBI/GeneBank data base (http://www.ncbi.nlm.nih.gov/).

**Fig. 3.** Schematic representation of the amaranth GBSSI structural gene. Box and lines represent exons and introns. Underline shows 7 positive genomic fragments amplified by genomic PCR.

**Fig. 4.** The structure of amaranth GBSSI protein. Three conserved boxes (positions 93–98, 479–486 and 502–509) are indicated as Box 1, 2 and 3, respectively. An arrowhead indicates predicted cleavage site of the transit peptide from the mature protein.

**Discussion**

Detection of GBSS activity

We reported the identification of a gene encoding active GBSSI protein from grain amaranth (*A. cruentus*). The species had two types of starch, i.e., waxy and nonwaxy, and the perisperm of nonwaxy amaranth contained 67 kDa of active GBSSI protein (Fig. 1). *A. hypochondriacus*, the related species of *A. cruentus*, was also known to have two types of starch and 68 kDa of GBSSI protein in the nonwaxy type (Konishi *et al.* 1985). Moreover, we suggested the segregation pattern in the F$_2$ population of the cross between the nonwaxy and waxy perisperm is a good fit to the expected 3 : 1 ratio for monogenic inheritance (Table 3). This result was also accorded with the previous report by *A. hypochondriacus* which has been reported as completely segregated nonwaxy and waxy type in a 3 : 1 ratio in the next generation (Okuno and Sakaguchi 1982). Therefore, our results suggest that the amylose content in the perisperm starch of *A. cruentus* is determined by a single major structural gene, *Waxy* (*GBSSI*).
Fig. 5. Comparison of the deduced amino acid sequence of GBSSI plant species and E. coli glycogen synthase. The three highly conserved regions are indicated as boxes 1, 2, 3. Bold bar highlight the N-terminal KTGGL motif involved in substrate binding and the C-terminal KTGGL ‘looks-like’ motif. The information of all amino acid sequence shows from Table 3.
1986), rice (Okagaki 1992), foxtail millet (Fukunaga et al. 2002), potato (van der Leij et al. 1991), and sweet potato (Kimura et al. 2000). All five-plant species contain 14 exons, with exons 2–14 (13 exons) contributing to the coding sequence. In this study, 5′UTR, including exon 1 and intron 1, was not detected. However, amaranth GBSSI gene contained 13 exons (coding sequence) and this structure was similar to those in five plant species. Therefore, we need to analysis more about the region of 5′UTR and the future studies will clarify this important issue.

Comparison with GBSSI proteins from other plants species

Alignment of the deduced amino acid sequence showed a widespread similarity between GBSSI of various plant species. The regions of full-length protein were 56.9%–68.7% identical between amaranth and other plant species (Table 4). Furthermore, comparison of the primary structure of the mature proteins of amaranth to those of the GBSSI proteins from other plant species showed that the mature protein regions were more than 70% identical among dicots. In particular, the amaranth GBSSI protein was more than 75.3% identical to that of a cassava (Table 4). These results were reflected in the relationships among species in phylogenetic tree (Fig. 6). The scale bar represents an evolutionary distance. Bootstrap values were shown at each node.

![Fig. 6. The phylogenetic tree of the GBSSI protein of plant species and glycogen synthase protein of E. coli. The tree constructed by the neighbor joining (NJ) method. The scale bar represents an evolutionary distance. Bootstrap values were shown at each node.](image)

The GC contents of the GBSSI genes of dicots including cassava, potato, sweet potato, pea, perilla and arabidopsis were lower than those of monocots such as rice, sorghum, wheat, foxtail millet, barley and maize (Table 3). This result was consistent with the general characteristics of different genes of dicot and monocots (Campbell and Gowri 1990). A comparison of GBSSI primary amino acid sequences in amaranth and other plant species revealed conserved regions with strong similarity. Three highly conserved regions of amino acid sequences from GBSSI in the other plant species such as E. coli glycogen synthase (Kumar et al. 1986, van der Leij et al. 1991) were also found within the coding region of amaranth GBSSI (Fig. 5). The first conserved region (position 91–106) contains the consensus motif KTGGL, which is believed to be the ADP-Glc binding site (Furukawa et al. 1990) found in amino acids 94–98 of the deduced GBSSI protein. A second conserved region is located at residues 478–486. Within this region, the sequence PSRFEPCGL is identical between dicots, rice, and E. coli sequences and shares one conservative substitution compared with the other monocot plant sequences. A third conserved region (position 502–509) contains the KTGGL “look-alike” motif, (Dry et al. 1992) STGGL. This motif sequence and the tail is one of the most important specific properties of GBSSI gene determined by a C-terminal region (Edwards et al. 1999). The KTGGL “look-alike” motif is so named because it resembles the N-terminal KTGGL motif identified as a site of ADP/ADP glucose binding, although it lacks the lysine residue that is thought to interact with the polyphosphate group of ADPG and ADP in bacterial glycogen synthase (Furukawa et al. 1993). Apart from these conserved regions, the amaranth GBSSI proteins share only around 30.4% identity with the E. coli glycogen synthase, the remaining identical residues being spread throughout the protein sequence (Table 4).

![Fig. 6. The phylogenetic tree of the GBSSI protein of plant species and glycogen synthase protein of E. coli. The tree constructed by the neighbor joining (NJ) method. The scale bar represents an evolutionary distance. Bootstrap values were shown at each node.](image)
Transit peptide prediction

The TargetP suggested that the deduced protein sequence of GBSSI contained the chloroplast transit peptide with a score of 0.950. It suggested that the peptide was located in the chloroplast as well as GBSSI of other plant species. The ChloroP predicted the presence of the transit peptide and the cleavage site of the amaranth GBSSI (Fig. 2 and Fig. 4). The consensus cleavage site sequence proposed for the chloroplast transit peptides is I/VXA/C↓A (Gavel and von Heijne 1990). In the dicots, the first amino acid after the cleavage site sequence is G instead of A as in monocots (Salehuzzaman et al. 1993). For example, in pea and cassava, the cleavage site sequence of GBSSI is IVC↓G and IIC↓G (Dry et al. 1992, Salehuzzaman et al. 1993). In the case of amaranth, its sequence showed no significant homology to the transit peptide of other plant species. It starts with ME and the candidate cleavage site is FIR↓S. However, the nearest neighbor residue on the C-terminal side of the region is G, and the IRSG sequence becomes similar to the other cleavage site sequences of dicots.

Therefore, it is possible that the lengths of the transit peptide could be 77 amino acids downstream from the ATG initiation codon with a molecular mass of 8.7kDa. The peptide was similar in size to the GBSSI transit peptide of sweet potato (Kimura et al. 2000), potato (van der Leij et al. 1991), and cassava (Salehuzzaman et al. 1993). The transit peptide of amaranth has 43.7% identity over amino acids with that of the sweet potato but much less with other plant species (14.1~32.9%) (Table 4). These results show that the sequence of the transit peptide in amaranth was generally low identity when compared with that observed in the sequence of mature protein.

Conclusion

This study suggested the amylase content in the perisperm starch of *A. cruentus* was determined by a single major structural gene, *Waxy* (GBSSI). In addition, we provided the complete sequence of the GBSSI cDNA and genomic structure, clarified the characterization of the GBSSI protein including the predicted presence of the transit peptide and cleavage site. This was revealed the strong similarity among conserved regions in the comparison of GBSSI primary amino acid sequences in amaranth and other plant species. Its conserved regions also showed similarity to the *E. coli* glycogen synthase. However, some regions (cleavage site) of the amaranth sequence had different amino acids sequences than that found in the other dicots.

This study will provide more useful information for understanding the structure and function relationship of this protein. Moreover, these results will be helpful to identify the structure of GBSSI mutations and properties for further development of GBSSI amaranth varieties.

Acknowledgment

We express our sincere thanks to Dr. D. Brenner of USDA-ARS-MWA-PIRU; Iowa State University, USA, for providing the accessions used in this experiment.

Literature Cited


