Mol Gen Genet (1993) 237:10–16

MGG

Molecular analysis of the gene encoding a rice starch branching enzyme

Tsutomu Kawasaki[1], **Kouichi Mizuno**[2], **Tadashi Baba**[2], **and Hiroaki Shimada**[1]

[1] Mitsui Plant Biotechnology Research Institute, Sengen 2-1-6, Tsukuba 305, Japan
[2] Institute of Applied Biochemistry, University of Tsukuba, Tsukuba 305, Japan

Received June 26, 1992 / Accepted August 28, 1992

Summary. The sequence of a rice gene encoding a starch branching enzyme (*sbe1*) shows extreme divergence from that of the rice gene, that is homologous to bacterial glycogen branching enzyme (*sbe2*). *sbe1* is expressed abundantly and specifically in developing seeds and maximally in the middle stages of seed development. This expression pattern completely coincides with that of the *waxy* gene, which encodes a granule-bound starch synthase. Three G-box motifs and consensus promoter sequences are present in the 5′ flanking region of *sbe1*. It encodes a putative transit peptide, which is required for transport into the amyloplast. A 2.2 kb intron (intron 2) precedes the border between the regions encoding the transit peptide and the mature protein, and contains a high G/C content with several repeated sequences in its 5′ half. Although only a single copy of *sbe1* is present in the rice genome, Southern analysis using intron 2 as a probe indicates the presence of several homologous sequences in the rice genome, suggesting that this large intron and also the transit peptide coding region may be acquired from another portion of the genome by duplication and insertion of the sequence into the gene.

Key words: Gene expression – Nucleotide sequence – Starch branching enzyme – Rice seed development – Promoter

Introduction

Starch is synthesized in plants through the combined activity of several enzymes, including starch synthase (E.C. 2.4.1.21) which produces amylose, an essentially linear polysaccharide, and starch branching enzyme (α-1,4-glucan: α-1,4-glucan-6-glucosyltransferase (E.C. 2.4.1.18), which catalyzes the synthesis of α-1-6 branching linkages characteristic of amylopectin. The physical properties of a grain of rice are strongly affected by the relative abundance of amylose and amylopectin in that grain. The amylopectin content of starch granules in rice endosperm is estimated to be more than 70%.

Several different starch branching enzymes have been identified in photosynthetic and also in nonphotosynthetic tissues of maize and pea (Boyer and Preiss 1978; Dang and Boyer 1988; Smith 1988; Smith et al. 1990). Bhattacharyya et al. (1990) have shown that the wrinkled-seed character of pea, which was first observed by Mendel, is caused by lack of a starch branching enzyme. This enzyme has a high degree of homology to the bacterial glycogen branching enzyme. Three different branching enzymes (BE-I, BE-IIa, and BE-IIb) have been identified in the starchy endosperm of maize (Boyer and Preiss 1978). Recently, we cloned the BE-I cDNA from maize, and found that the deduced amino acid sequence is highly divergent from those of the bacterial branching enzymes and the pea enzyme (Baba et al. 1991). This result suggests that more than two different types of branching enzymes exist in plant tissues.

We have cloned two genes from rice, *sbe1*, and *sbe2*, which correspond to these coding for maize BE-I and the bacterial enzyme, respectively. In this report, we describe the structure and expression patterns of the rice *sbe1* gene. We also present comparisons with the genes encoding other branching enzymes and discuss the evolutionary implications of these findings.

Materials and methods

Plant DNA, plasmids and rice genomic library

Plasmids pMB9, pRB13, and pWX15A containing cDNA clones encoding maize BE-I (Baba et al. 1991), rice BE-I, and the rice *waxy* protein, respectively, were used for hybridization probes. Isolation and characterization of the rice cDNAs will be described elsewhere. The *waxy* gene encodes a granule-bound starch synthase expressed in the starchy endosperm (Klösgen et al. 1986). A rice genomic library was purchased from Clontech

Correspondence to: H. Shimada

Laboratories (Palo Alto, Calif., USA). Rice (*Oryza sativa* L. Japonica, cv. Nipponbare) and *Arabidopsis thaliana* (Columbia) genomic DNAs were prepared from leaves according to Rogers and Bendich (1985).

Southern blot hybridization

Plant genomic DNA was separated on 0.8% agarose gels, then blotted onto nylon membranes (Hybond-N, Amersham, UK) after digestion with a series of restriction enzymes. DNA probes were prepared from pRB13. Hybridization was carried out according to the procedure described by Ausubel et al. (1987). Positive bands were detected using a BAS2000 Bio-image Analyzer (Fuji Photo Film Co. Ltd., Tokyo).

Isolation of genomic DNA encoding the rice branching enzyme

Phage clones containing DNA fragments encoding the branching enzyme were identified in a rice genomic library by plaque hybridization with the maize BE-I cDNA. Plaque hybridization was carried out according to Benton and Davis (1977). A physical map of the lambda clone was made with *Sal*I and *Eco*RI. The *Sal*I fragments and the overlapping *Eco*RI fragments from the clone were subcloned into pBR322 or pUC119.

Isolation by PCR of a genomic fragment homologous to a gene encoding a bacterial branching enzyme

The polymere chain reaction (PCR) was carried out according to Innis and Gelfand (1990). Three forward PCR primers: B1, AGTTCAATTATTGGGATGG; B2, GGATTTGGGAGTTCTTTAT; B4, GATGGTTCCTGGGGTTATCA, and two reverse primers: B7, AGCATCCAGCCTAAGTTCCA and B8, TGCCATGCATCACCTGGCAT, were synthesized using a DNA synthesizer (Applied Biosystems, Foster, Calif., USA). The amplified fragment was cloned into the *Hinc*II site of pUC118 after blunting the ends of the fragment using the Klenow enzyme.

Nucleotide sequence analysis

The dideoxy chain-termination method (Sanger et al. 1977) was used for nucleotide sequence determination in pUC118/119 (Takara Shuzo Co. Kyoto, Japan) and in the Bluescript vectors (Stratagene, La Jolla, Calif., USA). Computer-aided analysis of the nucleotide sequences was carried out using GENETYX program (Software Development Co., Tokyo).

Northern blot analysis

Developing rice seeds were harvested daily between 3 and 30 days after flowering. Leaves, stems, and roots collected from mature plants were stored at −80 °C immediately after being frozen in liquid nitrogen. Total RNA was extracted from each of these tissues by previously described methods (Baba et al. 1991). After electrophoresis of RNA samples on formaldehyde agarose gels (up to 1.1% formaldehyde), Northern blot analysis was carried out on a nylon membrane (Hybond-N, Amerhsam). The hybridization conditions were according to the method described of Ausubel et al. (1987).

Results

Number of genes encoding the rice branching enzyme

The number of genes in the rice genome that show homology to BE-I was determined by Southern blot analysis with the rice BE-I cDNA. As shown in Fig. 1, only one or very few genomic fragments hybridized to the BE-I cDNA. Since the cDNA contains several *Eco*RI and *Bam*HI sites, we expected to find several fragments in genomic digests with these enzymes. When a short fragment prepared from either end of the cDNA was used as the hybridization probe, only a single positive band was obtained after digestion with any of these enzymes. All these bands corresponded to those predicted

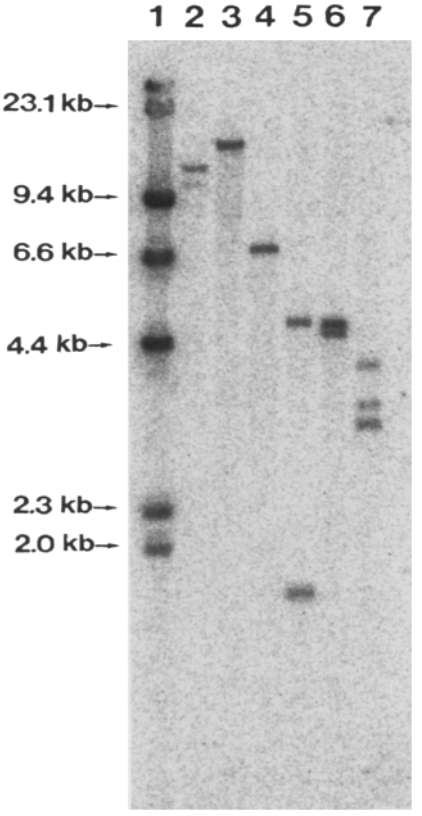
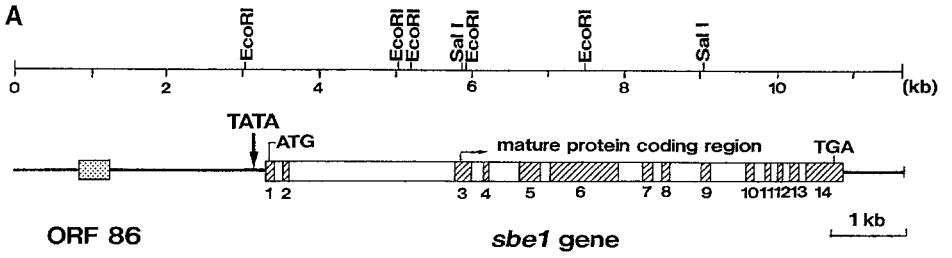


Fig. 1. Southern blot analysis of genomic rice DNA. Lane 1 contains lambda DNA digested with *Hin*dIII as size markers. The remaining lanes contain rice genomic DNA digested with: lane 2, *Bam*HI; lane 3, *Bgl*II; lane 4, *Dra*I; lane 5, *Eco*RI; lane 6, *Eco*RV; and lane 7, *Hin*dIII, respectively. The rice BE-I cDNA was used as a probe, and was prepared by partial digestion with *Eco*RI



B

```
3010      3020      3030      3040      3050      3060      3070      3080      3090      3100
AACAGTGTCAGCATAGAAATCTCATTACCTTTCGCTTTTGTCTTGTATTATGATTTACAGGAATTCGGTTTGTATTTATTTGCGTCTGGAAAAAGAAAAG

3110      3120      3130      3140      3150      3160      3170      3180      3190      3200
GAAGGAGAGACACGTGAAGGCCCATGGCAATTGGCCAAAGGCTCCTGGGCACCTCCTGGCCGTCCACGTGGCAGGTGTCCACGTCAGCACTTTGGCTTTG

3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
TTTTCTCCTTTTTTTTTCTCCCAATTTTCACTCCACTGCTGCACAAGCTTTTCCGTGCTTCCTCGCCGCCTCCGGCCTCCGCTCCGGCGCTATAAATCGC

3310      3320      3330      3340      3350      3360      3370      3380      3390      3400
CGCCGATTTCGAAGCTGTGGAAATGGGAGTCGCCTCCACGGCCACCGACATCCGCCGCAATGCTGTGTCTCACCTCCTCTTCCTCCTCCGCGCCCGCTCC
                                                  M  L  C  L  T  S  S  S  S  S  A  P  A  P

3410      3420      3430      3440      3450      3460      3470      3480      3490      3500
GCTCCTTCCCTCTCTCGCTGATCGACCGAGCCCGGGAATCGCGGTCAGTCAATATATAGAAACTTCTATATATGATATGATTAGTTTTCCCTCTCTCTCT
 L  L  P  S  L  A  D  R  P  S  P  G  I  A                                                              Intron 1

3510      3520      3530      3540      3550      3560      3570      3580      3590      3600
TGTTTTGTTAGGTGTTTGGGTTATTGATGGTGTGGTGGTCCGCAGGGCGGGGGTGGCAATGTTCGCCTGAGCGTGGTTTCTTCGCCGCGCCGGTCGTGGC
                                              G  G  G  G  N  V  R  L  S  V  V  S  S  P  R  R  S  W  P

3610      3620      3630      3640      3650      3660      3670      3680      3690      3700
CTGGAAAAGGTAAGATACTGTTGCAGGATATCTATCGAATTAGGAATTTGATGATGGTGATAATTAATTGTGGGGAAAACAAGTAAGTCTGAATTTCTTTG
 G  K

3710      3720      3730      3740      3750      3760      3770      3780      3790      3800
GCCACAGGCGACAGCCCCAAAACCCCGACACCCCACCGCCGACCACCTCGCCGCCGGCCACCGCTCCCCTTTGCCGCCGTCGGCTCTACCTCCTTCCCAC

3810      3820      3830      3840      3850      3860      3870      3880      3890      3900
TCTCCCGTCCTTCTCCTAGCCTCCCTCCTCTCTCTCTCTCTCTCAATACTTTTTTCTACCTTTCATACTACCTTCCATGTTCTTGTTCCGAATCTGGGTT

3910      3920      3930      3940      3950      3960      3970      3980      3990      4000
CCCAAGCCGCTGATGCGCCCCTTCCCTGATTCTTCTGATGGAACTAGGGGAGGCTGTGTGGCCATTTTTCCCGTTGGAGGGTTTCGTCTAGATCTGTCGG

4010      4020      4030      4040      4050      4060      4070      4080      4090      4100
GTGTGGGACATGCGGATTGCAGGTGCTGTCGGTTGTGTTGGCGGCGGCGGGTCCTGCCGGGATAGTTGGCCGCCGACGCGCCGCTTGGCTGTTGGGTTGC

4110      4120      4130      4140      4150      4160      4170      4180      4190      4200
ACGGTGTGTGCTGGCTGGTAGCGAGGATGGTTTTAGGGTGTTGGGCGAAAGCTCTGTCCGACTCATAGCCGGCCTGACGGCGATGAACGTCCTTGGACAT

4210      4220      4230      4240      4250      4260      4270      4280      4290      4300
CATGCAATGCCCCTCCTGGAGGCGTCGTCGCAAGAGCATCTCCAGTAGAGACCCTAAATACAATTCCTAAACAGTTTTAGGTGCTAAGGACAAAAAATA

4310      4320      4330      4340      4350      4360      4370      4380      4390      4400
AACTCCAGCAAAACCCATACTACAGGTCCTAAAATAGGAAGGACCTCAAATACCCCTCCGCAGTCCCTAGGCCTGGGGGCTGTAGACCGAGGCCCTATCG

4410      4420      4430      4440      4450      4460      4470      4480      4490      4500
CCGTTTTTCTACGCGGGAGGAAATTTCCTGACGTGTGGTGTCTGTCTTCCCTCCCGCGGAATCGCTGCCACGGCGCCGATCTTCGCCAGCTCGCTGTTCC

4510      4520      4530      4540      4550      4560      4570      4580      4590      4600
GCCGCTCGTGGCCGACGGTGCGACCATCCAGTACCTCCACCGGCCACTGCTTGTCGTCCGCGTGCCCACTTGCTTGTTTTTTCGTGGTCCTTGATCAGTT
                                                                                                      Intron 2
4610      4620      4630      4640      4650      4660      4670      4680      4690      4700
CGCACACTGATGCACTATATGGTAGACAAGAATGTTCTGAAATTCATGACCATCAGAAACATGTTCTAAACAATCCTGCTCTCGATTGGTTTATGGCTAA

4710      4720      4730      4740      4750      4760      4770      4780      4790      4800
CTGTGGTTCTAAACGATCATGGCATAAAAATTATTGTTCTGTTCCTTTAAAGTTTGTGGTGCTTGGTAGGTTGAGACAATTAGGCTGCTTGCAATTATGC

4810      4820      4830      4840      4850      4860      4870      4880      4890      4900
AGTAGTTCCTTCAAAGATTATTCTGCAGTGTTGTTCTTTTGTGTCAGTTGTGAGTTGAAGTTTAACTTCAAGGTTTTTTTTTTCTAGGAGGATTTAAGCT

4910      4920      4930      4940      4950      4960      4970      4980      4990      5000
CTTTCTGAAGTTTCTCAGATAGATTAGATTGGAAAAGGTATAGAGTTAATTTTATCTATTGATTATAGTTCTTATTTAATTGAACTACGTAGTGTCTTGA

5010      5020      5030      5040      5050      5060      5070      5080      5090      5100
ATACTTGCCGGTAGGATTTCACTCCCATGTTTGAGAATTTTGAATTTGAATTATGGTATTTAAAATTATGGATTTGAATACAATTGAATTCTATACATTA

5110      5120      5130      5140      5150      5160      5170      5180      5190      5200
GAAATATTCGTATTTGAATTATTACTATGTTAAACTAGGTGTAAGCATAGAGTATAATCAGAAATACAAGAGAAAAAGAAATGGGGGCTAAGAAATAGGG

5210      5220      5230      5240      5250      5260      5270      5280      5290      5300
TCTGCTGGTAGAGTTGGAGGTAATTTTTGAATTCTTAGAAAATAGGGACAGCCCTCATTCAACCTTTGAGGACTCTAAAATAGGGACTACTGCTGGAGAT

5310      5320      5330      5340      5350      5360      5370      5380      5390      5400
GCTCTAACACCCTGTTCCCCCTTGCTGCTGGGTGAAAACCCTTTCCAGTCTCCTGTTTATGCGATGGTGGCGTCCTTTCCGACGTCGTCACCTTCTTCAA

5410      5420      5430      5440      5450      5460      5470      5480      5490      5500
GGCATCGTTTTTGGAGAAACCCTGCAACCAGTCCCCCTGCTTTCCCATCCTTCTCCCCTATTCCATCCCCTCCTCCTCCCCTTTTCTTCTGTCAAGGGCT

5510      5520      5530      5540      5550      5560      5570      5580      5590      5600
CCTATGCTTGGAAACTCTCATGTATCTCTTCTCTGTAATATATTCAGGTGGGGAAATGTTGGATTTTTATTGATTGGAATACTGTATTGGGTCATCTCGG

5610      5620      5630      5640      5650      5660      5670      5680      5690      5700
TGACACCAAAGCTGTACTTTGGTGGAGTAGCAATCTTTGCCCTTATTGACCGGATAGGATTTTGGTTAAATTTATCTACGTTTTTGTTTGCGGTTCATCT

5710      5720      5730      5740      5750      5760      5770      5780      5790      5800
TTTTTCCTACCAGTCTTATACAAGATGGTACAGTTTAGCAATGATTGTTACATTGCAATATATAAATCGAAGTGATAGAAGCCACCTCAAGTAAATCTAA

5810      5820      5830      5840      5850      5860      5870      5880      5890      5900
CTATTGTTCATAATTCAAAGGTCAAGACCAATTTCTCAGTTCCTGCGACTGCGCGAAAAAACAAAACCATGGTGACTGTTGTGGAGGAGGTCGACCACCT
                    V  K  T  N  F  S  V  P  A  T  A  R  K  N  K  T  M  V  T  V  V  E  E  V  D  H  L

5910      5920      5930      5940      5950      5960      5970      5980      5990      6000
TCCTATATATGATCTGGACCCTAAGTTGGAGGAATTCAAGGATCACTTCAACTATAGGATAAAAAGATACCTCGACCAGAAATGCCTGATTGAAAAACAT
 P  I  Y  D  L  D  P  K  L  E  E  F  K  D  H  F  N  Y  R  I  K  R  Y  L  D  Q  K  C  L  I  E  K  H
```

Table 1. List of introns and sequence of exon/intron borders

Intron number	exon/	intron	/exon	GC content (%)
	(bp)			
1	GAATCGCG	GTCAGTCA.....	86... GTCCGCAG	GGCGGGGG 36.2
2	CTGGAAAG	GTAAGATA...	2196... ATTCAAAG	GTCAAGAC 45.7
3	TTCTAAAG	GTTAAGTT.....	99... TTTTGTAG	GCTATTTG 33.0
4	GCTGCACA	GTAAGTTC....	418... GCATATAG	AGAAGCAC 35.7
5	TGTGAAAG	GTCCTCTA.....	92... CCTTACAG	GTACGTGT 38.8
6	ATGATCAG	GTATATAG....	296... TTTTTCAG	TCCATTGT 34.9
7	TCCAAAAG	GTTATTTT....	141... TATTTCAG	ATGATTCA 35.7
8	GCAATGAG	GTAATATC....	421... TCTTTTAG	TTTGGCCA 35.9
9	GATACAAG	GTTGTGCC....	452... TTGAACAG	TATATGAA 36.3
10	AAGATAAG	GTAATGGC....	149... CTTTTCAG	GTTATTGT 36.3
11	TACAAGGG	GTAACTAA.....	65... ATGTGCAG	TTACAAAG 30.8
12	ATGGAAGA	GTAAGCAG....	101... TACCACAG	GTTGGCCA 38.5
13	CCTGTGTG	GTAAAGTC.....	66... ACTATCAG	GCTTACTA 36.6

Consensus sequences in the 5′ and 3′ end of introns are *underlined*

from the genomic sequence which is to be described elsewhere. Therefore, it appears that all these signals were derived from a single genetic locus, indicating that a single copy of the BE-I gene is present in the rice genome.

Isolation and analysis of a rice BE-I genomic clone

Thirteen positive phage plaques were obtained from a rice genomic library after screening with the maize BE-I cDNA. The physical maps of the genomic inserts indicated that they were all derived from a single genetic locus, consistent with the findings of genomic Southern blot analysis. We designated this gene *sbe1*. Figure 2A shows that the *sbe1* promoter and coding region lie within three *Sal*I fragments of 8.1 kb, 3.2 kb, and 6.1 kb.

Structure of sbe1

The region of the rice genome containing *sbe1* (12152 bp) was sequenced. The entire *sbe1* gene is approximately 8.5 kb in length. Alignment with the cDNA sequence shows that the gene is comprised of 14 exons and 13 introns (Fig. 2A). The cDNA sequence is identical to the corresponding genomic sequence. The borders between exons and introns were confirmed by comparison with the cDNA and are shown in Table 1. The consensus sequences, GT and AG, were found at 5′ and 3′ ends of all introns, following the "GT...AG" rule of the eukaryotic introns proposed by Breathnach and Chambon (1981). G/C content in the introns ranged between 30 and 40% except for intron 2, which had a G/C content of 45.7%. This is slightly higher than the G/C content of the cDNA (45.4%).

Introns 1 and 2 occur in the transit peptide-encoding region of *sbe1*. Intron 2 is the largest (2212 bp) and, in addition to its high G/C content, has a number of peculiar features. Figure 2B shows that this intron contains several direct as well as inverted repeat sequences and that stable secondary structures can occur in the 5′ half of the intron. The 3′ half of the intron is AT-rich, such that the G/C content of the 3′ half of the intron is less than 37.3%, whereas the G/C content of the 5′ half of the intron is more than 54.2%. Pyrimidine-rich sequences occur in both regions of the intron (Fig. 2B). Several fragments of the rice genome apart from the *sbe1* gene hybridized to a fragment of intron 2 in genomic Southern experiments (Fig. 3). This result indicates that although the rice genome contains only a single *sbe1* gene, it contains several regions with homology to intron 2 of this gene.

The sbe1 *promoter region*

As shown in Fig. 2B, sequences that are similar to the TATA and CAAT boxes, and are identical to consensus promoter sequence, are present in the 5′ flanking region of *sbe1*. Primer extension analysis indicated that transcription started from a position located 32 bp down-

Fig. 2A, B. Structure of the rice *sbe1* gene. **A** Physical map of the area around the rice *sbe1* gene. The *sbe1* gene and ORF86 are *boxed*. Exons and introns in the gene are indicated by *hatched boxes* and *open boxes*, respectively; the exons are *numbered* consecutively from the ATG initiation codon. A *shaded box* indicates ORF86. The coding region for the mature protein is depicted by a *bent arrow*. The *vertical arrow* indicates the position of the TATA sequence. **B** Nucleotide sequence of 5′ portion of *sbe1*. The protein coding regions are *boxed* and delineated by amino acid sequences *below* the nucleotide sequence. Putative CAAT and TATA sequences are *double underlined* and G-box like sequences are indicated by *wavy lines*. Direct and inverted repeat sequences are shown by *arrows*. Pyrimidine-rich sequences lying in the 2nd intron are indicated by *broken underlines*. A putative processing site for the transit peptide is depicted by an *open vertical arrow*, and the N-terminus of the mature proteins are shown by *closed vertical arrows*. The ATG codon near the border between the transit peptide coding region and the mature protein coding region is *boxed*

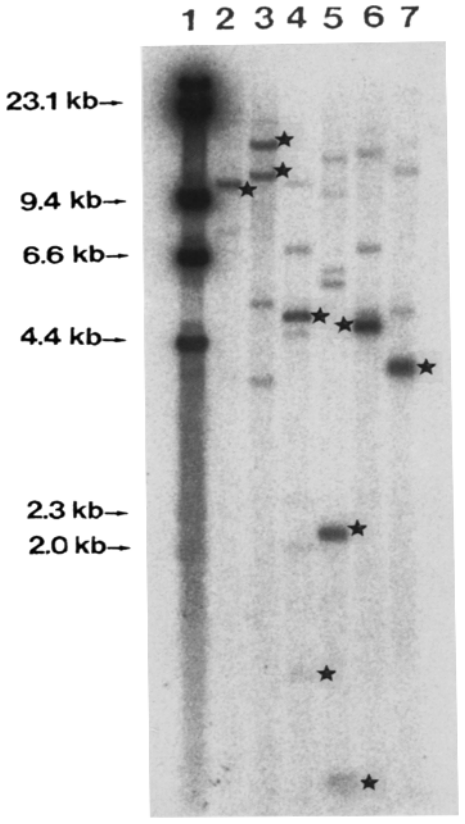


Fig. 3. Southern blot analysis of the rice genome with intron 2 of *sbe1*. Lane 1 contains lambda DNA digested with *Hin*dIII as size markers. The other lanes contain rice genomic DNA digested with: lane 2, *Bam*HI; lane 3, *Bgl*II; lane 4, *Dra*I; lane 5, *Eco*RI; lane 6, *Eco*RV; and lane 7, *Hin*dIII. Bands predicted from the nucleotide sequence of *sbe1* are indicated by *stars*. A fragment containing the entire region of intron 2, which was prepared by polymerase chain reaction, was used as a probe

stream from the TATA sequence (data not shown). Therefore, this TATA sequence may be a functional element of the *sbe1* promoter. The GC-rich repeated sequences, TCGCCGCC and CTCCGGC are also present near the TATA sequence. Upstream of the CAAT sequence, a G-box motif sequence is repeated. The second of these is completely identical to the consensus G-box sequence, CCACGTGGC (Giuliano et al. 1988; see Fig. 2B). An open reading frame, ORF86, is present in the region upstream of *sbe1* (Fig. 2A). Northern blot analysis revealed that a 1.5 kb mRNA was independently transcribed from this region at early stages of seed development (data not shown).

Cloning of a homologue of a bacterial branching enzyme gene

The genes encoding starch branching enzymes have been cloned from pea and potato (Bhattacharyya et al. 1990; Koßmann et al. 1991) and share homology with the genes encoding bacterial branching enzymes (Baecker et al. 1986; Kiel et al. 1990). The corresponding sequence was isolated from rice via PCR using primers based upon the amino acid sequences of the bacterial branching enzymes. Initially, several different primers were synthesized and PCRs were carried out using various combinations of these primers. Of these, only one pair of primers, B4 and B7, efficiently amplified a specific 0.6 kb fragment in both the rice and *Arabidopsis* genomes. The nucleotide sequences of these amplified fragments from rice and *Arabidopsis* were analyzed and compared with each other. The rice fragment shares 64% homology with the *Arabidopsis* gene, and is also 62% homologous to the *Escherichia coli* branching enzyme gene. Despite the high degree of homology between this rice branching enzyme gene and the genes in *E. coli* and *Arabidopsis*, there is very little homology between this rice branching enzyme gene and *sbe1*. This result indicates that at least two very different branching enyzme genes, and probably more, are present in the rice genome. We have designated the gene homologous to that for the bacterial branching enzyme *sbe2*.

Comparison of the branching enzyme genes

The deduced amino acid sequences of the *sbe1* and *sbe2* genes were compared with those of the bacterial branching enzymes. As shown in Fig. 4, the *sbe2* gene products from rice and *Arabidopsis* are highly conserved with respect to the bacterial genes, suggesting that they are similar to the pea branching enzyme. In contrast, the only region of homology shared between the *sbe1* gene product and the *E. coli* branching enzyme lies in the central region. The total homology between them is less than 25%, while the level of homology between *sbe1* and *sbe2* gene products is even lower.

```
Genes                   Deduced amino acid sequence

Rice sbe2               YGRNEVRSFLISNAYYWLKEFHIDG-LRVDAVASMLYLDYSRKDGEWLPNAYG
Arabidopsis sbe2        .T.N..HGYM.AS.LHC.KH.....PTMN........R.....A...V..RH.
Synechococcus glgB      YG.H..RNFLA.N..FWFDKY....-IR-........L..N..E...I..EY.
E.coli glgB             ...R..S...VG..LY.IERFG..A-L.V......I.R..S......I...F.
                                     |                     |
                                    390                   410

Rice sbe2               GRENLDVIEFLKQLNVMVDGE-RFPGVTIAEESTAQPAVSRPVYAGGL
Arabidopsis sbe2        .....EA.D..RH..-D..D.LET..A.V........G..QSTQQ..PGLFL
Synechococcus glgB      ....I..A....QV.-HLIFS-YF....S.....S..M..WP.YV..L.FN.
E.coli glgB             ....L..IE...NT.RI.-GEQVS-.AVTM....DF.G..RPQDM.....WY
                         |                   |                   |
                        430                 450                 470
```

Fig. 4. Alignments of the *sbe2* gene products from rice and *Arabidopsis* with the *Synechococcus* and *Escherichia coli* glycogen branching enzymes. Amino acid residues identical to the corresponding residues in rice *sbe2* are indicated by *dots*. *Hyphens* indicate points were a sequence has been shifted to maximize homology. *Numbers* at the bottom of the sequences indicate the amino acid residue number in the *E. coli* enzyme

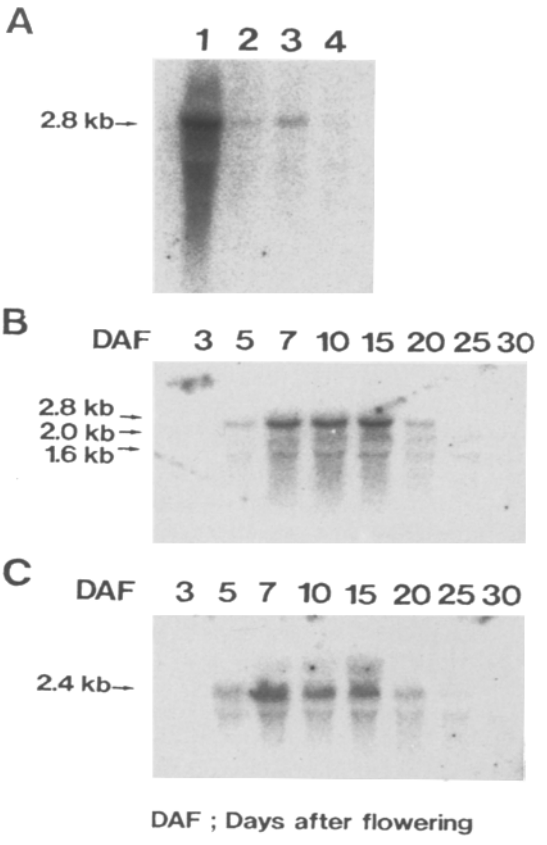


Fig. 5A–C. Northern blot analysis of the *sbe1* gene. **A** Tissue-specific expression of the *sbe1* gene. Lanes 1–4 contain 20 μg of total RNA isolated from developing seeds, leaves, stems, and roots, respectively. These were probed with the entire *sbe1* cDNA, prepared as described in Fig. 1. **B** Expression of *sbe1* during seed development. Each lane contains 10 μg of the total RNA isolated from seeds 3–30 days after flowering. The probe used was the same as in **A**. **C** Expression of the rice *waxy* gene during seed development. RNA was isolated and blotted as in **B**. The *waxy* cDNA was used as a probe. The *numbers* above the lanes in **B** and **C** indicate the number of days after flowering. The numbers to the left in **A**–**C** indicate the size of the mRNA detected

Tissue specificity of sbe1 *gene expression*

Northern blot analysis was performed to determine the tissue-specific expression of *sbe1* in rice. A 2.8 kb mRNA corresponding to the full-length *sbe1*-cDNA was detected in all organs tested. Extremely strong signals were obtained from the developing seeds. Signals approximately twentyfold weaker than those observed in seeds were found in leaves and stems, and far less in the roots (Fig. 5A). In addition to the 2.8 kb mRNA, mRNAs of 2.0, and 1.6 kb were also present in developing seeds (Fig. 5B). When fragments from either the 5′ or 3′ end of the cDNA were used for Northern blot analysis, only the 2.8 kb fragment was detectable (data not shown). Therefore, it appears that the 2.0 and 1.6 kb fragments may be products of mRNA degradation. When the intron 2 region was used for Northern blot analysis as a probe, no distinct mRNA corresponding to the intron was observed (data not shown), suggesting that this region is spliced efficiently.

Expression of sbe1 *during seed development*

Figure 5B shows *sbe1* gene expression in developing rice seeds. As mentioned above, three distinct bands of 2.8 kb, 2.0 kb, and 1.6 kb mRNA were also apparent throughout seed development. The *sbe1* mRNA level gradually increased after flowering and reached a maximum in the middle stage of seed development, between 7 and 15 days after flowering (DAF). By 20 DAF, *sbe1* mRNA began to disappear rapidly and was no longer detectable by 25 DAF. This expression pattern is very similar to that of the *waxy* gene, which encodes a granule-bound starch synthase (Fig. 5C).

Discussion

Rice plants contain more than two different starch branching enyzmes. Among them, BE-I, which produces amylopectin during seed development, is the major enzyme. Northern blot analysis revealed extremely low levels of expression of BE-I in leaves, stems, and roots in contrast with the abundant level of expression of this enzyme in developing seeds (Fig. 5A). It appears that this gene is constitutively expressed at low levels in most tissues of the plant and then is strongly induced during seed development. Both *sbe1* and *waxy* gene expression were maximal during the mid stage of seed development (Fig. 5B, C), consistent with the physiological observations reported by Asaoka et al. (1985).

We also isolated a rice gene homologous to the bacterial branching enzyme (*sbe2*). Expression of this gene was not detected in developing seeds or in leaves or stems; therefore, it may be expressed only at very low levels or is no longer expressed. Analysis of starch metabolizing enzymes in developing rice seeds has revealed many functional branching isozymes (Mizuno et al. 1992). However, no other gene with homology to *sbe1* or *sbe2* has been detected. Therefore, it appears that the other isozymes share little homology with *sbe1* and *sbe2*.

The 5′ flanking region of *sbe1* contains a number of promoter consensus sequences, which probably function as the promoter of this gene. In addition to these sequences, a number of repeated sequences and also G-box motifs are present in the promoter region. This suggests that *trans*-acting factors are involved in the regulation of *sbe1* gene expression. The *sbe1* gene contains 14 exons and 13 introns. Intron 2 is extremely large (Fig. 2A). This structural feature is common among genes involved in starch biosynthesis within amyloplasts, such as the *waxy* gene which encodes a granule-bound starch synthase and the *agpp* gene which encodes an ADP-glucose pyrophosphorylase (e.g. Rhode et al. 1988; Wang et al. 1990; Hirano and Sano 1991; Anderson et al. 1991). The first or second introns in these genes have been shown to be large. A large intron, similar to intron 2 of *sbe1*, is present in the gene encoding an anther-specific protein in sunflower, which also contains a transit peptide (Domon et al. 1991). Such large introns may be common in genes that contain transit peptide-coding regions. Intron 2 of *sbe1* has a GC-rich sequence that can form stable secondary structures. Since a stable secondary

structure in the intron may inhibit its splicing (Goodall and Filipowicz 1991), other factors may be required for efficient splicing.

Intron 2 of *sbe1* is located in the region immediately preceding the border between the transit preptide-coding region and the segment encoding the mature protein. Since several regions homologous to intron 2 are present in the rice genome (Fig. 3), the region including intron 2 may have been acquired via duplication and insertion of the sequence from another part of the genome. Nuclear genes encoding proteins that function in the plastid are thought to have originated in the plastid DNA, and to have relocated in the nucleus over the course of evolution (Gantt et al. 1991). Since proteins that are encoded in the nucleus and transported efficiently into the plastid must also have acquired a transit peptide during plant evolution (Nugent and Palmer 1991), a similar event might have occurred in this gene, and during this process the large intron might have been created near the border between the region encoding the transit peptide and the region encoding the mature protein.

When the coding sequences of rice and maize BE-I genes are aligned, it appears that the putative processing site of the transit peptide of the rice BE-I is located between alanine and arginine residues. However the N-terminus of rice BE-I may be processed further, since the mature protein begins with either a threonine or valine residue, which are -1 and $+1$ residues, respectively, from the methionine residue encoded by the ATG codon in the 3rd exon (Fig. 2B). The peptide preceding this methionine may not be essential for branching enzyme function in rice BE-I. This ATG codon might be a prokaryotic remnant of an ancestral gene, which predates acquisition of the region encoding the transit peptide, and may at one time have been a plastid-encoded gene.

Acknowledgements. We thank Dr. M. Sugiura for valuable suggestions and Drs. T. Tsuchiya, M. Yoshida, Y. Tada, and T. Fujimura for assisting our experiments. Rice plants were a gift of Life Science Laboratory of Mitsui Toatsu Chemicals, Inc. The nucleotide sequence data reported here have been submitted to the DDBJ, EMBL, and GenBank nucleotide sequence databases (accession nos. D10838 and D10839 for rice *sbe1* and *sbe2*, respectively).

References

Anderson JM, Larsen R, Laudencia D, Kim WT, Morrow D, Okita TW, Preiss J (1991) Molecular characterization of the gene encoding a rice endosperm-specific ADP-glucose pyrophosphorylase subunit and its developmental pattern of transcription. Gene 97:199–205

Asaoka M, Okuno K, Sugimoto Y, Fuwa H (1985) Developmental changes in the structure of endosperm starch of rice (*Oryza sativa* L.). Agri Biol Chem 49:1973–1978

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman TG, Smith TA, Struhl K (1987) Current protocols in molecular biology. Wiley, New York

Baba T, Kimura K, Mizuno K, Etoh H, Ishida Y, Shida O, Arai Y (1991) Sequence conservation of the catalytic regions of amylolytic enzymes in maize branching enzyme-I. Biochem Biophys Res Commun 181:87–94

Baecker PA, Greenberg E, Preiss J (1986) Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* 1,4-α-D-glucan 6-α-D-glucano)-transferase as deduced from the nucleotide sequence of the *glgB* gene. J Biol Chem 261:8738–8743

Benton WD, Davis RW (1977) Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196:180–182

Boyer CD, Preiss J (1978) Multiple forms of (1,4)-α-D-glucan, (1,4)-α-D-glucan-6-glycosyl transferase from developing *Zea mays* L. kernels. Carbohydr Res 61:321–334

Bhattacharyya MK, Smith AM, Ellis THN, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell 60:115–122

Breathnach R, Chambon P (1981) Organization and expression of eukaryotic split genes coding for proteins. Annu Rev Biochem 50:349–383

Dang PL, Boyer CD (1988) Maize leaf and kernel starch synthases and starch branching enzyme. Phytochemistry 27:1255–1259

Domon C, Evrard J-L, Pillay DTN, Steinmetz A (1991) A 2.6 kb intron separates the signal peptide coding sequence of an anther-specific protein from the rest of the gene in sunflower. Mol Gen Genet 229:238–244

Gantt JS, Baldauf SL, Colio PJ, Wooden NF, Palmer JD (1991) Transfer of *rpl22* to the nucleus greatly preceded its loss from chloroplast and involved the gain of an intron. EMBO J 10:3073–3078

Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik PA, Cashmore AR (1988) An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. Proc Natl Acad Sci USA 85:7089–7093

Goodall GJ, Filipowicz W (1991) Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J 10:2635–2644

Hirano H, Sano Y (1991) Molecular characterization of the *waxy* locus of rice (*Oryza sativa*). Plant Cell Physiol 32:989–997

Innis MA, Gelfand DH (1990) Optimization of PCRs. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR protocols a guide to methods and applications. Academic Press, San Diego, Calif., pp 3–12

Kiel JAKW, Boels JM, Beldman G, Venema G (1990) Nucleotide sequence of the *Synechococcus* sp. PCC7942 branching enzyme gene (*glgB*): expression in *Bacillus subtilis*. Gene 89:77–84

Klösgen RB, Gierl A, Schwarz-Sommer Z, Seadler H (1986) Molecular analysis of the *waxy* locus of *Zea mays*. Mol Gen Genet 203:237–244

Koßmann J, Visser RGF, Müller-Röber B, Willmitzer L, Sonnewald U (1991) Cloning and expression analysis of a potato cDNA that encodes branching enzyme: evidence for co-expression of starch biosynthetic genes. Mol Gen Genet 230:39–44

Mizuno K, Kimura K, Arai Y, Kawasaki T, Shimada H, Baba T (1992) Starch branching enzymes from immature rice seeds. J Biochem 112:643–651

Nugent JM, Palmer JD (1991) RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell 66:473–481

Rhode W, Becker D, Salamini F (1988) Structural analysis of the *waxy* locus from *Hordeum vulgare*. Nucleic Acids Res 16:7185–7186

Rogers SO, Bendich AJ (1985) Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol Biol 5:69–76

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Smith AM (1988) Major differences in isoforms of starch-branching enzyme between developing embryos of round- and wrinkled-seeded peas (*Pisum sativum* L.). Planta 175:270–279

Smith AM, Neuhaus HE, Stitt M (1990) The impact of decreased activity of starch-branching enzyme on photosynthetic starch synthesis in leaves of wrinkled-seeded peas. Planta 181:310–315

Wang Z-y, Wu Z-1, Xing Y-y, Zheng F-g, Guo X-1, Zhang W-g, Hong M-m (1990) Nucleotide sequence of rice *waxy* gene. Nucleic Acids Res 18:5898

Communicated by R.G. Herrmann