Update section

Sequence

Isolation and sequences of rice sucrose synthase cDNA and genomic DNA

Wei-Ping Yu¹, Ai-Yu Wang¹, Rong-Huay Juang¹, Hsien-Yi Sung¹ and Jong-Ching Su^{1,2,*}

¹ Department of Agricultural Chemistry, National Taiwan University, 1 Roosevelt Road Section 4, Taipei 10764, Taiwan (*author for correspondence); ² Institute of Biological Chemistry, Academia Sinica, P.O. Box 23–109, Taipei 10798, Taiwan

Received 5 August 1991; accepted 7 August 1991

Genomic DNA and poly(A)⁺ RNA were extracted and purified [1, 7] from excised leaflets and roots of 7- to 10-day old etiolated seedlings and milk-ripe grains, respectively, of rice (*Oryza sativa* cv. Tainong 67). Genomic and cDNA libraries were constructed by ligating *Sau* 3A digest [2] and synthesized cDNA [3] on the *Bam* HI site of λ EMBL3 and the *Eco* RI site of λ gt10 or pUC18 vectors, respectively [4, 7]. Screening of libraries was done by using a synthetic 17-mer (5'-TACATCTTCTTGAACTC-3', a sequence selected from exon 13 of maize *Sh1*) and maize SS *Sh1* (p17.6, on pBR322) and *Ss2* (p21.2, on pUC9) clones and probes [10]. Positive clones were subjected to restriction map analyses [8], subcloned in pUC18, pUC19, or pBluescript, subjected to unidirectional deletion to obtain overlapping clones [5, 6], and sequenced bidirectionally by the dideoxy termination method [9].

Among several positive cDNA clones that were sequenced, one from the λ gt10 library (pRSS2) had the feature of a full-length cDNA. It was 2890 bp long and contained a 2448 bp open reading frame (Fig. 1). From the homology analyses on the nucleotide and deduced amino acid sequences, it could be characterized as *RSs2*, the counterpart of maize *Ss2*. Five positive genomic

clones were analyzed for their restriction maps and sequenced. Among them, two had a complete overlap of about 1.1 kb, and together made up a sequence of 9.2 kb. It had 15 exons, one less than those of maize *Sh1*, that altogether formed an open reading frame completely identical with that of *RSs2* cDNA. Exon 15 of this rice gene appears to be the fusion product of exons 15 and 16 of maize *Sh1*. This feature has also been found in maize *Ss2* (L.C. Hannah, personal communication). The non-coding exon 1, intron 1, and the 1.1 kb upstream of exon 1 altogether had 2.66 kb, in which were found several unique structural features usually found in the regulatory gene. These sequences and other introns are presented in Fig. 2, and their insertion sites in the ORF are indicated in Fig. 1. The organization of *RSs2* gene is presented in Fig. 3.

Acknowledgement

This work was supported by grants from the National Science Council, Republic of China. The authors are especially grateful to Professor L.C. Hannah, University of Florida, Gainesville, for his generous gifts of maize clones p21.2 and p17.6.

ACCCCGTTGTTTTATTTCGTTTACTTCGTTTCTTGGGAAGGCACTTCCTTCCTCTCTAGGAAGGCTTGAGGAGCCAAGAGGAAGAGGAGCAATG
 M
 220
 GGGGAAGTGCCTGCGCAGCCGCTCTGAGCGCCTCCACAGCGTCAGGAGCGCATCGGCGACTCCCTCTCGGCCACCCCAATGAGCTCGTCGCGCTTCCACAGGCTGGTTAACCTC
 G E A A G D R V L S R L H S V R E R I G D S L S A H P N E L V A V F T R L V N L
 340
 GGAAAGGAATGCTTCAGGCCCAAGATCATTGCTGAGTACAACAACGCAATCTCTGAGGCTGATCGTGAGAAGCTGAAGGACGGTCTTTGAGGATGCTGAGGAGTGACACAG
 G K G M L Q A H Q I I A E Y N N A I S E A D R E K L K D G A F E D V L R S A Q E
 460
 GGGATTGTTATCTCCCATGGGTTGCCCTTGCATCCGCCGAGGCCGCTGTCTGGAGATGTGAGGGTCAATGTGAGCGAGCTCGCTGTTGAGTGTGACAGTCCCCGAGTACTTG
 G I V I S P W V A L A I R P R P G V W E Y V R V N V S E L A V E L L T V P E Y L
 580
 CAGTTCAAGGAACAGCTTGTGGAAGAAGCCACCAACAACACTTTGTGCTTGAAGTGGATTCGAGCCATTCAATGCCCTCTTCCCTCGTCTCTCTGTGCAAGTCCATTGGCAATGGT
 Q F K E Q L V E E G T N N N F V L E L D F E P F N A S F P R P S L S K S I G N G
 700
 GTGCAGTTCTCAACAGGATCTGTCTGCAAGCTCTTCCATGACAAGGAGAGCATGTACCCACTGCTCAACTTCTTGTGACACACAACAGGAGCATGACCATGTTGAATGAC
 V Q F L N R H L S S K L F H D K E S M Y P L L N F L R A H N Y K G M T M M L N D
 820
 AGAATCCGAGCTCTCAGTGCTCTCAAGTGCTGTGAGGAAGGCTGAGGAGCATCTTCTGGTCTTTCAGCAGACACCCCTTACTCAGAATTCACCAACAGGTTCCAGGAACCTGGTCTG
 R I R S L S A L Q G A L R K A E E H L S G L S A D T P Y S E F H H R F Q E L G L
 940
 GAGAAGGGTGGGGTACTGTGCCAAGCGTCCAGGAGACCATTCACCTCCTTTGAGCCTTCTTGAGGCCCTGATCCGTCACCCCTGAGAAGTTCCTTGAACAATCCCAATGGTG
 E K G W G D C A K R S Q E T I H L L L D L L E A P D P S T L E K F L G T I P M V
 1060
 TTCAATGTTGCTATTATGTCCACACGGTTACTTCGCGAAGCAATGTCTGGGTTACCTGACACTGGAGGGCAGTTGTCTACATTTTGACCAAGTCCGCTGCTATGGAGATGAG
 F N V V I M S P H G Y F A Q A N V L G Y P D T G G Q V V Y I L D Q V R A M E N E
 1180
 ATGCTTCTGAGGATCAAGCAGCAAGGCTCAACATTACACCAAGCATCTTATTGTCCAGGTTGCTCCCTGATGCAACTGGCACCACCTGCGGTGAGCTGTTGAGAAGTCTCTGGC
 M L L R I K Q Q G L N I T P R I L I V T R L L P D A T G T T C G Q R L E K V L G
 1300
 ACCGAGCACACCCACATCTACGTGTGCTTTGAGAAGTGAATGTTGCAAGTGGATCTCAGGTTTGAAGTCTGGCGTACCTGGAAACATTCACTGATGATGTCGGGAC
 T E H T H I L R V P F R T E N G I V R K W I S R F E V W P Y L E T F T D D V A H
 1420
 GAAATGCTGGAGAGCTCCAGGCAATCTGACTTGATCATCGGAACTACAGTGACGAAACCTTGTGATGCTGCTTGACACAGAAGTGGGTGTACCATTTGACCATTTGCCAT
 E I A G E L Q A N P D L I I G N Y S D G N L V A C L L A H K M G V T H C T I A H
 1540
 GCGCTTGAGAAACCAAGTCCCCCACTCTGACCTCTACTGGAAGAAGTTGAGGATCACTACCCTTCTCATGCCAGTTCAACACTGACTTGATTGCGATGAACACGCTGACTTCATC
 A L E K T K S P N S D L Y W K K F E D H Y H F S C Q F T T D L I A M N H A D F I
 1660
 ATCACCAGTACCTTCAAGAGATTGCCGAAACAAGGACACTGTTGGTCAAGTCAAGTCTCAGTGGCAATCACCATGCTGCTGACCGTGTGTCCAGGATGATGTTTTGAC
 I T S T F Q E I A G N K D T V G Q Y E S H M A F T M P G L Y R V V H G I D V F D
 1780
 CCGAAGTTTAACTTGTCTCGCTGGTGGGACATGTCATCTACTTCCCTTACTCTGAGTCTCGTAAGAGGCTCACCTCCCTCCACCCAGAGATTGAGGAGTGTCTACAGCGAAGTT
 P K F N I V S P G A D M S I Y F P Y S E S R K R L T S L H P E I E E L L Y S E V
 1900
 GACAACAACGAGCACAAGTTTATGCTGAAGGACAGGAACAAGCAATCATCTTCCATGGCTCGTCTTGACCGTGTCAAGAAGTCTGACTGGTCTGGTTGAGCTGTATGGTGCACACCT
 D N N E H K F M L K D R N K P I I F S M A R L D R V K N L T G L V E L Y G R N P
 2020
 CGCTGCAAGAGCTGGTTAACTTGTGGTGTCTGTTGTTGACCATGGCAACCATCTAAGGACAAGGAGGAGCGGTGAGTTCAAGAAGATGTTGACCTTATTGAGCAATACAACTG
 R L Q E L V N L V V V C G D H G N P S K D K E E Q A E F K K M F D L I E Q Y N L
 2140
 AATGGCCACATCCGCTGATCTCCGCGAGATGAACCGTTCGCGAAGGAGGAGTCTACCGTTACATCTGCGACACCAAGGGTCTTTTGTGACGCTTCTACGAGGATTCGGT
 N G H I R W I S A Q M N R V R N G E L Y R Y I C D T K G A F V Q P A F Y E A F G
 2260
 CTCAGTGTGTTGAGTCCATGACCTGTGGTCTCCGACATTTCGCAACGCGCTATGGTGTCCAGTGAGATCATCGTGAACGAGTGTCTGGCTCCACATTGAECCTACACGAGGTGAC
 L T V V E S M T C G L P T F A T A Y G G P A E I I V N G V S G F H I D P Y Q G D
 2380
 AAGGCTCGGCTGCTCGTTGAGTCTTTGAGAAGTGCAGGAAGACCCGACCATGGACCAAGATCTCGAGGCGGTCTTCAGCGTATCAGGAGAAATACACCTGGAGGCTCTAC
 K A S A L L V E F F E K C Q E D P S H W T K I S Q G G L Q R I E E K Y T W K L Y
 2500
 TCTGAGAGGCTGATGACCTCACCGGCTCTACGATCTGGAAGTACGTCTCGAACCTGGAGAGGCGTGAACCCGCGCTACCTTGAGATGCTGTATGCCCTCAAGTACCGACCATG
 S E R L M T L T G V Y G F W K Y V S N L E R R E T R R Y L E M L Y A L K Y R T M
 2620
 GCTAGCACCGTCCCGTGGCGTGGAGGCGAGCGCTGCAACAAGTGAAGTGGCCCAACATCTCAGGCTGAGACTCTGAAAGAGGAGCATGGAGTTTCTTTGCTTGTTCATCTCTC
 A S T V P L A V E G E P S N K *
 2740
 TGGTTTCAGAGATGAAAGAGAGAGATATTCAGGATGCTTGCATTCCTAGGCACTGCTCTTTTCTTAAGAAAGTAAGGTCGATTGAATTTGGGAGTTTCTAGTTGATTAG
 2860
 TTGGGGTGAGTTCAGTGAATGTGGTGTGGTGTGCTGTCTCTCTGTTATTGGAGTGTACCTAGTAATAAGAAACCGGCTTGGTCCGTTTCTGCAATAATATATTGTA
 2875
 GCATCCTGTTTCCA

Fig. 1. The complete assemblage of *RSs2* exons. An arrow accompanied with a number indicates the site where an intron is present. The number is the intron number as shown in Fig. 2.

Intron1 GTAATAAAGAAGACCTCCTCCTCCATGGATAGATAGATCAATGGGAGCTCTGCATTCCCTTTCTCGTCTCTGTT
CTGTTCTTGGGTGTGTGTTCAGTGTCTTCTGCTCACCTAGAACTGTGGTGGTGTCTGCTGCATACCGAAATGCCGGTAA
TGGCGATGATAATATGAGCGCTCATGCTCATCCATGCTGCCGGAAGGAACACGATGACGGAATTTTCTTGGACTCGCGA
GTGATTCCGATCATGGGGGAGGTCTGCCAAATTTTCTCCGGTTCATCGGATTCAGAGGGGGAGGTGCTGCCCTCTGA
ATGCAATTGAGTGCACCTGAAACAGAAAAAGAAAAAGAACTCTGCTTCCCTTTTTCTCTCTCTCATGCTCGTCTCT
GCGTTAGGGGATAATAAAGCCGAGATCTAGTGAATATTGGGGGGTTAGCGAACATGATCATCGATAACGACATGGT
GAGATCTAGTGGTGTGATCTTGGCTTCTGGATCTGGGGGTTAGTTTCAGCAAGCCTGAAATTCACCTGGGTTAAGGT
TTCTTTTATTTTCTTTTTCATCAGATTGAAGTGGATAGAATGCTAGATCTAGTAATGTCCTATGATTGTTCTCTTTC
ACCTGAATTTTTTAAAGCATTTTGTGATGATTTCTCTCTCTCTCTCTCGGCTGACAAAGTTTATGCTTTATCCGTTCC
TCTGATGTGCCCTCTGCTCGGAGCTCGGCGACAGACAGTCTGACAGTTTGGGTCTGCTGCTAATGGCAGGGGTTAAA
TTTATTTACCTTCATAGTATTAATCTTCTTGTACTACCACTTAAATCATATTTTAAAGTAAGATTTAATTAATTAACG
CCGTGTCAACGTTAATGACCTCCTCCTTTGCCATGCCAACAAGAAACCCCAATCCATCCCATGCTGGCCCTGCTGCTCCG
GAGGAGAGGAGAGGAGAGGAGAGGAGCTTGAATGGCTCACTCTTCTGTTCCAGCTTTGCTGCTGCTAGCTTAGCTTCGAT
TCAACGGCTATGGGTGAGCTGGTGCAGCAGAAGCGACGATCTTTTTCGTTGAAACCATCGGATTGTTTGGTGGTTTCT
CTCCTCAGATCTGCATAAATACGCGCTACCCCGCTTCAACCCAGTTCCCTCGGCTCAGCCTGGAGCATCATGTTCTCTTC
CTCCCTCCCTTCTCCATTGTTGATTCCTGTATGCTTCTCAGTCAATTATAGCTGCTATGCTAGTCTTTTCTTTTCTTT
TCTTCTAGATTGCTTGTGAAAAAGCTGGGATTTTCTTTTCTTCCAGTGTGTCAGAGGCGTGGTTTTTGGTGGGGTT
TATGGCTTGACTTTTTTTTTTCTGGAGTGATTGTGTTGACTGGGTCTGGATGTATCTGATGATGGTGGTGGTGTGCA
CGCAG

Intron2 GTTCGTGCTGATTTAGTGATTTCCAGCATTAGATTTTGTGGTTTCTAGTCTACTGCCCTCAGATGTTACTGTATTTTC
TTTTAGAAGGAGATGTTTCATATAGGATCTTTGTTGTTGAGATTGTTAGATCTGGCCAGGAATGGCTCATATTTACTGAAT
TGGATGCAATCATTTTGTAGTCACTTTTTTTTAAAGTTTCTGATTAGAATGAATATTTAAGTGGGCTTCTGCAGCCAA
GATTTTGTACAAACCTAGTACTACTGAATAATGATGAATATACAAATGATGTTTGGATTACTGTGGACTGGTAGTGTCT
AGATCTGACTGCATGTGCATGTTATTTATATATATACGGTTTACAACTGAATACAGTAATGAATTTCTGCATGGT
ACAGATGCTTGTGTTGGTAGCAAGTTTACAAAAAATAAAAAACCTACATCTTACTAGATCTATTGGCGAGCGCGTA
GATCTGATTATCGCGCATATTTTCAATTAAGTCCAATTAATGGTCAAACTAATCATTTTCATATCTACATGAATTTTAA
ATTCTATCTCAATGCAACAGATCATATATGGTCTTTTAAAGTGGCTAATAGCAATTTTCTTATTATGCGCAATGCTCA
AGTGCTAAAAATTATCTTATTGAGATATTATAGGCTGAGTTGATAGATCTGGCCTGATATTTTGTGACACTGTTATTT
GCGCAG

Intron3 GTTTGAATCAGATTTCAGATTTTCATTCATCACAGAGATCCATCTTACTCTACCGCTTGCTCTAACTTAACCTGTAATTG
TTTTTATCATGCAG

Intron4 GTATGAAATCTGCATAATGGGATACTAAAAACATATATTCTTAAATTTAAACTTAATTTTATTATTTTCTTTTATC
GATATATATGTCATTGTAATATCTCTGATGATATTTTCAG

Intron5 GTAAGGCAGCTTCATACTATTTCTCATCTCAGATTTTGTACTATGGCATATTAACATGAATTGAATTGATTTTGAGAT
GGAAGTATAAATCTGAATGGATGGAATATATATTTTCTGCAG

Intron6 GTACTGATAATTATTTTACATATTATGAACCTAGAAATGATTTTGTATGAAACACCGAAGATATTTTGGCAGTAT
TGTTGTGACTAAAAATGACCATTGATTGAAAAATACAG

Intron7 GTAGAATTCTTAATTTTGACATTTGTTTCATACTTAAAAAATTAACCTGTTCTCATTATTCTTTGTGCTATGATTACCG
AAAG

Intron8 GTAAGTGGGATACCTGATATGGTGAATACAACCTGTGCAGCTATAGCAATATGATATTAACCTCTTTGGTCTTGAACA
G

Intron9 GTAGATACCTAACCTCGCTGTACACCTTCTTGGTTATTCTATTATTAAGATAAACTTCATATTGACTCTTTCCCATGG
ATGTACAG

Intron10 GTACGCATTTTGACTACTCTGATCAGAAGTGTTTTAACTACATATTGCTCATTCTCATGTTGCCCTGTTCCATGATATG
CAG

Intron11 GTAATATTTCTATACAACCTAAAGTTTCTGGGTCACATCCGTCAAGTTATGTCGTAAAAATGCAATTACCTTTCCATGATC
TAATGGCATTTTCTGGTTTTTCAG

Intron12 GTAAGAAGCGGTTTTTACATAGCTCGATCTACATTCTGATGATGTCGCCCTCATTATCTGACTAGAATACTCTCTG
CAACAG

Intron13 GTAACCTTAATGATGTTCTCTCATTTGCAAAAAATCATGGTTGAATACTTGAATTGAAATCTAACTGTTCTCAAAATTTAT
CTGTCTTGGCTTTCCAG

Intron14 GTATGCAAAACACATCCAGTAATCTGTGTTTCTGAATACGCATTTCTGCTTTCTTGCAGTGCTAACATTATGTTTCTCC
AACCTTTGTTTCAG

Fig. 2. Nucleotide sequences of introns (this page) and upstream promoter-enhancer region (next page). Some structural features, such as A/G cluster (AG), CAAT-like sequence (CAT), GC box (GC), TATA box (TA), tandem repeat (TR), potential hairpin loop (HL), in the putative promoter-enhancer region have been marked.

```

      10      20      30      40      50      60      70      80      90      100
5' GCGAGCCCTG GCGAGGCGA GCGCGGCGG CCACTTGGG AGGAGTACGA GCGCGGCGG CCGGTGAGCG GCGCGGCGG ACAACAGCG AGTGTAGGAG
      TR1 GC
      110      120      130      140      150      160      170      180      190      200
GAGCGTCGGG AGTGGGCGAG CCTGGGACAA GCGGCGAGG TGAAGGCGG CAGAGAAAG AGGAGCAGCG GCAAGCGAG GCGGCGATGA CAACTAAGT
      GC
      210      220      230      240      250      260      270      280      290      300
GCGCGGCGG GCGGCGGAGG TGAAGGCGG CCGTTGAGG TAGAGAGAGA AGAGAGAAAG AATGAGAGG GGTCAAAAG ATAAGGAGA AAAAGAGAAA
      TR2 GC AG
      310      320      330      340      350      360      370      380      390      400
AAATCAAG GGTAGAACCG TACTTACAGA GCGTTTCTC TCTCAATTC TCCAGATAT AATAAAATA TCTGTGGGT GTCCAGAAC TAATTACTAT
      TR3
      410      420      430      440      450      460      470      480      490      500
TTTTTCCGG TAATTAGTAG CCGCGCTGAC GCGCGAGCTT AGGAGGACTA CCGTGTGTA GCAAGTTGG GAGGAUCCAG GCGCGCGCGG CAACCTGAGA
      TR1 GC
      510      520      530      540      550      560      570      580      590      600
GGACTAGGAG TGACGATCGT GCGTACTGC GCGGCGGGA GCGGCGAGCA GCGGCGAGCG CAGGAGGAGG TCGCGGCGCA GCGGCGGAGG TTCAAGCGCT
      GC TR2
      610      620      630      640      650      660      670      680      690      700
CGGTTGAGAA AGAATGAGAG AGGAGGAGAA GGATAAGCAA GAAAAAGAA AAAAAATCA AAGGAGTAA CCGTACTTAC ACTGCGGTTT CTCTCTACTG
      AG TR3
      710      720      730      740      750      760      770      780      790      800
TCCAGCACT AATTACTAGT TTITTACGCT GCGCTGAGC AGAAGCGCT TACGCTGTCT CTAGCTAAT TACACATTT TTTATATCG GTAGCAAAAT
      810      820      830      840      850      860      870      880      890      900
TTACCAAAAC AAAAGGATG ATCTTTCTTT AAAAAAAGG AAAAAAAGG AACAAGCAA AATCATACGA TCGAGCGCT CCGTTGCTAG TACATGAGCG
      910      920      930      940      950      960      970      980      990      1000
CAACAATCCA ACATATACAT GCAATAAAT CTATCCAAA ACCGCGACGA AATCAAAAGA AAAATTATCT GCTTGAGCTC ACCAACCAT TTTATATCGG
      CAT AG
      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
CAATGGCGGA TGGGGGTAA CCGACCGCG AGCGGAGCC ACCTCGGCC CCTTCCTCGT GTATAAATA GCGCGCGCGG CATGGCAATC CTCGCACAAA
      CAT HL CAT HL TATA HL CAT
      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
TCCATTCCAC CTCGGTTAAC CCGCTTCGTT TTATTTCGCT TTACTTCGTT TCTTGGGAG GCACTTCCTT CCTTCTCTTA GGAAGGTAAT AAAGAAGAGC 3'

```

Fig. 2. Continued.

Fig. 3. The organization of *RSs2* gene. The number indicates the length of DNA in bp.

References

- Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18: 5294-5299 (1979).
- Frischauf AM, Lehrach H, Poustka A, Murray N: Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* 170: 827-842 (1983).
- Gübler U, Hoffman BJ: A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263-269 (1983).
- Hanahan D: Techniques for transformation of *E. coli*. In: Glover D (ed) *DNA Cloning: A Practical Approach*, vol. 1, pp. 103-109. IRL Press, Oxford (1985).
- Henikoff S: Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28: 351-359 (1984).
- Hoheisel J, Poly FM: Simplified preparation of unidirectional deletion clones. *Nucl Acids Res* 14: 3605 (1986).
- Maniatis T, Fritsch EF, Sambrook J: *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
- Maraña C, García-Olmedo F, Carbonero P: Different expression of two types of sucrose synthase-encoding genes in wheat in response to anaerobiosis, cold shock and light. *Gene* 88: 167-172 (1990).
- Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467 (1977).
- Zack CD, Ferl RJ, Hannah LC: DNA sequence of a *shrunk* allele of maize: Evidence for visitation by insertional sequence. *Maydica* 31: 5-16 (1986).