

Update section

Sequence

Nucleotide sequence of a long cDNA from the rice *waxy* gene

Ron J. Okagaki

Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, USA; current address: Vegetable Crops Department, University of Florida, Gainesville, FL 32611, USA

Received 23 November 1991; accepted in revised form 31 January 1992

The *waxy* (*Wx*) gene in rice and other grasses encodes the major starch granule-bound ADP-glucose glycosyl transferase responsible for the presence of amylose in the endosperm and pollen. There has been a great deal of interest in this gene due to its importance in agriculture and genetics, and for these reasons the *Wx* genes from maize [2] and barley [4] have been cloned and sequenced. Recently, the sequence of a genomic clone containing the rice *Wx* gene has been reported and the sequence of the coding region deduced by comparison with the maize gene [6]. Here the sequence of a long rice cDNA is reported confirming the coding region predicted from the genomic *Wx* sequences.

A rice cDNA library was produced using poly(A)⁺ RNA from immature seeds. Procedures for poly(A)⁺ RNA isolation, northern blot analysis, and other techniques have been previously described [3]. Poly(A)⁺ RNA was isolated from the rice strain Nato, CI 8998, and cDNA was made using an Amersham kit. Inserts were ligated to λ gt10 arms using *Eco* RI linkers and packaged with the Gigapack packaging extract (Stratagene). This library was screened with a radiolabeled *Eco* RI-*Sal* I fragment containing the 5' region of the gene from the rice genomic clone [3]; the longest clone isolated was subcloned into pUC 119. The cDNA clone and portions of a

previously described genomic clone [3] were sequenced using Sequenase (United States Biochemical) according to supplied directions.

The 2.5 *Wx* cDNA clone isolated is longer than the 2.4 kb *Wx* transcript [3]. However Nato rice, unlike most rice strains examined, has a second *Wx* transcript detectable under stringent conditions. This second transcript is approximately

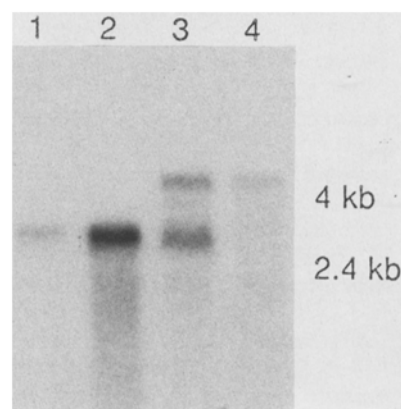


Fig. 3. Northern blot showing the *Wx* transcripts. One microgram of poly(A)⁺ RNA was fractionated, blotted, and hybridized as described previously [3]. Filters were probed with a 6 kb *Eco* RI fragment containing the rice *Wx* gene and washed at 65 °C in 0.1 × SSC, 0.1% SDS. Lane 1, maize; lane 2, Bluebonnet 50 rice; lane 3, Nato rice; lane 4, PI 291667 rice.

gaattcagtgtaggaatagattctcttcaaaacaatttaattcattcatctgatctgtcaaagctctgtgcacatctccgggtgcaacggccaggatatt
 100.
 tattgtgcagtaaaaaaatgcatatcccctagccaccaagaactgtccttaagtccttataagcacatatggcattgtaatatatatgtttgagtt
 200.
 ttagcgacaatttttttaaaacttttggctctttttatgaacgttttaagtttcactgtcttttttttgcgaattttaaatgtagcttcaaatcttaatt
 300.
 ccccaatccaaattgtaataaaacttcaattctcctaattaacatcttaattcattttattgaaaaccagttcaaattcttttaggtctccaaacctta
 400.
 aacaattcaattcagtgagagatcttccacagcaacagctagacaaccaccATGTCGGCTCTCACCAGTCCAGCTCGCCACCTCGGCCACCGCTTC
 500.
 M S A L T T S Q L A T S A T G F
 GGCATCGCCGACAGGTGCGCGCGCTCGCTGCTCCGCCACGGGTTCAGGGGCTCAAGCCCCGAGCCCGCGCGCGGCGAGCGAGCTCGCTCAGCG
 600.
 G I A D R S A P S S L L R H G F Q G L K P R S P A G G D A T S L S V
 TGACGACCAGCGCGCGCGAGCGCCCAAGCAGCAGCGGTGCGTGCAGCGTGGCAGCGGAGGTTCCCTCCGTCGTCGTGTSCGCCACCGCGCGCGCAT
 700.
 T T S A R A T P K Q Q R S V Q R G S R R F P S V V V Y A T G A G M
 GAACGTCGTGTTGTCGTCGGCGCGAGATGGCCCCCTGGAGCAAGACCGCGCGCTCGGTGACGTCCTCGGTGGCTCCCGCTGCCATGGCTGCGAATGGC
 800.
 N V V F V G A E M A P W S K T G G L G D V L G G L P P A M A A N G
 CACAGGGTCATGGTGTCTCTCTCCGTCAGCAGTACCAAGGACGCTTGGGATACAGCGTGTGGCTGAGATCAAGGTTGCAGACAGGTACGAGAGGG
 900.
 H R V M V I S P R Y D Q Y K D A W D T S V V A E I K V A D R Y E R V
 TGAGGTTTTCCATTGCTACAAGCGTGGAGTCGACCGTGTGTTTCATCGACCATCCGTCATTCTCGGAGAAGGTTGGGAAAGACCGGTGAGAATCTA
 1000.
 R F F H C Y K R G V D R V F I D H P S F L E K V W G K T G E K I Y
 CGGACCTGACATGGAGTTGATTACAAAGACAACAGATGCGTTTCAGCCTTCTTTGCCAGGCGAGCTCGAGGCTCCTAGGATCTTAACCTCAACAAC
 1100.
 G P D T G V D Y K D N Q M R F S L L C Q A A L E A P R I L N L N N
 AACCCTACTTCAAGGAACCTTATGGTGGAGTGTGTGTTGCTCTGCAACGACTGGCAGCTGGCCCACTGGCGAGCTACCTGAAGAACACTACAGC
 1200.
 N P Y F K G T Y G E D V V F V C N D W H T G P L A S Y L K N N Y Q P
 CCAATGGCATCTACAGGAATGCAAGGTTGCTTTCTGCTCCCAACATCTCTTACCAGGGCGGTTTCGCTTTTCAGGATACCTGAGCTGAACCTCTC
 1300.
 N G I Y R N A K V A F C I H N I S Y Q G R F A F E D Y P E L N L S
 CGAGAGGTTAGGTCATCTCTGATTTCATCGACGGGTATGACACGCGGTGGAGGGCAGGAAGATCAACTGGATGAAGGCGGAATCTGGAAGCCGAC
 1400.
 E R F R S S F D F I D G Y D T P V E G R K I N W M K A G I L E A D
 AGGGTGCTACCGGTGAGCCCGTACTACCGGAGGAGCTCATCTCCGGCATCGCCAGGGGATGCGAGCTCGACAACATCATGCGGCTCACCGGCATCACCG
 1500.
 R V L T V S P Y Y A E E L I S G I A R G C E L D N I M R L T G I T G
 GCATCGTCAACGGCATGGACGTCAGCGAGTGGGATCTAGCAAGGACAAGTACATCACCGCCAAGTACGACGCAACACCGCAATCGAGGCGAAGCGCT
 1600.
 I V N G M D V S E W D P S K D K Y I T A K Y D A T T A I E A K A L
 GAACAAGGAGCGGTTGACGGCGGAGGCGGGTCTTCCGGTTCGACAGGAAATCCCACTGATCGCGTTCATCGGCAGGCTGGAGGAAAGAAGGGCCCTGAC
 1700.
 N K E A L Q A E A G L P V D R K I P L I A F I G R L E E Q K G P D
 GTCATGGCCCGCCCATCCCGGAGCTCATCGAGGAGCTCCAGATCGTTCTTCTGGTACTGGAAAGAAGAAGTTCGAGAAGCTGCTCAAGACATGG
 1800.
 V M A A A I P E L M Q E D V Q I V L L G T G K K K F E K L L K S M E
 AGGAGAAGTATCCGGGCAAGGTGAGGCGGTGGTGAAGTTCAACGCGCGCTTGTCTCATCTCATATGGCCGAGCGGAGCTGCTCGCGTCCCGGCGCG
 1900.
 E K Y P G K V R A V V K F N A P L A H L I M A G A D V L A V P S R
 CTCGAGCCCTGTGGACTCATCCAGCTGCAGGGGATGAGATACGGAACGCCCTGTGCTTGGCGCTCCACCGGTGGGCTCGTGGACACGGTCATCGAAGGC
 2000.
 F E P C G L I Q L Q G M R Y G T P C A C A S T G G L V D T V I E G
 AAGACTGGTTTCCACATGGGCGCTCTCAGCGTCGACTGCAAGGTGGTGGAGCCAAGCGACGTGAAGAAGGTGGCGCCACCTGAAGCGCGCATCAAGG
 2100.
 K T G F H M G R L S V D C K V V E P S D V K K V A A T L K R A I K V
 TCGTGGCAGCGCGGCTACGAGGAGTGGTCAAGAACTGCATGAACAGGACCTCTCTGGAAGGGGCTGCGAAGAACTGGGAAGTGTGCTCTCTGGG
 2200.
 V G T P A Y E E M V R N C M N Q D L S W K G P A K N W E N V L L G
 CCTGGCGCTCGCGGCGAGCGCGCGGGGATCGAAGGCGAGGATCGCGCGCTCGCCAAGGAGAAGTGGCTGCTCTTgaagagcctgagatctacat
 2300.
 L G V A G S A P G I E G D E I A P L A K E N V A A P
 atggagtgattaataatagcagtatatggatgagagacgaatgaaccagtggtttgtttgtgtgtagtgaattttagctatagccaattatagggc
 2400.
 taataagtttgatgttgactcttctgggtgtgcttaagtatcttatcgaccctgaatttatgtgtgtggttattgccaataatattaagtaataaag
 2500.
 ggtttattatattaatataatattgttatattatactaaaaa
 2541.

Fig. 2. Sequence of the rice *Wx* cDNA. The coding region is presented in capital letters and begins at nucleotide 453. Two potential poly(A) addition sequences and five ATG trinucleotides upstream of the predicted translation start site are underlined, and the 5' *Eco* RI site is present in the genomic sequence. The amino acid sequence is presented in the single-letter code.

4 kb in length, and the cDNA clone was apparently derived from this transcript (Fig. 1). In the *Wx* mutant strain PI 291667 a 4 kb transcript is the only transcript detected. Southern blot analysis suggests that the presence of this transcript is not associated with an insertion or deletion as no unique differences distinguished Nato and PI 291667 from Bluebonnet 50 and other *Wx* alleles [3]. It is also unlikely that Nato rice is heterozygous for two *Wx* alleles because rice primarily self-pollinates. In wheat a second *Wx* transcript of approximately this size has also been found (C.C. Ainsworth, personal communication).

Sequencing the cDNA clone identified a long open reading frame with extensive amino acid identity to the maize waxy protein (Fig. 2). Overall the sequence identity is 82.7%, however this similarity is largely confined to the mature protein. The first 72 amino acids of the maize polypeptide encodes a transit peptide that directs the polypeptide's transport into the amyloplast [2], and previous work had suggested the existence of a transit peptide in rice [3]. Between transit peptides from nuclear-encoded chloroplast proteins amino acid sequence similarity is limited to short blocks [5], and this pattern holds for the predicted amyloplast transit peptides from the maize, barley, and rice *Wx* genes.

As mentioned above this cDNA clone is probably derived from the 4 kb transcript in Nato rice; the clone is longer than the normal 2.4 kb *Wx* transcript. If this is true, then the 4 kb transcript may represent an aberrant processing event where the first intron is retained in the processed transcript. The 3' end of the cDNA clone appears to terminate at the proper site. The cDNA and genomic sequences diverge at the beginning of a short poly(A) tail and within 60 nucleotides of the poly(A) tail are two consensus poly(A) addition sequences, AATAAT and AATAAA [1]. This suggests that the 2.4 kb and 4 kb transcripts differ at their 5' ends upstream of the translation start codon.

Translation of the maize and barley *Wx* genes begins in exon two [2, 4] making it likely that there is an untranslated first exon in the rice *Wx*

gene. In this cDNA clone the sequence upstream of the indicated start codon is identical to the rice genomic clone except for isolated single nucleotide differences; the genomic clone was isolated from the strain Labelle, CI 9708 [3]. It is unlikely that this sequence is part of the normal *Wx* transcript as there are five potential start codons in this sequence, and therefore part of this sequence may well be the first intron. This sequence is A/T-rich which is a common characteristic of plant introns [7], and the AG dinucleotides upstream of the indicated start codon could be a 3' splice site. However, the possibility that the 4 kb transcript is the result of an upstream transcription start site has not been excluded. If the 4 kb transcript reflects the lack of splicing of the first intron, then additional studies of this intron in Nato and PI 291667 rice strains could provide insight into sequences necessary for the proper splicing of plant introns.

Acknowledgements

I thank Dr Oliver E. Nelson Jr for allowing me to complete this work in his lab, Dr Susan R. Wessler in whose lab the cDNA clone was isolated, Dr Ko Shimamoto for sharing unpublished data, the late Dr George Baran for help with cDNA cloning, Dr Curt Hannah for comments, and the Rockefeller Foundation for supporting work done in Dr Wessler's lab. RJO was supported by a NIH training grant to the Laboratory of Genetics, University of Wisconsin. This is paper 3220 from the Laboratory of Genetics, University of Wisconsin, Madison.

References

1. Dean C, Tamaki S, Dunsmuir P, Favreau M, Katayama C, Dooner H, Bedbrook J: mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. *Nucl Acids Res* 14: 2229–2240 (1986).
2. Klosgen RB, Gierl A, Schwarz-Sommer Z, Saedler H: Molecular analysis of the *waxy* locus of *Zea mays*. *Mol Gen Genet* 203: 237–244 (1986).
3. Okagaki RJ, Wessler SR: Comparison of non-mutant and

- mutant *waxy* genes in rice and maize. *Genetics* 120: 1137–1143 (1988).
4. Rohde W, Becker D, Salamini F: Structural analysis of the *waxy* locus from *Hordeum vulgare*. *Nucl Acids Res* 16: 7185–7186 (1988).
 5. Schmidt GW, Mishkind ML: The transport of proteins into chloroplasts. *Annu Rev Biochem* 55: 879–912 (1986).
 6. Wang Z, Wu Z, Xing Y, Zheng F, Guo X, Zhang W, Hong M: Nucleotide sequence of the rice *waxy* gene. *Nucl Acids Res* 18: 5898 (1990).
 7. Wiebauer K, Herrero J-J, Filipowicz W: Nuclear pre-mRNA processing in plants: Distinct modes of 3'-splice-site selection in plants and animals. *Mol Cell Biol* 8(5): 2042–2051 (1988).