

Plant Gene Register

Genomic Nucleotide Sequence of a Wild-Type Shrunken-2 Allele of *Zea mays*¹

Janine R. Shaw and L. Curtis Hannah*

2211 Fifield Hall, University of Florida, Gainesville, Florida 32611

The maize (*Zea mays*) endosperm enzyme, ADP-glucose pyrophosphorylase, which produces ADP-glucose from ATP and glucose-1-phosphate, is an important enzyme in the synthesis of starch. ADP-glucose arising from the action of this enzyme is the major, if not sole, donor of glucose for starch biosynthesis. The enzyme is composed of two dissimilar subunits encoded by the two unlinked genes, *Shrunken-2* (*Sh2*) and *Brittle-2* (*Bt2*) (1, 2). The enzyme is allosterically activated by 3-phosphoglyceric acid and inhibited by phosphate (3). Although it remains an open question whether these allosteric properties are physiologically relevant or whether they simply reflect the evolutionary history of the two structural genes (4), there does exist, nevertheless, much interest in determining whether genetic modification of this enzyme could lead to increased rates of starch biosynthesis in the maize endosperm. Because endosperm starch content comprises approximately 70% of the dry weight of the seed, alterations in starch biosynthesis would clearly affect corn yield.

As a first step toward such experiments, we have isolated and sequenced genomic clones of *Sh2*. The structure of the gene (Table I, Fig. 1) is based on sequence analysis of three overlapping clones isolated from a Black Mexican Sweet genomic library. Exonic sequences were defined by comparison with the cDNA sequenced previously (2) and further sequencing of cDNAs subsequently isolated. These sequences

¹ This work was supported by the U.S. Department of Agriculture competitive grants 88–37234–3917 and 90–37262–5638. This is University of Florida (Agricultural Experiment Station) Journal series No. R-01954.

correct for the fact that the original cDNA contained some non-*Sh2* sequences at its extremities. The exonic genomic and cDNA sequences are nearly 100% identical. The few differences may be due to DNA polymorphisms because these clones were isolated from different corn lines.

Placement of the start of the first exon is based on primer extension experiments. In the absence of dideoxynucleotides, four major bands, differing by one nucleotide, were observed. In the presence of dideoxynucleotides, the three large bands were seen in all four sequencing tracks. In each case, the largest band occurred at the same distance from the primer. It is unknown whether this reflects heterogeneity within the normal population of *Sh2* transcripts or some laboratory artifact. Nevertheless, the start of transcription can be placed within three base pairs.

ACKNOWLEDGMENTS

We would like to thank Jung-Myung Bae for the 5' mRNA sequence. We thank the ICBR DNA Synthesis Core, University of Florida, for the oligonucleotides used for sequencing.

LITERATURE CITED

1. Bae JM, Giroux M, Hannah L (1990) Cloning and characterization of the *Brittle-2* gene of maize. *Maydica* 35: 317–322
2. Bhave MR, Lawrence S, Barton C, Hannah LC (1990) Identification and molecular characterization of shrunken-2 cDNA clones of maize. *Plant Cell* 2: 581–586
3. Dickinson DB, Preiss J (1969) Presence of ADP-glucose pyrophosphorylase in *Shrunken-2* and *Brittle-2* mutants of maize endosperm. *Plant Physiol* 44: 1058–1062
4. Hannah LC, Nelson OE (1975) Characterization of adenosine diphosphate glucose pyrophosphorylase from developing maize seeds. *Plant Physiol* 55: 297–302

Table I. Characteristics of a Shrunken-2 Wild-Type Gene from *Zea mays* L.**Organism:***Zea mays* L. var Black Mexican Sweet**Location on Chromosome:**

3L

Gene Product; Pathway:

A subunit of ADP-glucose pyrophosphorylase; starch biosynthesis

Gene Designation:*Sh2***Techniques:**

λ EMBL3 *Sau*3A partial genomic library screened with *Sh2* cDNA resulting in three overlapping clones; restriction fragment subcloning into pUC19 and pSPORT; double-stranded plasmid dideoxynucleotide sequencing of the overlapping clones (both strands) using various subclones and synthetic oligonucleotide primers; primer extension with and without dideoxynucleotides to determine transcription start.

Methods of Identification:Exon sequences nearly 100% identical to *Sh2* cDNA**Expression Characteristics:***Sh2* is expressed only in the endosperm of the maize kernel.**Gene Structure:**The coding region of 1913 bp^a is interrupted by 15 introns ranging in size from 68 bp to 1821 bp

Promoter Region:	putative TATA box	TATATAAA at -33 bp to -26 bp
	putative RY repeats	CATGCATG at -926 bp to -919 bp
		CATGCATA at -510 bp to -503 bp
	similarity to animal enhancer	GTGGAAAC at -911 bp to -904 bp
Intron Region:	similarity to animal enhancer	GTGGATAG in 5' part of largest intron

Codon Usage:

43.2% G + C in coding region

No obvious bias in codon usage

Protein Characteristics:

The 1551 bp open reading frame gives rise to a 516 amino acid peptide; *M*, 57,178.98; ADP-glucose pyrophosphorylase is allosterically regulated by phosphoglyceric acid (activation) and phosphate (inhibition).

Antibodies:

Antibodies made against fusion protein expressed in *Escherichia coli* recognizes an endosperm-specific protein. Size and amount of this protein is altered in *sh2* mutants (M. Giroux, L.C. Hannah, unpublished).

Subcellular Localization:

Reported to be in the amyloplasts.

GenBank Accession No.:

M81603

^a Base pairs.

-1020 TGATGCTTTTCTGGCCAGGAGAGCTATGAGACGTATGCTCCTCAAAGCCACTTTGCAT
 -960 TGTGTAAACCAATATCGATCTTTGTTACTTCACTCATGCATGAACATTTGGGAAACATC
 -900 TAGCTTACAAGCATTAGTGACAGCTCAGAAAAAGTTATCTCTGAAAGGTTTCATGTGTA
 -840 CCGTGGGAAATGAGAAATGTTGCCAACTCAAAACCCCTCAATATGTTGTTGGCAGGCAA
 -780 CTCTTCTGGAAGAAAGGTGCTAAAACTATGAACGGTTACAGAAAGGTATAAACCCAGC
 -720 CTGTGCAATTTGGAAGTATCATCTATAGATGCTGTGTGAGGGGAAAGCCGATCGCCAAAC
 -660 TTATTTACTCAGAAACAGCTTCAACACACAGTGTGTCTGCTTTATGATGGCATCTCCACCG
 -600 AGGCCCCACCATCACTATTCACTATCTCTGTCGCTGTTTATTTTCTGTCCTTCT
 -540 GATCATAAAAAATCATTAAAGGTTTGCACCAATGCATAGGCATATCAATATGCTCATTTA
 -480 TTAATTTGCTAGCAGATCTCTCTCACTCTTACTTTATTTATTTGTTGAAAAATATG
 -420 CCTGCACTAGGAGCTCGTATACAGTACCAATGCATCTTCAATTAATGTAATTTTCAGA
 -360 AAGGAATAGGAACTATGAGAGTATTTTCAAATTAATAGCGGCTTCTATTATGTTG
 -300 ATGACAAAGGCCAAGGCCAAATCGGAACACTAATGATGGTTGGTTGCATGAGCTGTGCG
 -240 ATTACTTGAAGAAATGTAACCTTTGTTTCTGTCGCTGGGCATAAAACAAACAGCTTCT
 -180 AGCCTCTTTTACGGTACTTGCATTTGCAAGAAATGTAACCTTTTCTTTTCTGTATGT
 -120 GGCATAATGTCACCAAGCTCCAGGCTTTTTCATGGTTGATGCTTTTACACAGTTCAT
 -60 CTCCACAGTATGCCCTCATACTCTATATAAACACATCAACAGCATCGCAATTAGCC
 1 ACAAGTCACTTCGGGAGGCAAGTGTGATTTGCGACTTGCAGCCACCTTTTGGTCTG
 61 TTgtaagtatacttccctcaactcttctctgttagtttaattgtaattgggaagta
 121 ttatgggaaagagatgagatgctatctctatctgactctgcaaatgcaatctgagcttat
 181 atggctgtctcctataaattggaattgctccatctctgcccacaataattgcaaggtat
 241 atgcttagtccatcaaaagttctgtttttctctctctctctctctctctctctctctctct
 301 tttgtccatgaggaagaaactctgtttgtttgtctctgaggtgcaatctctcat
 361 atgtccagttttatggaagtaataaactctcaagtttggtcataaagatgcaatataaagtt
 421 caaacatatactcaagttcaactctgtaaatgctcccttttctctctctctctctctctct
 481 ctcaattattgagttgagctgtagtctgtagtggaggagatattgcaatctgcaactctg
 M Q F A L
 541 CATTGGACACGAACCTCAGTCTCACCAGATAAGATCTTGTGAGGCTGATGGGATGGACA
 A L D T N S G P H Q I R S C E G D G I D
 601 GGTGGAATAAATTAAGTATTTGGGGCAGAAAGCAGGAGAAAGCTTTGAGAAATAGTGTCT
 R L E K L S I G G R K Q E K A L R L N R C
 661 TTGTTGGTAGAGTTGCTGCAACTACAAATGATTTACTTACCTGACGCTGCTTCCGTA
 F G G R V A A T T Q C I L T S D A C P E
 721 CTCTTgtaagtataccactcaattattactctacatggtggttactcttactgcttctgct
 T L
 781 tttcaaggaaatttactgtatttttgtgtttgtggaggtctctactctctgctgtggac
 841 tggttatgtaagattttgtcaaatagggtcactctataaattggttgaactctgggaac
 901 tgtgttttcaactgctgctcaggaagaaaggttaattattggttactgcatgaataactatag
 961 aaatagaccttagagttgctgctagatatacaaaactgctgcaagatctctataata
 1021 gttctctcgaacctgcaatatacaataaactgcaactcctagttgctgctcaaaaaaa
 1081 atgcaactcttagaagctcaccaggttaactcttctgtaattggtatttaattgagctat
 1141 atgcaactctgtaactcttagagtaaatgtaactcaactctgagccgcttatattgca
 1201 gCACTTCAAACACAGTCTCTGAGAAAAATATGCTGATGCAACCCGATATCTGCTAT
 H S Q T Q S S R K N Y A D A N R V S A I
 1261 CATTTTGGGCGGACCTGATCTCAGCTCTTCTCCTGACAAAGCACAGAGGACTACGCC
 I L G G G T G G C T Q L F P L T S T R A T P
 1321 TGCTgtaaggataaacactgaaactgcaactggttactctattatagatattatacaga
 A
 1381 ctgacttttogaattattcttagttttctacaattatttagtgattctctctcattttca
 1441 agatacacaattgatacctaatacgaagtggtatgtaagacagtgagttcaaaagattata
 1501 ttttgggagactccagcaaaattttctagaagtttttttggctcagatgctcaataa
 1561 gtcgcgcttctcactctttttaaatttttaatttgggtgcaactattagTACCTGTTGGA
 V P V G
 1621 GGATGTTACAGCCTTATGATATCCCTATGAGTAACTGCTTCAACAGTGGTATAAATAAG
 G C Y R L I D I P M S N C F N S G I N K
 1681 ATATTTGTGATGAGTCACTCAATTTCTACTTCCGCTTAAACCCATATCTCATCTGATAC
 I F V M S Q F N S T S L N R H I H R T I
 1741 CTTGAAGGCGGATCAACTTTGCTGATGGATCTGACAGTgatttaccctcatctgtgtg
 L E G G I N F A D G S V Q
 1801 atgtgtaactctgtaattaggagtagatttgtgtggagagaataataaacagatgcccag
 1861 atctctttctaaaagctctagatccaaagcattgtggttcaaaacactatggaactctc
 1921 catttattgcaactcttgccttaattgctcaatggaatggggcaaaattattgattctca
 1981 agtgtttaaataaaaactaattgttctcactgagGATTAGCGGCTACACAAATGCCTG
 V L A A T Q M P
 2041 AAGAGCCAGCTGGATGTTCCAGGATCAGCAGACTCTATCAGAAAAATTTATCTGGGTAC
 E E P A G W F Q G T A D S I R K F I W V
 2101 TCGAGgttagttgatattttctgctttatgaatgctcactcactctctgtagcatttt
 L E
 2161 tcttttgaattttgagttctcctgtatttctttagGATTATTACAGTCAAAATCCATT
 D Y S H K S I
 2221 GACAACATTGTAATCTTGTAGTGGCGATCAGCTTTATCGGATGAATTACATGAACTTTG
 D N I V I L S G D Q L Y R M N Y M E L V
 2281 CAGgttagttctcctgctcctcctcaatgctcactgctgctgattttgatttaacca
 Q
 2341 actacttttggcatgattattttccagAAACATGTCGAGGACGATGCTGATATCACTATA
 K H V E D D A D I T I
 2401 TCATGTGCTCTGTTGATGAGAGTaatcagttgtttatatacctcaataatgaaatgt
 S C A P V D E S
 2461 catcttggatccaacagagatgcataggttcaatctgcttctcttttttcccttc
 2521 ggaagCCGAGCTTCAAAAATGGGCTAGTGAAGATTGATCATACTGGACGTGACTTCAA
 R A S K N G L V K I D H T G R V L Q
 2581 TTCTTTGAAAAACCAAGGGTCTGATTTGAATTTCTATGgttagaattcctgtgtaatt
 F F E K P K G A D L N S M
 2641 ccaactctttgttttct
 2701 ataactgtactgtacttattcagaatgatttctatttctctctctctctctctctctctctct
 2761 aagctaagtagcactgattgctgagAGTGTGAGCAACCACTCTGAGCTATGCTATAGAT
 R V E T N F L S Y A I D
 2821 GATGCACAGAAATCCATACCTTGCATCAATGGGCAATTTATGCTTCAAGAAAGATGCA
 D A Q K Y P Y L A S M G I Y V F K K D A
 2881 CTTTATAGACCTTCTCAgtaacttctctgtagtacttctctctctctctctctctctctct
 L L D L L K
 2941 acccttaacagtgtaacttcttagTCAAAATATACTCAATTACATGACTTTGGATCTG
 S K Y T Q L H D F G S
 3001 AAATCTCCCAAGAGCTGACTAGATAGTGTGCAAGTgagttctctctctctctctctctct
 E I L P R A V L D H S V Q
 3061 atgttctgttaactgtaactgtaactgtaactgtaactgtaactgtaactgtaactgtaactgta
 3121 ctaccaatgcaagatttattgattttctctctctctctctctctctctctctctctctctctct
 3181 ttactgttagcttagGCACTGATTTTACCGGCTATTTGGGAGGATTTGGAACATCAA
 A C I F T G Y W E D V G T I K
 3241 ATCATTCTTTGATGCAAACTGGCCCTCACTGAGGACTctctctctctctctctctctctct
 S F F D A N L A L T E Q
 3301 tgcataatattacctggaactcaactgcaactgattgtagacctcttagttccatcc
 3361 gttttctcaatagcttct
 3421 cctaattattgttct
 P S K F D F Y D P K T P F
 3481 CTTCACTGCACCCGATGCTTCCGACCAATTTGGACAAAGTCAAGGactatgattctct
 F T A P R C L P P T Q L D K C K
 3541 actgagcacaattgttactgagcaagattttgtgctactgactgttctctctctctctctctct
 3601 TGAAATATGATTTATCTCAGATGCTTCTGACTAGAGAAATGCAACTCAGGAACTTCTG
 M K Y A F I S D G C L L R E C N I E H S
 3661 TGATGGAGTCTGCTCAGCTCAGCTCTGGAETGAACTCAAGgttacatactctgccc
 V I G V C S R V S S G C E L K
 3721 tgtatctactctgtagtataccatttcaacacacacacacacacacacacacacacacacac
 3781 agcaacaagacctttttagtcccaagcaatttaggttagctgaggtgaaactcaaca
 3841 acaaaatgcaagctctcactcaggttagtttttccatgcaactctattttagaactaa
 3901 tttttgggtatacactcactcactcactcactcactcactcactcactcactcactcactcact
 3961 cccacttactctcct
 4021 ggcactgttcaaacactcactcactcactcactcactcactcactcactcactcactcactcact
 4081 ctaacactcactcactcactcactcactcactcactcactcactcactcactcactcactcact
 4141 atgcaacatacagcattttagcaacttattgtagttagcactcactcactcactcactcactcact
 4201 acatattgcaacatacaatgtagcttagcactcactcactcactcactcactcactcactcact
 4261 ctatgttagtctcctctgcaacttagaacaactatgctgtagtctcactcactcactcactcact
 4321 tgctttgattctctgctcaactctcactcactcactcactcactcactcactcactcactcact
 4381 taaatattggaataacattcttctgggcaacttagcactcactcactcactcactcactcactcact
 4441 gctccttcttctgtgttagtagtaccggaagctcactcactcactcactcactcactcactcact
 4501 ctaagtcccggttctgactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct
 4561 tgcgtgtgctccgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct
 4621 ttgcaagatgagcagttgtagcctgctgtagttagttagttagttagttagttagttagttagt
 4681 cggccgggacattgttctgct
 4741 gctgagtttttagttctaccagcttcaaacctttggacttagctcctcctcctcctcctcctcct
 4801 tcaacactctgatttctatttctctcactcactcactcactcactcactcactcactcactcact
 4861 tgtaaaaagcacaacaaatgtaactcctctgtagtctcctgtaattatccatcaca
 4921 agaaaaaagtaaggtcaaaagttagctttgataatgctcactcactcactcactcactcactcact
 4981 atctgatactctcct
 5041 tgagttcaacgtaatactctgtagatttctgcaaacctcctcactcactcactcactcactcact
 5101 cactctcactgatttaacctttggcagtaaaacacacagaaagttgttctctctctctctct
 5161 taagatccacacaaaagagttgtagtcaaaagactgtagtcaactgtagtcaactgtagtca
 5221 aatcctttgtgttttcttaaacctgaatattcctcactcactcactcactcactcactcactcact
 5281 tcaatctgttttagcaagtgaaacttctgttctgttctgttctgttctgttctgttctgttctgt
 5341 cccccccggagggcagatatacagcagatgaatacaagaatattgaaacccagactgata
 5401 gttgtttgtagctgtgaaactggtgcaactcatttctgttattgctgcttctgataaac
 5461 gacagGACTCCGTGATGATGGAGGGGACCTATGAACTGAAGAAAGGCTTCAAAGC
 D S V M M G A D T Y E E A S K
 5581 TACTGTAGCTGGGAAGGTCCTGATTTGAAATGGAAGAAACAAAGATTAAGgttagatL
 L L L A G K V P V G I G R N T K I R
 5641 ggtatggaaccccggttagttcccaaaatcactcactcactcactcactcactcactcactcact
 5701 ctgattattttcagGAACCTGATCATTGACATGAATGAGTGGGAAAGGCTGGTG
 N C I I D M N A R I G K N V V
 5761 ATCACAACAGTAAGgtgagcagcactcactcactcactcactcactcactcactcactcactcact
 I T N S K
 5821 tctcctaattcggtaattctcactcactcactcactcactcactcactcactcactcactcactcact
 5881 ctctgtgacagGGCTCAAGAGGCTGATCACCAGGAAAGGCTACTCACTAAGGCTGTG
 G I Q E A D H P E E G Y Y I R S
 5941 GAATCTGGTGATCTGGAAGAATGCAACCATCAACGATGGGCTCTGATATAGATCGGCT
 G I V V I L K N A T I N D G S V I -
 6001 GCGTGTGCTCTACAAAACAAGAACCTCAATGGTATGTCATCGATGGATCGTGAACCT
 6061 TGATGTTGTAAGAGCCGCTTGACAGAAAGTGGAGCTTCCGGCAGAGCCTGATGCTGGC
 6121 ATGCTGCTCTGACCAATTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT
 6181 GCAGCAAACTTTTATGAACTTTGATTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
 6241 GTCATCTATATATTAATAATTTTACGCTGTTTTCTAAATTCGGGCTGCTTTGGGAT

Figure 1. Nucleotide sequence and deduced amino acid sequence of a *Zea mays* wild-type *Sh2* allele. Base number 1 is the transcription start site. Introns are indicated by lowercase letters. Arrow indicates 3' end of cDNA. Putative TATA, RY dyad, and enhancer sequences are underlined.