

The complete nucleotide sequence of soybean glycinin A₂B_{1a} gene spanning to another glycinin gene A_{1a}B_{1b}

Yoshiaki Kitamura, Masaomi Arahira, Yoshifumi Itoh and Chikafusa Fukazawa

Genetic Engineering Laboratory, National Food Research Institute, Ministry of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki 305, Japan

Submitted June 19, 1990

EMBL accession no. X53404

We here present the complete nucleotide sequence encoding a glycinin A₂B_{1a} subunit precursor (obtained from soybean variety Bonminor), of which 5'-flanking region reaches to 3'-end of another glycinin subunit gene A_{1a}B_{1b} which locates 2806 bp upstream of the cap site (defined as +1) of this gene. The transcribed regions are underlined and the translated regions are

shown in capital letters. The boundary region between two glycinin genes contains a unique motif of 'ATAT repeat' (-1920 to -1969; shown with the bold letters) which is composed of 25 repeats of the 'TA' pair. The functions of the 5'-flanking region for the expression of a glycinin A₂B_{1a} subunit gene will be presented elsewhere.

```

..... ..AAGTTCCT GGTTCACCT CAGGAGTCTC AGAAGAGAGC TGTGGCTtag agcccttttt -3001
gtagtgcta ccccaatttt gctcttttgg caatagtgct agcaaccaat aaataataat aataataatg aataagaaaa caaaggcttt agcttgccctt ttgttcactg taaaataata -2881
atgtaagtac tctctataat gagtcacgaa acttttgcgg gaataaaagg aqaataatcca atgagttttc tgtcaaatct tctttgtct cctctctct ccttttttt ttctttctt -2761
ctgagctct tgcaaaaaca aaggcaaca ataacgattg gtccaatgat agttagcttg atcgatgata tctttaggaa gtgttggac gacaggacat agatgtagaa gactaaaat -2641
gaaagtattg cagaccaaat agttgaagat taactttaag aatgaagacy tctttatcagg tctttcatga cttggagctc aaccacaact gaaaagtctg agagtattg gaccattgtg -2521
ctttgtgtct tcaaacataa aacatcgctc caaatttaac atggggacta aaaaatgtgt ttttctggga ttttaatttc caacagatgc aaggatgtgt ttgcatatga tgcctttagt -2401
tccattgtcc acactaaaat gatattggct tcaagaataa ttaatttcat ttcatgact tccaattcat aaaccttaaa cgaatataa ttaaaactat acctcaaat gataaattg -2281
aaaaaaaaat accccaatc ggtaatttga ctcacaagtt agttagtga tattttgaag cttgaaattc agacatggac atcacacaca tatgagcaca gacactctgc catagctaat -2161
gtgtaaaaaca tagaatgaca ggacatcaca tatattttta cacacacaaa aaaaagaact taataaaaaa atatgtgtag cttaacaaat atataaattg atggtaataa atttactttt -2041
taaaattcat ctatgttttt ttatatgata acaaacataa aaaaagtgag aaacattaag tattttaata ttatatatat atatatatat atatatatat atatatatat -1921
aatatattat attttaagtt ggcctacaaa atgatgataa ctttttttaa taatattagt gtgcgtctct ttaagttaa ggtgctaatt taagtgtta aatatataa agatgataac -1801
ttaactattc tttgaataaa ttgaattatg ttgtgtgctc actacttatt tcttgcaaaa tgtgcaatca ttcccgtata ttgggtgctc atggccctaa attttagctt cgagctctca -1681
ttctcattct tgtccatttc ttatttctcg tgtgtgagga gctacaagtt cttaatgctt tgatagttct taatcattgc atagaactac aattaacta tccagattgt tgcattaatt -1561
gactcgaata accaagaaac ttttggcagc catggacagt gggcacaana tgtgcaatct tccctgaaac aataagactg aggtttatgt gaaagcaaaa gaaaaagaa gaaagaaat -1441
ttaacagctc aacatttggg caacacaact aatttgacag aaaaatttaa tttctgtatt ttagaanaaa gaggaattaa cgggtctcaat ttttaaatct taaaactaaa aaaccaaat -1321
tatggatttg aataaatagg aggatacaaaa tacatttaatt tttttgtttt aattgattct gacagctttc aaactgtcac aaatgtaac ttattttta atgaaataa -1201
ggtagtagct agggctgagc ataaaccgca gagaagctcg aaccaaaccc acactaaga gaaataaaa aagacacaaat cggggcggtt caaatgtttt attcggacta tacgggtta -1081
aaaattgctt actctgggtt taataactcg taataaactt taataaactt cctccggaga agaagaagt gtttgggccc gcaaaaagaa ccaagccaca acgagattg gaacacacc -961
aagaatccgt tgcattcact aagcccgaac ctgacaaaat cgagaggcta atgagttaga gaaactccag atcagaaata ccaacagttt ttaactctg actttgtaat -841
ttgtaattg atgatatgt cactactaac ctacctatga taagagctca aataaattt ttaagataa aagaactcaa tacaagaaa aaaaaattaa aatagaact caaaaaaaa -721
ttaatcatat ttattataga ctaaaacat ataatgttt ttgtttaact aataataattg ttttaaacct agcattttc ttcttatttt ttcttatttt ttcttatttt tgacaactca -601
aacgaggata atcaaaaaa caaatgtcc ctaatatata atttctctat ccaagaat attatatata cagtacaat tattctccac taatctttat gatagttaa ttaaagtatg -481
attaattatt tccbaattaa tttaaaaaa attgagtaac tagtatcaca taagaactct tcaaaaacct gaatagttag aatttcataa cactataaa ttctctatc ctagagtga -361
ggctattttt ggacgcccct ttctctcact ttgaggtccc catcattgtc acacggtga ttcatatgat ttcatatgtc ataaaccgca gaacatgaaa tgaaccatg gtccccctt -241
ccaccaccgt tttctggcaa ttgatcgaa tacaacaca ctgtgtattt gtcacatgat gttgatgtc aactgtcgaa gccacctac accatgaac ttaagtgtt gtaacacaca -121
aggttccat agccatgat actgagaagt gtctcaagct cagcaccaca ctctgtgac gtgcccctca cccacccttc tctcttctct ataaataacc agcagcagg ttctccgct -1
cacaacaca acatttctct cctgtctttt gaacactcat caccATGGCC AAGCTTGTC TTCCCTTTG TTCCCTTCTT TTAGTGGCTT GCTTCGCCTT GAGAGAGGAG GCACAGCAAA +120
ATGAGTGCCA GATCCAAAAG CTGAATGCC TCAAAACCGA TAACCGTATA GAGTCGGAGG GTGGGTTTCAT TGAGACTCGG AACCTAACA ACAAGCCATT CCAGTGTGCC GGTGTGCC +240
TCTCTCGTG CACCCTTAAC CGCAATGCC TTCGTAGACC TTCCTACACC AACGGTCCC AGGAAATCTA CATACAACA Ggtccataat ctgttttaatt ttttttttta tcaagatgta +360
atggttagtt tttgttagta aaattgattg aaactgtgac tttgtttttt cttcttaacc gtccaaccta tcttatata tcaaatat aataacttgt tcaatttctt cagttttgtg +480
aaaccatgca ttttgaagg aaatgggta tgtaaatgtt ttttttttct ttgtgtgtgt gtgtaataa tatgaatag TAATGGTATT TTTGGCATGA TATTCCCGG TTGTCTAGC +600
ACTTATCAAG AGCCGCAAGA ATCTCAGCAA CGAGGACGAA GCCAGAGGCC CCAAGACCGT CACCAAAAGG TACATCGCTT CAGAGAGGGT GATTTGATCG CAGTGCCTAC TGGTGTGCA +720
TGGTGGATG ACAACAATGA AGACACTCCT GTTGTGCCG TTCTATTAT TGACACCAAC AGCTTGGAGA ACCAGCTCGA CCAGATGCCCT AGGgtgagac cacatagcaa tattagat +840
tataattctt taaaggtta aatatcattt tagttcgtg agttgactt tctaattag tacctataga ttaaaatag ccaattgaa ccttatagtt gtgttttttt atccaattt +960
gttctgtct tgaatacat ggacaatatt gtagctgata aaaaaaggaa actggactac atgttaactg taagattaga attcttaagt tctaacta gctggttacg gattgacaac +1080
tattgttttt gacaattctt gqacAGATT CTATCTGCT GGAACCAAG AGCAAGAGTT TCTAAAATAT CAGCAGCAGC AGCAAGGAGG TTCCAAAAGC CAGAAGGAAA AGCAACAAGA +1200
AGAAGAAAAC GAAGGAAGCA ACATATTGAG TGGCTTCGCC CTGAATTCT TGAAGAAGC GTTCCGGCTG AACATGCAGA TAGTGAGAAA CCTACAAGGT GAGAACGAAG AGGAGGATAG +1320
TGGAGCCATT GTGACAGTGA AAGGAGGCT AAGAGTCACA GCTCCAGCCA TGAGGAAGCC ACAGCAAGAA GAAGATGATG ATGATGAGGA AGAGCAGCCA CAGTGGCTGG AGACAGACAA +1440
AGGTTCGCCA CGCCAAAAGCA AAAGGAGCAG AAATGGCATT GATGAGACCA TTTGCACAAI GAGACTTCGC CAAAACATTG ATCAGAAATC ATCACCAACC CTCAAGCTGG +1560
TAGCATCAA ACCGCCACCA GCCTTGACTT CCCAGCCCTC TGGCTTCTCA AACTCAGTGC CCAGTATGGA TCACCTCCGA AGTctactac atcatatctg ttaaaatata gctcaacaa +1680
atgttaaaaa atgtaaatag ttatatatag taattaaaat gtgtgacgac aaagtattct tagcacttgc atatgcatta ttcatgtgct tttttttat atatagaat gcccgaagt +1800
atgaagtatg cacagtttct tgtgtcattt ttactaatct aattaactc tcatgagat gtaattatgc atgtgttatt aatttttgtt acatgtgata caattagtag ataatttgtt +1920
cttttagata tatgactcat gagttgaaa ttttttcaa ccttgaaaaa ttgtttttg acttttaacc aaaaaact ttgattaaa aaaaagaaaa atattttgt gccaanaata +2040
ttctgattca caaaaacata tttttctatg agactatata ttgacatgag ataaaatacc tgaataatct ttcatcaga ttaggagaaa aaggtataag atagtgagat aattaatla +2160
tatcatgatg aaaaagataa aggaataata taatatata aaaaattgg atgaataaa ttgttgtatt aactatagct tgcatactg tgcagata ttgatgatg ttgacagATG CTATGTTCTG +2280
GCCACACTAC ACCCTGAAGC CGAACAGCAT AATATACGA TTGAATGGCC GGGCATTGGT ACAAGTGGT AATGCAATG GTGAGAGGT GTTGTATGGA GAGCTGCAAG AGGGAGGGT +2400
GCTGATCGTT CCACAAAAC TTGCGGTGGC TGCAAAATCC CAGAGCGATA ACTTTGAGTA TGTGTCATT AAGACCAATG ATAGACCTC GATCCGAAAC CTTGAGGGG CAAACTCAT +2520
GTTGAAGCA TTGACAGAGG AAGTATCCA GCACACTTTT AACCTAAAAG CAGCAGCAGC CAGCAGGTTG AAGAACACA ACCCTTTCAG CTTCTTGTG CCACCTCAGG CTCACTCAGG +2640
GAGAGCTGTG GCTtagagag cccttttga tgtgtacca cacttttgc ttttggctc tagctagtgc tgtgtacgt aataagaat gataaaaaa acaaaagct tagatgccc +2760
ttgttaagt cagaataacg ggtatgtaa ctatgtacc ttgtgtaag agtactcaa gtacctatg gggaaaaqa taaataaaq aqaatattc aataaattt ctgtccaatc +2880
tcttttctt tcttttatt ctttctatc ctcttgac acaaaaataa ggaatattg ttgatac
    
```