

Nucleotide sequences of two genes encoding the light harvesting chlorophyll a/b binding protein of rice

Sheng Luan and Lawrence Bogorad

The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
 Submitted February 3, 1989
 EMBL accession nos X13908, X13909

Southern blots suggest that the rice genome contains about four genes coding for light harvesting chlorophyll a/b binding protein (LHCP). From a genomic library kindly provided by Dr. Susan Wessler at the University of Georgia, we have cloned and sequenced two of the four genes. Here we report the nucleotide sequences of the coding regions and some of 3' and 5' flanking regions of these genes, *cab1R* and 2R. The coding regions for mature proteins are very similar to one another and also to those of LHCP genes from wheat(1) and pea(2). The presumed transit sequences and flanking sequences of the two genes differ substantially.

```

      1      10      20      30      40      50      60      70
1R GCACATCAGCTCCCAATCTCCCGCCCCAGGAGGCAATCCCGCCTCGCTTCCCGCGCTATTTAAACTCCCGCGC
2R cggCccatcCagCCAATCcgacCGacaaAaatgtTCCCaagTgGCCaCGaGatCTATTaATaCgCCaagGt
                                     [cgctctccacg]
1R CAACTCCCAACTCACACTCGCTCGCTCATGGCCATCTCTCTGAGCTCTCACAGCTCACTGCATCA*****
2R gttGgtcgcCaTgACACagcagcacacaCaCGccagCagcagcagCagcgAgCagctgagcItgaagctgctga

1R *****
2R gcgaggtagac

1R ATG GCC GCG GCC ACC ATG GCG CTC TCC TCC CCG GTG ATG GCC CGC GCG GCG CCG TCG
   M  A  A  A  T  M  A  L  S  S  P  V/a M/l  A  R/g A/k  A  P/a S/a
2R ATG GCC GCc GCC ACC ATG GCc CTC TCg TCC CCG GcG cTG GCC gGC aaG GCc gcc gCG

1R ACC TCC TCC GCG CTC TTC GGC GAG GCG CCG ATC ACC ATG CGC AAG ACC GCC GCG AAG
   T  S/k S  A  L  F  G  E  A/g R  I  T  M  R  K  T/s A  A  K
2R *** aag *** *** GtG TTC GGC GAG GgG CgC ATC ACC ATG CGC AAG tCg GCg GCG AAG

1R CCC AAG CCG GCG GCG TCG TCG GGG AGC CCG TGG TAC GGC GCC GAC CGC GTC CTC TAC
   P  K  P  A  A  S  S  G  S  P  W  Y  G  A  D  R  V  L  Y
2R CCC AAG CCc GCc GCG TCG TCG GGG AGC CCG TGG TAC GGC GCC GAC CGC GTg CTC TAC

1R CTC GGC CCG CTC TCG GGC CGC GAG CCG CCG AGC TAC CTC ACC GGC GAG TTC CCG GGC
   L  G  P  L  S  G  R  E  P  P  S  Y  L  T  G  E  F  P  G
2R CTC GGC CCG CTC TCc GGC CGC GAG CCG CCG AGC TAC CTg ACC GGC GAG TTC CCc GGC

1R GAC TAC GGG TGG GAC ACC GCG GGG CTC TCC GCC GAC CCG GAG ACG TTC GCC AAG AAC
   D  Y  G  G  W  D  T  A  G  L  S  A  D  P  E  T  F  A  K  N
2R GAC TAC GGG TGG GAC ACC GCG GGG CTC TCC GCC GAC CCG GAG ACG TTC GCC AAG AAC

1R CCG GAG CTG GAG GTG ATC CAC TCC CCG TGG GCG ATG CTG GGC GCG CTC GGC TGC GTC
   R  E  L  E  V  I  H  S  R  W  A  M  L  G  A/g L  G  C  V
2R CCG GAG CTG GAG GTG ATC CAC TCC AGG TGG GCG ATG CTc GGC GgG CTg GGC TGC GTg

1R TTC CCG GAG CTC CTC GCC CGG AAC GGC GTC AAG TTC GGC GAG GCC GTG TGG TTC AAG
   F  P  E  L  L  A  R  N  G  V  K  F  G  E  A  V  W  F  K
2R TTC CCG GAG CTC CTC GCC CGc AAC GGC GTC AAc TTC GGg GAG GCg GTG TGG TTC AAG
    
```

```

1R GCG GGC TCG CAG ATC TTC AGC GAG GGC GGG CTC GAC TAC CTC GGC AAC CCG AGC CTG
   A G S Q I F S E G G L D Y L G N P S L
2R GCG GgG TCG CAG ATC TTC AGC GAG GGC GGG CTC GAC TAC CTC GGC AAC CCG AGC CTG

1R ATC CAC GCG CAG AGC ATC CTC GCC ATC TGG GCG GTG CAG GTG GTG CTC ATG GGC GCC
   I H A Q S I L A I W A V Q V V L M G A
2R ATC CAC GCG CAG AGC ATC CTC GCC ATC TGG GCG GTG CAG GTG GTG CTC ATG GGC GCC

1R GTC GAG GGC TAC CGG ATT GCT GGC GGG CCG CTC GGC GAG GTC GTC GAC CCG CTG TAC
   V E G Y R I A G G P L G E V V D P L Y
2R GTC GAG GGC TAC CGG ATT GCc GGC GGG CCG CTC GGC GAG GTC GTC GAC CCG CTC TAC

1R CCC GGC GGC GCC TTC GAC CCG CTC GGC CTC GCC GAT GAC CCC GAG GCG TGC GCG GAG
   P G G A F D P L G L A D D P E A C/f A E
2R CCg GGC GGC GCC TTC GAC CCG CTC GGC CTC GCC GAT GAC CCg GAG GCc TtC GCG GAG

1R CTC AAG GTG AAG AAG ATC AAG AAC GGC CGC CTC GCC ATG TTC TCC ATG TTC GGC TTC
   L K V K K/e I K N G R L A M F S M F G F
2R CTC AAG GTG AAG gAG ATC AAG AAC GGC CGC CTC GCC ATG TTC TCC ATG TTC GGC TTC

1R TTC GTC CAG GCC ATC GTC ACC GGC AAG GGC CCC CTC GAG AAC CTC GCC GAC CAC CTC
   F V Q A I V T G K G P L E N L A D H L
2R TTC GTC CAG GCC ATC GTC ACC GGC AAG GGC CCC CTC GAG AAC CTC GCC GAC CAC CTC

1R GCC GAC CCC GTC AAC AAC AAC GCC TGG GCC TAC GCC ACC AAC TTC GTC CCC GGC AAG
   A D P V N N N A W A Y A T N F V P G K
2R GCC GAC CCC GTC AAC AAC AAC GCC TGG GCC TAC GCC ACC AAC TTC GTC CCC GGC AAG

1R TGA
   -
2R TGA
   940
1R AGTGGGGACCGTAGCTTAGCAGTGGTTAATGTGGTTGGATGGATTTGTGGCCAGCGAGTTCGTTGCTTTGGGT
2R gcgccGccgCcgcggtgctgccatGgcgAcgcategccttcaGcTaaGctagCtagGttgaCGacGctg'gcGtc
   964
1R TGGGGAAGATGGGTTTAGTGCACGAGATGATGATCGAGTTGGTGGTTGTGTACACTAAGAAGATGAAGAAGAAG
2R TctGcAgGAgGtgTgcGTGtGtgtAcgcGgTGcagtAGaTGtacGggGgaTctgCgcccagctcatctacAtgcc
   1113

```

Comparing cab1R with 2R: The CAAT and TATA sequences are underlined. Small letters indicate amino acid/nucleotide difference. *** represents presumed deletions.

ACKNOWLEDGEMENTS

This research was supported in part by a research grant from the Rockefeller Foundation. S.L. is supported by K.C. Wong Education Foundation.

Reference:

1. Lamppa, GK, Morelli, G, Chua, NH (1985) Mol. Cell. Biol. 5,1370-1378
2. Cashmore, AR (1984) Proc. Natl. Acad. Sci. USA 81, 2960-2964