

The soybean chloroplast genome: complete sequence of the *rps19* gene, including flanking parts containing exon 2 of *rpl2* (upstream), but lacking *rpl22* (downstream)

A.Spielmann, E.Roux, J.-M.von Allmen and E.Stutz

Laboratoire de Biochimie Végétale, Université de Neuchâtel, CH-2000 Neuchâtel, Switzerland

Submitted January 6, 1988

Accession no.X06429

We mapped both IR-LSC junctions (1). A nucleotide sequence (1024 bases) straddling the IRB-LSC junction is presented. It contains (5'→3') the 3'terminal part (52 bases) of the intron of the *rpl2* gene and exon 2 (C-terminal part: 123 codons), the intergenic spacer (57 bases), the complete *rps19* gene (93 codons) and 206 additional positions which lack the *rpl22* gene, which usually maps in this region (2,3). The sequencing results support hybridization results (4) suggesting that leguminous chloroplast DNA lacks the *rpl22* gene. The decoded soybean S19 aminoacid sequence is to 92% identical with that of spinach (2) and tobacco (3). 3'terminal intron consensus sequence is underlined. Arrow marks a possible splice site.

```

      10      20      30      40      50      60      70      80      90     100
      ↓ D M P L G K A I H N I E I T L G
GTACAGTTTGGGAAGAGGTTTTGATTGATCAAAAAGAAGAATCTACTTCACCGATATGCCCTTAGGCAAGGCAATACATACATAGAAATCACACTCGG
K G G Q L A R A A G A V A K L I A K E G K S A T L K L P S G E V R
AAAGGTGGACAATTAGCTAGAGCAGCGAGGTGCTGTAGCGAACTAATTGCAAAGAGGGGAAATCGGCCACATTAATAATACCTTCTGGGGAGGTCCGT
L I S K N C S A T V G Q V G N V G V N Q K N L G R A G S K C W L G
TTGATATCAAAAACCTGCTCGGCAACAGTCGGCAAGTGGGAAATGTTGGAGTAAACCAGAAAAATTTAGGTAGAGCCGGATCTAAATGTTGGCTAGGTA
K R P V V R G V V M N P V D H P H G G G E G R A P I G R K K P A T P
AGCGTCTCTAGTAAAGAGGAGTAGTTATGAACCCGTAGACCATCCGCATGGGGGTGGTGAAGGGAGGGCCCCAATGGTAGAAAAAACCCGCAACTCC
W G F P A L G R R S R K R K K Y S D N L I L R R R T K *
TTGGGTTTTCTGCCTTGGGAAGAAGTAGAAAAAGGAAGAAATATAGTGATAATTGATTCTCTCGTCGTACTAAATAGTAGAGAAAAATAGAAA
M T R S L K K N P F V A N H L L R K I N
TTGTTTCTCGTCTTTACAAGAAAAAGGAGTAATTAACGTGACAGTTCACATAAAAAAATCCTTTGTAGCGAATCATTATTAAGAAAAATAATA
K L N T K A E K D I I I T W S R A S T I I P T M I G H T I A I H N G
AACTTAACACAAAAGCGAAAAAGATATAATAATAACTTGGTCCAGAGCATCTACCATTATACCTACAATGATGGCCATACCATTGCTATCCAACCGG
IRB ← LSC
K E H L P I Y I T D R M V G H K L G E F S P T L N F R G H A K N D
AAAAGAGCATTACCTATTTATATAACAGATCGTATGGTAGGCCATAAATTAGGAGAATTTTCACCTACTCTAAATTTCCGAGGACATGCGAAAAATGAT
N R S R R *
AATAGATCTCGTGAAGTAAATTTTAAACATAAATTTTAAATCGTAATTTATAACTAAATATTACAAAATGGAATCTTATTAATAATGAAATTCAAA
TCATTTCTTTTATAAAAAAAAATTCITTTAATTAATATTTAATGAAATTAATATTTAAATTAATAAAAATTAGCTGTGTTTTATTTCGATTTTTTG
AATTGGTTTTTTTTTACAGTAGCT
    
```

References : (1) A. Spielmann and E. Stutz (1983) NAR 11: 7157-7167. (2) G. Zurawski, W. Bottomley and P.R. Whitfeld (1984) NAR 12:6547-6558. (3) M. Tanaka, T. Sugita, K. Shinozaki and M. Sugiura (1986) PNAS 83:6030-6034. (4) J.D Palmer B. Osorio and W.F. Thompson (1988) Curr. Genet. in press.