

Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*

Inge Unfried and Peter Gruendler

Institut für Botanik der Universität Wien, Abteilung Cytologie und Genetik, Rennweg 14, A-1030 Vienna, Austria

Submitted May 9, 1990

EMBL accession no. X52320

From a genomic library of *Arabidopsis thaliana* (Brassicaceae) constructed in the phage lambda vector EMBL3 we have isolated and subcloned a complete rRNA gene unit. The intergenic region (IGR; 1) and the region coding for 18S rRNA (2) were published previously. Here we present the nucleotide sequences of the 5.8S rRNA and the 25S rRNA genes as well as of the internal transcribed spacers (ITS).

The 5.8S coding region shows 97.0% homology with the 5.8S gene from *Sinapis alba* (Brassicaceae; 3) and 96.9% homology with the 5.8S gene from tomato (4). The 25S coding region has 93.3% homology with the analogous gene from tomato (5).

ITS1 and ITS2, which are regarded as not being evolutionarily conserved, are 75.7% (ITS1) and 81.1% (ITS2) homologous between the two Brassicaceae mustard and *Arabidopsis*, whereas the IGR's of the two Brassicaceae *Raphanus sativus* and *Arabidopsis* show no significant homology. This may indicate

that most of the ITS's could be important for post-transcriptional processing.

ACKNOWLEDGEMENTS

This work was supported by grants P6577B and P7279BIO from the Austrian 'Fonds zur Förderung der wissenschaftlichen Forschung'.

REFERENCES

- Gruendler, P., Unfried, I., Pointner, R. and Schweizer, D. (1989) *Nucl. Acids Res.* 17, 6395-6396.
- Unfried, I. and Gruendler, P. (1989) *Nucl. Acids Res.* 17, 7513.
- Rathgeber, J. and Capesius, I. (1989) *Nucl. Acids Res.* 17, 7522.
- Kiss, T., Kiss, M., Abel, S. and Solymosy, F. (1989) *Nucl. Acids Res.* 17, 2127.
- Kiss, T., Kiss, M. and Solymosy, F. (1989) *Nucl. Acids Res.* 17, 796.

```

1  gaattctag taagcggag teatcagct gqgttgact cytccctgc cttgtacac acgcggcgt gctctaccg attgatgat ccggtgaat
101  gttccgagc cggcgcctg ggtggttgc cgcgcgcgc gtccgcgaa gtccactaa octtatait tagcgaagc agaatcgtc acaagtttc
201  cgtagtgaa cctgaggag gatcattgc GATACCTGT CAAAACAGAA CGACCCCGA ACCAAGATC ACCACTCTG GTGGCCGGT TTCTTAGCC
301  ATTCCTTGC CCGCGATCC GGTGTTTGC GTATGGCAT GATCGGAGC TTTTATCTG GCTCTGTGT GCGCTGTCT TCCGGATAT ACAAAAACC
401  GGACGAAA GTGTCAAGG ACATGC AACGCGCTG CATTCGCTC CCGGAGAGC GACTGTGGC GATGTCTGT CTGGAACTG AACTTAAA
501  GACTCTGCG CACGGATAT CTGCGCTGC GCATCGATG AGAACGTAG GAAATGGAT ACTTGTGTG ATTTAGGAA TCCCGTGAAC CATCGACTC
601  TTGAGCCAA GTGGCGCCC AAGCTCTGT GCGAGGGCA CTTCTGCTG GTGTCAAAA ATGTGTCTC CTCACCAAT TTCTGTGAT GCGGAGGAA
701  CTTGTCTGC CTTGTCTGC AGCTGTAGT GACCCAAAT CCAATATGC ACCCAAGTC AGCGGGATT ACCCGTGTG TTTAGCATA TCAATATGC
801  ATGCGTCTC CTTGTCTGC AGCTGTAGT GACCCAAAT CCAATATGC ACCCAAGTC AGCGGGATT ACCCGTGTG TTTAGCATA TCAATATGC
901  GAGGAAAGC AACTACAGC GATTCCTTA GTAAAGGCA GCGAACCGC AAGAGCCAG CTTGAAATC GGAAGCTTC GCGTGTGAA TTGTGTCTG
1001  GAGAGCTC CTCAGCAC GACCGGCTT AAGTTCCTG GAAAGGGGC CAGAGAGGC TCGAGAGCC TGTGTCCCG ACCCTGTGC ACCAGAGGC
1101  GCTGTCTAG AGTGGGTTG TTTGGAAAT CAGGCCAAT CCGGGGTAA ATTCCTTCA AGCTTAAAT GCGGAGAGC ACCGATAGC AACAGTACC
1201  CCGAGTAAA GATGAAAGC ACTTTGAAA GAGACTTAA GATGCTTGA AATTCCTGC AGGAGAGCG ATGGGGGCG GCGAGGCTC CTGTGTGAT
1301  CCGGAGCGA GCAATCGCT CCGCGCACG ATTCGGGGC TCGAACGAC CCGATTAGC TGGCGCTTA AGCGGGCTT TTTGATAGC TTGTGTGAC
1401  GTGCTGGCG TGAATGTGT CTCAGCACG CCGTAAAGC CCGCTTGGC CACTAGGCT CCGCGGCTC GCGCTGTGT GCTGCCATT CGACCGTCT
1501  TGAACAGCG ACCAGGAGT CTCAGCACG TCGAGTCAA CCGTGTGTA AACCCATAG CCGCAGGAA GCTGATGCG GCGAGCTCT CCGGTGACG
1601  GCGACAGCG CTTGATCTT TCGAGAGGT GCGAGTCTA GCGTGTGTA CCGGACCGA AAGATGTGA ACTATGCTG AGCGGGTAA AGCTGAGGA
1701  AACTGTGCG GAGCCGCGA GCGATCTGA CCGTAAATC GTTCTGTG CTTGGTATA GCGGGAAGC ACTATGCTA CCACTAGTA CTGTGTCTC
1801  TCGGATGTT CCGTCAAGT AGCTGTAGT CCGAGCGAG TCTATGCG TAAAGCATT GATTTAGCG ATTTGGGCG CAGCGCTGT ACTATGCTC
1901  AACTTAAA TAGTAGGAC CTTGTGCTG CTTGTGCTG CCGTCAAGC GATGTGAGC CTCACAGTC GCGATTTTG TTAGCAGAA CTGGGATGC
2001  CCGATGACC CAGACCGCG TTGGTGTCT CACTCGCGT CTAAGTAGA ACCCAAGG GATTTGTCT GATTTAGCA CAGGAGGCT GCTATGGA
2101  GTGAAATC CTAAGGAGT CTTATCAC CACTGTGCG AATCACAG CCGCCAAAT GATTTGTCT TAAAGCGAC CTAAGCGCG CCGTGGGCG
2201  AAGAGCAG CCGTGTGAG TAGAGGGCG CCGCGTGGC TCGAAACTT AGCGGCGAG CCGCGAGCG CCGTGTGCG CAGATCTGG TGTATGAGC
2301  AATATTCGA ATGAGACTT TGAAGCGCA AGAGGGAAA GATTCAAGT GAAAGCGCT TCGATGTGG TTATGTGAT CTAAGAGTC GCGGAAACC
2401  GCTGTAGAC CTTTAAAGC ACTTGTGAG GCGATCGGT TAAATTCGC GAAAGCGCT GCGAGGTTG AGCGAAGT TAAAGGATC GCGAGCTGC
2501  CCGGGGCTT CCGGAGAGT TATCTTTCT GTTAAAGC CTCGCCACC TCGAAAGCG TCGAGCGGAG TTAGGTTCA GCGCTGTGA GAGCACTGC
2601  CCGCGGCTG TGTGTGCTG CCGCGGCGC GCGTTGAAA ATCGGAGCA CCGAGTGGC CTCAGCGCG GTGTACTCA TAAAGCGAT AGCTGTCAA
2701  GTTAAAGC CTTGTGTGA TCGAACATG TAGCAGAGC AATGTGCGA AATGATCGC TAACTGTGG AAGAGGATG CTTGTGAGG CTTGGCTGG
2801  GCGTCCAGT TCGAACCGC TCGCTGTCA GCGACTGCT CAGACTGCT CCGCGGAGC AGCGGTCGC CCGTGTGCG CCGGGAGCG ACTGTGAGC
2901  CTTGTGTG GACTTTTCC CCGCGTGA ACAGTCTCT CAGACTGCT CAGACTGCT CAGACTGCT CAGACTGCT CAGACTGCT CAGACTGCT
3001  CCGTGTGAG CTAAGCGAA TGTATTTCT CCGACTGCT CTAAGTGA AACTGAGAA ATTCACCAA CCGCGGATA ACCCGGAGC TAACTAGCA
3101  TCTCTAAGC TAGCAGAGC CCGTGTGAG TAAATGTA CCGCAGAA TCGATTAGC AGATTCCAC TGTGTCTG TACTATGAG AAGAACTCA
3201  CCGAAGGAA CCGCTTGGC AAGATCAGC GCGAAGAGC ACCCTTGA CTTTACTCT ATCTGACTT TGTGAAAGA CTTGAGAGT GTAGGATAG
3301  TGGAGCTC GCGCAGAG AATACACT ACTTTAAG TTTTACTT TACTGTGA ATCGAGGCG GCGTACAC CCGTGTGTT GTCCCGTAG
3401  CCGCTGCG CCGTGTGAG CCGCGGAGC AATTTCTAG TCGGGAGTT TCGCGGCG CCGCAGCTG TTAAGATA ACCAGGCTT CTAAGTGA
3501  CCGTACAGC AATGAAATC TCGTGTGAA CAAAAGGTA AAGCTGCT TGAATCTAT TTTCTGAGC AATGAGAGC TCGAAAGCT GCGTATGCA
3601  TCGTGTGAG TCGGATTT GAGCTAGAG GTGTGAGAA AGTACAGC GCGATCTG CCGTGTGCA CCGAGGCTT CTAAGGAGC TCGCTTTG
3701  ATCTGTGAT GTGCTCTT CTAATCTAG TCGAGCAGA TTAAGGAT GTTGTGAT TCGAACCA ATAGGAGC TCGAGGAGC TCGAGGAGC TCGAGGAGC
3801  GTGAGCAGC TTAGTTTAC CTAATCTAG CCGCGGAGC GATGATAT CAACTAGTA CCGAGGAGC CCGTGTGCG CAGATGAGC CATCGGCTC
3901  GCGTGTGAG CAGTGTGCG GAGCTAGC TCGCTGTG TCGAGTGA CCGCTTAG TCGAGTGC GCGTGTGAG CCGAGGAGC GCGAGGAGC CCGCGGCTC
4001  GATGTGAG CCGTGTGAG AGCTTAGCT CCGAGGAGC GTGCTGTG CTAAGTGT TCGCGGAGC GCGTGTGAG ACCCGTGA ATTAATATTA
4101  CCGAGGAGC GCGTGTGAG TCGTGTGAG AGCTTAAA TCGCGAGCG GTATTGTA CCGAGGAGC GCGTGTGCG CCGAGTGA CTAAGTGA
4201  GCGTGTGCT GATGATTC GCGTGTGAG taactcact caaaaaaac aatcccaat tctacacag tttttctca ctaacaaag acaagctct
4301  taacgaattc

```

Figure 1. Nucleotide sequences of the 5.8S and 25S coding regions and of ITS1 and ITS2. ITS1 and ITS2 are in upper case, the 5.8S coding region is in upper case and underlined, and the 25S coding region is in upper case and bold.