# cDNA nucleotide sequence of *Sn*, a regulatory gene in maize

Gabriella Consonni, Angelo Viotti[1], Stephen L.Dellaporta[2] and Chiara Tonelli*
Dipartimento di Genetica e di Biologia dei Microrganismi, Università degli Studi di Milano, Via Celoria 26, 20133 Milan, [1]Istituto di Biosintesi Vegetali CNR, Via Bassini 15 Milan, Italy and [2]Department of Biology, Yale University, PO Box 6666, New Haven, CT 06511, USA

Submitted October 3, 1991 EMBL accession no. X60706

The R gene family determines the timing, distribution and amount of anthocyanin pigmentation in maize. This family comprises a set of regulatory genes, consisting of a cluster of several elements at the *R* locus, located on the long arm of chromosome 10, the *Lc* and *Sn* gene lying about two units *R* distal and *B* on chromosome 2 (1). Each gene determines a tissue specific pigmentation of different parts of the seed and plant. The proposed duplicated function of *R, Sn, Lc* and *B* loci, implicated as related genes with similar regulatory roles in distinct tissues, is reflected in DNA sequence similarity. DNA crosshybridization has been shown for the *R, B, Lc*, and *Sn* genes (2, 3, 4, 5). The expression of the structural genes of the anthocyanin pathway is mediated by the presence of an active form of these genes suggesting that they encode structurally related proteins that act as transcriptional activators of the genes encoding the biosynthetic enzymes.

We now report the complete cDNA sequence of the *Sn:bol3* allele and the comparison with the cDNA sequences of *Lc* and of the *S* component of the *R* complex, referred as *R-S*. The sequence was determined in two independent cDNA clones (2) and the polymorphisms with *Lc* and *R-S* have been also controlled in two different *Sn* genomic clones. The start of transcription was determined by primer extension (2). The 5′ end of *Sn* cDNA starts 19 bases from the transcription start site. The clones were sequenced on both strands by the dideoxy method using oligodeoxynucleotides as primers.

The sequence allows the detection of some features specific of the eukaryotic genes. It contains a 616 amino acids long open reading frame beginning with the fifth ATG start codon at nucleotide position 382 and ending with a TGA stop codon followed by 340 nucleotides containing putative polyadenylation signals. In the leader sequence two short open reading frames of 38 and 15 amino acids are also present. The putative protein encoded by the 616-amino acid open reading frame has features similar to those of transcriptional activators. It contains a large acidic domain (amino acids 188−324) with 35 acidic and 8 basic amino acids for a cumulative negative charge of −27 and a basic region (amino acids 421−514) part of which was found to have similarity to the *Myc* family of oncogene proteins. This region has been proposed as a novel DNA binding and dimerization motif called *helix-loop-helix* (6). A comparison between the three cDNA sequences of *Sn, Lc* and the *S* component of *R* (*R-S*) (3, 7) discloses a very high degree of homology particularly in the translated region. In addition there are specific features that make each clone distinguishable from the others. Two insertions, present in the *Sn* and *R-S* clones, are not found in the *Lc* gene. The first is a 146 bases insertion (starting at position 230) present in the leader sequence (the reported *R-S* clone starts inside this insertion at position 237). The second is a small 18 nucleotide insertion present in the translated sequence at position 1028 located inside the acidic domain causing with the addition of two acidic aminoacids a more negative charge of the *Sn* and *R-S* putative proteins in comparison to the *Lc* one. Three other small insertions, the first two of three bases and the third of six bases at position 1144, 1520 and 1635 respectively, are present in the *Sn* and *Lc* cDNAs in comparison to the *R-S* gene. All these insertions however do not cause frame shift mutations. Other minor polymorphisms are present among the three different clones, most of them confined to the trailer region. These findings support the hypothesis that *Sn, Lc* and *R* genes encode functionally related proteins that act as transcriptional activators of the genes encoding the enzymes of the anthocyanin biosynthetic pathway. The distinct pattern of pigmentation determined by these genes may thus reflect differences in the promoter region rather than functional differences in their product.

## ACKNOWLEDGEMENTS

Research supported by funds of the 'Bridge Program' of the Commission of the European Communities (Contract BIOT-0164-C EDB) to C.T. and of the National Research Council of Italy, Special Project RAISA, Sub-project 2, paper N.259.

## REFERENCES

1. Robbins,T.P., Walker,E.L., Kermicle,J.L., Alleman,M. and Dellaporta,S.L. (1991) *Genetics* **129**, 271−283.
2. Tonelli,C., Consonni,G., Faccio Dolfini,S., Dellaporta,S.L., Viotti,A. and Gavazzi,G. (1991) *Mol. Gen. Genet.* **225**, 401−410.
3. Ludwig,S.R., Habera,L.F., Dellaporta,S.L. and Wessler,S.R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7092−7096.
4. Chandler,V.L., Radicella,P.J., Robbins,T.P., Chen,J. and Turks,D. (1989) *Plant Cell* **1**, 1175−1183.
5. Dellaporta,S.L., Greenblatt,I., Kermicle,J.L., Hicks,J.B. and Wessler,S.R. (1988) In Gustafson J.P. and Appels,R. (eds), *Chromosome Structure and Function: Impact on New Concepts*. Plenum Press, New York, pp. 263−281.
6. Murre, C., Schonleber McCaw,P. and Baltimore,D. (1989) *Cell* **56**, 777−783.
7. Perrot,G.H. and Cone,K.C. (1989) *Nucleic Acids Res.* **17**, 8003.

* To whom correspondence should be addressed