Southern blot analyses show that the soybean genome contains many U1 snRNA genes, both in tandem arrays and as isolated genes. We have cloned and sequenced two soybean U1 snRNA genes, U1a and U1b, which represent distinct multicopy tandem repeat elements. Sequence analysis of the first 46 nt of soybean embryo protoplast U1 snRNA by primer extension shows that sequences corresponding to U1a and U1b are present in approximately equal proportions. Soybean U1a and U1b snRNA sequences, shown aligned to maximize homology, are both very similar to common bean U1 snRNA (1). Plus signs indicate sequences common to all three genes; dots indicate bases identical between the two soybean genes. The predicted secondary structure and sequences of the single-stranded 5' end, loops (indicated by Roman numerals in brackets), and Sm binding site are virtually identical to those of other U1 snRNAs (1).

The 5' flanking regions of the two soybean U1 snRNA genes exhibit 71% homology, mostly concentrated in blocks of identity. A "TATAAA" sequence occurs 31-26 bases 5' to both genes. Blocks of sequences shared by common bean U1 and soybean U1a and U1b snRNA genes (indicated by + in the figure) occur as far as 228 bases 5' to the U1 snRNA gene. One shared sequence (TCCCACATCG, underlined in figure) also occurs 5' to several Arabidopsis thaliana U2 snRNA genes and is required for their expression (2). However, we have been unable to detect transcription of the cloned soybean U1a and U1b genes when introduced into soybean embryo or tobacco leaf protoplasts.

REFERENCES