Isolation of tobacco SSU genes: characterization of a transcriptionally active pseudogene

J.K.O'Neal, A.R.Pokalsky, K.L.Kiehne and C.K.Shewmaker*

Calgene, Inc., 1920 Fifth St., Davis, CA 95616, USA

Received August 10, 1987; Accepted September 9, 1987

ABSTRACT

Genomic clones containing three genes for the small subunit (SSU) of ribulose biphosphate carboxylase were isolated from tobacco. Detailed analysis was performed on two of these clones to give a clearer picture of this multigene family in tobacco. This analysis demonstrated that one of the clones contained a pseudogene that was unusual in that it was transcriptionally active. This is the first transcriptionally active pseudogene that has been reported in plants. In addition, another clone was found to contain coding sequences which are 100% homologous to a previously-cloned tobacco SSU gene (Mazur, B.J. and Chiu, C-F. [1985] *Nuc. Acids Res.* 13, 2372-2386), indicating that gene duplication and/or gene conversion may have played a role in the evolution of the tobacco SSU family.

INTRODUCTION

One of the most abundant proteins in nature is the enzyme ribulose-1,5-bisphosphate (RuBP) carboxylase. This multimeric enzyme which catalyzes the first step in the Calvin cycle accounts for approximately 50% of the soluble protein in leaves¹. The approximately 550,000 MW holoenzyme is composed of 8 large subunits and 8 small subunits². The 53,000 MW large subunit (LSU) is encoded and synthesized in the chloroplast where the holoenzyme functions³. The 14,000 MW small subunit (SSU) is encoded in the nucleus⁴ and synthesized on cytoplasmic ribosomes as a 20,000 MW precursor. This precursor is transported into the chloroplast with removal of an amino terminal transit peptide^{5,6}.

The level of SSU protein in leaves has been shown to be regulated by light⁷. The SSU mRNA levels increase in light and this increase has been shown to be modulated by phytochrome⁸⁻¹². SSU protein is encoded by only a few nuclear genes. Numbers of 6-10 copies per genome have been reported^{11,13,14}. Conversely, the LSU gene encoded by the chloroplast genome is present at several thousand copies per cell¹⁵. For this reason, the sequences of SSU genes that allow for light regulation and the coordinate regulation with LSU gene expression are of great interest.

We report here the isolation of SSU genes from *Nicotiana tabacum*. Of the two genes characterized, one is a transcriptionally active pseudogene, the first such gene reported for plants. *N. tabacum* is an amphidiploid that arose from *N. sylvestris* and *N. tomentosiformis*¹⁶. The SSU proteins from these species differ slightly in amino acid sequence¹⁷ and thus their genes can be distinguished. This offers the opportunity to trace the fate of genes from both parents in the resultant amphidiploid.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Enzymes

The lambda phage Charon 32¹⁸ was grown in either LE392 (F⁻ hsdR514 supE44 supF58 lacY1 or lac(I-Y)6 galK2 galT22 metB1 trpR55 λ⁻)¹⁹ or K802 (F⁻ metB1 Δ(lacI-Y)6(lac-3) or lacY1 galK2 galT22 λ⁻ supE44 hsdR2)²⁰. *E. coli* 71-18 (Δ(lac-proAB) supE thi F' lacI^qZ M15 proA⁺B⁺))²¹ was routinely used for transformations with phages M13 mp18 and M13 mp19²². The *E. coli* strain 71-18 was routinely grown on 2YT medium²³ while LE392 and K802 were grown on NZYM media²⁴. Plasmid pSS15¹³ contains an 850 bp pea SSU cDNA and plasmid pSTV 34²⁵ contains an 185 bp *N. sylvestris* SSU cDNA.

Restriction enzymes and *Ba*31 were obtained from Bethesda Research Laboratories, T4 ligase from Promega Biotech, T4 kinase from PL Biochemicals, and AMV reverse transcriptase from Life Sciences. All enzymes were used in accordance with supplier's instructions. General cloning procedures are as described²⁴. Oligo dT cellulose was from Collaborative Research. All ³²P-labelled radionucleotides were from New England Nuclear.

Isolation and Mapping of Tobacco SSU Genes

A tobacco (*N. tabacum* 'Samsun') *Eco*RI partial library in Charon 32 was kindly provided by Dr. Robert Goldberg (UCLA). The library was screened for members of the RuBP carboxylase SSU gene family with a ³²P-labelled pea SSU cDNA probe (see probe E below). The phage library was plated using either LE392 or K802 as a host. Phage DNA was transferred to Gene Screen Plus (New England Nuclear) filters using the method of Benton and Davis²⁶ except that the lysis solution was 0.5M NaOH, 1.5M NaCl. Filters were neutralized in a solution of 0.5M Tris-HCl pH7.5, 1.5M NaCl and rinsed in 2x SSC, 0.1% SDS. Prehybridization, hybridization and washing of filters were done at 33°C as previously described²⁷.

Three positive SSU clones selected after autoradiography (designated TSSU3-1, TSSU3-2 and TSSU3-8) were replated and rescreened until plaque

purified. Phage stocks were made and large scale DNA preparations performed as described²⁴.

SSU coding sequences for each of the three isolates were located by restriction mapping as described by Maniatis et al²⁴ using the ³²P-labelled pea SSU cDNA (probe E). 5' and 3' orientations of the SSU genes within the insert were determined by hybridization at 42°C with a ³²P-labelled *N. sylvestris* SSU cDNA that contains only 3' coding sequences (probe A).

Subcloning and Sequencing

A 3.4 kb *Eco*RI fragment containing the TSSU3-8 gene was subcloned into M13 mp18. From these original subclones various smaller fragments were generated by restriction digestion or *Ba*/31 deletion and subcloned as needed to complete the sequencing. For the TSSU3-2 gene, a 4.0 kb *Kpn*I fragment from the original lambda vector was subcloned into M13 mp18 and M13 mp19. Smaller subclones were generated as needed from a variety of restriction fragments. In addition, overlapping clones derived from the method described by Dale et al²⁸, were constructed as necessary to complete the sequencing.

Both TSSU3-8 and TSSU3-2 were sequenced using the dideoxy chain termination procedure²⁹. The Maxam and Gilbert method of chemical cleavages was employed where necessary³⁰. Both strands of the DNA were sequenced and all sequences in both strands were overlapped. Sequence analysis was aided by computer programs from Intelligenetics, Inc.

Probe Description and Preparation

Oligomers. Synthetic oligomers were synthesized on an Applied Biosystems (Foster City, CA) 380A DNA Synthesizer. Oligomers were detritylated and those used in primer extension analysis were utilized without further purification. Those used as probes were gel purified in accordance with the instructions provided by Applied Biosystems. Oligonucleotides used as probes for Northern and Southern analysis were 5'-end labelled by a procedure modified from Berent et al³¹. For each reaction, 300 ng of oligomer was added to 10 µl of 10x kinase buffer³¹, 21 µl [γ -³²P] ATP (10 µCi/µl), and 2 µl T4 polynucleotide kinase (10 units/µl). Water was added to a final volume of 100 µl. The solution was incubated at 37°C for 20 minutes. An additional 1 µl of kinase was added and incubation allowed to continue for an additional 20 minutes at 37°C. Unincorporated [γ -³²P] ATP was separated from the labelled oligonucleotides by chromatography on a Sephadex G-50 column in 10 mM Tris-HCL pH 8.0, 1 mM EDTA. Specific activities of 1-3x10⁹ cpm/µg were obtained.

Synthetic oligomers used were:

- a. Probe 1 - a 30 nucleotide oligomer of sequence 5'-TGTTAATTACACTTTAAGACAGAAAGATTT-3' derived from the

5'-untranslated region of gene TSSU3-8 immediately upstream of the initiation codon.

- b. Probe 4 - a 30 nucleotide oligomer of sequence 5'-TGTTGAAAGTAATTGATTAGCTTAAAGCTA-3' corresponding to an area in the 5'-untranslated region of gene TSSU3-2 immediately upstream of the initiation codon.
- c. Probe 10 - a 30 nucleotide oligomer of sequence 5'-CCAGTGAAAGGTGCAACCATGTTAGCTTGA-3' corresponding to an area in the first exon of gene TSSU3-8.

Isolated DNA fragments. DNA fragments used as probes were isolated from gels and labelled by nick-translation as previously described²⁷.

Probe A: This is a *Bam*HI fragment of pSTV34, a SSU cDNA clone from *N. sylvestris*²⁵. It contains 185 bases of the 3' coding region.

Probe B: Nucleotides +349 to +464 of TSSU3-8 isolated as a *Hae*III-*Eco*RI fragment. This probe lies entirely within the coding region of the gene.

Probe C: Nucleotides -1374 to +73 of TSSU3-8. Contains 5'-flanking and 5'-untranslated regions but no coding sequences.

Probe D: Nucleotides -411 to +73 of TSSU3-8. Derived from probe C.

Probe E: This is a 740 base *Pst*I fragment from pSS151¹³, a pea SSU cDNA clone.

DNA Isolation and Southern Analysis

Tobacco leaf tissue (*N. tabacum* 'Xanthi nc', *N. tabacum* 'Samsun', *N. sylvestris*, and *N. tomentosiformis*) for DNA analysis was quick frozen in liquid nitrogen and stored at -70°C until needed. Plant DNA was isolated using the modified cetyltrimethylammonium bromide (CTAB) procedure of Taylor and Powell³². Southern analysis was done as previously described²⁷ with slight modifications; agarose gels were 0.8%. Prehybridization and hybridization were at 42°C with washes at 55°C.

RNA Isolation and Northern Analysis

RNA was prepared from tobacco leaf tissue (*N. tabacum* 'Xanthi nc') by a slight modification of the guanidine thiocyanate procedure of Colbert et al³³ as previously described³⁴. PolyA⁺ RNA was purified over oligo (dT) cellulose as described²⁴. RNA denaturing gels were run, and blotted as previously described³⁴. Northern blots probed with synthetic oligomers were prehybridized and hybridized at 37°C in 0.9 M NaCl, 6 mM EDTA, 20 mM Tris-HCl pH 8.0, 1x Denhardt's, 10% dextran sulfate, 100 µg/ml denatured herring sperm DNA (as modified from Torczynski et al³⁵). After hybridization, filters were washed in 5x SSC, 0.1% SDS, 2x 30 minutes at 37°C, then 2x 20 minutes at 45°C.

Primer Extension Analysis

The procedure for the primer extension reactions was modified from Lee and Luse³⁶. For each reaction, 1.0 µg of polyA⁺ leaf RNA and 0.2 µg of the appropriate synthetic oligomer were ethanol precipitated. The nucleic acids were dissolved in 10 µl of 0.1 M NaCl, 20 mM Tris-HCl pH 8.0, 0.1 mM EDTA, denatured at 100°C for 2 minutes, and incubated at 60°C for 6 hours. Two µl of 400 mM Tris-HCl pH 7.0, 50 mM MgCl₂, 20 mM dithiothreitol was added. After 5 minutes on ice, 1.5 µl H₂O, 2 µl nucleotide mix (2 mM each of dATP, dCTP, dGTP, and TTP with 20 mM Tris, pH 7.5), 4 µl [γ -³²P] dCTP (10 µCi/µl, New England Nuclear, Boston, MA) and 0.5 µl AMV reverse transcriptase (10 units/µl) was added. Reactions were incubated at 37°C for 30', extracted with phenol:chloroform (1:1), and again with chloroform before ethanol precipitation. To provide accurate size standards, the identical primers used for the primer extension reactions were used to prime dideoxy sequence reactions²⁹ on single stranded template from M13 mp18 or M13 m19 phages containing 5' regions of the appropriate gene. Primer extension reaction products and the dideoxy sequence standards were run on 12% (w/v) polyacrylamide/urea sequencing gels as described²⁹. Gels were subjected to autoradiography.

RESULTS

Preliminary characterization of genomic isolates

Screening of an *N. tabacum* cv. 'Samsun' genomic library was performed with SSU cDNAs from both pea and *N. sylvestris*. The pea cDNA clone, pSS15¹³, designated probe E, contains the entire mature coding sequences plus some of the sequences for the transit peptide. The *N. sylvestris* cDNA, pSTV34²⁵, designated probe A, contains only the 3' half of the mature coding region. Three phage isolates, TSSU3-1, TSSU3-2, and TSSU3-8, were obtained. Maps of these clones are shown in Figure 1. Clones TSSU3-1 and TSSU3-2 contain entire genes while clone TSSU3-8 contains only the 5' half of a gene. (The rest of the gene was presumably not cloned in the *Eco*RI partial digest that gave rise to TSSU3-8.) Previous studies in *Lemna*, pea, and petunia^{14,15,37,38} have indicated the SSU genes can be closely linked. The restriction pattern for these clones indicated that this was not the case for the particular clones isolated here. This was confirmed by performing Southern hybridizations between the three clones (data not shown). Comparisons of the maps of TSSU3-1, TSSU3-2, and TSSU3-8 with those of the two tobacco SSU genes previously cloned³⁹ also indicate that these phage contain SSU genes different than those previously reported.

Figure 1. Schematic representation of the three small subunit genes. E = *EcoRI*; S = *SphI*; B = *BamHI*; K = *KpnI*. ■ = SSU gene regions; □ = tobacco DNA cloned in Ch32; — = Ch32.

Sequence analysis

The SSU genes and flanking regions in TSSU3-2 and TSSU3-8 were sequenced. Using sequences of other SSU genes, the leader peptide, mature coding regions and introns were located. In Figure 2, the sequence of these two genes is shown and compared to that of a previously sequenced tobacco SSU gene, NtSS23³⁹.

As mentioned, clone TSSU3-8 contains only the 5' half of the coding region and ends at an *EcoRI* site frequently found in the second exon of SSU genes^{11,39,40}. The position of the first intron is the same as that found for other small subunit genes^{11,39}. Comparison of TSSU3-8 to NtSS23 shows a striking degree of homology. The sequences from the ATG to the *EcoRI* site, including the intron, are identical. This is unexpected as intron sequences frequently vary more between analogous genes than do coding sequences⁴¹. That TSSU3-8 and NtSS23 are indeed different genes can be confirmed by looking at the sequences upstream of the ATG. Between -400 and the ATG there is approximately 90% homology between the two genes. Further upstream, the homology decreases.

The homology between TSSU3-2 and NtSS23 is less than that between TSSU3-8 and NtSS23. This is true for all regions examined; coding regions, introns, and flanking sequences. Within the coding region alone, there are 60 nucleotide differences between NtSS23 and TSSU3-2. These variations in the coding region lead to eleven amino acid changes between TSSU3-2 and NtSS23. One of these produces a translational stop codon in the leader sequence, two amino acids before the processing site (see Figure 2). Thus, TSSU3-2 could not produce SSU protein and is therefore a pseudogene. Most dicot SSU genes sequenced to date have two introns at conserved sites^{11,15,39} and TSSU3-2 has

pairs upstream of the ATG, but beyond that no significant homology is found. The mature SSU protein of *N. sylvestris* differs at three amino acids from that of *N. tomentosiformis*¹⁷. *N. sylvestris* mature SSU has Ile-Asn at positions 7-8 and a His at position 48, whereas *N. tomentosiformis* has Tyr-Gly and Arg, respectively, at these positions. Thus TSSU3-2 originated from the *N. sylvestris* parent as did TSSU3-8 and NtSS23³⁹.

Determination of Transcriptional Initiation Sites

The TATA box, a region located between -35 to -30 base pairs from sites of RNA initiation, has been shown to be necessary for RNA transcription in eukaryotes. Both TSSU3-8 and TSSU3-2 have putative TATA boxes (see Figure 2), located 109 and 78 base pairs from their respective initiation ATGs. The TATA boxes of SSU genes published to date show a remarkable similarity and a consensus sequence for the TATA boxes of SSU genes has been proposed⁴². The TATA of TSSU3-8, CATTATATATAG, matches the consensus exactly. The TATA of TSSU3-2, CACTATATATAG, differs only at one nucleotide. Thus both genes appear able to function transcriptionally, although TSSU3-2 has a stop codon in the transit peptide coding region. The functionality and transcription initiation sites of TSSU3-8 and TSSU3-2 were determined by primer extension analysis. Two 30 base oligomers, complementary to the first 30 nucleotides upstream of the ATG in TSSU3-8 (probe 1) and TSSU3-2 (probe 4) were used. Due to the similarity between TSSU3-8 and NtSS23 in this region, probe 1 will hybridize to at least both TSSU3-8 and NtSS23 transcripts but not to TSSU3-2 transcripts. To determine the 5' end of TSSU3-8, probe 1 was annealed to tobacco polyA⁺ RNA and extended in the presence of ³²P-dCTP and unlabelled dATP, dTTP, and dGTP. The results are seen in Figure 3. The band indicated by the upper arrow corresponds to a 5' terminus at 5' CATC. This start site is 33 base pairs from the TATA box. A set of bands 11 to 13 base pairs below the first band is also seen in the primer extension lane. As mentioned earlier, probe 1 would also hybridize to transcripts from NtSS23; the only difference in this 30 base pair region being a two base pair insertion in TSSU3-8. Using 33 base pairs as a common distance between TATA and cap site, NtSS23 transcripts would be 12 base pairs shorter than those of TSSU3-8. Thus, these lower bands could be attributed to NtSS23 and possibly other SSU genes. The 5' end of TSSU3-2 (Figure 3) was determined in the same manner. The arrow corresponds to a band, indicating the site of initiation is at AAGG, 32 base pairs from the beginning of the TATA box. Smaller bands beneath the arrow are believed to be due to impurities in the synthetic primer.

Figure 3. Primer extension analysis of genes TSSU3-8 and TSSU3-2. The PE lanes contain 0.2 μ g of appropriate primer hybridized to 1 μ g of polyA⁺ tobacco leaf RNA. The T, C, G, A lanes display sequencing reactions using these primers and M13 templates within the region of interest.

(↓) indicates the initiation site; — indicates the probe used; — indicates the TATA box; and ■ the ATG.

Determination of levels of expression

Relative levels of expression of the SSU genes were determined by Northern analysis using synthetic probes (Figure 4). Probe 1, as mentioned above, will hybridize to at least TSSU3-8 and NtSS23. Probe 4, as determined by Southern analysis, (data not shown) hybridizes exclusively to TSSU3-2. Probe 10, (also a 30 base oligomer) is complementary to sequences in the first exon of TSSU3-8, and because of conserved sequences in this region, should hybridize to transcripts of the entire SSU family. That TSSU3-2 is a transcriptionally active pseudogene is demonstrated by the detection of a transcript, albeit a low level one, by probe 4 (Figure 4, lane 4). Comparisons of relative levels of expression among TSSU3-8, NtSS23 and the rest of the SSU gene family is clouded somewhat by

Figure 4. Northern analysis of tobacco leaf polyA⁺ RNA using synthetic oligomers. Lane 1 = ³²P-λ Hind III, lanes 2-4 contain 1 μg of tobacco leaf polyA⁺ RNA. Probes used were: lane 2, probe10 (complementary to sequences in first exon of TSSU3-8); lane 3, probe 1 (complementary to sequences in the 5'-untranslated region of TSSU3-8 and NtSS23); and lane 4, probe 4 (complementary to sequences in the 5'-untranslated of TSSU3-2).

the lack of proven exclusiveness of probe 1 to TSSU3-8 and NtSS23. It can be said, however, that transcripts detected by probe 1 make up a major portion of the entire SSU gene family transcription, as determined by probe 10 (Figure 4, lanes 2 and 3). Conversely, transcription attributed to TSSU3-2 probably constitutes less than 1% of the SSU total (Figure 4, lanes 2 and 4).

Southern analysis

Southern analyses were performed on genomic DNA cut by *Eco*RI from tobacco cultivars 'Samsun' and 'Xanthi nc', as well as tobacco parental species *N. sylvestris* and *N. tomentosiformis*. All tobacco SSU genes characterized to date have an *Eco*RI site in the second exon (TSSU3-8, TSSU3-2, this paper; NtSS23³⁹). This gives a high probability of being able to distinguish between members of the gene family due to the variance of surrounding *Eco*RI sites. The totally different patterns obtained by probing the *Eco*RI cut tobacco DNA with probes from 5' (Figure 5A, lane 2) and 3' (Figure 5B, lane 4) regions argue that almost all tobacco

Figure 5. Southern analysis of small subunit genes using coding region probes. Panel A: Hybridization to a probe from the 3' coding regions of SSU genes. Lane 1 = ^{32}P - λ *HdIII* DNA, Lanes 2-5; 10 μg of genomic DNA restricted with *EcoRI*. The specific DNAs were: lane 2, *N. tabaccum* 'Samsun'; lane 3, *N. tabaccum* 'Xanthi'; lane 4, *N. sylvestris*; and lane 5, *N. tomentosiformis*. The probe (probe A) contains 185 bp from the 3' coding region of a *N. sylvestris* SSU cDNA clone, pSTV34. Panel B: Hybridization to a probe from the 5' coding regions of SSU genes. Lane 1, ^{32}P - λ *HdIII* DNA; lanes 2-5, lambda and genomic DNAs restricted with *EcoRI*. Specific DNAs were: lane 2, 250 pg of phage TSSU3-2; lane 3, 250 pg of phage TSSU3-8; lane 4, 10 μg of *N. tabaccum* 'Samsun' DNA; and lane 5, 10 μg of *N. sylvestris* DNA. The probe (probe B) contains 115 bp from the 5' coding region of TSSU3-8.

SSU genes are cleaved by *EcoRI* and support the use of this enzyme as a way of distinguishing the genes.

A variety of probes were used for hybridization. Probe D, possessing only 3' regions from *N. sylvestris*, is used in Figure 5A. In the tobacco cultivar 'Samsun' (lane 2), six distinct bands are detected, suggesting a SSU gene family of at least six members. In 'Xanthi nc' (lane 3), only five bands are detected, indicating one member of the gene family has been lost, or rearranged so as to be masked in this analysis. *N. sylvestris* and *N. tomentosiformis* both show four bands, having two bands in common, both smaller than 0.5 kb. As indicated by sequence analysis,

TSSU3-2 has a 3' *Eco*RI fragment slightly smaller than 0.5 kb, that is from the *N. sylvestris* parent. Due to the commonality of the sub 0.5 kb band, the parentage of TSSU3-2 cannot be substantiated by this analysis. Also, the presence of these small bands in *N. sylvestris*, *N. tomentosiformis*, and both tobacco cultivars would tend to indicate that several SSU genes possess 3' ends of this size. This suggests the possible divergence of the gene family from a few common ancestors, and makes estimates of gene copy number a minimum figure only. All of the bands in the tobacco cultivars can be traced to one of their diploid ancestors, with the exception of that band unique to the 'Samsun' cultivar. Its uniqueness suggests it may be a rearrangement of an ancestral gene. The largest band from *N. tomentosiformis*, approximately 6.0 kb, is the only band not also found in *N. tabacum*.

When probed with a 115 base fragment from the 5' coding region of TSSU3-8 (probe B), at least eight bands are indicated in *N. tabacum* 'Samsun' (Figure 5B, lane 4). *N. sylvestris* shows four bands (Figure 5B, lane 5). The pattern for *N. tabacum* 'Xanthi nc' is identical to the one obtained for *N. tabacum* 'Samsun' (data not shown). The parentage of TSSU3-8 is apparent in this blot since a band of the proper size (Figure 5B, lane 3) occurs in both *N. tabacum* 'Samsun' and *N. sylvestris*. A band of the appropriate size for NTSS23 (4.7kb)³⁹ is seen in *N. tabacum* 'Samsun' and *N. sylvestris*. TSSU3-2 however, is present in *N. tabacum* 'Samsun,' but no band of corresponding size is present in *N. sylvestris*. The possibility exists that since this gene was rendered a pseudogene, possibly after the amphidiploid *N. tabacum* was formed, other rearrangements were then allowed to occur.

Because of similarities in the sequence of the 5' upstream regions of TSSU3-8 and NtSS23, we probed 'Samsun' DNA with probes of various lengths from the 5' upstream region of TSSU3-8 (Figure 6). The 1.4 kb of untranslated region of probe C (Figure 6A) selects a multitude of bands and is therefore not exclusive to SSU genes. Probe D, having only the 485 bases immediately 5' of the coding region, selects far fewer bands (Figure 6B). The bands selected by probe D are a subset of those selected by the 115 base 5' coding region probe (probe B, Figure 5B, lane 4) indicating an exclusiveness of probe D to SSU genes. The 3.4 kb band here corresponds to 3-8 (lanes 2 and 4), while the 4.7 kb band corresponds to NtSS23³⁹. The other two bands (3.1 kb and .7 kb) are found with the 115 bp 5' coding region probe in *N. tabacum* 'Samsun' but not in *N. sylvestris*.

DISCUSSION

In this report, three clones for tobacco SSU genes have been isolated, and two of these have been sequenced and characterized. One, TSSU3-8, contains

Figure 6. Southern analysis of SSU genes using probes containing 5'-untranslated and 5'-untranscribed regions. Lanes 1, ^{32}P - λ *HdIII* DNA; lanes 2-4, lambda and genomic DNA restricted with *EcoRI*. The specific DNAs are: lane 2, 250 pg of phage TSSU3-8; lane 3, 250 pg of phage TSSU3-2; and lane 4, 10 μg of *N. tabaccum* 'Samsun'. In panel A, the probe (probe C) contains 73 bp of the 5'-untranslated and 1373 bp of the 5'-untranscribed region of TSSU3-8. In panel B, the probe (probe D) contains 73 bp of 5'-untranslated and 411 bp of the 5'-untranscribed region of TSSU3-8.

sequences highly homologous to those of an already published tobacco SSU gene, NtSS23. The degree of homology, 100% in coding region and intron, is so far unique within members of SSU gene families. Although high degrees of homology have been found among SSU genes from the same species (see for example pea⁴³ and soybean⁴⁴), none have been reported with this high degree of homology. The three tobacco SSU genes isolated here have not been shown to be in close proximity to each other on the chromosome, although it has been shown that 3 petunia SSU genes and some pea SSU genes^{14,15} can be mapped closely to each other. It is possible that NtSS23 and TSSU3-8 represent two SSU genes in the tobacco genome whose similarities could arise from gene duplication and/or gene conversion. Since NtSS23 and TSSU3-8 are present (by genomic Southern analysis) in both tobacco and *N. sylvestris*, it is possible that such a gene duplication event, if it occurred, took place before the hybridization event that gave

rise to tobacco. Tobacco is thought to have arisen relatively recently^{45,46}. As gene conversion between duplicated genes can make them appear more similar than would be predicted from the time elapsed since duplication^{47,48}, this lends credence to the idea that gene conversion may have also played a role in the similarities observed between NtSS23 and TSSU3-8. Gene conversion has been used to explain the identical coding regions but different flanking and non-coding sequences of two γ -interferon genes⁴⁹.

Although the similarity between NtSS23 and TSSU3-8 is less upstream of the ATG, there is still significant homology in this region. The 405 bp upstream of the ATG show 93% homology. Southern analysis of tobacco DNA with a probe covering this region hybridized to 4 out of 6 bands found with a probe covering the 5' coding region. Two of these four bands correspond to NtSS23 and TSSU3-8, and are present in *N. sylvestris*, whereas the other two are not. Sequences in this upstream non-coding region may well be important in maintaining levels of expression and light inducibility. A short conserved sequence surrounding the TATA box in pea SSU has been shown to confer light inducibility⁴². However, sequences further upstream of these sequences in pea SSU have been shown to be necessary for high levels of expression⁵⁰. A pea SSU fragment spanning -327 to -48 has been found to confer good levels of light inducibility⁵¹. The sequences in the 5' flanking region of these four tobacco SSU genes may have been conserved over time to allow effective control of expression. Both TSSU3-8 and NtSS23, which have these sequences, are highly expressed. The pseudogene, TSSU3-2, does not share significant homology in this region and is not highly expressed.

The multitude of bands seen on a genomic Southern with a larger (1.4 kb) 5' probe indicates that the tobacco SSU gene TSSU3-8 lies adjacent to a repeat which is present in many places in the genome. The role of this repeat cannot be ascertained from the data presented here.

The sequence of TSSU3-2 indicates it is a pseudogene; however Northern analysis and primer extension studies indicate it is transcribed. Pseudogenes have been found previously in multigene families; several examples of this can be seen in the globin gene family⁵². Several pseudogenes have also been found in plants. In petunia, two fragments have been found that contain only the 3' end of SSU genes¹⁴. Similarly in tomato, a fragment containing sequences homologous to the 3' end of CAB genes has been found⁵³.

Most pseudogenes isolated to date are not able to function transcriptionally and specifically in plants no transcriptionally active pseudogenes have been reported. However, transcriptionally active human interferon, *Xenopus* 5S and

goat globin pseudogenes have been found^{54,55,56,57}. Two of these are active only *in vitro* due to defective polyadenylation or termination regions^{56,57}. The fact that the pseudogene, TSSU3-2, can still function transcriptionally *in vivo* suggests that it has been a pseudogene for a relatively short time. It is possible that it became a pseudogene after the hybridization event that gave rise to tobacco as a species. The relatively large amount of nucleic acid and amino acid changes observed in TSSU3-2 is consistent with the fact that pseudogenes diverge more rapidly than non-pseudogenes^{58,59}. More information on the evolution of this gene will be obtained if SSU genes from *N. sylvestris* are cloned and sequenced.

ACKNOWLEDGMENTS

We wish to thank John Caton and Caty DeJesus for help with the sequence analysis. We are grateful to David Stalker and Robert Goodman for helpful discussion. We are also grateful to N-H. Chua for pSS15 and to J. Fleck for pSTV34.

*To whom correspondence should be addressed

REFERENCES

1. Ellis, R. J. (1979) Trends in Biochem. Sci. **4**, 241-244.
2. Jensen, R. G. and Bahr, J. T. (1977) Ann. Rev. Plant Physiol. **28**, 379-400.
3. Coen, D. M., Bedbrook, J. R., Bogorad, L. and Rich, A. (1977) Proc. Natl. Acad. Sci. **74**, 5487, 5491.
4. Kawashima, N. and Wildman, S. G. (1972) Biochim. Biophys. Acta **262**, 42-49.
5. Highfield, P. E. and Ellis, R. J. (1978) Nature **271**, 420-424.
6. Chua, N-H. and Schmidt, G. W. (1978) Proc. Natl. Acad. Sci. **75**, 6110-6114.
7. Mohr, H. (1977) Endeavor, New Series **1**, 107-114.
8. Tobin, E. M. (1978) Proc. Natl. Acad. Sci. **75**, 4749-4753.
9. Bedbrook, J. R., Smith, S. M. and Ellis, R. J. (1980) Nature **287**, 692-697.
10. Smith, S. M. and Ellis, R. J. (1981) J. Mol. Appl. Gen. **1**, 127-137.
11. Berry-Lowe, S. L., McKnight, T. D., Shah, D. M. and Meagher, R. B. (1982) J. Mol. Appl. Gen. **1**, 483-498.
12. Stiekema, W. J., Wimpee, C. F., Silverthorne, J. and Tobin, E. M. (1983) Plant Physiol. **72**, 717-724.
13. Coruzzi, G., Broglie, R., Cashmore, A. and Chua, N-H. (1983) J. Biol. Chem. **258**, 1399-1402.
14. Dean, C., Van den Elzen, P., Tamaki, S., Dunsmuir, P. and Bedbrook, J. (1985) Proc. Natl. Acad. Sci. **82**, 4964-4968.
15. Cashmore, A. R. (1983) In Kosuge, T., Meredith, C. P. and Hollaender, A., (eds), Genetic Engineering of Plants, an Agricultural Perspective, Plenum, New York, pp. 29-38.
16. Gray, J. C., Kung, S. D., Wildman, S. G. and Sheen, S. J. (1974) Nature **252**, 226-227.
17. Müller, K-D, Salnikow, J. and Vater, J. (1983) Biochim. Biophys. Acta **742**, 78-83.
18. Loenen, W. A. M. and Blattner, F. R. (1983) Gene **26**, 171-179.

19. Murray, N.E., Brammer, W.J. and Murray, K. (1977) *Molec. Gen. Genet.* **150**, 53-61.
20. Wood, W. B. (1966) *J. Mol. Biol.* **16**, 118-133.
21. Messing, J., Groenenborn, B., Muller-Hill, B. and Hofschneider, P. H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3642-3646.
22. Norrander, J., Kempe, T. and Messing, J. (1983) *Gene* **26**, 101-106.
23. Messing, J. (1983) In Wu, R., Grossman, L. and Moldave, K. (eds), *Methods in Enzymology: Recombinant DNA*, Academic Press, New York, New York, pp. 20-77.
24. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
25. Pinck, L., Fleck, J., Pinck, M., Hadidane, R. and Hirth, L. (1983) *FEBS Lett.* **154**, 145-148.
26. Benton, W.D. and Davis R.W. (1977) *Science* **196**, 180-182.
27. Shewmaker, C. K., Caton, J. R., Houck, C. M. and Gardner, R. C. (1985) *Virology* **140**, 281-288.
28. Dale, R. M. K., McClure, B. A. and Houchins, J. P. (1985) *Plasmid* **13**, 31-40.
29. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
30. Maxam, A. M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560-564.
31. Berent, S. L., Mahmoudi, M., Torczynski, R. M., Bragg, P. W. and Bollon, A. P. (1985) *BioTechniques* **3**, 208-220.
32. Taylor, B. and Powell, A (1982) *Focus* **4**, 4-6.
33. Colbert, J.T., Hershey, H.P. and Quail, P.H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2248-2252.
34. Facciotti, D., O'Neal, J. K., Lee, S. and Shewmaker, C. K. (1985) *Bio/Technology* **3**, 241-246.
35. Torczynski, R. M., Motohiro, F. and Bollon, A. P. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6451-6455.
36. Lee, D. C. and Luse, D. S. (1982) *Focus* **4**, 1-3.
37. Wimpee, C. F., Stiekema, W. J. and Tobin, E. M. (1983) In Goldberg, R.B. (ed), *Plant Molecular Biology*, UCLA Symposium on Molecular and Cellular Biology, New Series, Alan R. Liss Inc., New York, Vol. 12, pp. 391-401.
38. Polans, N. O., Weeden, N. F. and Thompson, W. F. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5083-5087.
39. Mazur, B. J. and Chui, C-F (1985) *Nuc. Acids Res.* **13**, 2373-2386.
40. Pichersky, E., Bernatzky, R., Tanksley, S. D. and Cashmore, A. R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3880-3884.
41. Perler, F., Efstratadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) *Cell* **20**, 555-566.
42. Moreli, G., Nagy, F., Fraley, R. T., Rogers, S. G. and Chua N-H. (1985) *Nature* **315**, 200-204.
43. Fluhr, R., Moses, P., Morelli, G., Coruzzi, G. and Chua, N-H. (1986a) *EMBO Journal* **5**, 2063-2071.
44. Grandbastien, M.A., Berry-Lowe, S., Shirley, B. W. and Meagher, R. B. (1986) *Plant Mol. Biol.* **7**, 451-465.
45. Janick, J., Schery, R.W., Woods, F.W. and Rutton, V.W. (1974) *Plant Science: An Introduction to World Crops*, W.H. Freeman and Co., San Francisco.
46. Gerstel, D.U. (1976) In Simmonds, N.W. (ed), *Evolution of Crop Plants*, Longman Inc., New York, pp. 273-277.
47. Slightom, J. L., Blechl, A. E. and Smithies, O. (1980) *Cell* **21**, 627-638.

48. Shen, S-H., Slighton, J. L. and Smithies, O. (1981) *Cell* **26**, 191-203.
49. Todokoro, K., Kiouisis, D. and Weissman, C. (1984) *EMBO Journal* **3**, 1809-1812.
50. Timko, M. P., Kausch, A. P., Castresana, C., Fassler, J., Herrera-Estrella, L., Van den Broeck, G., Van Montagu, M., Schell, J. and Cashmore, A. R. (1985) *Nature* **318**, 579-582.
51. Fluhr, R., Kuhlemeier, C., Nagy, F. and Chua, N-H. (1986b) *Science* **232**, 1106-1112.
52. Vanin, E. F., Goldberg, G. I., Tucker, P. W. and Smithies, O. (1980) *Nature* **286**, 222-226.
53. Pichersky, E., Bernatsky, R., Tanksely, S.D., Breidenbach, R.B., Kausch, A.P. and Cashmore, A.R. (1985) *Gene* **40**, 247-258.
54. Miyata, T. and Hayashida, H. (1982) *Nature* **295**, 165-168.
55. Goeddel, D.V., Leung, D.W., Dull, T.J., Gross, M., Lawn, R.M., McCandliss, R., Seeburg, P.H., Ullrich, A., Yelverton, E. and Gray, P.W. (1981) *Nature* **290**, 20-26.
56. Miller, J.R. and Melton, D.A. (1981) *Cell* **24**, 829-836.
57. Shapiro, S.G. and Lingrel, J.B. (1984) *Mol. Cell. Biol.* **4**, 2120-2127.
58. Miyata, T. and Yasunaga, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 450-453.
59. Miyata, T. and Hayashida, H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5739-5743.