Mol Gen Genet (1987) 209:247–256

MGG
© Springer-Verlag 1987

Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato

Mamoru Sugita[1,*], Thianda Manzara[1], Eran Pichersky[2], Anthony Cashmore[3], and Wilhelm Gruissem[1]

[1] Department of Botany, University of California, Berkeley, CA 94720, USA
[2] Department of Biology, University of Michigan, Ann Harbor, MI 48109, USA
[3] Department of Biology, University of Pennsylvania, Philadelphia, PA 19134, USA

Summary. We have cloned and sequenced all five members of the gene family for the small subunit (rbcS) of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato, *Lycopersicon esculentum* cv. VFNT LA 1221 cherry line. Two of the five genes, designated *Rbcs-1* and *Rbcs-2*, are present as single genes at individual loci. Three genes, designated *Rbcs-3A, Rbcs-3B* and *Rbcs-3C,* are organized in a tandem array within 10 kb at a third independent locus. The *Rbcs-2* gene contains three introns; all the other members of the tomato gene family contain two introns. The coding sequence of *Rbcs-1* differs by 14.0% from that of *Rbcs-2* and by 13.3% from that of *Rbcs-3* genes. *Rbcs-2* shows 10.4% divergence from *Rbcs-3.* The exon and intron sequences of *Rbcs-3A* are identical to those of *Rbcs-3C,* and differ by 1.9% from those of *Rbcs-3B.* Nucleotide sequence analysis suggests that the five rbcS genes encode four different precursors, and three different mature polypeptides. S_1 nuclease mapping of the 5′ end of rbcS mRNAs revealed that the mRNA leader sequences vary in length from 8 to 75 nucleotides. Northern analysis using gene-specific oligonucleotide probes from the 3′ non-coding region of each gene reveals a four to five-fold difference among the five genes in maximal steady-state mRNA levels in leaves.

Key words: Tomato – Nucleotide sequence – Ribulose-1,5-bisphosphate carboxylase – Multigene family – Differential expression

Introduction

All components of the photosynthetic apparatus in higher plants are composed of proteins encoded by the nuclear or chloroplast genomes. Of these, the stromal enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase is a widely studied example for the coordinate regulation of nuclear and plastid genes during chloroplast development and assembly of photosynthetic proteins (Miziorko and Lorimer 1983). The enzyme consists of eight large subunits of 55000 M_r and eight small subunits of 14000 M_r (Blair and Ellis 1973). The expression of the gene for the large subunit (rbcL), located in the large single copy region of the chloroplast genome (Bedbrook et al. 1979), has been analyzed at the transcriptional, post-transcriptional and translational levels in several plants (Berry et al. 1985; Inamine et al. 1985). The gene for the small subunit (rbcS) is present as a multigene family in the nuclear genome of most higher plants (Berry-Lowe et al. 1983). The DNA sequence of at least some of the rbcS genes has been determined in pea (Timko et al. 1985; Fluhr et al. 1986b), soybean (Berry-Lowe et al. 1982), petunia (Dean et al. 1985; Tumer et al. 1986), wheat (Broglie et al. 1983), tobacco (Mazur and Chui 1985), duckweed (Stiekema et al. 1983) and tomato (Pichersky et al. 1986). The expression of the rbcS genes appears to be controlled primarily by light during chloroplast development (Tobin and Silverthorne 1985), but tissue-specific and developmental regulation of this gene family has also been reported (Dean et al. 1985; Fluhr et al. 1986b; Sheen and Bogorad 1986).

We have initiated a systematic study of regulatory mechanisms that control the expression of nuclear and plastid genes in different organs and during plastid differentiation in tomato (*Lycopersicon esculentum*) (Piechulla et al. 1985, 1986; Gruissem et al. 1987). As a first step in an attempt to elucidate the DNA regions and *trans*-acting elements involved in the regulation of rbcS expression, we have cloned and sequenced the complete tomato rbcS gene family. We present here the organization and DNA sequence of all five tomato rbcS genes. In addition, we have used the DNA sequence information to construct gene-specific synthetic oligonucleotides. The gene-specific probes were employed to estimate the relative amounts of mRNA for each of the rbcS genes in tomato leaves.

Materials and methods

Plant material. Seeds of *Lycopersicon esculentum* cv VFNT LA1221 cherry line were germinated in soil; seedlings grown hydroponically for 3 weeks were used as root and leaf material.

Isolation of tomato DNA and RNA. High molecular weight DNA was isolated from frozen root tissue essentially according to the procedure of Bendich et al. (1979). Total RNA was prepared as described by Piechulla et al. (1986) from leaves harvested between 10 a.m. and 12 noon.

Construction of tomato genomic libraries. The libraries were constructed following the general procedure of Maniatis et al. (1982). Two different tomato genomic libraries were

* *Present address:* Department of Botany, Hokkaido University, Sapporo 060, Japan

Offprint requests to: W. Gruissem

constructed to ensure isolation of all tomato rbcS genes. For the first library, high molecular weight DNA was partially digested with *Sau*3A and size-fractionated through a 10%–40% sucrose gradient. DNA fragments larger than 15 kb (0.4 μg) were ligated to lambda EMBL3–*Bam*HI digested arms (Frischauf et al. 1983), packaged in vitro using a packaging extract (Gigapack, Vector Cloning Systems) and amplified in *E. coli* LE 392. The yield was 1.5×10^6 phage per μg insert DNA, and the total size of the library was 9.4×10^5 phages, this represents a nearly complete genomic library, as the genome size in tomato is 3.54×10^9 bp (Bennett and Smith 1976). For the second library, tomato genomic DNA was digested to completion with *Hin*dIII, and DNA fragments larger than 10 kb were size-fractionated. To prepare arms of lambda Charon 35 (Loenen et al. 1983), Charon 35 DNA was digested sequentially with *Hin*dIII and *Bam*HI. The arms were annealed and separated from the stuffer DNA on a sucrose gradient. Size-fractionated tomato DNA fragments (0.4 μg) hybridizing to the tomato rbcS cDNA probe were cloned using the Charon 35 arms (1 μg) and a packaging extract (Gigapack). When approximately 18 kb DNA fragments were used for cloning, 7.2×10^5 phages were obtained per μg insert DNA. Efficiencies were lower with insert DNA fragments larger or shorter than 18 kb.

Isolation of clones containing the rbcS gene from the tomato libraries. The libraries were screened by plaque hybridization according to the procedure of Maniatis et al. (1982). The probe used was a 0.7 kb *Pst*I fragment containing the entire coding sequence of the tomato rbcS cDNA clone 3–91 (Bernatzky and Tanksley 1986). Filters were incubated at 42° C for at least 1 h in prewashing solution (50 mM Tris-HCl pH 8, 1 M NaCl, 1 mM EDTA, and 0.1% SDS), and incubated at 65° C for 3–4h in 5 × SSPE (1 × SSPE = 0.75 M NaCl, 50 mM NaH_2PO_4, pH 7.4 and 5 mM EDTA), 5 × Denhardt's (1 × Denhardt's = 0.1% each of Ficoll, polyvinylpyrrolidone and BSA), 0.1% SDS and 100 μg/ml denatured salmon sperm DNA. Hybridization was carried out at 65°–68° C for 20 h. The filters were washed three times in 2 × SSC (1 × SSC = 0.15 M NaCl and 0.015 M sodium citrate, pH 7) and 0.1% SDS at room temperature for 10 min each, and then washed twice for 1 h in 1 × SSC and 0.1% SDS at 65°–68° C. Positive plaques were picked, replated and rehybridized until a single phage had been isolated. Phages were isolated using CsCl discontinuous gradients, and phage DNAs were purified by phenol extraction (Maniatis et al. 1982).

Subcloning and sequencing of fragments containing the rbcS gene. DNA fragments containing the rbcS gene were subcloned into phage M13mp18/19, pUC18/19 and pUC118/119 (Vieira and Messing 1987). DNA fragments used for sequencing were as follows: 1.5 kb *Sau*3A fragment (*Rbcs-1*); 1.2 kb *Sau*3A, 3.8 kb *Hin*dIII-*Sau*3A and 2 kb *Eco*RI fragments (*Rbcs-2*); 1.3 kb *Sau*3A and 1.55 kb *Hin*dIII–*Eco*RI fragments (*Rbcs-3A*); 2.9 kb *Sau*3A–*Hin*dIII fragment (*Rbcs-3B*); 1.75 kb *Sau*3A fragment (*Rbcs-3C*). The templates for DNA sequencing were prepared essentially as described by Messing (1983). DNA sequencing was performed using the dideoxy chain termination method and ^{35}S-dATP (Biggin et al. 1983). Synthetic oligonucleotide primers for sequencing were 5′CAAGCTAGCATGGT, 5′GGATGGGTTCCTTGCTTGG, 5′GGGTGCACTGATGCAAC and their complementary sequences. These primers are located in the protein coding regions. Oligonucleotides 5′GTAGTTACTGACCAC, 5′GAGTTCATCAATTCTAG and 5′GAACATCACTAGTC were used for sequencing the introns of *Rbcs-2*. Both strands of DNA fragments containing the rbcS gene were sequenced. Synthetic oligonucleotides were prepared as described previously (Gruissem and Zurawski 1985).

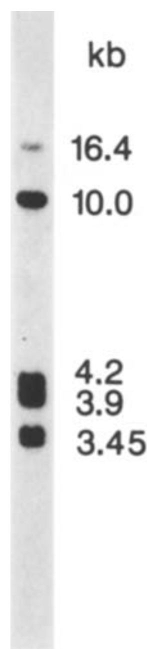

Fig. 1. Southern blot analysis of tomato genomic DNA. Tomato nuclear DNA (10 μg) was digested to completion with *Hin*dIII, separated on 0.8% agarose gel and transferred to a nylon membrane filter. The filter was hybridized with the tomato rbcS cDNA probe at 65° C for 20 h in 5 × SSPE, 5 × Denhardt's, 0.1% SDS and 100 μg/ml denatured salmon sperm DNA. The final wash was at 65° C in 1 × SSC, 0.1% SDS

Southern blot analysis. For the genomic Southern blot analysis high molecular weight DNA (10 μg) was digested to completion with *Eco*RI, *Eco*RV and *Hin*dIII, and separated on a 0.8% agarose gel. The gel was incubated in 0.25 M HCl for 12 min at room temperature. Denaturation and neutralization of the gel, and transfer of DNA to a nylon membrane filter (Hybond-N, Amersham) were according to Southern (1975). The filter was prehybridized in 5 × SSPE, 5 × Denhardt's and 0.1% SDS at 65° C for 1 h, then hybridized for 20 h before washing twice with 2 × SSC at 65° C for 20 min, and then with 2 × SSC and 0.1% SDS for 30–60 min. For the analysis of cloned DNA fragments, the same procedure as described above was used except that incubation of the gel in 0.25 M HCl was omitted. The tomato rbcS cDNA clone and a 5′-specific DNA fragment, the 362 bp *Pvu*II – *Eco*RI from phage 20B (Pichersky et al. 1986), as well as five gene-specific oligonucleotides, were used as hybridization probes.

5′ end S_1 mapping rbcS mRNA. To prepare gene-specific fragments for S_1 mapping, the coding region primer, 5′ACCATGCTAGCTTG (+55 to +68) was labeled with ^{32}P-ATP using T4 polynucleotide kinase (Maniatis et al.

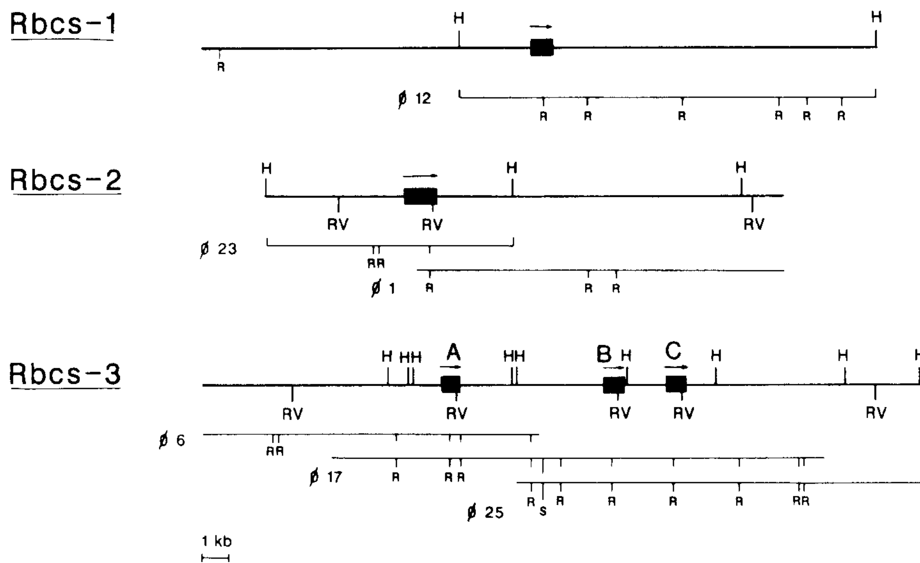

Fig. 2. Restriction maps of the tomato rbcS genomic clones. *Thick lines* represent the region of cloned genomic DNA, and *thin lines* indicate the insert of the respective recombinant phage. Restriction sites are indicated as *H* (*Hin*dIII), *R* (*Eco*RI), *RV* (*Eco*RV) and *S* (*Sal*I). *Rbcs-1, Rbcs-2* and *Rbcs-3* indicate the loci of the rbcS genes. The three *Rbcs-3* genes are indicated as *A, B* and *C*. *Horizontal arrows* indicate the transcription direction of each gene

1982). The end-labeled primer was annealed to the recombinant M13 clones containing the non-coding strands of the respective tomato rbcS gene and then incubated at 37° C for 30 min in 30 mM Tris-HCl, pH 8.3, 36 mM NaCl, 5 mM DTT, 6 mM MgOAc, 0.08 mM each of dATP, dCTP, dGTP, dTTP, and with AMV reverse transcriptase (0.06 units, Life Sciences) and Klenow fragment (0.43 units, BRL) according to the procedure of Zaug et al. (1984). After primer extension, the reaction mixture was ethanol-precipitated and digested with restriction enzymes, *Hae*III (*RbcS-1* and *RbcS-3C*), *Fok*I (*Rbcs-2*), *Sau*3A (*Rbcs-3A*) or *Dde*I (*RbcS-3B*). The 5′-end-labeled coding strands were separated by electrophoresis on 5% polyacrylamide gels (Maxam and Gilbert 1977). S_1 mapping was carried out as described previously (Sugita and Sugiura 1984). The 5′-end-labeled coding strands were hybridized with total tomato leaf RNA (20 μg) at 50° C or 52° C for 6 h in 15 μl 70% formamide, 0.04 M PIPES-NaOH, pH 6.7, 0.5 M NaCl and 1 mM EDTA. The hybridization mixture was then diluted with 150 μl 30 mM Na-acetate, pH 4.6, 1 mM $ZnSO_4$, 0.25 M NaCl and 20 μg/ml denatured salmon sperm DNA and digested with 700 units/ml S_1 nuclease at 24° C for 60 min (*Rbcs-1*) or at 37° C for 30 min (other genes). The S_1-protected fragments were ethanol-precipitated with 5 μg carrier yeast tRNA and analyzed on 6% sequencing gels in parallel with the sequencing ladder of the coding strand primer-extended with ^{32}P-dATP and Klenow fragment.

Northern hybridization. RbcS gene-specific oligonucleotide probes (20 mers) were labeled with ^{32}P-ATP using T4 polynucleotide kinase. Unlabeled primer was added to adjust the specific activity. Total leaf RNA was electrophoresed in a 1.2% agarose gel containing formaldehyde (Maniatis et al. 1982), transferred to a nylon filter (Hybond, Amersham), and hybridized with the 5′ end-labeled gene specific probe at 10° C below the T_m (melting temperature) of each probe (Wallace et al. 1979) for 20 h in 5 × SSPE, 5 × Denhardt's, 0.1% SDS and 100 μg/ml denatured salmon sperm DNA. After hybridization, the filters were washed twice at room temperature for 15 min in 6 × SSC, and then at 5° C below the T_m of each probe for 30 min.

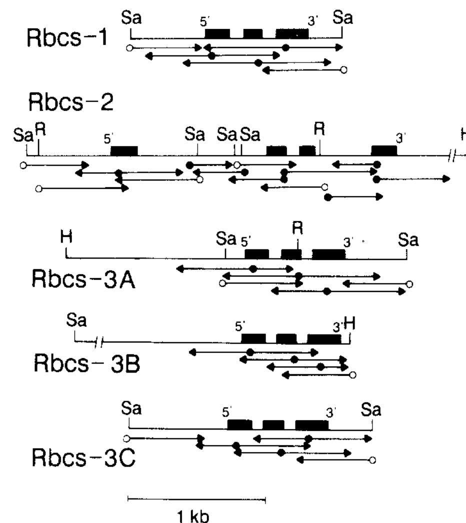

Fig. 3. DNA fragments containing the tomato rbcS gene and sequencing strategies. Restriction sites used for subcloning are indicated as *H* (*Hin*dIII), *R* (*Eco*RI) and *Sa* (*Sau*3A) (see Materials and methods). The protein-coding (exon) regions are indicated by *filled boxes* and their ends are marked 5′ and 3′. The *closed* and *open circles* indicate the positions of the synthetic oligonucleotide primers and M13 primers, respectively. The *arrows* indicate the direction and extent of sequencing

Results

Southern blots of tomato genomic DNA

Tomato nuclear DNA was digested to completion with *Eco*RI, *Eco*RV and *Hin*dIII, and subjected to Southern blot hybridization with the insert DNA fragment from tomato rbcS cDNA clone 3–91 (Bernatzky and Tanksley 1986). Five discrete *Hin*dIII fragments from 3.45 to 16.4 kb hybridized to the rbcS probe (Fig. 1). In addition, at least six *Eco*RI fragments, ranging from 1.7 to 13 kb, and seven *Eco*RV fragments, ranging from 2.45 to 20 kb, gave a positive hybridization signal (data not shown).

Rbcs-1

AGGGCGTAAAATATTTTTTTAAAAAAGGCACTTAGCTCCAATTTCTTACCTTTCATGTGGCCATTAAACTTTGTAATATATCAACAACCAAATCCAATGGTCGCCCTTCATCTAAGATGAGGCTTCTTTTGTTTCTATCC -141

GTTAGATTTTAAAAAACGTCTAAAACCTTATCATTATATAAAGGGACGATACCACTGTGTAATAAGCATCTTTAAAAAAAATATAGTTTCTTTGAAATTAAAAAAAAAAAAAACATTTTATAGCTAAGTAAGTAAACGCA -1

M A S S I V S S A A A A T R S N V A Q A S M V A P F T G L K S A A S F

ATG GCT TCC TCA ATT GTC TCA TCG GCA GCC GCT GCT ACC CGT AGC AAT GTT GCT CAA GCT AGC ATG GTC GCA CCC TTC ACC GGA CTC AAA TCC GCC GCT TCT TTT 105

P V T K K N N N V D I T S I A S N G G R V R C M Q

CCC GTT ACC AAG AAG AAC AAC AAC GTT GAC ATT ACC TCC CTT GCT AGC AAT GGT GGA CGC GTT AGA TGC ATG CAG GTTATTCAAAAATAAGTGATTCAATTATCAGTATG 215

V W P P I N M K K Y E T L S Y L P D L

TATATATAATTTAAATTCAACGCGATTATTTTATAATCGTGGAATGTTTGACTAATGTAG GTG TGG CCA CCA ATC AAC ATG AAG AAA TAC GAG ACA TTG TCA TAC CTT CCT GAC TTG 332

S D E Q L L S E I E Y L L K N G W V P C L E F E T E

TCC GAT GAG CAA TTG CTT AGC GAA ATT GAG TAT CTT TTG AAA AAT GGA TGG GTT CCT TGC TTG GAA TTC GAG ACT GAG GTCAATATCTCGTATTTTATTTAATGATTTAT 442

R G F V Y R E N N S S P G Y Y D G R

CTAATTATTTGAAGTTATAAAAATAATTTGAACTGTCGTTTGTTTGTAATATATGTATGTAACAG CGC GGA TTT GTG TAC CGA GAG AAC AAC AGT TCC CCT GGA TAC TAC GAT GGT AGA 561

Y W T M W K L P M F G C T D A T Q V L A E V Q E A K K A Y P Q A W V R

TAC TGG ACC ATG TGG AAG TTA CCT ATG TTT GGG TGC ACT GAT GCA ACA CAG GTG TTG GCT GAG GTT CAA GAG GCG AAG AAG GCG TAC CCA CAA GCC TGG GTT CGT 666

I I G F D N V R Q V Q C I S F I A Y K P E G F ·

ATT ATC GGA TTC GAC AAC GTT CGT CAA GTC CAG TGC ATC AGC TTC ATT GCC TAC AAG CCA GAA GGA TTC TAA ATTT(CCAAACATATAATCAAAATG)TTTGTTTGATTTGTAC 778

CTAAGTCTTTAATTTTTTTTCTTCAATTTTTGTTTTGCATTTTGTTCCTATTTGTACCGGTGTATTTTCTTCGATTCCGACCAAGTTATGAGAAATTAATAATGATGAATCGGTGGTTATATGTTTAATCATGTGTTCTT 918

TTTTTAAAAAAATTTGTGTGGTGTGAATGAGTTTTATTATCCAAAATAAAAATCTCACTTTAAATATGAAAATGACATATGAACTACTTTTGTCCGGGCAACTTTAGCAGAGATTAACCATAGATATATGAGATATATAAT 1058

Rbcs-2

TTACATAAAGTTGAGGACAATATAAGCAAGCAAACAAGTACTCTAGCTATCAATTTACTTTGGACTACTATATGATAAATATTTCAACCCCCCTCCTCCCCCACCAAAAAAAAATTAAAATAACAAGTTGAAGGACTCAA -281

AAAAAAAAAAAAAAAAAACTCAAAACCAACCTCAATCATACATTCATATCCTCTTCCTACCCCCATCTTGGATGAGATAAGATTAACGAGGTGCTTACACGTGTCACCTCTATTGTGGTGACTTAAAAAAAATTCCAACC -141

TTTCATATGTAGATATTAAGTAATTGTATAATGTTATCAAGAACCACATAACATATCAAAAACCTTATCATTTCATTATATAAAAGGATAGTGGACATCAAAAGGTTCATATTGAACCAAAAAAAGAGAGAAGAAGCAAT -1

M A S S V I S S A A V A T R S N V T Q A S M V A P F T G L K S T A S F

ATG GCT TCC TCT GTC ATT TCT TCA GCA GCT GTT GCC ACA CGC AGC AAT GTT ACA CAA GCT AGC ATG GTT GCA CCT TTC ACT GGT CTC AAA TCT TCA GCC ACT TTC 105

P V T K K Q N L D I T S I A S N G G R V S C▼M Q

CCT GTT ACA AAG AAG CAA AAC CTT GAC ATC ACT TCC ATT GCT AGC AAT GGT GGA AGA GTT AGC TGC ATG CAG GTAACTCACTCATCGGTTTTCCGAAATTGATATTTTCTTAGT 219

ATGTTGTAACATGTAATCTGTCTTATTTCTCGATTCAATTATCTAAGGGTTATATGTATGTTAGCTACTCTCTCTGGCTTAATTTATGTGACGGTAGTACTATTTTATTTTAGTTAGTATAAAAAAGAGTGATTTACTTT 359

GTTTCAGTTTGTTTGTTTAATCTTAATTTAACATAACTTTTAAGAAAATTAAAAAGTCTTTTGAATTTGTAGTTTTAAATTAAAGTTACGTAGAATGTATTAAAATATCCTACATTTTGTGATTTCAAACATGTTAACTG 499

TCACGTGAAAAATTAGAATTAGATAATTATCAAAAAAAAAAAAAGAGACTTTAGAAGAGTAATATAAACAAATTGAAACATGAGTATATGTAGTTACTGACCTCTGATCTAGTTTATGCCATTATTATTTTTCTTTTTAA 639

GTCTGTTTTAAAAGAATGACATATTTTTATATTTAATAGCAGTATTGATTTAAGTTTGAAAAAAATTTCTATTTTATCTTTAATGAGATGATTTATAATTACATAAGTATTTATATTATCTTAGACCATAAACTACGAAA 779

ATCTCTGTTTATTCTTAAAATTCCATTTACCATTAAGTACCTCCACATAAGAATTAAAAAAAGAGAGTATCTTTCAAATTATCTGATAATTCATAGAAGTTAAAAGATCCAATTGCTTAACTTATCTTTGATTATATTAA 919

ACACATGAAAAGTGTATCCTACGTATAATCTCTACGTTCAAAATTAAAGACAAATTTAAGATCTAGAATTGATGAACTCAAAATATCTCTATCTTATTATTATTATGTATATATACAACTTAATTTTTCGAAATGTTGAT 1059

V W P

GAATGTATGTTAAATCCTTTACAATAGTGCATTTTTAGATAATTCGATAAGAGTGTAACAATATTTTCAGAAAGTCGATAACTTTTTTTTCTTTTTTTTTTTTGGATATGTGTATAG GTG TGG CCA 1185

P I N M K K Y E T L S Y L P D L S D E Q L L S E I E Y L L K N G W V P

CCA ATT AAC ATG AAG AAG TAC GAG ACA CTC TCA TAC CTT CCT GAT TTG TCT GAC GAG CAA TTG CTT AGT GAA ATT GAG TAC CTT TTG AAA AAT GGA TGG GTT CCT 1290

C L E F E T E H G

TGC TTG GAA TTT GAG ACT GAG GTCAATATAATATTTACTTTTGTTTTATCGAAAAATGATATATTTTTATATTTAGATACGATTTATATTTAACACTTCAATGTTTTTTTTATGTGTAG CAC GGA 1415

F V Y R E N N K S P G Y Y D G R

TTT GTC TAC CGT GAG AAC AAC AAG TCA CCA GGA TAC TAT GAT GGA AG GTCAGTAGCGGATTCATAAGTTAAAGAATTCATTCATACTTTTTTTTAGTTAAATCGTAGGTATAGGTACGAGA 1536

ATCATGAAAATTAATATAAATATCATGTTTCGAACTCGTTTATTAAAAAAAAAAAAACTCATAAAACATACTATGAACATCACTAGTCTGTATAAAAAGGGAAAAAAAAGAAGATAAATATACCCTTAGAATATCATAAGT 1676

GGTATGCAGATATCTTCGGTCATAGTTTTGGGACATTGGTACTTCTGCCGTCCAAAAACTAGAACGTTGATATATATGTTGTAGTTTTTGGACCGCAGAAGCATCAATGTCCCAAAAATATGATGAAAAATATCTGTATA 1816

Y W T M W K L P M F G C T

CTATTTACGATAGTTCGGATATATTTGTCCCTTTTTTCTAAGTTTTATAAAGTTAAAAAGTTAAAATTTTGAATCCGCAG G TAC TGG ACC ATG TGG AAG TTG CCT ATG TTT GGG TGC ACT 1937

D A T Q V L A E V Q E A K K A Y P Q A W V R I I G F D N V R Q V Q C I

GAT GCA ACC CAA GTG TTG GCT GAG GTT CAA GAG GCT AAA AAG GCA TAC CCA CAA GCA TGG GTC AGA ATC ATT GGA TTC GAC AAT GTG CGT CAA GTG CAG TGT ATC 2042

S F I A Y K P E G Y

AGT TTC ATT GCC TAC AAG CCA GAA GGC TAC TAA GTT(TCATATTAGGAACAAAATTG)TCTTTAGGGACACTTTGTTTTAAATGCTACTTAGGTCTTTTTTCTTTTTTGTCTTAATTCCAACAAA 2166

CTCTTTGTGTCTTGTACTATTCGGTTTATGTTTTGGATTTATGAGTACCTAATTATATGATAATGATTTGGTGCTTTGTTTGTAAATTTTGATTCTTGTGGTTTATATGACTTTTTGTGTTTTATAACATCTTCAATCCT 2306

Rbcs-3A

TCTTGTGGCTTAATTAATATATCTAATTATTATTATAGCTCACCCACCCTCCACGCCCAAATTAATGTCATTAAGATGGAGTTATAATTCTACTTAATAGATTCGATAAAATTCTACTTTTGAAATGTGAACAAGGGCAT -141

GATCCAATGGTTACAAATGGGTTGGTTAATTTGTGTCGTTAGATGGGAAAGTTAAGTGAAACCTTATCATTATATATAGAGGGAGAGACTAGAAAGCAATAACCCTCTTGAGTTCAAGATAAGCACTTGGTTTTCAGCA -1

M A S S V M S S A A V A T R G N G A Q A S M V A P F T G L K S T A S F

ATG GCT TCT TCA GTA ATG TCC TCA GCA GCT GTT GCC ACC CGC GGC AAT GGT GCA CAA GCT AGC ATG GTT GCA CCC TTC ACT GGA CTC AAG TCC ACC GCT TCT TTC 105

P V S R K Q N L D I T S I A S N G G R V S C▼M Q

CCT GTT TCA AGG AAG CAA AAC CTT GAC ATT ACC TCC ATT GCT AGC AAC GGT GGA AGA GTC AGT TGC ATG CAG GTTTGTGTGTGTATATATATATACGTACAACAAAATTCATTGA 220

V W P P I N M K K Y E T L S Y L P D L S D E

CTATAATGTTATACTCGATTAGCTAATTTAACTATTTATAATTGTATAG GTG TGG CCA CCA ATT AAC ATG AAG AAG TAC GAG ACT CTG TCG TAC CTT CCT GAT TTG TCC GAC GAG 335

Q L L S E I E Y L L K N G W V P C L E F E T E

CAA TTG CTC AGC GAA ATT GAG TAC CTA TTG AAA AAT GGA TGG GTT CCT TGC TTG GAA TTC GAG ACT GAG GTCAACATCTATCTCCTCTGTTTTTAAAAATTTACTAGCTAGTATGT 450

H G F V Y R E N H K S P G Y Y D G R Y W T M W K

TGATATGTCGTGTTAACAGTGTTGTGGGATATCATGTGCAG CAC GGA TTT GTG TAC CGT GAG AAC CAT AAG TCA CCA GGA TAC TAC GAT GGC AGA TAC TGG ACC ATG TGG AAG 563

```
 L   P   M   F   G   C   T   D   A   T   Q   V   L   A   E   V   Q   E   A   K   K   A   Y   P   Q   A   W   V   R   I   I   G   F   D   N
TTG CCC ATG TTC GGG TGC ACT GAT GCA ACC CAG GTC TTG GCT GAG GTG CAG GAG GCA AAG AAG GCT TAC CCA CAG GCA TGG GTC CGT ATC ATC GGA TTC GAC AAT    668

 V   R   Q   V   Q   C   I   S   F   I   A   Y   K   P   E   G   Y
GTT CGT CAA GTG CAG TGC ATC AGT TTC ATC GCT TAC AAG CCC GAA GGA TAC TAA  (GTTCTAGAAAAACTAATTGC)CCTATGTTTATAGGGACAGTTTGTTTGAATTCTCCTTTGGGTTT    786

TCCCTTGAGAAATTGTATTCATTTTCCCTTGTTTTCCATCTTTTTATTCCGTGTATGATTAATTCGAATTTCAATCAAGTTTACAAGAACTATATTAATGATATTTGTTTCTTTACTAATCTCTTTCTGAAAAATTGCTT    926
AATTTGTGGTTTATTTATATAGCTAGTGAAGAAACCTTACCCAGAACACGAACCCAGGTTCGTGTAGTAAAGATACAAACACTTATTGAATTAAAAACCTTCCTCACTCAAGGAAGGAAAAACCTTGTTTTATTAATTCA  1066
```

Rbcs-3B

```
AGAAATTGTCAAAGACACATACCTCTATGAGTTTTTTCATCAATTTTTTTTTCTTTTTAAACTGTATTTTTAAAAAAATATTGAATAAAACATGTCCTATTCATTAGTTTGGGAACTTTAAGATAAGGAGTGTGTAATT  -141
TCAGAGGCTATTAATTTTGAAATGTCAAGAGCCACATAATCCAATGGTTATGGTTGCTCTTAGATGAGGTTATTGCTTTAGGTGAAACCTTATCATCATTATATATATACAAGGGGATACTAGAGACCAATTATTGTCAACA    -1

 M   A   S   S   I   V   S   S   A   A   V   A   T   R   G   N   G   A   Q   A   S   M   V   A   P   F   T   G   L   K   S   T   A   S   F
ATG GCT TCC TCT ATA GTT TCT TCA GCT GCT GTT GCC ACC CGC GGC AAT GGT GCA CAA GCT AGC ATG GTT GCA CCC TTC ACT GGA CTC AAG TCC ACT GCT TCT TTC    105

 P   V   S   R   K   Q   N   L   D   I   T   S   I   A   S   N   G   G   R   V   S   C ▼ M   Q
CCT GTT TCA AGG AAG CAA AAC CTT GAC ATT ACC TCC ATT GCT AGC AAC GGT GGA AGA GTC AGT TGC ATG CAG   GTTTGTGTGTGTATATATATATATATGTACAACAAAATTCATTGA  220

                                                   V   W   P   P   I   N   M   K   K   Y   E   T   L   S   Y   L   P   D   L   S   D   E
CTATAATGCTATACTCGATTAGCTAATTTAACTATTTATAATTGTATAG  GTG TGG CCA CCA ATT AAC ATG AAG AAG TAC GAG ACT CTG TCG TAC CTT CCT GAT TTG TCC GAC GAG    335

 Q   L   L   S   E   I   E   Y   L   L   K   N   G   W   V   P   C   L   E   F   E   T   E
CAA TTG CTC AGC GAA ATT GAG TAC CTA TTG AAA AAT GGA TGG GTT CCT TGC TTG GAA TTC GAG ACT GAG     GTTAACCTCTGTTTTTAAAATTTATATATTAGCTAGTGTGT    445

                                              H   G   F   V   Y   R   E   N   H   K   S   P   G   Y   Y   D   G   R   Y   W   T   M   W   K
TGATATGTTGTGTTAATAGTGTTGTGGGATATCATGTGCAG    CAC GGA TTT GTG TAC CGT GAG AAC CAT AAG TCA CCA GGA TAC TAC GAT GGC AGA TAC TGG ACC ATG TGG AAG    558

 L   P   M   F   G   C   T   D   A   T   Q   V   L   A   E   V   Q   E   A   K   K   A   Y   P   Q   A   W   V   R   I   I   G   F   D   N
TTG CCC ATG TTT GGG TGC ACT GAT GCA ACC CAG GTC TTG GCT GAG GTG CAG GAG GCA AAG AAG GCT TAC CCT CAG GCA TGG GTC CGT ATC ATC GGA TTC GAC AAT    663

 V   R   Q   V   Q   C   I   S   F   I   A   Y   K   P   E   G   Y
GTT CGT CAA GTG CAG TGC ATC AGT TTC ATC GCT TAC AAG CCC GAA GGA TAC TAA     AATGCCAATTTTTAA(TTATGTAATGTATAACTGAC)CCTATGTATTTAGGGGAAGCTT    774
```

Rbcs-3C

```
TCAGAAAGAAAAAGCAAAAGCAACAAAAAAATTGCAAGTATTTTTTAAAAAAGAAAAAAAAAACATATCTTGTTTGTCAGTATGGAAGTTTGAGATAAGGACGAGTGAGGGGTAAAATTCAGTGGCCATTGATTTTGTA  -141
ATGCCAAGAACCACAAAATCCAATGGTTACCATTCCTGTAAGATGAGGTTTGCTAACTCTTTTTGTCCGTTAGATAGGAAGCCTTATCACTATATATATACAAGGCGTCCTAATAACCTCTTAGTAACCAATTATTTCAGCA    -1

 M   A   S   S   V   M   S   S   A   A   V   A   T   R   G   N   G   A   Q   A   S   M   V   A   P   F   T   G   L   K   S   T   A   S   F
ATG GCT TCT TCA GTA ATG TCC TCA GCA GCT GTT GCC ACC CGC GGC AAT GGT GCA CAA GCT AGC ATG GTT GCA CCC TTC ACT GGA CTC AAG TCC ACC GCT TCT TTC    105

 P   V   S   R   K   Q   N   L   D   I   T   S   I   A   S   N   G   G   R   V   S   C ▼ M   Q
CCT GTT TCA AGG AAG CAA AAC CTT GAC ATT ACC TCC ATT GCT AGC AAC GGT GGA AGA GTC AGT TGC ATG CAG   GTTTGTGTGTGTATATATATATATACGTACAACAAAATTCATTGA  220

                                                     V   W   P   P   I   N   M   K   K   Y   E   T   L   S   Y   L   P   D   L   S   D   E
CTATAATGTTATACTCGATTAGCTAATTTAACTATTTATAATTGTATAG   GTG TGG CCA CCA ATT AAC ATG AAG AAG TAC GAG ACT CTG TCG TAC CTT CCT GAT TTG TCC GAC GAG  335

 Q   L   L   S   E   I   E   Y   L   L   K   N   G   W   V   P   C   L   E   F   E   T   E
CAA TTG CTC AGC GAA ATT GAG TAC CTA TTG AAA AAT GGA TGG GTT CCT TGC TTG GAA TTC GAG ACT GAG   GTCAACATCTATCTCCTCTGTTTTTAAAATTTACTAGCTAGTATGT  450

                                               H   G   F   V   Y   R   E   N   H   K   S   P   G   Y   Y   D   G   R   Y   W   T   M   W   K
TGATATGTCGTGTTAACAGTGTTGTGGGATATCATGTGCAG     CAC GGA TTT GTG TAC CGT GAG AAC CAT AAG TCA CCA GGA TAC TAC GAT GGC AGA TAC TGG ACC ATG TGG AAG  563

 L   P   M   F   G   C   T   D   A   T   Q   V   L   A   E   V   Q   E   A   K   K   A   Y   P   Q   A   W   V   R   I   I   G   F   D   N
TTG CCC ATG TTC GGG TGC ACT GAT GCA ACC CAG GTC TTG GCT GAG GTG CAG GAG GCA AAG AAG GCT TAC CCA CAG GCA TGG GTC CGT ATC ATC GGA TTC GAC AAT    668

 V   R   Q   V   Q   C   I   S   F   I   A   Y   K   P   E   G   Y
GTT CGT CAA GTG CAG TGC ATC AGT TTC ATC GCT TAC AAG CCC GAA GGA TAC TAA  A(TGTGTATATGTCAACAGTGA)GAAACTGTTCGCATTTTCCGTTTTGCTTCTTTCTTTCTATTCAA    787

TGTATGTTGTTGGATTCCAGTTGAATTTATTATGAGAACTAATAATAATAGTAATAATCATTTGTTTCTTTACTAATTTGCATTTTCACATATGATTTCTGGTGCATATCATAATTTTCATTCCACCAATATTAATTTCC   927
```

Fig. 4. Nucleotide sequences of all five tomato rbcS genes. The non-coding strands including the protein coding and the flanking regions are shown. The amino acid sequences deduced are indicated above the nucleotide sequences and a *triangle* indicates the putative cleavage site of the transit peptide. *Numbers* are nucleotides relative to the 5′-end of the protein coding region. *Arrows* indicate the positions of the major transcription start sites. The conserved sequences including "TATA" and "CAAT" boxes, an enhancer-like sequence identified by Fluhr et al. (1986a), the sequence 5′-AAA/GCCTTATC, which is similar to a petunia sequence identified by Dean et al. (1985), and the putative polyadenylation signals are *underlined.* The sequences in *parentheses* in the 3′ flanking regions represent the sequences used as gene-specific probes for detection of rbcS mRNA

Genomic organization of the tomato rbcS genes

There were 15 positive clones isolated from a partial *Sau*3A phage library, 9 of which contained different inserts. Of these 9, phages 1, 6, 17 and 25 are shown in Fig. 2. Several positive clones were obtained from libraries of size-fractionated tomato nuclear DNA digested with *Hin*dIII. Two clones, phage 12 and 23 were chosen for further analysis (Fig. 2). The location and the orientation of the rbcS genes in the insert of each clone were determined as described in Materials and methods.

The 16.4 kb *Hin*dIII DNA fragment in phage 12 contains a single rbcS gene, located on two *Eco*RI fragments of 1.7 kb and 13 kb. Phage 1 and 23 overlap each other and encompass a 20 kb region of genomic DNA. Detailed restriction enzyme analysis of the cloned fragments and DNA sequence analysis reveal that only one rbcS gene is located in this 20 kb region. The inserts in phage 6, 17 and 25 span a 28 kb region that contains three rbcS genes. The three rbcS genes are located on 3.9 kb, 4.2 kb and 3.45 kb *Hin*dIII fragments and are organized in a tandem array within a 10 kb region (Fig. 2).

Tomato rbcS genes have been previously mapped to three different loci, *Rbcs-1* (chromosome 2) *Rbcs-2* (chromosome 3) and *Rbcs-3* (chromosome 2) using genetic and Southern blot analysis (Vallejos et al. 1986). The rbcS gene on the 16.4 kb *Hin*dIII insert of phage 12 corresponds to *Rbcs-1,* and the gene in phage 1 and phage 23 to *Rbcs-2.* The three tandemly arranged genes in phage 17 correspond to *Rbcs-3.* We have designated the locus 3 rbcS genes on the 3.9 kb, the 4.2 kb and the 3.45 kb *Hin*dIII fragments as *Rbcs-3A, Rbcs-3B* and *Rbcs-3C,* respectively. All *Eco*RI, *Eco*RV and *Hin*dIII fragments carrying the rbcS genes (Fig. 2) correspond in size exactly to all fragments detected in the genomic Southern analysis and additional fragments could not be detected (data not shown). We therefore conclude that the tomato rbcS multigene family has a total of five members.

Structure of the tomato rbcS genes

RbcS-gene-specific DNA fragments were subcloned and sequenced using the strategies shown in Fig. 3. Intron positions were assigned by computer analysis of open reading frames and comparison with the sequence of the tomato rbcS cDNA 3–91 (Fig. 4). The *Rbcs-1* and the three *Rbcs-3* genes each have two introns, while the *Rbcs-2* gene has three introns. Introns in *Rbcs-1* and *Rbcs-3,* as well as the first two introns of *Rbcs-2,* are at identical positions. In addition, the first and the third introns in *Rbcs-2,* which are 999 bp and 435 bp long, respectively, are significantly longer than introns in the genes from loci 1 and 3. Both introns in *Rbcs-1* and *Rbcs-3* genes are similar in size (95 bp and 97 bp in *Rbcs-1,* 92 bp and 87 bp in *Rbcs-3A* and *Rbcs-3C,* 92 bp and 82 bp in *Rbcs-3B;* Fig. 4). The structure of the *Rbcs-3* genes is highly conserved. The exon and intron sequences of *Rbcs-3A* and *Rbcs-3C* are identical, and *Rbcs-3B* differs by eight nucleotides in the DNA sequence for the transit polypeptide and by two nucleotides in the mature protein. The exon sequences of the *Rbcs-1* gene differ by 13.3% from those of the *Rbcs-3* genes, and by 14.0% from those of *Rbcs-2. Rbcs-2* exons show a 10.4% divergence relative to *Rbcs-3A* and *Rbcs-3C,* and a 10.2% divergence relative to *Rbcs-3B.* The intron sequences are less conserved than the exon sequences between genes at the different loci.

Predicted structure of tomato rbcS proteins

The protein predicted from the DNA sequence of *Rbcs-1* contains 181 amino acid residues, including a putative transit peptide of 58 residues. Putative transit peptides were deduced from the amino acid sequence of mature proteins from several plants (Muller et al. 1983). *Rbcs-2* and *Rbcs-3* genes code for precursor proteins of 180 residues, containing transit peptides of 57 residues and mature polypeptides of 123 residues. The mature polypeptides encoded by the three *Rbcs-3* genes are identical. The transit peptides encoded by *Rbcs-3A* and *Rbcs-3C,* however, differ by two amino acid residues from *Rbcs-3B.* The predicted primary amino acid sequence of the mature proteins from *Rbcs-1, Rbcs-2* and *Rbcs-3* genes differ by a few residues (*Rbcs-1/Rbcs-2* three, *Rbcs-1/Rbcs-3* four and *Rbcs-2/Rbcs-3* one amino acid). Thus, the five tomato rbcS genes code for four different precursors and three different mature polypeptides.

The 5′ and 3′ flanking sequences of tomato rbcS genes

The 5′ ends of the tomato rbcS mRNAs were determined by S_1 nuclease mapping as described in Materials and methods. The 5′ end of the *Rbcs-1* mRNA is located approximately 75 nucleotides upstream from the ATG initiation codon (Fig. 5). Multiple S_1-protected fragments are most likely the result of incomplete protection at the termini due to the AT-rich sequence of the *Rbcs-1* non-translated mRNA sequence. The 5′ end of *Rbcs-2* mRNA maps to a T 28 nucleotides upstream from the ATG. The S_1 protected fragments of *Rbcs-3A, Rbcs-3B* and *Rbcs-3C* are located 39(A), 8(T) and 17(A) nucleotides 5′ to the ATG, respectively. Sequences 5′TATATAAA (*Rbcs-1* and *Rbcs-2*) and TATATATA (*Rbcs-3*), which correspond to the eukaryotic "TATA-box" (Breathnach and Chambon 1981), are present 30–36 nucleotides upstream from the transcription start sites. A sequence ATCCAATGGT is conserved in the 5′ region of *Rbcs-1* and all three *Rbcs-3* genes (Fig. 4). This sequence includes the "CAAT box" motif, which is found 100–120 nucleotides upstream from transcriptional start sites of other eukaryotic genes (Breathnach and Chambon 1981). The sequence TCCAA, which is conserved for all five rbcS genes, is present 120 nucleotides upstream from the transcription start site of *Rbcs-2.* The 5′ flanking sequences are moderately conserved between the *Rbcs-3* genes, but are substantially different between the genes of the three loci except for the TATA-like and CAAT-like sequences, which are common to all rbcS genes in tomato. In addition, sequences homologous to the viral and mammalian enhancer sequence GTGG/AT/AT/AT or TA/TA/TA/CCAC (Gillies et al. 1983; Weiher et al. 1983) are located in the 5′ upstream region. Similar sequences have been found upstream of rbcS genes in other higher plants (Mazur and Chui 1985; Fluhr et al. 1986a).

The 3′ flanking sequences immediately following the translational stop codon TAA are less conserved between members in the different loci and also between the genes within locus 3. A sequence AATAAT, which resembles the animal consensus sequence AATAAA present about 20 nucleotides upstream from the polyadenylation site (Proudfoot and Brownlee 1976), appears to be the most conserved plant polyadenylation signal (Dean et al. 1986). Sequences homologous to the polyadenylation signal consensus sequence are present 110–140 nucleotides downstream from the TAA codon (Fig. 4).

Expression of each member of tomato rbcS genes in leaves

To compare the relative amounts of steady-state mRNA for all five rbcS genes, total tomato leaf RNA was subjected to Northern analysis using rbcS gene-specific probes. We used 20-base oligonucleotides complementary to the 3′ non-translated regions of the rbcS mRNAs and positioned within 35 nucleotides downstream from the TAA stop codon (see Fig. 4). The specificity of each probe was determined by Southern hybridization to the genomic clone DNAs. In every case, the probe hybridized specifically to its respective gene (data not shown). Specific radioactivity and hybridization efficiency of the probes used in these experiments were nearly identical (Sugita and Gruissem 1987). All probes were well within the transcribed 3′ portions of the genes and hybridized to a single RNA species of approximately 0.8 kb (Fig. 6). Based on the Northern analysis and

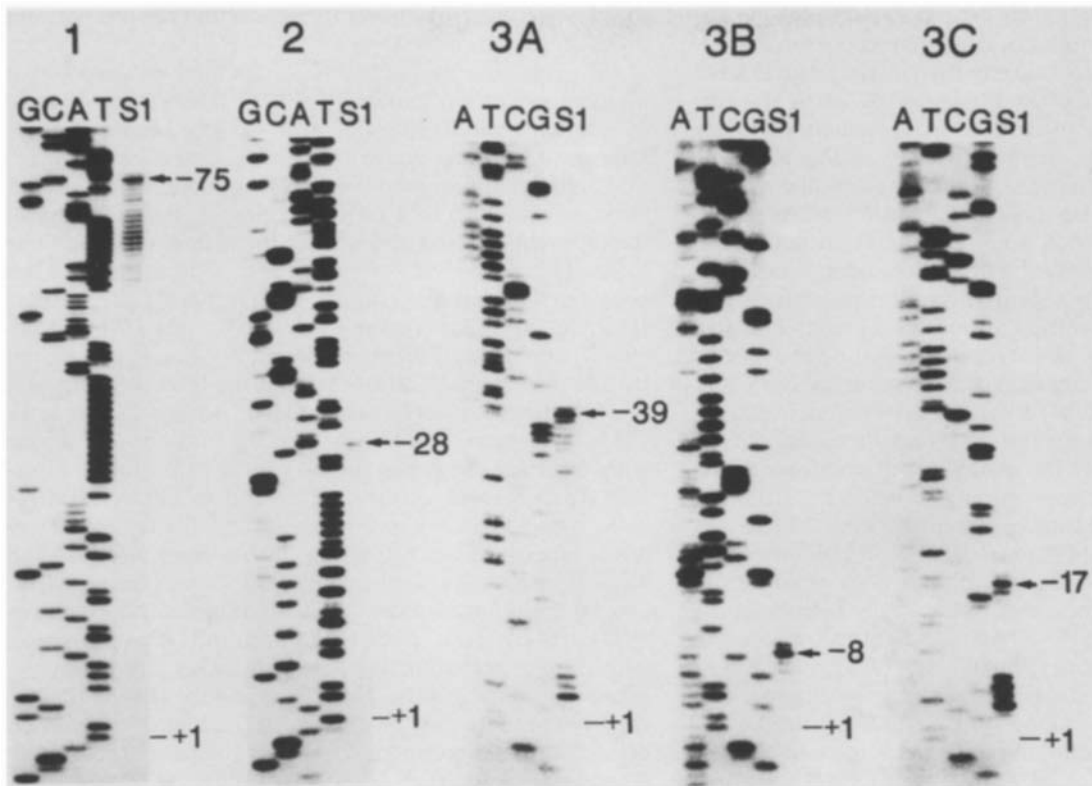

Fig. 5. Determination of the 5′-ends of the tomato rbcS mRNAs (*1–3c*) by S_1 nuclease mapping. The following coding-strand probes were synthesized using the 5′-labeled coding region primers as described in Materials and methods: a 288 b *Hae*III-primer (*RbcS-1*, +68 to −220), a 262 b *Fok*I-primer (*Rbcs-2*, +68 to −194), a 204 b *Sau*3A-primer (*Rbcs3A*, +68 to −136), a 146 b *Dde*I-primer (*Rbcs-3B*, +68 to −78) and a 205 b *Hae*III-primer (*Rbcs-3C*, +68 to −137). The single-stranded DNA probes were hybridized to total tomato leaf RNA (20 µg). *Arrows* indicate positions of the major S_1-protected fragments relative to the 5′-end (+1) of the protein-coding regions. The protected fragments around the ATG codon (*3A* and *3C*) result from S_1 digestion of hybrids formed between the respective probe and homologous regions of mRNAs from both genes. The sequence ladders of the coding strand (G, C, A and T) are shown

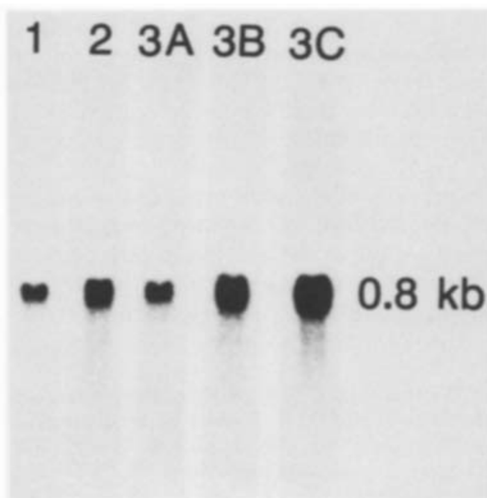

Fig. 6. Northern hybridization of the tomato rbcS gene-specific probes to total tomato leaf RNA. Total leaf RNA (15 µg) was fractionated on 1.2% agarose-formaldehyde gel, transferred to a nylon filter and hybridized with 5′-end-labeled oligonucleotide probes specific to the respective members of the rbcS gene family (1–3c). The gene-specific probes located in the 3′-flanking regions are shown in Fig. 4

S_1 mapping we conclude, therefore, that all tomato rbcS genes are expressed in tomato leaves. The *Rbcs-3C* probe showed the highest level of hybridization in leaves, followed by a slightly lower level of mRNA hybridization to the *Rbcs-3B* probe. The *Rbcs-2* mRNA level is approximately half the level of that of *Rbcs-3B*. The levels of *Rbcs-1* and *Rbcs-3A* transcripts are 4- to 5-fold lower than those of *Rbcs-3B* and *Rbcs-3C* genes.

Discussion

Genomic Southern and detailed restriction map analysis of the cloned DNA fragments containing the tomato rbcS genes demonstrate that tomato has five members in the rbcS multigene family. Previously published studies have proposed that the *Rbcs-2* locus contains two genes, and six rbcS genes were therefore postulated for the small subunit protein in tomato (Pichersky et al. 1986; Vallejos et al. 1986). This discrepancy is most probably the result of the unexpectedly long intron between the first two exons of the *Rbcs-2* gene. The number of rbcS genes is at least nine in petunia (Dean et al. 1985; Tumer et al. 1986), five in pea (Fluhr et al. 1986b), and ten in soybean (Berry-Lowe et al. 1982). Pseudogenes are evidently not present in the tomato rbcS gene family. Our study is the first report to

describe the genomic organization and structure of the entire set of rbcS genes in tomato. DNA sequences for *Rbcs-1*, *Rbcs-3A* gene and a cDNA derived from *Rbcs-2* gene have been reported previously (Pichersky et al. 1986). In this study, selection has been suggested as a mechanism to explain, at least in part, the highly conserved coding information in the portion of the rbcS genes encoding the mature polypeptides. It was also proposed that *Rbcs-1* diverged from the other tomato rbcS genes first, and that the *Rbcs-2* and *Rbcs-3* loci split from each other only later. Since only the *Rbcs-2* gene contains a third intron, it is possible, therefore, that this is the result of an insertion event. This hypothesis is supported by the observation that no rbcS genes in species other than *Solanaceae*, where three rbcS loci seems to be the ancestral situation, have been found to contain this third intron (Tanksley and Pichersky, to be published). In contrast to the moderate nucleotide sequence divergence (10%–14%) among tomato rbcS genes from the three different loci, the three genes in the *Rbcs-3* locus are identical or nearly identical to each other at the nucleotide level. Similar sequence conservation has been reported for some members of the pea rbcS gene family (Timko et al. 1985; Fluhr et al. 1986b), which are tightly linked on a unique chromosomal locus (Polans et al. 1985). Such a high nucleotide homology could be explained by recent gene duplication or gene conversion events. However, unlike the situation in the pea rbcS sequences, the strong homology of the nucleotide sequences of the tomato *Rbcs-3* genes does not extend to the 5′ or 3′ flanking regions. Thus, in the tomato *Rbcs-3* locus, gene conversion is a more likely explanation for the following reasons: The nucleotide sequences of *Rbcs-3A* and *Rbcs-3C* are identical throughout the exon and intron sequences, but this homology begins at the eighth nucleotide upstream from the first ATG codon and terminates abruptly directly downstream from the TAA stop codon. Proximal to these borders, the homology is minimal except for short DNA sequence motifs, which are conserved in the 5′ regions of all five genes. A recent gene duplication event cannot be excluded, but seems implausible, since it is unlikely that a duplication will selectively involve the coding region. In the case of *Rbcs-3B*, the evidence for gene conversion is even stronger because here the 5′ end of the homologous regions are within the genes, approximately 30 nucleotides downstream from the first ATG. This may indicate the position where the gene conversion process terminated. Consequently, there are nine substitutions in the first 30 nucleotides relative to *Rbcs-3A* and *Rbcs-3C*, but only nine additional subsitutions in the rest of the gene. Thus, our data strongly suggest the occurrence of gene conversion within the *Rbcs-3* locus, of which a *Rbcs-3A*/*Rbcs-3C* conversion appears to be the most recent event. In this case the nucleotide sequence homologies within this locus cannot be interpreted as an indication of the time of origin of the gene duplications which gave rise to each of the three rbcS genes in the locus. *Rbcs-1* and *Rbcs-3A* genes sequenced previously were derived from the tomato line T6, and a cDNA clone for the *Rbcs-2* gene was synthesized from RNA of the VF36 cultivar (Pichersky et al. 1986). These and the corresponding rbcS genes from the tomato cherry VFNT line reported in this study are identical in their protein-coding regions, but differ by several nucleotides in the intron and the flanking regions. This indicates that significant sequence constraints operate for the tomato rbcS protein-coding regions, but comparatively rapid divergence may occur in non-coding regions between different tomato lines.

S_1 nuclease mapping and Northern analysis suggest that all members of the tomato rbcS gene family are expressed in tomato leaves. At the mRNA level, *Rbcs-3B* and *Rbcs-3C* appear to be most active; the other genes are expressed at a significantly lower level. Relative amounts of mRNA indicate a 4- to 5-fold difference among the members of the different loci and also among the members of the same locus. Different levels of expression of members in rbcS gene families have previously been reported for petunia (Dean et al. 1985; Tumer et al. 1986), pea (Fluhr et al. 1986b) and maize (Sheen and Bogorad 1986). A correlation between the number of introns and the level of expression of rbcS genes has been suggested in petunia (Dean et al. 1985). It is interesting to note that the tomato *Rbcs-2* gene with three introns is not highly expressed in leaves. However, variations in intron number and length between tomato rbcS genes may not be responsible for their different levels of expression. Messenger RNAs from *Rbcs-3B* and *Rbcs-3C*, which are highly expressed in leaves, are undetectable in fruits or etiolated seedlings (Sugita and Gruissem 1987). In contrast, *Rbcs-1*, *Rbcs-2*, and *Rbcs-3A* genes, which are expressed at low levels in leaves, are highly expressed in other organs. Thus, each member of the tomato rbcS gene family may be controlled differently in different organs. In addition, there is a differential response to light among the tomato rbcS genes within the same tissue (Sugita and Gruissem 1987). Several studies have provided examples for organ-specific (Dean et al. 1985; Fluhr et al. 1986b), light-dependent (Tobin and Silverthorne 1985), or chloroplast-coordinated expression of rbcS genes (Shinozaki et al. 1982). Sequences involved in the light-regulated and/or organ-sepcific expression of rbcS genes have been located in upstream regions of some rbcS genes (Morelli et al. 1985; Fluhr et al. 1986a). Fluhr et al. (1986a) have identified a sequence which is conserved among rbcS genes and is contained within an upstream region involved in light induction. This sequence is also present in the upstream regions of the tomato rbcS genes. Additionally, a sequence conserved among the upstream regions of several petunia rbcS genes (Dean et al. 1985) is found in the five tomato rbcS genes. Whether these conserved sequences are required for organ-specific and/or light-regulated expression is unclear.

The unique expression of the rbcS genes in various organs, and during tomato fruit development and chromoplast differentiation, as well as the analysis of each member of the rbcS gene family in transgenic tomato plants, will allow us to study in detail the regulatory components required for the control of their expression.

Acknowledgements. We thank Dr G. Zurawski for the preparation of synthetic oligonucleotides and Dr R. Fischer for kindly providing us with phage Charon 35. We also thank Drs J. Narita, B. Piechulla, and D.B. Stern for valuable suggestions. This work was supported by NIH grant GM33813 to W.G. and by Research Aid from the Inoue Foundation for Science, Japan, to M.S.

References

Bedbrook JR, Coen DM, Beaton AR, Bogorad L, Rich A (1979) Location of the single gene for the large subunit of ribulose bisphosphate carboxylase on the maize chloroplast chromosome. J Biol Chem 254:905–910

Bendich AJ, Anderson RS, Ward BL (1979) Plant DNA: long, pure and simple. In: Leaver CJ (ed) Genome organization and expression in plants. Plenum Press, New York London, pp 31–33

Bennett MD, Smith JM (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc Lond 274:227–274

Bernatzky R, Tanksley SD (1986) Majority of random cDNA clones correspond to single loci in the tomato genome. Mol Gen Genet 203:8–14

Berry JO, Nikolau BJ, Carr JP, Klessig DF (1985) Transcriptional and post-transcriptional regulation of ribulose-1,5-bisphosphate carboxylase gene expression in light- and dark-grown amaranth cotyledons. Mol Cell Biol 5:2238–2246

Berry-Lowe SL, McKnight TD, Shah DM, Meagher RB (1982) The nucleotide sequence, expression, and evolution of one member of a multigene family encoding the small subunit of ribulose-1,5-bisphosphate carboxylase in soybean. J Mol Appl Genet 1:483–498

Biggin MD, Gibson TJ, Hong GF (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 80:3963–3965

Blair GE, Ellis RJ (1973) Protein synthesis in chloroplasts I. Light-driven synthesis of large subunit of Fraction I protein by isolated pea chloroplasts. Biochim Biophys Acta 319:223–234

Breathnach R, Chambon P (1981) Organization and expression of eukaryotic split genes coding for proteins. Annu Rev Biochem 50:349–383

Broglie R, Coruzzi G, Lamppa G, Keith B, Chua N-H (1983) Structural analysis of nuclear genes coding for the precursor to the small subunit of wheat ribulose-1,5-bisphosphate carboxylase. Biotechnology 1:55–61

Dean C, Elzen PVD, Tamaki S, Dunsmuir P, Bedbrook J (1985) Differential expression of the eight genes of the petunia ribulose bisphosphate carboxylase small subunit multigene family. EMBO J 4:3055–3061

Dean C, Tamaki S, Dunsmuir P, Favreau M, Katayama C, Dooner H, Bedbrook J (1986) mRNA transcripts of several plant genes are polyadenylated at multiple sites in vivo. Nucleic Acids Res 14:2229–2240

Fluhr R, Kuhlemeier C, Nagy F, Chua N-H (1986a) Organ-specific and light-induced expression of plant genes. Science 232:1106–1112

Fluhr R, Moses P, Morelli G, Coruzzi G, Chua N-H (1986b) Expression dynamics of the pea rbcS multigene family and organ distribution of the transcripts. EMBO J 5:2063–2071

Frischauf A-M, Lehrach H, Poustka A, Murry N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827–842

Gillies SD, Morrison SL, Oi VT, Tonegawa S (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33:115–123

Gruissem W, Callan K, Lynch J, Manzara T, Meighan M, Narita J, Piechulla B, Sugita M, Thelander M, Wanner L (1987) Plastid and nuclear gene expression during tomato fruit formation. In: Nevins D (ed) Tomato biotechnology. Liss, New York (in press)

Gruissem W, Zurawski G (1985) Identification and mutational analysis of the promoter for a spinach chloroplast transfer RNA gene. EMBO J 4:1637–1644

Inamine G, Nash B, Weissbach H, Brot N (1985) Light regulation of the synthesis of the large subunit of ribulose-1,5-bisphosphate carboxylase in peas: Evidence for translational control. Proc Natl Acad Sci USA 82:5690–5694

Loenen WAM, Blattner FR (1983) Lambda Charon vectors (Ch32, 33, 34 and 35) adapted for DNA cloning in recombination-deficient hosts. Gene 26:171–179

Mazur BJ, Chui C-F (1985) Sequence of a genomic DNA clone for the small subunit of ribulose bis-phosphate carboxylase-oxygenase from tobacco. Nucleic Acids Res 13:2373–2386

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101:20–78

Morelli G, Nagy F, Fraley RT, Rogers SG, Chua NH (1985) A short conserved sequence is involved in the light-inducibility of a gene encoding ribulose-1,5-bisphosphate carboxylase small subunit of pea. Nature 315:200–204

Miziorko HM, Lorimer GH (1983) Ribulose-1,5-bisphosphate carboxylase-oxygenase. Annu Rev Biochem 52:507–535

Muller K-D, Salnikow J, Vater J (1983) Amino acid sequence of the small subunit of D-ribulose-bisphosphate carboxylase/oxygenase from *Nicotiana tabacum*. Biochim Biophys Acta 742:78–83

Piechulla B, Imlay KRC, Gruissem W (1985) Plastid gene expression during fruit ripening in tomato. Plant Mol Biol 5:373–384

Piechulla B, Pichersky E, Cashmore AR, Gruissem W (1986) Expression of nuclear and plastid genes for photosynthesis-specific proteins during tomato fruit development and ripening. Plant Mol Biol 7:367–376

Pichersky E, Bernatzky R, Tanskley SD, Cashmore AR (1986) Evidence for selection as a mechanism in the concerted evolution of *Lycopersicon esculentum* (tomato) genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc Natl Acad Sci USA 83:3880–3884

Polans NO, Weeden NF, Thomspon WF (1985) Inheritance, organization, and mapping of rbcS and cab multigene families in pea. Proc Natl Acad Sci USA 82:5083–5087

Proudfoot N, Brownlee GG (1976) 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature 263:211–214

Sheen J-Y, Bogorad L (1986) Expression of the ribulose-1,5-bisphosphate carboxylase large subunit gene and three small subunit genes in two cell types of maize leaves. EMBO J 5:3417–3422

Shinozaki K, Sasaki Y, Sakihama T, Kamikubo T (1982) Coordinate light-induction of two mRNAs, encoded in nuclei and chloroplasts, of ribulose-1,5-bisphosphate carboxylase/oxygenase. FEBS Lett 144:73–76

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–51

Stiekema WS, Wimpee CF, Tobin EM (1983) Nucleotide sequence encoding the precursor of the small subunit of ribulose-1,5-bisphosphate carboxylase from *Lemna gibba* L. G-3. Nucleic Acids Res 11:8051–8061

Sugita M, Gruissem W (1987) Developmental, organ-specific and light-dependent expression of the tomato ribulose-1,5-bisphosphate carboxylase small subunit gene family. Proc Natl Acad Sci USA (in press)

Sugita M, Sugiura M (1984) Nucleotide sequence and transcription of the gene for the 32,000 dalton thylakoid membrane protein from *Nicotiana tabacum*. Mol Gen Genet 195:308–313

Timko MP, Klausch AP, Hand JM, Cashmore AR, Herrera-Estrella L, Van den Broeck G, Montagu MV (1985) Structure and expression of nuclear genes encoding polypeptides of the photosynthetic apparatus. In: Steinback KE, Bonitz S, Arntzen CJ, Bogorad L (eds) Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 381–396

Tobin EM, Silverthorne J (1985) Light regulation of gene expression in higher plants. Annu Rev Plant Physiol 36:569–593

Tumer NE, Clark WG, Tabor GJ, Hironaka CM, Fraley RT, Shah DM (1986) The genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase are expressed differentially in petunia leaves. Nucleic Acids Res 14:3325–3342

Vallejos CE, Tanksley SD, Bernatzky R (1986) Localization in the tomato genome of DNA restriction fragments containing sequences homologous to the rRNA (45S), the major chloro-

phyll a/b binding polypeptide and the ribulose bisphosphate carboxylase genes. Genetics 112:93–105
Vieira J, Messing J (1987) The production of single-stranded plasmid DNAs. Methods Enzymol (in press)
Weiher H, Koenig M, Gruss P (1983) Multiple point mutations affecting the simian virus 40 enhancer. Science 219:626–631
Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K (1979) Hybridization of synthetic oligodeoxyribonucleotides to phi X 174 DNA: the effect of single base pair mismatch. Nucleic Acids Res 6:3543–3557
Zaug A, Kent JR, Cech T (1984) A labile phosphodiester bond at the ligation junction in a circular intervening sequence RNA. Science 224:574–578

Communicated by R.B. Goldberg

Received March 3, 1987