Structure, Expression, and Evolution of the 5-Enolpyruvylshikimate-3-phosphate Synthase Genes of Petunia and Tomato*

(Charles et al., 1986; Duncan et al., 1987) and have been sequenced to determine the amino acid sequences of the corresponding enzymes. We have previously described the isolation of a full-length cDNA for wild type EPSP synthase from a glyphosate-resistant Petunia hybridra cell line and have shown that high level expression of this enzyme in transgenic petunia plants confers tolerance to glyphosate (Shah et al., 1986). In plants, EPSP synthase is localized primarily in chloroplasts and other plastids (Mousdale and Coggins, 1985). The sequence of the petunia cDNA shows that the enzyme is synthesized as a cytoplasmic precursor with an N-terminal transit peptide sequence. The transit peptide is required for the translocation of the precursor enzyme to the stroma of the plastids where it is cleaved to form the mature protein (della-Cioppa et al., 1986).

In this report we present the complete nucleotide sequences of the petunia (Petunia hybridra) and tomato (Lycopersicon esculentum) EPSP synthase cDNAs. A comparison between these plant EPSP synthase sequences and those from fungal and bacterial sources indicates that this enzyme is highly conserved in evolution.

We find that there are only one or two EPSP synthase genes in both tomato and petunia and have isolated the petunia gene corresponding to our cDNA. We present the complete exon-intron map of this gene and an analysis of EPSP synthase expression in both petunia and tomato. Surprisingly, there are dramatic differences in expression patterns between the two plants. Since petunia showed organ-specific variability in expression, we mapped EPSP synthase transcript initiation sites in several petunia organs. The organs which express EPSP synthase mRNA at low levels utilize multiple start sites, showing some preference for one start site. This preferred site is the only transcription start site in petunia petals where this gene is most highly expressed.

MATERIALS AND METHODS

Isolation of RNA—Total RNA was isolated as previously described (Rochester et al., 1986). Poly(A+) RNA was isolated by oligo(dT)-cellulose chromatography (Aviv and Leeder, 1972). Floral organs were collected for RNA isolation when the flowers were just beginning anthesis.

Isolation and Sequencing of EPSP Synthase cDNA Clones—The isolation of petunia EPSP synthase cDNA clones has been described previously (Shah et al., 1986). Tomato cDNA libraries were constructed in Agt10 (Huynh et al., 1985) from RNA isolated from anthers or pistils by a modification of the procedures of Huynh et al. (1985) and Guibler and Hoffman (1983). The libraries were plated on BNI02 (Huynh et al., 1985) and filter replicas were made by standard techniques (Davis et al., 1980). The filters were hybridized with the petunia EPSP synthase insert from pM0NE6140 (della-Cioppa et al., 1987) that had been labeled with 32P by the random priming method of Feinberg and Vogelstein (1983). Hybridizing plaques were subcultured and screened again with the same probe.

Charles S. Gasser‡, Jill A. Winter§, Cathy M. Hironaka, and Dilip M. Shah

From Plant Molecular Biology, Monsanto Company, St. Louis, Missouri 63198

The abbreviations used are: EPSP, 5-enolpyruvylshikimate 3-phosphate; kb, kilobase (pairs).

5-Enolpyruvylshikimate-3-phosphate (EPSP) synthase is an enzyme of the shikimate pathway which is located in the chloroplasts in higher plants. This enzyme is the target of the nonselective herbicide glyphosate. We have isolated and sequenced cDNA clones encoding EPSP synthase from petunia and tomato. The deduced amino acid sequences of the two enzyme precursors show 93% identity in the mature protein regions and 58% identity in the transit peptides. The sequences of the plant enzymes show significantly greater similarity to bacterial EPSP synthases than to fungal EPSP synthases. A genomic clone containing an EPSP synthase gene was isolated from a library of petunia DNA and was shown to contain seven intervening sequences. This gene is expressed ~25-fold higher in flower petals than in other organs of petunia. Transcription of this gene is initiated at multiple sites in petunia leaves and in a glyphosate-tolerant petunia cell line overproducing EPSP synthase mRNA. In petals, however, transcription of this gene is initiated almost entirely from only one of these sites. In contrast to petunia, the levels of EPSP synthase mRNA in different organs of tomato do not differ significantly.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby advertised in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) J03227.

* The abbreviation used are: EPSP, 5-enolpyruvylshikimate 3-phosphate; kb, kilobase (pairs).

‡ Present address: Dept. of Biochemistry, University of Missouri, Columbia, MO 65211.

§ Present address: Dept. of Biochemistry, University of Missouri, Columbia, MO 65211.

The abbreviations used are: EPSP, 5-enolpyruvylshikimate 3-phosphate; kb, kilobase (pairs).

C. S. Gasser, manuscript in preparation.
DNA was isolated from the hybridizing phage and was subcloned into pUC8 (Viera and Messing, 1982) or pUC118 (a gift of Jeff Viera) by the method of Crouse et al. (1983).

The cDNA clones were sequenced by the method of Maxam and Gilbert (1980) or by the dideoxy chain termination method (Sanger et al., 1977) after subcloning into M13 phage (Messing, 1983) or pUC118. Sequencing reactions were performed using the University of Wisconsin Genetics Computer Group programs by the methods of Smith and Waterman (1981) and Needleman and Wunch (1970).

**Isolation and Characterization of Petunia Genomic Clones—**High molecular weight DNA was isolated from the phytophthora-resistant MP4-G cell line (Steinman et al., 1986) as described by Shurer et al. (1983). This DNA was digested with BamHI and the fragments were cloned into the phage vector, M1G4, as described previously (Ramahadran et al., 1985). MG14 was obtained from Dr. M. Olson, Washington University Medical School, St. Louis, MO. The library was screened with a 32P-labeled petunia EPSP synthase cDNA insert from pMON9556 (Shah et al., 1986) (Fig. 1A). A genomic clone, F10, which hybridized strongly to the probe, was identified and purified. Three BglII fragments of F10 spanning the EPSP synthase gene were subcloned into pMON5850, and portions were sequenced using the dideoxy chain termination method (Sanger et al., 1977) as described above. pMON9556, a gift of H. Kliewer (Monsanto Company), was derivatized with pUC18 (Yamisch-Perron et al., 1985) in which a portion of the multilinker sequence has been replaced with a synthetic DNA fragment which includes a BglII site.

**DNA and RNA Filter Hybridizations—**The restriction digests of the high molecular weight petunia DNA were electrophoresed through 0.7% agarose gels and blotted onto nitrocellulose (Southern, 1975). The filters were hybridized with 32P-labeled (Rigby et al., 1977) cDNA inserts and washed according to the conditions outlined by Fedoroff et al. (1983). Tomato DNA was isolated by the method of Kislev and Rubenstein (1980). Restriction digests of tomato DNA were blotted onto a ZetaBind (AME-Cuno) membrane using 20 mM sodium phosphate, pH 6.8. Prehybridization was carried out in 2 × SSPE (1 × SSPE is 150 mM NaCl, 10 mM sodium phosphate, 1 mM EDTA, pH 7.4), 4 × Denhardt’s solution (1 × Denhardt’s solution is 0.02% each Ficoll, bovine serum albumin, and polyvinylpyrrolidone), 100 µg/ml denatured salmon sperm DNA at 68 °C. Hybridization was performed in a fresh solution with 1-2 × 10⁶ cpm/ml of 32P-labeled tomato cDNA insert. Filters were exposed to Kodak XAR-5 film with intensifying screens at −70 °C.

Total cellular RNA, or poly(A) RNA, was fractionated on 1.3% agarose gels containing 6.5% formaldehyde, as described elsewhere (Maniatis et al., 1982). Gels were blotted onto GeneScreen (Du Pont-New England Nuclear) without any pretreatment, using 10 × SSC (1 × SSC is 15 mM sodium citrate, 150 mM NaCl, pH 7.0). One µg/ml of ethidium bromide was included in gels to visualize the 18 S and 26 S ribosomal RNAs on the filter by ultraviolet light illumination, after transfer (Anderson and Axel, 1985). Their positions were marked with a pencil and served as internal molecular weight standards. Blots were hybridized using 32P-labeled probes according to the conditions supplied with the GeneScreen membrane.

**Primer Extension Analysis—**Primer extension analysis was performed by the method of Day et al. (1987). The products were electrophoresed in 6% polyacrylamide gels containing 8 M urea. The gels were fixed in 10% methanol, 10% acetic acid (v/v), dried, and exposed as above.

**RESULTS**

**Isolation and Characterization of EPSP Synthase cDNA Clones—**We have previously described the isolation of a full-length cDNA to petunia EPSP synthase from a glyphosate-resistant cell line containing multiple copies of the cDNA, and from petunia EPSP synthase (Shah et al., 1986). The cDNA was isolated as two nonoverlapping clones by hybridization with a mixture of oligonucleotides based on a partial amino acid sequence of petunia EPSP synthase. The two clones abut at an EcoRI site internal to the cDNA and together account for the entire coding sequence of petunia EPSP synthase (Fig. 1A).

The complete sequence of the petunia cDNA was determined and is shown in Fig. 1B. An open reading frame of 1548 base pairs is apparent in the correct orientation for the polyadenylate tract to represent the poly(A) tail of the message. Amino acids 73–85 of the deduced amino acid sequence (Fig. 2) align precisely with the amino acid sequence determined from the purified enzyme (Shah et al., 1986). Since EPSP synthase has been shown to be present in the chloroplast (Moisl and Coggins, 1985; della-Cioppa et al., 1986), the first 72 amino acids of the open reading frame which are not present in the enzyme isolated from petunia cells represent the transit peptide which directs the enzyme to the chloroplast (della-Cioppa et al., 1986; Shah et al., 1986, della-Cioppa et al., 1987).

Tomato EPSP synthase cDNA clones were isolated from phage libraries made from RNA of mature tomato pistils1 and mature tomato anthers (McCormick et al., 1987) by hybridization with a petunia cDNA probe. Two independent clones were isolated from each library. The cDNAs were subcloned into plasmid vectors for further analysis. The restriction map of the longest clone, from the pistil library, which contains the entire EPSP synthase coding sequence, is shown in Fig. 1A.

Portions of three of the cDNA clones were sequenced (Fig. 1A) allowing the assembly of a complete composite sequence (Fig. 1B). With two exceptions the clones were identical in all of the regions sequenced. Clone P1 differed from clone A1 in that the region coding for the third third of the enzyme contained a base substitution and a single base deletion. This deletion would lead to a frameshift causing early termination of translation, eliminating the last 55 amino acids of the enzyme (data not shown). We believe these differences to be an artifact of the cDNA cloning procedure. Such artifacts are to be expected due to the imprecision of reverse transcriptase (Gopinathan et al., 1979). A vector for expression of mature petunia EPSP synthase in Escherichia coli has been previously described (Padgette et al., 1987). Construction of a similar vector with the tomato cDNA sequence, using the PstI to BamHI fragment of clone P1, and the BamHI to HindIII fragment of clone P2 led to production of active EPSP synthase in E. coli, confirming that the illustrated sequence is that of tomato EPSP synthase. The tomato sequence contains an open reading frame of 1560 base pairs. The reading frame is longer than that of petunia due to the addition of four codons to the transit peptide region (Fig. 2A).

The mature petunia and tomato enzymes are very similar (Fig. 2B). They differ in only 30 of 447 amino acids, a 93% identity. In contrast, the transit peptides of the enzymes from the two species are quite diverged (Fig. 2A). The tomato sequence has four insertions in the transit peptide relative to the petunia, and 90 of 72 amino acids are different (58% identity). The pattern of homology in the cDNA sequences shows a less extreme variation, there being 74% identity in the transit peptide regions and 87% identity in the regions coding for the mature peptides.

Comparison of the plant EPSP synthases with those of other organisms could indicate the functionally critical regions of the enzyme by showing which portions evolve most slowly. A comparison of the predicted amino acid sequence for the mature petunia and tomato enzymes with those of the E. coli and Aspergillus nidulans enzymes is illustrated in Fig. 2B. In Aspergillus, EPSP synthase activity is associated with a large multifunctional protein, the AROM complex (Charles et al., 1986). The region of the protein responsible for EPSP synthase activity, identified by its homology to the E. coli enzyme, is used for comparison in Fig. 2B.

The E. coli enzyme is 17 amino acids shorter than the mature petunia enzyme. Excluding the amino acid “deletions,” the bacterial enzyme is 54% identical to petunia EPSP synthase (293 of 435 amino acids, DNA identity = 60%). The

1 C. S. Gasser, unpublished data.
Petunia and Tomato EPSP Synthase Genes

EPSP synthase region of the Aspergillus AROM complex is less similar. It includes nine insertions and four deletions relative to the petunia enzyme. Excluding the deletions and insertions, only 162 of 427 amino acids are conserved (38%, DNA identity = 44%). Notably, 83% of the amino acids conserved between petunia and Aspergillus are also present in the E. coli sequence despite the fact that the E. coli enzyme is only 38% identical to the Aspergillus enzyme (overall DNA
Petunia and Tomato EPSP Synthase Genes

The precise locations of the introns were determined by sequencing the exons and the intron-exon junctions. Regions of the BgIII clones which hybridized to the petunia cDNA clones were subcloned into M13-derived phage (Messing, 1983) and sequenced according to the illustrated strategy (Fig. 3A). Comparison of the sequence with that of the cDNA clone indicates that the gene is interrupted by seven introns (Fig. 3A). Two of these introns are large: 2.3 and 1.3 kb. The sequence of each exon is identical to the corresponding region of the cDNA clone. Each intron has a short region of identity between the acceptor and donor sequences (Fig. 3B) making it impossible to precisely define the intron-exon boundaries. However, in each case an assignment can be made which conforms with the GT/AG rule (Fig. 3B) (Aebi et al., 1986; Brown et al., 1986).

To demonstrate that the gene we had isolated from MP4-G cells was not rearranged, Southern blotting patterns of petunia leaf DNA (Fig. 4, A and B, lanes 1–3) were compared to the patterns of λF10 (Fig. 4, A and B, lanes 4–6). We had previously shown that the MP4-G cell line contained multiple restriction fragments which hybridized to the EPSP synthase probe (Shah et al., 1986). Only a subset of these sequences were amplified in this cell line relative to the glyrophosate-sensitive progenitor MP4 cells. The patterns of hybridization in the petunia leaf DNA (Fig. 4, A and B, lanes 1–3) are identical to the previously reported hybridization patterns of DNA from the glyrophosate-sensitive MP4 cells. Comparison of this pattern with that of λF10 (Fig. 4, A and B, lanes 4–6) shows that the phage clone corresponds to the subset of the bands that were shown to be amplified in the MP4-G cell line. The internal fragments of the phage clone align with corresponding fragments in the genomic lanes. Thus, the gene we have cloned from the MP4-G cells is present in normal petunia tissues and represents one of at least two sets of sequences that can hybridize with the EPSP synthase cDNA.

Similar hybridization experiments were performed on restriction enzyme digests of tomato DNA. Fig. 4C shows the pattern of hybridization when the complete coding sequence of tomato EPSP synthase was used as a probe. In each case a set of strongly hybridizing and a set of weakly hybridizing fragments can be seen. This filter was eluted and rehybridized with isolated 5' and 3' regions of the tomato EPSP synthase gene.6 If the locations of the introns are conserved between petunia and tomato then these experiments indicate that there are at least two EPSP synthase genes in tomato. If tomato has only a single EPSP synthase gene then the gene must have several additional introns and be greater than 30 kb in length.

**Differential Expression of Petunia EPSP Synthase**—The different organs of a plant may have different requirements for aromatic amino acids. Products of the shikimate pathway also feed into pathways for synthesis of aromatic compounds such as lignin and various flavonoids. One might, therefore, hypothesize that the EPSP synthase gene could be differentially expressed in different organs of a plant. To determine if this was the case, the level of EPSP synthase mRNA in several organs of petunia plants was compared by Northern blot analysis. The results shown in Fig. 5A demonstrate that this gene is developmentally regulated in petunia, the expression being lowest in leaves and pistils, higher in stems, roots, and stamens, and very much higher in petals (~50-fold over leaves). To determine if this pattern was common in the Solanaceae, a similar experiment was performed on RNA from tomato organs. The results of this experiment (Fig. 5B) show that in tomato the level of expression in the petals is similar to that of the vegetative and other reproductive organs, varying

---

**Fig. 2. Comparison of EPSP synthase protein sequences.** A, alignment of transit peptides of petunia and tomato EPSP synthase as deduced from the sequence of the cDNAs. The entire petunia sequence is shown. Only those amino acids of the tomato sequence which differ from the petunia sequence are shown. Periods indicate amino acids which are identical to the petunia sequence, and dashes indicate insertions/deletions. B, alignment of mature enzymes of petunia and tomato (predicted from the cDNA sequences) with the sequence of E. coli EPSP synthase (Duncan et al., 1984), and with a region of the Aspergillus nidulans (A. nidu) AROM complex (Charles et al., 1986) which has homology to EPSP synthase. Only amino acids which differ from the petunia sequence are shown. Other symbols are as in A. The Aspergillus sequence is numbered with the first amino acid of the E. coli EPSP synthase protein as reference. A and B, lanes 1–3 are identical to the previously reported hybridization patterns of DNA from the glyrophosate-sensitive MP4 cells. Comparison of this pattern with that of λF10 (Fig. 4, A and B, lanes 4–6) shows that the phage clone corresponds to the subset of the bands that were shown to be amplified in the MP4-G cell line. The internal fragments of the phage clone align with corresponding fragments in the genomic lanes. Thus, the gene we have cloned from the MP4-G cells is present in normal petunia tissues and represents one of at least two sets of sequences that can hybridize with the EPSP synthase cDNA. Similar hybridization experiments were performed on restriction enzyme digests of tomato DNA. Fig. 4C shows the pattern of hybridization when the complete coding sequence of tomato EPSP synthase was used as a probe. In each case a set of strongly hybridizing and a set of weakly hybridizing fragments can be seen. This filter was eluted and rehybridized with isolated 5' and 3' regions of the tomato EPSP synthase gene. If the locations of the introns are conserved between petunia and tomato then these experiments indicate that there are at least two EPSP synthase genes in tomato. If tomato has only a single EPSP synthase gene then the gene must have several additional introns and be greater than 30 kb in length.

**Differential Expression of Petunia EPSP Synthase**—The different organs of a plant may have different requirements for aromatic amino acids. Products of the shikimate pathway also feed into pathways for synthesis of aromatic compounds such as lignin and various flavonoids. One might, therefore, hypothesize that the EPSP synthase gene could be differentially expressed in different organs of a plant. To determine if this was the case, the level of EPSP synthase mRNA in several organs of petunia plants was compared by Northern blot analysis. The results shown in Fig. 5A demonstrate that this gene is developmentally regulated in petunia, the expression being lowest in leaves and pistils, higher in stems, roots, and stamens, and very much higher in petals (~50-fold over leaves). To determine if this pattern was common in the Solanaceae, a similar experiment was performed on RNA from tomato organs. The results of this experiment (Fig. 5B) show that in tomato the level of expression in the petals is similar to that of the vegetative and other reproductive organs, varying

---

### A

<table>
<thead>
<tr>
<th>Petunia</th>
<th>Tomato</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>HMKNNHGLQ</td>
</tr>
<tr>
<td>2.</td>
<td>TIFSFHWNQ</td>
</tr>
<tr>
<td>3.</td>
<td>STFIFSFHWNQ</td>
</tr>
<tr>
<td>4.</td>
<td>PFHWNQ</td>
</tr>
<tr>
<td>5.</td>
<td>HMKNNHGLQ</td>
</tr>
</tbody>
</table>

---

### B

<table>
<thead>
<tr>
<th>Petunia</th>
<th>Tomato</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>KSFHWNQ</td>
</tr>
<tr>
<td>2.</td>
<td>TIFSFHWNQ</td>
</tr>
<tr>
<td>3.</td>
<td>STFIFSFHWNQ</td>
</tr>
<tr>
<td>4.</td>
<td>PFHWNQ</td>
</tr>
<tr>
<td>5.</td>
<td>HMKNNHGLQ</td>
</tr>
</tbody>
</table>

---

**Identity = 46%**. This is an indication that these amino acids are conserved for functional reasons.

**Isolation of a Genomic Clone for Petunia EPSP Synthase**—Total DNA was isolated from the petunia cell line line MP4-G which contains multiple copies of the EPSP synthase gene (Shah et al., 1986). This DNA was digested with BamHI and inserted into the vector λMGM14 (a gift of M. Olson). The library was screened with the petunia EPSP synthase clone pMON9543 (Shah et al., 1986), leading to the isolation of a phage, λF10. The region of the phage exhibiting hybridization to the EPSP synthase cDNA was subcloned as three BgIII fragments into λB400 (a gift of H. Klee, see "Materials and Methods"). The restriction map of the subcloned region spanning an entire petunia EPSP synthase gene is illustrated in Fig. 3A.

The hybridization of the cDNA clone, which contains no BgIII sites, to three BgIII fragments of a single genomic clone was an indication that the petunia gene contained introns.
FIG. 3. Petunia EPSP synthase gene. A, The 9-kb region of petunia genomic clone AF10 containing the coding sequence of EPSP synthase. The map was constructed by restriction analysis of the entire plasmid, and of three BglII subclones (designated pMON9560, 9561, and 9562 as illustrated). Fragments of the subclones were inserted into M13 phage (Messing, 1983) and sequenced according to the strategy indicated by the arrows below the map. Dark boxes below the map indicate coding sequences, light boxes indicate 5'- and 3'-untranslated regions which are present in mature EPSP synthase mRNA. Numbers below the map indicate the sizes of introns and exons in base pairs. B, the intron/exon junction regions of the petunia EPSP synthase gene. There is some ambiguity in assigning the location of the junctions in each intron. The intron/exon boundary could occur on either side of each of the underlined bases. In each case we illustrate the division so that the ends of the introns conform to the GT/AG rule. The consensus sequence at the bottom of the figure is from Brown et al. (1986) and is included for comparison (see "Discussion").

**TABLE 1.**

<table>
<thead>
<tr>
<th>Exon</th>
<th>Intron</th>
<th>3′ Exon</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GAA</td>
<td>GTAAGT</td>
</tr>
<tr>
<td>2</td>
<td>AAG</td>
<td>GTAGTA</td>
</tr>
<tr>
<td>3</td>
<td>AAG</td>
<td>GAAAA</td>
</tr>
<tr>
<td>4</td>
<td>CAA</td>
<td>GTAAGT</td>
</tr>
<tr>
<td>5</td>
<td>CAG</td>
<td>GTATGG</td>
</tr>
<tr>
<td>6</td>
<td>AGG</td>
<td>GTAAGG</td>
</tr>
<tr>
<td>7</td>
<td>GAG</td>
<td>GTAAGG</td>
</tr>
</tbody>
</table>

*Consensus:* GTAAGT

**DISCUSSION**

**Sequence and Evolution of Plant EPSP Synthases**—The great similarity between the nucleotide and amino acid sequences of EPSP synthase of petunia and tomato is not surprising given that they are both members of the family Solanaceae. This conservation is most pronounced in the region coding for the mature protein, with the transit peptides showing considerably less sequence conservation (Fig. 2). Although the transit peptide sequences are diverged, both the petunia and tomato sequences are very rich in positively charged amino acids, as has been observed for the transit peptides of other proteins (Schmidt and Mishkind, 1986).

When comparisons are made between the sequences of the EPSP synthase coding regions of fungi, bacteria, and plants, a consistent pattern emerges (Fig. 7). The amino acid identity between either of the plant enzymes, or either of the bacterial enzymes and the available fungal sequences is approximately 38%. The similarity between the plant and bacterial enzymes is much greater, with 54% identity at the amino acid level. Because plants and fungi are both eukaryotic organisms, we had expected that their EPSP synthase enzymes would show more similarity to each other than they would to the bacterial enzymes. These data, however, would indicate that the divergence between the plant and bacterial EPSP synthases occurred some time later than the divergence between the plant and fungal EPSP synthases. The AROM complex of fungi is localized in the cytoplasm, whereas the plant enzymes are localized in the plastids. One hypothesis for the origin of plastids is that they are the descendants of endosymbiotic procaryotes (Margulis, 1970). If this is the case, then it is possible that the plant EPSP synthase sequences that we have isolated represent genes that migrated from the plastids to the nucleus (Weeden, 1981). Such genes would be expected
to be homologous to those of prokaryotes. A similar hypothesis has recently been proposed to explain the evolutionary relationships between the glyceraldehyde-3-phosphate dehydrogenase enzymes of plants, animals, and fungi by Shin et al. (1986). In contrast to our results they find that the glyceraldehyde-3-phosphate dehydrogenase of Escherichia coli shares homology with the cytosolic enzymes of plants and other eukaryotes, while the plastidic enzymes show special homology to the enzymes of thermophilic bacteria. Reconciliation between these two sets of data will require the sequence of EPSP synthase from a thermophilic bacterium.

If the above hypothesis concerning the origin of the sequences we have characterized is correct, then it is possible that the second set of EPSP synthase hybridizing sequences present in the genomes of petunia and tomato (Fig. 4) represents the original eukaryotic genes, which may no longer function.

Amino acids which are conserved between the EPSP synthases of the four species compared in Fig. 2 are distributed along the entire sequence. There are, however, two regions where relatively long stretches of conservation between the plant, E. coli, and Aspergillus sequences are found. These are amino acids 20-35 and 95-107 as numbered in the petunia sequence (Fig. 2B). In the first region, 13 of 16 amino acids are conserved between the plant and E. coli sequences, or between the plant and fungal sequences. No data are available concerning the function of this region. In the second region, 15 of 16 amino acids are conserved between the plant and bacterial sequences, and 12 of the 16 amino acids are also conserved in Aspergillus. It is notable that a mutation in this region of the Salmonella EPSP synthase gene (which is identical to the E. coli enzyme in this region) is reported to alter the binding of glyphosate (Stalker et al., 1985). It is possible that these amino acids form a portion of a binding site for glyphosate.

Structure of a Petunia Gene—A petunia EPSP synthase gene was isolated from a genomic library made from a glyphosate-resistant petunia cell line containing multiple copies of the gene. Using Southern blot analysis of leaf DNA, we have confirmed that this gene is not rearranged (Fig. 4). The sequence of the exons is identical to the sequence of the cDNA that we have isolated, demonstrating that this is an active EPSP synthase gene in the petunia cell line. The gene is interrupted by seven introns, two of which are large (Fig. 3A). Although short regions of identity between the 5'- and 3'- splice sites preclude precise definition of the splice junction (Fig. 3B), there is only one assignment in each case that allows conformity with the GT/AG rule (Fig. 3B) (Aebi et al., 1986). The bases just outside the splice junctions of the introns are in close agreement with the published consensus for plants (Brown et al., 1986), while those more distal to the junctions often diverge significantly from this consensus.

Expression of EPSP Synthase—Northern blot analysis indicates that in petunia the level of EPSP synthase mRNA is much higher in petals than in other organs tested. Even in newly emerging leaves, there is a clear need for rapid synthesis of aromatic amino acids to provide material for growth, the message is almost undetectable in total RNA (Fig. 5A). In contrast, we find that the level of EPSP synthase mRNA in tomato plants varies only slightly, if at all, between

---

**Fig. 4.** Hybridization pattern of genomic DNA and genomic clone. A, petunia leaf DNA (5 μg/lane, lanes 1-5) and AF10 DNA (150 μg/lane, lanes 4-6) were digested with EcoRI (lanes 1 and 4), HindIII and NcoI (lanes 2 and 5), or BglII (lanes 3 and 6), electrophoresed on an agarose gel, and transferred to ZetaBind membrane. The filter was hybridized with 32P-labeled insert of pMON9556 (Shah et al., 1986) (Fig. 1A). This clone includes all of the cDNA except for the region of the first exon 5' of the initial EcoRI site in the coding sequence (Fig. 1A). Exposure times were 72 h for the genomic lanes, 24 h for the phage lanes. B, the same filter used in A was eluted and hybridized with 32P-labeled insert from pMON9531. This fragment contains almost the entire coding sequence of tomato EPSP synthase. The exposure time was 110 h. Numbers between panels indicate the sizes of marker DNA fragments in kb.

**Fig. 5.** Levels of EPSP synthase RNA in plant organs. A, total RNA (40 μg/lane) from the indicated organs of petunia seedlings or mature plants was electrophoresed on an agarose gel containing formaldehyde and transferred to GeneScreen membrane. The filter was hybridized with 32P-labeled insert from the petunia cDNA clone pMON9556 and exposed to x-ray film for 44 h. Leaf(NE), RNA from newly emerged leaves; Leaf(O), RNA from mature leaves. B, poly(A) selected RNA (2 μg/lane) from the indicated organs of tomato plants was electrophoresed and transferred as in A. The resulting filter was hybridized with the 32P-labeled tomato EPSP synthase cDNA. The filter was exposed to x-ray film for 72 h. Numbers between lanes indicate the migration of the 26 S and 18 S ribosomal RNAs.
FIG. 6. Start of transcription of petunia EPSP synthase mRNA. A, primer extension analysis. An oligonucleotide corresponding to bases 53–82 of the coding sequence of petunia EPSP synthase (within the transit peptide region) was end-labeled with \(^{32}P\), hybridized to RNA from the indicated source, and extended with reverse transcriptase. Sequencing reactions using \(^{35}S\)-labeled nucleotides were performed as described on an M13 clone of the 5'-end of the petunia transcriptase. The products extending to bases 53–82 of the coding sequence of petunia EPSP synthase (\(^{3}A\)), the noncoding strand; A, mRNA.

 FIG. 7. Phylogenetic tree of EPSP synthases. The percent identity of the EPSP synthase mature protein and DNA sequences of the illustrated organisms were determined from optimal alignments (Fig. 2B), and used to construct a phylogenetic tree. The numbers beside branch points indicate the degree of identity between the proteins of members of the two branches, and numbers in parentheses represent the degree of identity of the DNA sequences. The tree was arranged so that sequences with greatest similarity are closest together. The yeast sequence is from Duncan et al. (1987), the Salmonella sequence is from Stalker et al. (1985), and references for other sequences are as in Fig. 2B.

organs (Fig. 5B). Wild type petunia flowers are purple in color due to the accumulation of large quantities of anthocyanin pigments. It is possible that the high level of EPSP synthase expression seen in petunia petals is necessary for the production of these aromatic compounds. Tomato flowers do not accumulate these pigments. We hypothesize that increased expression of EPSP synthase in the petals is a property of plants with purple flowers, rather than a general characteristic of the Solanaceae. We find that varieties of petunia with white flowers still accumulate high levels of EPSP synthase message in the petals. These varieties probably arose as mutants from the normally purple petunia plants so that it is not surprising that they maintain the regulation of EPSP synthase expression.

Primer extension analysis (Fig. 6) reveals a complex pattern of transcription start sites in the MP4-G cell line. Transcription initiates primarily between bases -33 and -37, but minor bands can be seen on the gel, indicating the existence of transcripts with longer 5'-flanking regions. These bands are present even when the temperature or salt conditions used in the hybridization or extension reactions are altered significantly (data not shown). In addition, S1 nuclease protection experiments show a similar pattern of start sites, confirming the accuracy of the primer extensions (data not shown). In leaves, EPSP synthase mRNA appears to start nearly equally at several of the locations seen in the MP4-G cells. In contrast essentially all of the EPSP synthase mRNA in petals begins at the -33 to -37 start site. Even in long exposures there is no evidence of the longer primer extension products in the petal lanes (data not shown).

Since we find two sets of sequences in the petunia genome that hybridize to the EPSP synthase cDNA, it is possible that two genes are responsible for this pattern of transcripts. We consider this to be unlikely. Only one gene is amplified in the MP4-G cells, yet RNAs starting at all of the sites are present at elevated levels. In addition, the primer we have used hybridizes to the transit peptide region of the gene. This is the region that evolves most rapidly (see above) so that it is improbable that this primer would hybridize to the product of a diverged second gene which hybridizes only weakly to the entire cDNA (Fig. 4A). It appears, therefore, that the low level of expression in leaves results from a relatively imprecise initiation of transcription at several sites on a single gene. In petals, specific induction or derepression of transcription from one of these sites leads to accumulation of a high level of mRNA. The pattern in the MP4-G cells is intermediate between that seen in leaves and petals, although the minor bands are more intense due to the amplification of the EPSP synthase genes. It is possible that the relative utilization of the transcription initiation sites in this line is just an amplified version of the normal callus pattern. Alternatively, selection for resistance to glyphosate in this line may have led to the activation of this gene in addition to gene amplification.

Although we have made considerable progress in the characterization of EPSP synthase genes in plants, additional research will be necessary before we fully understand the evolution and expression of this essential gene. We are currently characterizing EPSP synthase genes from other plant families. Information gained from these studies may aid us in our evolutionary analysis. In addition, gene construction and
modification in conjunction with plant transformation should allow better understanding of the mechanism of regulation of the petunia gene.

Acknowledgements—We thank Robert Fraley and Ernest Jaworski for their support of this work, and Harry Klee and Guy della-Cioppa for helpful comments on the manuscript.

REFERENCES