Bot. Bull. Acad. Sin. (2001) **42:** 231-241

Potato and tomato *Forever Young* genes contain class-I patatin promoter-like sequences

Hongyong Fu[1,*], Ji Du[2,4], Junqi Song[3], Jiming Jiang[3], and William D. Park[2,*]

[1]*Institute of Botany, Academia Sinica, 128, Sec 2, Academy Rd., Nankang, Taipei, Taiwan 11529, Republic of China*

[2]*Department of Biochemistry and Biophysics, Texas A & M University, College Station, Texas 77843, USA*

[3]*Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706, USA*

(Received December 21, 2000; Accepted July 3, 2001)

Abstract. Class-I patatin genes from potato, encoding the major storage protein in tubers, contain promoter repeat sequences, which are critical for developmental and metabolic regulation. Previously, RFLP mapping indicated the patatin genes of potato and tomato reside in orthologous loci near the end of the long arm of chromosome VIII. Interestingly, a DNA sequence homologous to the promoter of the potato class-I patatin gene was mapped to the middle of chromosome III in the tomato and potato genomes [Ganal et al., Mol. Gen. Genet. 225, 501-509 (1991)]. We isolated and characterized genomic clones containing the class-I Patatin Promoter-Like sequences (*PPL*) from tomato, *S. bulbocastanum*, and *S. tuberosum*. Sequence analyses reveal that these *PPL* sequences are highly conserved, as compared to the repeat sequence of the patatin promoter, particularly with respect to the critical regulatory *cis*-elements. These *PPL* sequences are located upstream of the translational initiation site of *Forever Young* genes, which encode proteins belonging to a large family of oxidoreductases with conserved structural motifs. The *Forever Young* genes from tomato and *S. bulbocastanum* both contain eight exons and seven introns. Six of the introns are uniquely large, in the range of 418~2,092 bp. Sequence conservation suggests that the class-I patatin repeat and the *PPL* from *Forever Young* genes may have a direct phylogenetic relationship and may regulate expression of their respective genes through common *cis*- and *trans*-factors.

Keywords: *Forever Young*; Oxidoreductase; Patatin; Potato; Promoter evolution; Tomato.

Introduction

Patatin is a family of glycoproteins that have lipid acyl hydrolase and transferase activities (Andrews et al., 1988) and account for up to 40% of the total soluble protein in potato tubers (Paiva et al., 1983). In potato, the patatins are encoded by a gene family with ~10-15 members per haploid genome. These genes can be divided in about equal numbers into two classes, class-I and class-II, based on the presence or absence of a 22-bp insertion in the 5´ UTR. Whereas these two classes of genes have conserved coding regions, the 5´ flanking sequences are highly diverged (Pikaard et al., 1987; Mignery et al., 1988). The class-I genes encode the majority of tuber patatins and have an essentially tuber-specific expression pattern (Paiva et al., 1983). Besides being regulated developmentally, the class-I genes are also regulated by metabolic signal(s) derived from sucrose (Paiva et al., 1983; Rocha-Sosa et al., 1989; Wenzler et al., 1989; Jefferson et al., 1990). In contrast, patatins detected in roots and anther epidermal cells are primarily derived from class-II genes (Pikaard et al., 1987; Mignery et al., 1988). Unlike class-I genes, class-II genes are not sucrose-inducible (Köster-Töpfer et al., 1989; Liu et al., 1991).

Cis- and *trans*-factors responsible for developmental and metabolic regulation of the class-I patatin genes from potato have been examined extensively. An imperfect repeat in the 5´ region of the class-I patatin promoter contains separable *cis*-acting elements that are critical for "tuber-specific" and sucrose-inducible regulation. The sequence can be divided into two sub-regions (box A and box B), which are positioned from -552 to -514 and -626 to -592, respectively, for the distal repeat (in referring to the transcription start site as +1; Mignery et al., 1988). Motifs within these repeat sequences were found to interact with nuclear factors (Jefferson et al., 1990; Liu et al., 1990; Grierson et al., 1994).

Genetic and physical mapping analyses indicated the patatin genes show syntenic chromosomal organization in potato and tomato genomes and were mapped as a single cluster of loci near the end of the long arm of chromo-

---

[4]Current address: Emory University, Atlanta, GA 30322, USA.

[5]The amino acid and nucleotide sequence data reported for *S. bulbocastanum FEY*, *S. tuberosum FEY*, Tomato *FEY*, and Arabidopsis *FEY3*, will appear in the GenBank, EMBL and DDBJ databases under the accession numbers AF216836, AF250031, AF216835 and AF217275, respectively.

*Corresponding authors. H. Fu, Tel: 886-2-27823050 ext 422, Fax: 886-2-27827954, E-mail: hongyong@gate.sinica.edu.tw; W.D. Park, Tel: 979-845-8868, Fax: 979-862-4471, E-mail: wdpark@tamu.edu

some VIII (Ganal et al., 1991). In these loci, patatin genes exist at ~10-15 and ~3 copies in potato and tomato, respectively. Whereas the chromosome VIII patatin loci in tomato contain only class-II genes, the orthologous potato loci have further evolved additional class-I genes. Interestingly, a DNA sequence homologous to the class-I patatin promoter was mapped to the middle of chromosome III in both the potato and tomato genomes. As tuber-bearing potato species, including the cultivated potato, are generally believed to have evolved recently (Hawkes, 1990), an interesting possibility arises that the *PPL* sequence was recruited from chromosome III during the evolution of class-I patatin genes located in chromosome VIII in potato.

To examine this hypothesis, we isolated and characterized genomic clones containing the class-I patatin promoter-like sequence from tomato, a tuber-bearing wild potato species *Solanum bulbocastanum*, and *Solanum tuberosum*. Sequences are highly conserved for these *PPL*-containing genomic clones, except for two relatively diverged regions, including one located immediately upstream of the *PPL* sequence. This indicates these genomic clones are indeed from orthologous loci. The patatin promoter-like sequence contains regions that are highly conserved compared to the critical regulatory sequences of potato class-I patatin genes. Interestingly, the *PPL* is located immediately upstream of a gene encoding *Forever Young* oxidoreductase (*FEY*). Our results support the hypothesis that the patatin repeat-like sequences in *FEY* genes have a direct phylogenetic relationship with the class-I patatin regulatory elements and that closely related *cis*- and *trans*-factors participate in regulating the expression of both *FEY* genes and class-I patatin genes.

Materials and Methods

Construction and Screening of the Tomato λ-EMBL3 Genomic Library

The tomato genomic library was constructed using the λ-EMBL3 vector (Stratagene, La Jolla, CA). Genomic DNA was isolated from young tomato leaves (cv. VFN-8) according to the method described by Rogers and Bendich (1988). DNA was partially digested with *Sau*3AI and was size-fractionated by centrifugation on a sucrose gradient (Sambrook et al., 1989). Ligation of fractionated DNA (~15 kb in length) into the *Bam*HI site of the λ vector and packaging were performed according to the manufacturer's instructions (Stratagene). *Escherichia coli* CES200 was used as the host for the recombinant λ-EMBL3. Screening was performed as described by Sambrook et al. (1989) at reduced stringency (final washing at 60°C in 1 × SSC [0.1 M NaCl, 0.015 M sodium citrate]) by using a DNA fragment isolated from the class-I patatin promoter as a probe. The probe fragment is from -685 to -90 relative to the transcriptional start site of the patatin genomic clone pPS20 (Mignery et al., 1988). Polymerase chain reactions using the Vent polymerase (New England Biolabs, Beverly, MA) were performed to determine the sizes of upstream and downstream sequences that flank the patatin-like region for the isolated tomato clones. For upstream sequence amplification, the primer pair of the oligonucleotides, tn2 and a λEMBL3 vector-specific primer of either λL or λR, was used. For downstream sequence amplification, tn1 and either λL or λR was used. The 4.1-kb overlapping genomic fragment from T05, which was used to append the final contiguous 11,472 bp sequence, was amplified by PCR using the primer pair, tn15 and λL, and the high fidelity pfuTurbo™ DNA polymerase (Stratagene). The sequences from 5′ to 3′ for the oligonucleotides tn1, tn2, tn15, λR and λL are GTTACTGGTTCTACCAGCG GTATCGG, CCGATACCGCTGGTAGAACCAGTAAC, CACATGTAAACCATTCCATATG, CAGCGCACATGG TACAGCAAG, and CATGGTGTCCGACTTATGCCC, respectively.

Isolation of PPL-Containing Genomic Clones from Potato

The potato BAC library was constructed as described by Song et al. (2000) using genomic DNA isolated from a diploid (2n=2x=24) potato species *Solanum bulbocastanum* (clone PT29, provided by Dr. J. P. Helgeson at the University of Wisconsin-Madison). The library consists of 23,808 clones with an average insert size of 155 kb in length and is equivalent to approximately 3.7 haploid potato genomes. Screening of the library was performed at reduced stringency (final washing at 65°C in 2 × SSC) using an ~8.0 kb DNA fragment as a probe, which was amplified by PCR from the tomato genomic clone T17 using oligonucleotide primers tn1 and λL (Figure 1A). Four positive clones, 2J8, 2O9, 4M6, and 31A14 were obtained from screening the complete BAC library. The *PPL*-containing genomic fragment from a cultivated potato species *Solanum tuberosum* (cultivar Russet Burbank) was isolated by PCR using high fidelity pfuTurbo™ DNA polymerase and oligonucleotide primer pairs, which have sequences perfectly matched with the genomic sequences of both tomato and *S. bulbocastanum.* The 5′ primer P5′ is located 3,731 and 3,006-bp, respectively, upstream of the initiation codon for *FEY* from *S. bulbocastanum* and tomato, and the 3′ primer P3′ is located 84-bp downstream of the *FEY* start codon (Figure 1B). The sequences for the oligonucleotides P5′ and P3′ are: ggAAGCTTACC TTGGATGTGAATGTGAAAATATACC and GACCgg ATccAACCATCCTCGTAACCATTCCATCCATCC, respectively.

DNA Sequence Analysis

The 11,472 bp tomato sequence was determined and appended from a 10,832 bp sequence from T17 and an immediate downstream 640 bp sequence derived from T05. The upstream 10,832-bp sequence corresponds to a *Hind*III to *Pst*I region in T17 (Figure 1A). To determine the 10,832-bp upstream sequence, overlapping restricted fragments from the T17 λ-clone were subcloned into pBluescript (Stratagene). To determine the immediate downstream 640 bp sequence, a ~4.1 kb fragment overlap-

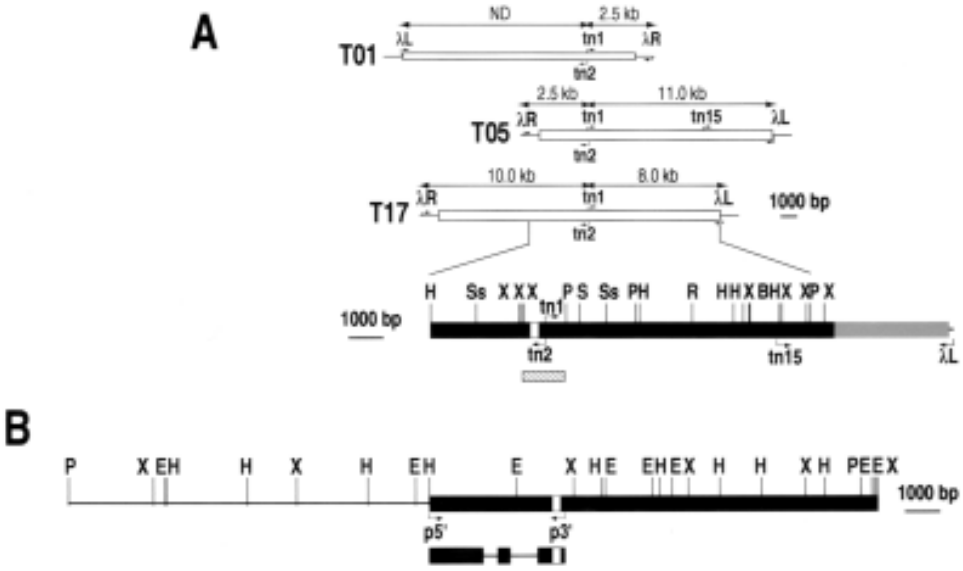


Figure 1. Genomic clones from tomato and potato containing patatin promoter-like sequences. (A) The restriction map of a tomato genomic fragment containing patatin promoter-like sequence. Three characterized overlapping λ genomic clones are shown above the map. The mapped fragment was derived from clone T17 and T05. The locations for the analyzed restriction enzymes were not determined downstream of the most 3′ *Xba*I site. Sizes in kilobase of regions flanking the *PPL* sequences for three genomic clones were determined by PCR and are indicated by double arrow lines. Positions and orientations of the primer oligonucleotides used are indicated by named bent arrows. The rectangular hatched box shown below the map is the location of the probe used previously for RFLP mapping (Ganal et al., 1991). ND, size is not determined. (B) The restriction map of a genomic fragment from *S. bulbocastanum* containing patatin promoter-like sequence. The fragment is appended from a ~23 kb *Pst*I fragment and an overlapping 2.0 kb *Xba*I fragment isolated from a BAC genomic clone. The schematic graph of black boxes and lines below the map represents a *PPL*-containing genomic fragment isolated by PCR from *Solanum tuberosum* (lines indicate deletion as compared to *S. bulbocastanum*). Arrows designate position and orientation of oligonucleotide primers used in the amplification reaction, and the names of the primers are indicated. The solid black and gray boxes indicate regions that have or have not been sequenced, respectively. The regions that are highly homologous to the patatin regulatory sequence are indicated as open boxes. B, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *Pst*I; S, *Sal*I; Ss, *Sst*I; X, *Xba*I.

ping with T17 was obtained from T05 by PCR using tn15 and λL as primers as well as the high fidelity pfuTurbo™ polymerase (Stratagene) and subcloned into pBluescript. A 1,612 bp sequence, which had an overlapped 972 bp region identical to T17, was determined for the ~4.1 kb fragment from T05. To determine the 12,755-bp potato sequence, the overlapping restricted fragments from the BAC clone 31A14 were cloned into pBluescript. The 2,646-bp *PPL*-containing genomic fragment from *S. tuberosum* was cloned into pGEM-T vector (Promega, Madison, WI). Sequences of both strands for each of the subcloned fragments were determined by primer walking using the dideoxy chain termination method and an automated DNA sequencer (ABI 373, Applied Biosystems, Foster City, CA) according to the manufacturer's instruction. Sequence assembly and analyses were performed using programs from the University of Wisconsin Genetics Computer Group (UW-GCG) software package (Deveraux et al., 1984).

Isolation of Arabidopsis Forever Young cDNA

The Arabidopsis *Forever Young* cDNA was isolated from a flower cDNA library (Lansberg erecta; ABRC stock number CD4-6) by PCR using the primers, AtFEY5′ and AtFEY3′ as well as the high fidelity pfuTurbo™ polymerase (Stratagene). The sequence for the Arabidopsis cDNA was determined by primer walking. The sequences from 5′ to 3′ for the oligonucleotides AtFEY5′ and AtFEY3′ are AG[cAtATG]AGTGACGAAACGACGTCATCTCC and TT[Gagctc]CTATTCGTGTTGTGCTCCATACCGGCATTG, respectively. Lower case-letters are mutated bases for engineering restriction sites to facilitate subsequent cloning. Bases in brackets are the engineered *Nde*I and *Sac*I site, respectively.

Results

Isolation of Tomato Genomic Clones Containing a Patatin Promoter-Like Sequence

We first screened for *PPL*-containing genomic clones from a tomato λ-EMBL3 library (1.0×10^6 plaque-forming units). Nine tomato genomic clones (including T01, T03-05, T07-08, T15-17) were isolated using a DNA probe derived from the promoter of a potato class-I patatin gene (pPS20; Mignery et al., 1988). The DNA probe contained both distal and proximal repeat sequences, critical for developmental and metabolic regulation of the patatin genes (Jefferson et al., 1990; Liu et al., 1990; Grierson et al., 1994). Results from restriction and sequencing analyses indicated that the isolated clones are overlapping and likely come from the same genetic locus (Figure 1A and data not shown). To select tomato clones for sequence analysis, PCR reactions were performed to determine the lengths of

sequences flanking the *PPL* sequence. Results for three genomic clones are shown in Figure 1A. The T17 clone has large 5′ and 3′ flanking regions, relative to the patatin promoter-like region, of 10 kb and 8 kb in length, respectively. In contrast, the T01 contains only ~2.5 kb 3′ flanking sequence, and the T05 contains only ~2.5 kb 5′ flanking sequence. To determine the patatin promoter-like sequence and to identify a possible adjacent gene, a contiguous 11,472 bp sequence was determined from T17 and T05. In this sequence, the 5′ 10,832-bp and the 3′ 640-bp were derived from T17 and T05, respectively.

Isolation of Potato Genomic Clones Containing the Patatin Promoter-Like Sequence

We isolated orthologous genomic clones from a diploid wild potato species, *Solanum bulbocastanum*, by screening a BAC genomic library using an 8 kb DNA probe derived from tomato T17 λ-clone. Four positive clones were obtained from screening a total of 23,808 clones with an average insert size of 155 kb, which are equivalent to approximately 3.7 haploid potato genomes (Song et al., 2000). One of the clones, 31A14, was further examined by DNA blot analysis to identify a ~23 kb *Pst*I fragment that cross-hybridized with the T17 probe. As shown in Figure 1B, the restriction map for the BAC clone 31A14 was partially determined, which includes the 23-kb *Pst*I fragment and a downstream overlapping 2.0-kb *Xba*I fragment. To determine the patatin promoter-like sequence and to identify a possible adjacent gene, the sequence of a 12,755-bp region was determined for the assembled *Pst*I-*Xba*I fragment of 31A14 (Figure 1B). A 2,646 bp *PPL*-containing genomic fragment was subsequently isolated by PCR from *S. tuberosum*, and its sequence was determined (Figure 1B).

PPLs are Highly Homologous to the Class-I Patatin Regulatory Sequences

The isolated genomic clones from tomato, *S. bulbocastanum* and *S. tuberosum* contain sequence regions of 250-, 234-, and 247-bp, respectively, which are highly homologous to a 160-bp region (the distal repeat) in the potato class-I patatin promoter (Figure 2A). Relative to the first base of the patatin translation initiation codon, the conserved region in class-I promoters is located between position -642 and -483. This region contains two short sequences (box B and A, underlined in Figure 2A), which also exist as an imperfect repeat (proximal repeat) between -269 and -195 (Bevan et al., 1986; Mignery et al., 1988). The two repeat sequences in the class-I patatin promoter were shown previously to be critical for developmental and metabolic regulation (Jefferson et al., 1990; Liu et al., 1990; Grierson et al., 1994). As described by Grierson et al. (1994), these repeats contain sequences that bind to nuclear factors (BBBF and BABF for Box B Binding Factor and Box A Binding Factor, respectively), and contain critical sequences for responding to sucrose induction (SURF for Sucrose Responsive Factor) (Figure 2A).

The *PPL* sequences have been divided into six sub-regions (I to VI), with regions III and V representing sequences not found in the 160-bp region of the class-I patatin promoter sequence. The sequence identity of *PPL* and the patatin promoter averages 93% when gaps are excluded for calculation and is 61% with gap residues in-

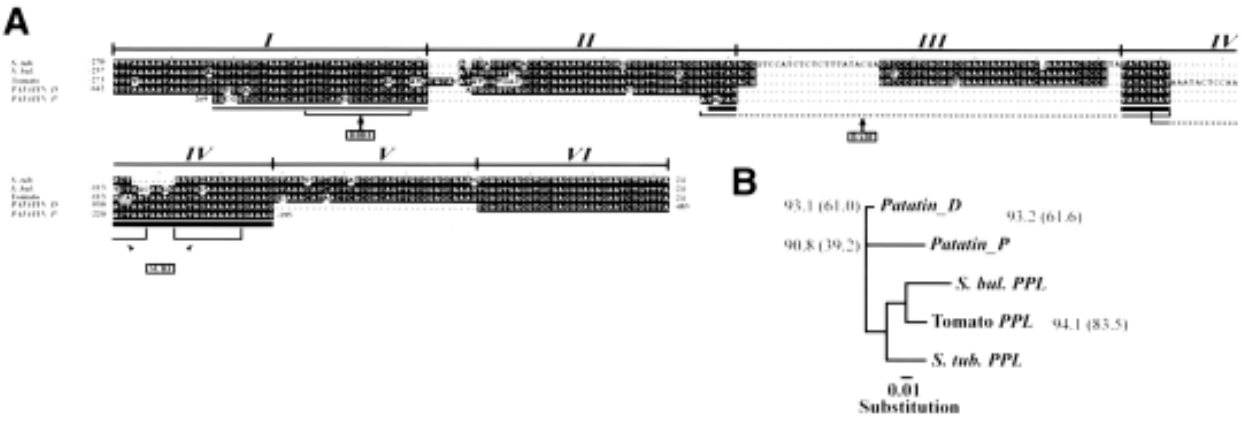


Figure 2. Comparison between patatin and the patatin promoter-like sequences. (A) Alignment of the class-I patatin 5′ region and patatin promoter-like sequences. The aligned sequences are sub-divided into six regions indicated as I through VI. The coordinates are referred to the translational initiation codons for either patatin or an adjacent gene containing the patatin promoter-like sequences (see below). The repeat boxes B and A are indicated by a double-line and a thick line, respectively (Bevan et al., 1986). Sequences involved in nuclear factor binding are indicated by brackets with arrows. Letters in reverse type are identical bases. BBBF, Box B Binding Factor; BABF, Box A Binding Factor; SURF, Sucrose Responsive Factor. Patatin_D and Patatin_P indicate the distal and proximal repeat sequences of the class-I patatin promoter, respectively. *S. tub*., *Solanum tuberosum*; *S. bul*., *Solanum bulbocastanum*. (B) A phylogram showing the sequence similarity among the patatin and patatin promoter-like sequences. The phylogram was generated by the UW-GCG computer programs Paupsearch and Paupdisplay, using the bootstrap analysis option. Sequence similarities between distal and proximal patatin repeats, among *PPL* sequences, are indicated to the right. Sequence similarity of the distal and proximal patatin repeat as compared to the *PPL* sequences are indicated to the left. Values indicate percent nucleotide identity when gap residues are excluded or included (in parentheses) for calculation.

cluded (Figure 2). *Cis*-elements for BBBF and SURF binding are highly conserved in the patatin promoter-like sequences although sequences for BABF binding are interrupted by insertion of region III (Figure 2A). An insertion of 11-bp or a deletion of 7-bp was also found in the distal SURF binding site for the *PPL* sequences from tomato and *S. tuberosum*, respectively. In addition to the homology of the patatin repeat sequence (box B and box A), adjacent sequences are also highly conserved as compared to the 160-bp patatin sequence (5′ part of region I, regions II and VI).

As shown in Figure 2B, phylogenetic analysis revealed that the patatin promoter-like sequences from tomato, *S. bulbocastanum* and *S. tuberosum* are more similar to one another than to either the 160-bp distal or the 75-bp proximal repeat from the potato class-I patatin promoter. The patatin promoter-like sequences average 94.1% identity when gap residues are excluded from calculation and 83.5 % identity when gap residues are included. In contrast, when compared to the distal 160-bp patatin repeat, the patatin promoter-like sequences average only 61.0% identity when gap residues are calculated and 93.1% identity when gap residues are excluded. Also, comparisons with the proximal 75-bp patatin repeat show the patatin promoter-like sequences average only 39.2% identity when gap residues are calculated and 90.8% when excluded.

Sequence Comparison of PPL-Containing Loci from Tomato, S. bulbocastanum, and S. tuberosum

Sequence similarity of the genomic regions containing the patatin promoter-like sequence from the *S. bulbocastanum* and tomato was analyzed by the dot-plot method. The 12,755-bp sequence, which is from *Hind*III to *Xba*I (Figure 1B), from *S. bulbocastanum* was compared to a corresponding region from the tomato genomic clone. As shown in Figure 3, the two sequences are highly similar. When gap residues are not included in the calculation, the two sequences are 91.3% identical. Only two regions show no sequence similarity. Interestingly, the 5′ dissimilar region is located immediately upstream of the patatin promoter-like sequence. The 3′ dissimilar region is primarily due to a ~700-bp insertion existing in the potato sequence. This insertion contains a 132-bp sequence that is highly homologous to tomato SINE transposable elements (SINE1-3) located in a gene encoding farnesyl-protein transferase β subunit (Rebatchouk and Narita, 1997). The 132-bp sequence is flanked by 15-bp direct repeats, which likely represent the duplicated target site. The insertion in the *S. bulbocastanum* sequence is located in the sixth intron of an adjacent gene (see below).

The 2,646-bp *PPL*-containing genomic fragment from *S. tuberosum* is also highly homologous to the genomic clones from both tomato and *S. bulbocastanum*. Compared to the *PPL*-containing genomic clone of *S. bulbocastanum*, the genomic fragment from *S. tuberosum* is 93.1% identical when the gap residues are excluded for calculation and is 60.9% identical when gap residues are included. The two sequences have co-linear sequence similarity except for two large deletions (364-bp and 785-bp) that exist upstream of the *PPL* in the *S. tuberosum* genomic fragment. When compared to the tomato *PPL*-containing genomic clone, the genomic fragment from *S. tuberosum* is 87.8% identical when the gap residues are excluded for calculation and is 65.1% identical when gap residues are included. The region of least similarity between the tomato and *S. tuberosum* genomic sequences is also immediately upstream of the *PPL*.

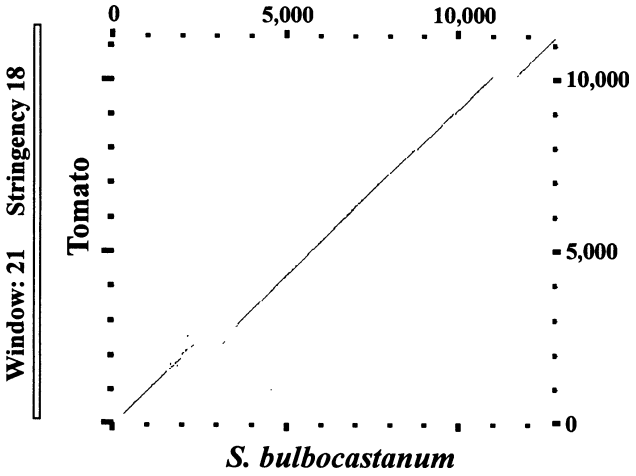


Figure 3. Sequence comparison of genomic sequences containing the patatin promoter-like sequence from *S. bulbocastanum* and tomato. Sequence similarity of corresponding genomic regions containing the patatin promoter-like sequence from *S. bulbocastanum* and tomato were analyzed by a dot-plot matrix using window and stringency of 21 and 18 residues, respectively. The 12,755 bp sequence of *S. bulbocastanum* (from *Hind*III to *Xba*I) was compared to a corresponding 11,228 bp region (also from *Hind*III to *Xba*I) from the tomato sequence.

High sequence similarity indicates that the *PPL*-containing genomic clones from tomato, *S. bulbocastanum*, and *S. tuberosum* are indeed from orthologous loci. In agreement, the *PPL* sequence was confirmed to be located in chromosome III by fluorescent *in situ* hybridization (data not shown) as predicted by previous RFLP mapping in both tomato and potato (Ganal et al., 1991).

The Patatin Promoter-Like Sequences Reside Next to the Forever Young Coding Sequence

The patatin promoter-like sequences from tomato, *S. bulbocastanum*, and *S. tuberosum*, may regulate expression of an adjacent gene. To identify the adjacent gene, the genomic sequences from potato and tomato were searched against the databases. The search identified closely related cDNA and genomic sequences from rice, *Medicago truncatula*, and Arabidopsis, which encode an oxidoreductase gene, designated as *Forever Young, FEY* (Callos et al., 1994; Wilson and Cooper, 1994). The *FEY* cDNAs from rice and Medicago were truncated at the 5′ end. Furthermore, the sequence of the only likely full-length *FEY* cDNA from Arabidopsis (designated here as

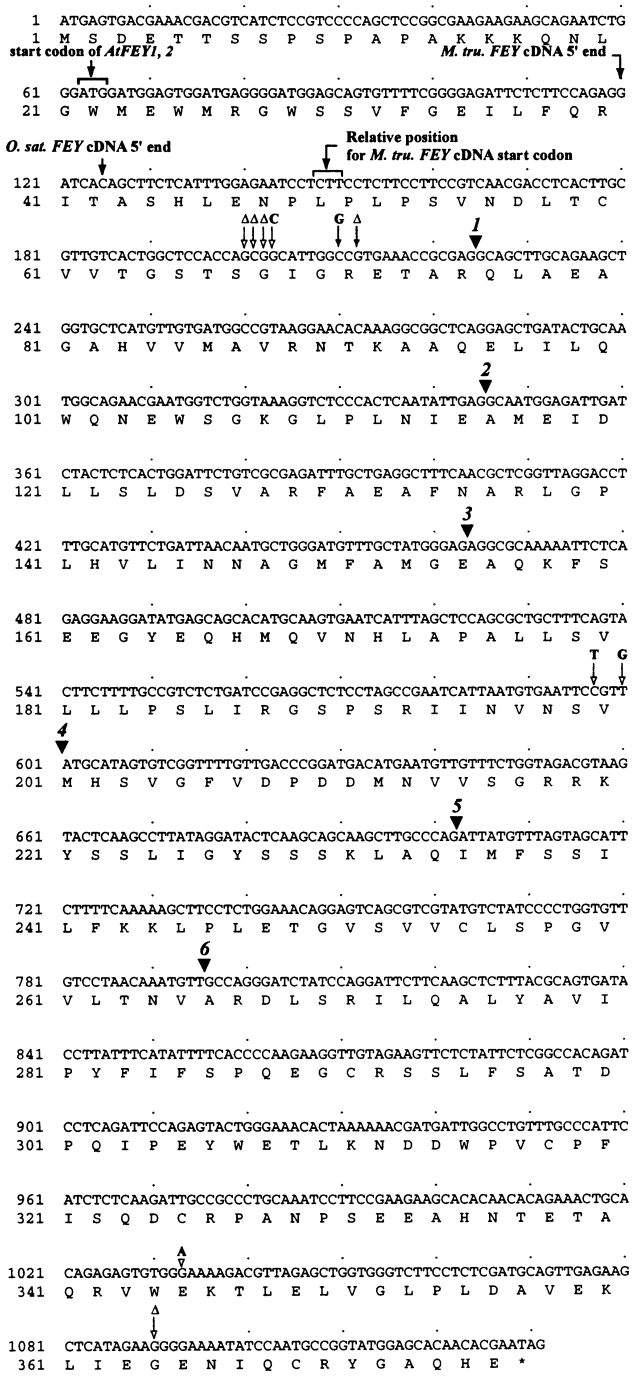


Figure 4. Nuceotide and deduced amino acid sequence of an Arabidopsis *Forever Young* cDNA, *AtFEY3*. The Arabidopsis *FEY3* cDNA was isolated from a cDNA library by PCR using a high fidelity polymerase. Mismatched residues as compared to *AtFEY1*, *AtFEY2* (Callos et al., 1994), or both are indicated by solid arrows, open arrows, and open arrow heads, respectively. The substituted bases or deletion (△) in *AtFEY1* and *AtFEY2* are indicated. The truncated positions of 5′ end for the rice and *Medicago FEY* cDNAs are indicated by arrows. Brackets with arrows indicate the wrong annotated start codons for the Arabidopsis *FEY1* and *FEY2* and for the Medicago *FEY* cDNA. The solid triangles indicate intron positions determined by comparison with a *FEY* genomic sequence (Callos et al., 1994).

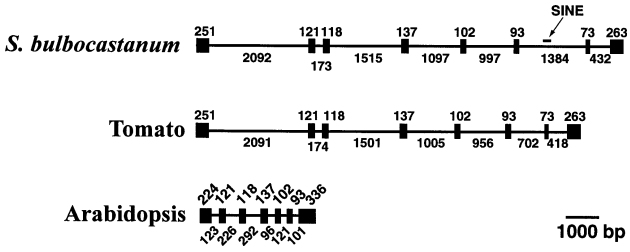


Figure 5. Exon and intron organization of *FEY* genes from *S. bulbocastanum*, tomato, and Arabidopsis. Exons and introns of *Forever Young* genes from *S. bulbocastanum*, tomato and Arabidopsis are drawn schematically to scale. Black boxes and lines indicate exons and introns, respectively. Numbers above or below corresponding exons and introns indicate lengths in bp. For the first and last exons, the lengths shown contain only the coding regions. The bar above the potato *FEY* indicates the position of the 132-bp SINE element.

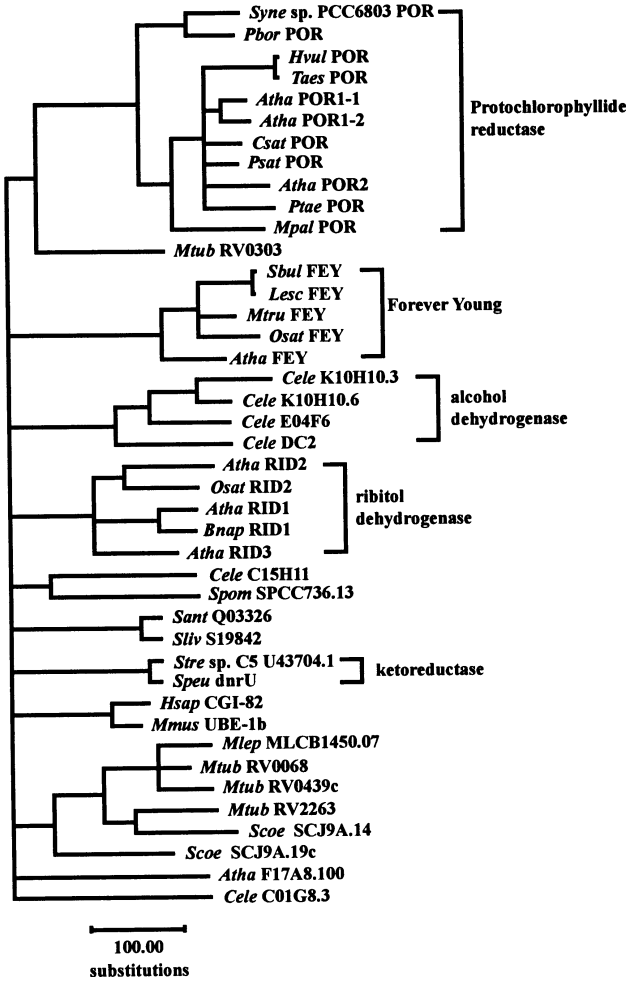
At*FEY1*; Callos et al., 1994) has a number of mismatched residues with two Arabidopsis *Forever Young* genomic sequences, designated here as At*FEY2* (Callos et al., 1994) and At*FEY3* (accession number, AL035602) from the same species. The primary concern is a single base pair deletion, which exists only in the Arabidopsis *FEY1* cDNA but not in *AtFEY2*, *AtFEY3* or any of the other characterized *FEY* genes from other species including rice, Medicago, tomato, *S. bulbocastanum* and *S. tuberosum*. A single base pair deletion was also found near the stop codon in *AtFEY2*. To clarify this, we isolated an apparently full-length Arabidopsis *FEY3* cDNA from a flower cDNA library (Figure 4). The sequence of the Arabidopsis *FEY3* cDNA is identical to the overlapped exons (exon 1 to 4) of the *FEY3* genomic sequence. The designated translational initiation codon for the At*FEY3* cDNA is very likely to be correct because it is the nearest ATG immediately upstream of the conserved sequence in Arabidopsis *FEY3* cDNA as compared to all other *FEY* sequences. In addition, an in frame stop codon was found immediately upstream of the putative start codon.

The junctions of all exons and introns for the potato and tomato *FEY* genes were easily identified, by comparing their sequences to that of the Arabidopsis At*FEY3* cDNA (Figure 5). Except for the last intron (intron 7), absent in Arabidopsis *FEY2*, the presence and location of all introns are completely conserved in *FEY* genes from tomato and *S. bulbocastanum*. Both sizes and sequences of the corresponding introns from *S. bulbocastanum* and tomato *FEY* genes are highly conserved. The only exception is the sixth intron of *S. bulbocastanum*, in which a ~0.7 kb insertion including a SINE element was found. The patatin promoter-like sequence is located 23-bp upstream of the translational initiation codon of the *FEY* genes from both tomato and *S. bulbocastanum*. Similarly, *PPL* in *S. tuberosum* also appears to reside 23-bp upstream of the *Forever Young* gene. The *PPL*-containing genomic fragment isolated from *S. tuberosum* contains most of the first exon of *FEY* gene and was found by PCR to have the rest

Table 1. Homology of the derived amino acid sequences of *FEY* genes from various species[a].

	Tomato	Arabidopsis	Rice	Medicago
S. bulbocastanum	98.4 (99.0)	74.5 (80.4)	72.1 (79.9)	79.8 (86.3)
Tomato		74.5 (80.4)	73.0 (80.5)	80.1 (86.6)
Arabidopsis			67.8 (76.2)	72.0 (81.3)
Rice				74.8 (82.6)

[a]Percent amino acid sequence identity or similarity (in parenthesis) was determined by UW-GCG Bestfit program.



of the *FEY* gene attached immediately downstream (data not shown).

High sequence conservation among the *FEY* genes from various species indicates these genes are likely homologs. As shown in Table 1, the *FEY* genes from tomato and *S. bulbocastanum* are 98.4% and 99.0% identical and similar in amino acid sequence, respectively. Also, the *FEY* genes from *S. bulbocastanum* and tomato are on average 75.7% and 82.4% identical and similar, respectively, when compared to the *FEY* genes from Arabidopsis, rice, and Medicago. In addition, *FEY* genes from Arabidopsis, Medicago and rice have an average of 71.5% and 80.0% sequence identity and similarity.

FEYs Belong to a Superfamily of Oxidoreductase Genes

A large set of sequences was identified from database searches for the derived amino acid sequences of *FEY* genes from both *S. bulbocastanum* and tomato. These sequences include protochlorophyllide oxidoreductases, alcohol dehydrogenases, ribitol dehydrogenases, and ketoreductases from various prokaryote and eukaryote species. Many of these sequences are derived from various genome projects. As shown in Figure 6, phylogenetic analyses place these sequences into distinct groups. Alignment analyses of a representative member from each branch identified six conserved structural motifs as shown in Figure 7. Five of these motifs are conserved with motifs of a known protein fingerprint designated as GDHRDH (for glucose/ribitol dehydrogenase; Yamada and Saier, 1987).

Figure 6. A phylogram showing the amino acid sequence relationships for a family of related oxidoreductases. A collection of related peptide sequences were identified from databases and analyzed by Paupsearch and Paupdisplay programs (UW-GCG). Abbreviations for species: *Atha*, *Arabidopsis thaliana*; *Bnap*, *Brassica napus*; *Cele*, *Caenorhabditis elegans*; *Csat*, *Cucumis sativas*; *Hsap*, *Homo sapiens*; *Hvul*, *Hordeum vulgare*; *Lesc*, *Lycopersicon esculentum; Mlep*, *Mycobacterium tuberculosis*; *Mmus*, *Mus musculus*; *Mpal*, *Marchantia paleacea*; *Mtru*, *Medicago truncatula*; *Mtub*, *Mycobacterium tuberculosis*; *Osat*, *Oryza sativa*; *Pbor*, *Plectonema boryanum*; *Psat*, *Pisum sativum*; *Ptae*, *Pinus taeda*; *Sant*, *Streptomyces antibioticus*; *Scoe*, *Streptomyces coelicolor*; *Sliv*, *Streptomyces lividans*; *Speu*, *Streptomyces peucetius*; *Spom*, *Schizosaccharomyces pombe*; *Stre*, *Streptomyces*; *Sbul*, *Solanum bulbocastanum*; *Syne*, *Synechocystis*; *Taes*, *Triticum aestivum*. Accession numbers for the analyzed sequences: *Atha* F17A8.100, CAB38642; *Atha* FEY, AF217275; *Atha* POR1-1, P21218; *Atha* POR1-2, AAC49043; *Atha* POR2, AAB97702; *Atha* RID1, AAC23625; *Atha* RID2, AAB63619; *Atha* RID3, CAA20464; *Bnap* RID1, S42651; *Cele* C01G8.3, AAB37640; *Cele* C15H11, CAB02732; *Cele* DC2; AAD14726; *Cele* E04F6, AAA68362; *Cele* K10H10.3, CAB05779; *Cele* K10H10.6, CAB05784; *Csat* POR, S20941; *Hsap* CGI-82, AAD34077; *Hvul* POR, P13653; *Lesc* FEY, AF216835; *Mlep*, RVMLCB1450.07, CAA16249; *Mmus* UBE-1b, BAA82657; *Mpal* POR, BAA31693; *Mtru* FEY, L22766; *Mtub* RV0068, CAA16249; *Mtub* RV0303, CAB09592; *Mtub* RV0439c, CAA17396; *Mtub* RV2263, CAA17300; *Osat* FEY, AF093628; *Osat* RID2, BAA83360; *Pbor* POR, BAA25993; *Psat* POR, S20941; *Ptae* POR, S30169; *Sant* Q03326, Q03326; *Scoe* SCJ9A.14, CAB53275; *Scoe* SCJ9A.19c, CAB53280; *Sliv* S19842, S19842; *Speu* dnrU, AAD04717; *Spom* SPCC736.13, CAA19277; *Stre* sp. C5 U43704.1, AAB08016; *Sbul* FEY, AF216836; *Syne* sp. PCC6803 POR, BAA10580; *Taes* POR, S39394.

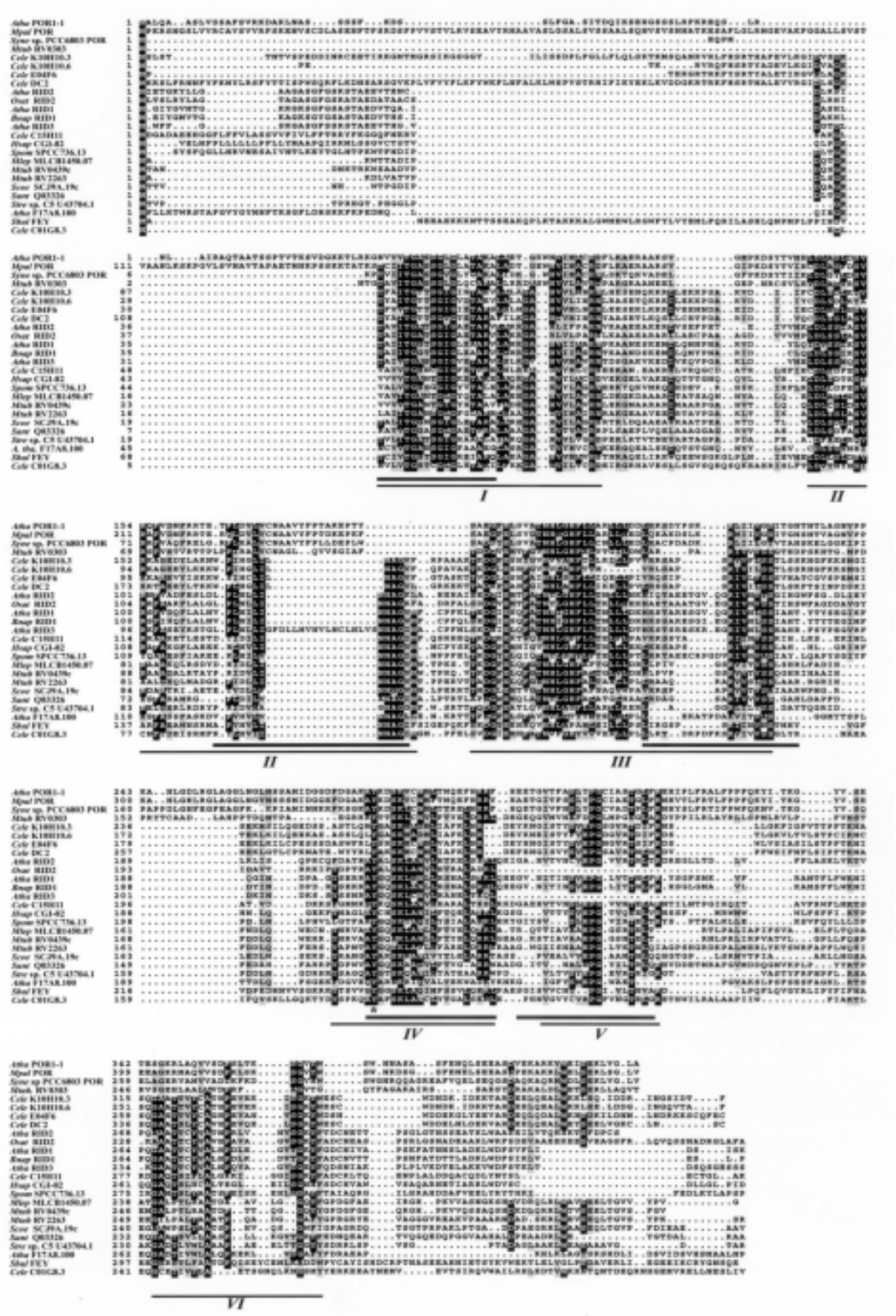


Figure 7. Alignment of amino acid sequences of a family of related oxidoreduactases. The amino acid sequences of *FEY* gene from *S. bulbocastanum* (*Sbul FEY*) are aligned with sequences of a representative member from each branch of the family of the oxidoreductase genes (Figure 6). The six conserved regions identified are underlined and indicated as motifs I through VI. The first five regions contain subsequences, which correspond to the first five elements of a previously identified protein fingerprint named GDHRDH (Yamada and Saier, 1987) specified by double lines. The right border of motif II includes a highly conserved NNAGI sequence. The Tyrosine residue in YXXXK of motif IV is important for subunit binding. Motif I contains critical residues, which may be involved in binding to the cofactor NADPH (Armstrong et al., 1995). Species abbreviation and accession numbers for the aligned sequences are the same as in Figure 6.

Discussion

The class-I patatin genes produce approximately 40% of the total protein in potato tubers, but are not normally expressed in other tissues. Since most close relatives of potatoes, such as tomato and pepper, do not have tubers, the evolution of the "tuber-specific" class-I patatin genes represents a very interesting biological question. A high degree of synteny has previously been shown between the chromosomes of potato and tomato and that all of the patatin coding sequences in both species map to a single region near the end of chromosome VIII (Ganal et al., 1991). This same region also contains class-II patatin promoter sequences in both tomato and potato. However, the location for the sequence homologous to the class-I promoter differs. In potato, sequences homologous to class-I patatin promoters mapped to the end of chromosome VIII at the same locus as patatin structural genes, as expected. However, sequences homologous to class-I patatin promoter (*PPL*) also mapped near the middle of chromosome III to a locus that lacked patatin structural genes. In tomato, sequences homologous to the class-I patatin promoter were found only at the corresponding locus on chromosome III. An interesting possibility is that *PPL* sequences on chromosome III involved in the regulation of other genes were duplicated and recruited during the evolution of the class-I patatin genes in potato. To examine this, we have isolated and characterized genomic clones containing patatin promoter-like sequences from tomato, *S. bulbocastanum*, and *S. tuberosum*. Fluorescent *in situ* hybridization (data not shown) confirmed the chromosome III location of the *PPL* sequences in potato. We found that the *PPL* sequences are highly conserved with the critical regulatory sequences of the class-I patatin genes from potato. The *PPL* sequences reside upstream of *Forever Young* genes, which share high sequence similarity to a large set of sequences encoding oxidoreductases, including protochlorophyllide oxidoreductase (e.g., Armstrong et al., 1995). Sequence conservation suggests the class-I patatin promoter and the *PPL* sequence have a direct phylogenetic relationship. In addition, the *PPL* may regulate expression of *Forever Young* through common *cis*- and *trans*-factors with the class-I patatin genes.

Since it is generally believed that the evolution of tuber-bearing potato species is a relatively recent event (Hawkes, 1990), the tuber-specific and sucrose-inducible class-I patatin genes are also likely to have evolved recently. Recruitment of the critical repeat sequences, which exist in 1-3 copies in characterized class-I patatin genes, is likely an important step during the evolution of the tuber-specific patatin genes. Phylogenetic analyses show that the *PPL* sequences located in *FEY* genes from tomato, *S. bulbocastanum*, and *S. tuberosum* are more similar to one another than to the critical repeat sequences in class-I patatin genes from potato. This indicates a duplication event of the ancestor sequence for *PPL*, and the class-I patatin repeat likely occurred before the separation of potato and tomato. The *PPL* locus may actually be the direct descendant of the ancestral sequence. A duplicated sequence for the class-I patatin repeat may have been derived from this locus. Interestingly, the sequence immediately upstream of the patatin promoter-like sequence is actually one of the two most diverged regions between the potato and tomato *FEY* genes. Furthermore, the second most diverged region in *FEY* genes contains a *SINE* transposable element (Figure 3 and 5).

The patatin promoter-like sequences in tomato, *S. bulbocastanum*, and *S. tuberosum* are located immediate upstream of sequences highly homologous to genes encoding *Forever Young* oxidoreductase from Arabidopsis, rice, and *Medicago truncatula* (Callos et al., 1994; Wilson and Cooper, 1994). High-level sequence similarity strongly indicates the tomato and *S. bulbocastanum* sequences encode *FEY* homologs and may be able to carry out a similar reduction reaction on unknown but specific substrate(s) in vivo. In Arabidopsis, interruption of *FEY* gene by a T-DNA insertion caused defects in development of shoot apical meristem (Callos et al., 1994). The in vivo substrate(s) of *Forever Young* reductase thus may play an important function in shoot meristem development. The related sequences and the conserved structural motifs identified (Figure 6 and 7) are likely to provide clues for future identification of the in vivo substrates for the *Forever Young* reductase.

Many of the sequence elements in the class-I patatin promoter, which are critical for binding to nuclear factors and for developmental and metabolic responses (Grierson et al., 1994), are conserved in the patatin promoter-like sequence of the *FEY* genes from tomato, *S. bulbocastanum*, and *S. tuberosum*. These include the two sub-repeat sequences, box A and box B, in class-I patatin promoter and the adjacent sequences (regions I, II, IV, and VI; Figure 2A), which contain many *cis*-elements with binding activities for nuclear factors. However, unique insertion sequences exist in the patatin promoter-like sequences (region III and V). The identification of conserved *cis*-elements in the *PPL* sequences from *FEY* genes suggests that these conserved sequences may mediate expression of the *FEY* genes through closely related transcription factors like the class-I patatin genes. However, a few regions in the *PPL* sequence, corresponding to *cis*-elements of class-I patatin promoter, are interrupted by insertion or deletion. The expected combinatorial action of these sequences and associated *trans*-factors may ultimately yield a different expression pattern for the *FEY* genes as compared to the class-I patatin genes in potato.

The class-I patatin promoter-like sequences in *FEY* genes from tomato, *S. bulbocastanum*, and *S. tuberosum* are located 23-bp immediately upstream of the translational initiation codon and very likely reside in the 5′ UTR regions. Alternatively, the sequence may reside in a leader intron located in the 5′ UTR. Important regulatory sequences located in regions other than 5′ flanking regions such as exons, introns, or 3′ flanking regions are not without precedents (e.g., Larkin et al., 1993; Fu et al., 1995a,b; Dickey et al., 1998). To address the actual functions of the *PPL* sequences, we have initiated *cis*-element shuf-

fling experiments between class-I patatin repeat sequences and the *Forever Young PPL* sequences using promoter-reporter gene fusion and transgenic approaches.

Acknowledgements. This work was supported by funds from the National Science and Technology Program for Agricultural Biotechnology from Academia Sinica and the National Science Council (8905A02), from a U. S. Department of Agriculture Competitive Grant No. 92-37301-7788, and from a Heritage Prize to H.F from the Li Foundation, CA, USA. We thank Mingchu Lee, Clarissa Hew, Yenfen Lee, and Yun-Chi Tang for technical support and Dr. Charles M. Papa for his critical reading of this manuscript.

Literature Cited

Andrews, D.L., B. Beames, M.D. Summers, and W.D. Park. 1988. Characterization of the lipid acyl hydrolase activity of the major potato (*Solanum tuberosum*) tuber protein, patatin, by cloning and abundant expression in a baculovirus vector. Biochem. J. **252:** 199-206.

Armstrong, G.A., S. Runge, G. Frick, U. Sperling, and K. Apel. 1995. Identification of NADPH:protochlorophyllide oxidoreductase A and B: A branched pathway for light-dependent chlorophyll biosynthesis in *Arabidopsis thaliana*. Plant Physiol. **108:** 1505-1517.

Bevan, M.W., R. Barker, A. Goldsbrough, M. Jarvis, T. Kavanagh, and G. Iturriaga. 1986. The structure and transcription start site of a major potato tuber protein gene. Nucl. Acids. Res. **14:** 4625-4638.

Callos, J.D., M. DiRado, B. Xu, F.J. Behringer, B.M. Link, and J.I. Medford. 1994. The *forever young* gene encodes an oxidoreductase required for proper development of the Arabidopsis vegetative shoot apex. Plant J. **6:** 835-847.

Deveraux, J., P. Haberberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387-395.

Dickey, L.F., M.E. Petracek, T.T. Nguyen, E.R. Hansen, and W.F. Thompson. 1998. Light regulation of Fed-1 mRNA requires an element in the 5´ untranslated region and correlates with differential polyribosome association. Plant Cell **10:** 475-484.

Fu, H., S. Kim, and W.D. Park. 1995a. High-level tuber expression and sucrose inducibility of a potato *Sus4* sucrose synthase gene require 5´ and 3´ flanking sequences and the leader intron. Plant Cell **7:** 1387-1394.

Fu, H., S. Kim, and W.D. Park. 1995b. A potato *Sus3* sucrose synthase gene contains a context-dependent 3´ element and a leader intron with both positive and negative tissue-specific effects. Plant Cell **7:** 1395-1403.

Ganal, M., M.W. Bonierbale, M.S. Roeder, W.D. Park, and S. D. Tanksley. 1991. Genetic and physical mapping of the patatin genes in potato and tomato. Mol. Gen. Genet. **225:** 501-509.

Grierson, C., J.-S. Du, M.T. Zabala, K. Beggs, C. Smith, M. Holdsworth, and M. Bevan. 1994. Separate *cis* sequences and *trans* factors direct metabolic and developmental regulation of a potato tuber storage protein gene. Plant J. **5:** 815-826.

Hawkes, J.G. 1990. The Potato: Evolution, Biodiversity, and Genetics Resources. Smithsonian Institution Press, Washington, D.C.

Jefferson, R., A. Goldsbrough, and M. Bevan. 1990. Transcriptional regulation of a patatin-I gene in potato. Plant Mol. Biol. **14:** 995-1006.

Köster-Töpfer M., W.B. Frommer, M. Rocha-Sosa, S. Rosahl, J. Schell, and L. Willmitzer. 1989. A class II patatin promoter is under developmental control in both transgenic potato and tobacco plants. Mol. Gen. Genet. **219:** 390-396.

Larkin, J.C., D.G. Oppenheimer, S. Pollock, and M.D. Marks. 1993. Arabidopsis *GLABROUS1* gene requires downstream sequences for function. Plant Cell **5:** 1739-1748.

Liu, X.J., S. Prat, L. Willmitzer, and W.B. Frommer. 1990. *Cis* regulatory elements directing tuber-specific and sucrose-inducible expression of a chimeric class I patatin promoter/GUS-gene fusion. Mol. Gen. Genet. **223:** 401-406.

Liu, X.-J., M. Rocha-Sosa, S. Hummel, L. Willmitzer, and W. B. Frommer. 1991. A detail study of the regulation and evolution of the two classes of patatin genes in *Solanum tuberosum* L. Plant Mol. Biol. **17:** 1139-1154.

Mignery, G.A., C.S. Pikaard, and W.D. Park. 1988. Molecular characterization of the patatin multigene family of potato. Gene **62:** 27-44.

Paiva, E.P., R.M. Lister, and W.D. Park. 1983. Induction and accumulation of the major potato tuber protein, patatin. Plant Physiol. **71:** 161-168.

Pikaard, C.S., J.S. Brusca, D.J. Hannapel, and W.D. Park. 1987. The two classes of genes for the major tuber protein, patatin, are differentially expressed in tubers and roots. Nucleic Acids Res. **15:** 1979-1994.

Rebatchouk, D. and J.O. Narita. 1997. Foldback transposable elements in plants. Plant Mol. Biol. **34:** 831-835.

Rocha-Sosa, M., U. Sonnewald, W. Frommer, M. Stratmann, J. Schell, and L. Willmitzer. 1989. Both developmental and metabolic signals activate the promoter of a class I patatin gene. EMBL J. **8:** 23-29.

Rogers, S. and A.J. Bendich. 1988. Extraction of DNA from plant tissues. *In* S.B. Gelvin, D.P.S. Verma and R.A. Schilperoort (eds.), Plant Molecular Biology Manual, Kluwer Academic Publishers, Dordrecht, pp. A6:1-10.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Song, J., F. Dong, and J. Jiang. 2000. Construction of a bacterial artificial chromosome library for potato molecular cytogenetics research. Genome **43:** 199-204.

Wenzler, H.C., G.A. Mignery, L.M. Fisher, and W.D. Park. 1989. Analysis of a chimeric class I patatin-GUS gene in transgenic potato plants: high level expression in tubers and sucrose-inducible expression in cultured leaf and stem explants. Plant Mol. Biol. **12:** 41-50.

Wilson, R.C. and J.B. Cooper. 1994. A nodulin cDNA with homology to protochlorophyllide reductase. Plant Physiol. **104:** 289-290.

Yamada, M. and M.H. Saier. 1987. Glucitol-specific enzymes of the phosphotransferase system in *E. coli*- Nucleotide sequence of the GUT operon. J. Biol. Chem. **262:** 5455-5463.

馬鈴薯及蕃茄 *Forever Young* 基因含有類似 class-I patatin 基因啟動子之序列

符宏勇[1] 杜杰[2,4] 宋俊岐[3] 蔣繼明[3] William D. Park[2]

[1] 中央研究院植物研究所

[2] Department of Biochemistry and Biophysics, Texas A & M University

[3] Department of Horticulture, University of Wisconsin

[4] Current address: Emory University, Atlanta

馬鈴薯塊莖主要儲藏蛋白 patatin 之 class-I 基因，其近端啟動子所含重複序列為該基因受發育及代謝調控之重要 *cis*-elements。先前限制酵素片斷多型性分析（RFLP）顯示馬鈴薯及蕃茄之 patatin 基因位於相似之基因座，即靠近第八條染色體長臂之末端。輿圖定位顯示馬鈴薯與蕃茄之第三條染色體中央含有與 *patatin* 基因近端啟動子重複序列極同源之基因座［Ganal *et al.*, Mol. Gen. Genet. 225, 501-509 (1991)］。我們自蕃茄、栽培種馬鈴薯（*S. tuberosum*）與野生種馬鈴薯（*S. bulbocastanum*）將此類似 class-I patatin 啟動子序列（*PPL* 代表 class-I Patatin Promoter-Like sequences）分離出來。序列比較顯示 *PPL* 含有與 patatin 基因上重要 *cis*-elements 極同源之序列。*PPL* 位於 *Forever Young* 基因轉譯起始點之上游，該基因轉譯出之蛋白屬於一含保守結構之氧化還原酵素群。蕃茄與 *S. bulbocastanum* 之 *Forever Young* 基因皆含有八個 exons 與七個 introns，其中六個 introns 皆格外的大，介於 418~2,092 bp 之間。序列之高度保守性指出 class-I patatin 基因之重複序列與 *Forever Young* 基因之 *PPL* 可能有直接之演化親緣關係，並可能透過類似之 *cis*- 及 *trans*- 因子來調控其相對應基因之表達。

關鍵詞： *Forever Young* ；氧化還原酵素； patatin ；馬鈴薯；啟動子演化；蕃茄。