

Molecular characterization and genetic mapping of DNA sequences encoding the Type I chlorophyll a/b-binding polypeptide of photosystem I in *Lycopersicon esculentum* (tomato)

E. Pichersky,^{1,2,5,*} N. E. Hoffman,² R. Bernatzky,³ B. Piechulla,⁴ S. D. Tanksley³ and A. R. Cashmore²
¹Laboratory of Cell Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA;
²Plant Science Institute, Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA;
³Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853, USA;
⁴Institut für Biochemie der Pflanze der Universität Göttingen, Untere Karspule 2, 34 Göttingen, Federal Republic of Germany;
⁵Permanent address: Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA (* author for correspondence)

Received 16 March 1987; accepted in revised form 12 May 1987

Key words: evolution, gene family, genomic clone, PSI CAB, RFLP mapping, Solanaceae

Abstract

We report the isolation and characterization of a tomato nuclear gene encoding a chlorophyll a/b-binding (CAB) protein of photosystem I (PSI). The coding nucleotide sequence of the gene, designated *Cab-6B*, is different at eight positions from that of a previously isolated cDNA clone derived from the *Cab-6A* gene, but the two genes encode identical proteins. Sequence comparison with the cDNA clone revealed the presence of three short introns in *Cab-6B*. Genetic mapping experiments demonstrate that *Cab-6A* and *Cab-6B* are tightly linked and reside on chromosome 5, but the physical distance between the two genes is at least 7 kilobases. *Cab-6A* and *Cab-6B* have been designated Type I PSI CAB genes. They are the only two genes of this branch of the CAB gene family in the tomato genome, and they show substantial divergence to the genes encoding CAB polypeptides of photosystem II. The Type I PSI CAB genes, like the genes encoding PSII CAB proteins, are highly expressed in illuminated leaf tissue and to a lesser extent in other green organs.

Introduction

The chlorophyll a/b-binding (CAB) proteins are constituents of the photosynthetic apparatus of higher plants. They form protein-pigment complexes which function in capturing light energy and in the distribution of excitation energy between photosystem I (PSI) and photosystem II (PSII) in the thylakoid membranes of the chloroplast [1]. The CAB polypeptides of both photosystems are structurally related to each other [6, 9]. At least two types of CAB polypeptides, designated Type I and

Type II, have been identified in tomato PSII, and each is encoded by multiple nuclear genes [17, 19, 20]. Genes for PSII Type I and Type II CABs have also been isolated and characterized in other plant species [5, 7, 8, 12–15, 24]. Nucleotide sequence comparisons demonstrate that the genes encoding Type I PSII CAB proteins diverged from the genes encoding Type II PSII CAB polypeptides at least before the divergence of monocotyledonous and dicotyledonous plants [19, 20, 22].

The isolation of genes encoding PSI CAB polypeptides has not yet been reported. Recently, we

isolated a cDNA clone encoding a tomato PSI CAB polypeptide which we designated "Type I" [11]. The overall homology of this protein to any tomato PSII CAB polypeptide is only 30%, but two segments within it, each about a quarter of its total length, show approximately twice this level of homology to the equivalent segments in the PSII CAB proteins. Here we report the isolation and characterization of a tomato gene encoding the Type I PSI CAB protein. In addition, we examined the mode of expression of Type I PSI CAB genes in various tomato organs and the number and chromosomal location of these genes in the tomato genome. The evolutionary relatedness of the Type I PSI CAB genes to genes in the other recognized branches of the CAB gene family is also discussed.

Materials and methods

Gene isolation and characterization

A ³²P-labelled tomato cDNA clone which encodes a PSI Type I CAB protein [11] was used as a probe in the screening of a tomato genomic library in the Charon 4 phage vector [17]. A single positively hybridizing phage was isolated and its tomato DNA insert was subcloned into pUC18 plasmids as previously described [17]. Southern blots [23] and nucleotide sequence determination [16] were carried out as described [17].

Genetic mapping

The position of the cloned DNA fragments on the genetic map of tomato was determined by the restriction fragment length polymorphism (RFLP) mapping technique [2, 3]. F₂s (40 individuals) of an interspecific cross between *L. esculentum* and *L. pennellii* were examined because of the relative lack of RFLPs among *L. esculentum* cultivars (see ref. [3] for detailed discussion of the methodologies used in setting up crosses, phenotype scoring and analyses of data).

Northern blots

mRNA was isolated from the different tissues and Northern blots were performed as previously described [21].

Results and discussion

Isolation and characterization of a tomato gene encoding a PSI Type I CAB polypeptide

We have previously reported the identification and characterization of five loci in the tomato genome which contain genes encoding PSII CAB polypeptides [17, 19, 20; 27]. These five loci, designated *Cab-1* through *Cab-5*, each reside on a different chromosome [17, 20, 27]. *Cab-1* and *Cab-3* are complex loci: they contain four and three genes, respectively, all encoding the Type I PSII CAB polypeptides [17]. *Cab-4* and *Cab-5* each contain a single gene encoding the Type II PSII CAB proteins [20]. The single gene in the fifth locus, *Cab-2*, which, unlike the genes in the *Cab-1*, 3, 4 and 5 loci, has not yet been physically isolated and characterized, also appears to encode a PSII CAB protein [27].

We have recently isolated a tomato cDNA clone encoding a PSI CAB protein. We have designated the gene from which it was transcribed *Cab-6A*, and the protein it encodes, CAB-6A, as a Type I PSI CAB [11]. We used the cDNA clone of *Cab-6A* to screen a genomic library of partially *EcoRI*-digested tomato DNA in the phage vector Charon 4 [17]. A single recombinant phage containing tomato DNA hybridizing to the probe was isolated. The nucleotide sequence of the single Type I PSI CAB gene present in this phage (see below) was determined and was found to differ from that of the sequence of the cDNA clone of *Cab-6A* (Fig. 1). We therefore designated the cloned gene *Cab-6B*. The two genes differ by eight nucleotides within the coding region but they encode identical proteins (Fig. 1). In addition, there is one nucleotide substitution within the 3' transcribed region, and one nucleotide substitution and two deletions/insertions

ATATCCTCAAAAATATCCACTTCATCTCCAGGTGGA CCAAT CACAAAACAGAAACCAGTCTCTATAG CCAA A 74
 ACAAATCTAAATTCAGATTTGCC TATA TCTCAAACTCCCTCCT TATA CATTCCACAATTACACCAATTTTTTTTC 146
 ATTCATATCCAAACTTTTTTTTGTACATCTTTTAAATACCAAAAAAAAAAAGAGGAAGAAGATATGGCTTCTAATACA 227
 * C - -- M A S N T
 L M S C G I P A V C P S F L S S T K S K F A A A M P
 TTGATGAGTTGTGAATTCAGCTGTTTGCCTCTTTCCCTCTCTTCTACCAAGTCCAAATTCGCCGCCGCGATGCCG 305
 EcoRI
 V S V G A T N S M S R F S M S A D W M P G Q P R P S
 GTGTCCGTTGGAGCTACCAACTCCATGTC TCGGTCTCCATGTCAGCTGATTGGATGCCCGGCAGCCCGTCCATLT 363
 C A
 Y L D G S A P G GT INTRON I AG
 TACCTCGACGGCTCAGCCCCGGTAAAGTAGTACTACTTCTCTGTTTATTATGTGATATGATGTATGTATGTATCTC 461
 D F G F D P L G
 ATACGATATTTAAGTTGTAGTATATTAAGTTGGATAACTTTATAAATGCAGAGATTTTGATTGATCCACTTGGT 539
 L G E V P A N L E R Y K E S E L I H C R W A M L A V
 TTAGGAGAAGTACCTGCAAAATTTGGAAAGATACAAGAATCTGAACCTATTCAATTGCAGATGGGCTATGCTTGTCT 617
 G T A A C T T C C T T A C C A T A A C A T A T A C G T T A T A T T T G A A C A T T C T T A A C G A A A T T G T T C T T C G C A A
GTAACTTCT INTRON II AG 695
 HindIII
 P G I I V P E A L G L G N W V K A Q E W A A I
 TTTTGTAGCCCGGAATCATAGTACCAAGGCCATTGGGCTTAGGTAATTGGGTCAGGCCCAAGAAATGGGCAGCCATT 773
 A
 P G G Q A T Y L G Q P V P W G T L P T I L A I E F L
 CCTGGTGGACAAGCTACATATTTGGGCCAGCCGT CCATGGG GAACACTTCTACTATTTTGGCTATTGAATTTCTA 851
 NcoI T
 A I A F V E H Q R S M E K D S E K K K Y P G G A F D
 GCCATAGCTTTTGTAGAACCACCAAAAGGAGTATGGAAAAAGATTCAGAAAAGAAAGTACCCCTGGTGGGGCTTTTGAC 929
 C T
 P L G Y S K D P A K F E E L K V K E I K N GT
 CCATTGGCTACTCTAAAGACCCCTGCTAAATTTGAAGAACCAAAAGTCAAGGAAATCAAGAAATGGT TAGAATATTAAC 1077
 T T
 INTRON III GT
 TCACCTCTCCTTATTATATTATGGATTTAAGTTATATATGCTGACAATATCAATCTCGATTATTTAGTTAGGT 1065
 R L A L L A F V G F C V Q Q S A Y P G T G P L E N L
 CGCTTGCATTGTGGCATTGTGGGATTTGTGTACAACAATCAGCATATCCAGGAACAGGACCATTGGAGAACTTG 1163
 A T H L A D P W H N N I G D V I I P K G I F P N
 GCAACTCACTTGGCTGACCCATGGCACAACAACATTGGGGATGTTATTATCCCTAAAGGCATTTCCCTAATTAAT 1241
 NcoI
 G T A C T A T T T A T C A T A T A G T A T A T T A G C C A A A A C A C C T T G T A A A A C T C C C A T G T A C T A A T A G A T T T C T A A T G G A A T 1319
 G G A C T T T C T G T G A T C C T T A G T C T A G T A G T G T T G T T G G T A A A G A G C T T T A A A T T A T T G G T A T A G A T G T G G T G G T G 1397
 T * DraI

Fig. 1. Nucleotide sequence of *Cab-6B* and the amino acid sequence of the encoded polypeptide. The sequence of the cDNA clone of *Cab-6A* [11] is shown below where it differs from that of *Cab-6B*. The asterisks indicate the positions of the first and last nucleotides of the cDNA clone. The positions of the introns are derived from comparison of the cDNA sequence with the genomic sequence. The first two nucleotides and the last two nucleotides (GT..AG) of each intron, and the location of some restriction sites referred to in the text are underlined. Probable regulatory sequences in the promoter region of the gene are boxed.

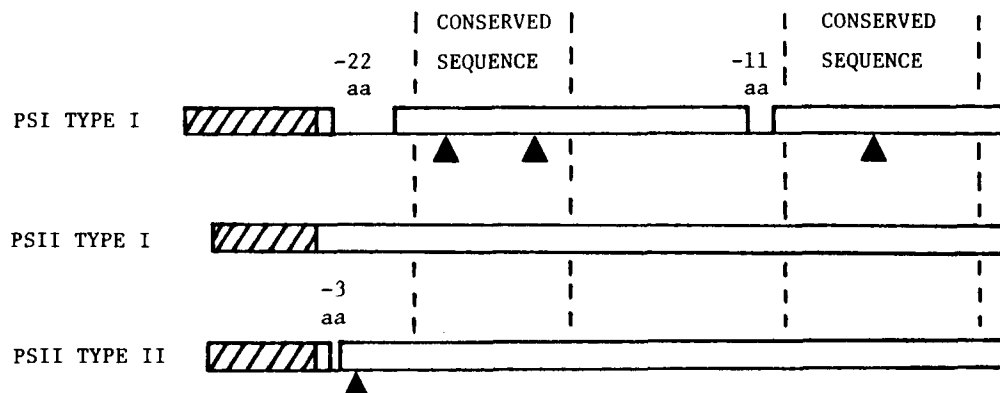


Fig. 2. A schematic diagram of the main structural features of the three types of CAB polypeptides whose genes have been characterized to date. The N-terminus is on the left. The lengths of the polypeptides range from 246 to 269 amino acids. The shaded area is the transit peptide (33–44 amino acids) which is removed after import into the chloroplasts. Segments which are strongly conserved in all three types are outlined. The locations of introns in the genes in respect to the protein sequences are indicated with triangles.

tions (of one and two nucleotides, respectively) in the 5' transcribed region (Fig. 1).

Comparison of the nucleotide sequence of *Cab-6B* with that of the cDNA clone for *Cab-6A* also revealed the existence in *Cab-6B* of three short introns of 107, 87, and 89 nucleotides, respectively (Fig. 1). The Type I PSII CAB genes do not contain introns [5, 13, 17], whereas the Type II PSII CAB genes contain a single intron [12, 24]. However, none of the introns in *Cab-6B* is located in the equivalent place of the single intron of the Type II PSII CAB gene (Fig. 2). All three introns in *Cab-6B* are found in regions of the gene encoding segments of the protein which have been found to be strongly conserved among PSI and PSII CAB polypeptides (Fig. 2; ref. [11]).

From consideration of protein structure and sequence homology, the cDNA sequence of *Cab-6A*, containing an open reading frame of 246 codons (ref. [11]; Fig. 1), was deemed complete. The completeness of the open reading frame is now confirmed by the observation that the nucleotide sequence of *Cab-6B* upstream from the first ATG of the 246-residue open reading frame (without the introns) contains no other ATG triplet in a 180-nucleotide stretch and does contain several in-frame stop codons (Fig. 1). Furthermore, two TATA boxes are found at nucleotides -97 to -94 and again at -117 to -114 upstream from the ATG

start codon, and two CCAAT boxes are found at nucleotides -145 to -141 and at -176 to -172 (Fig. 1). In addition, a third CCAAT sequence is found 15–20 nucleotides downstream to the downstream-most TATA sequence. Which of these sequences actually plays a role in the transcriptional activity of the *Cab-6B* gene remains to be determined. We note here that multiple TATA sequences have been found in the upstream regions of other tomato CAB genes [17] and also pea CAB genes [5, 26].

Southern blots of restricted tomato DNA using the cDNA clone of *Cab-6A* reveal two or more fragments hybridizing to the probe with every restriction enzyme used (Figs. 3A, 4A). The minimum number of fragments together with their minimum sizes indicate that the Type I PSI CAB protein is encoded by more than one gene in tomato, consistent with the isolation of the cDNA clone of *Cab-6A* and the genomic *Cab-6B* clone. Furthermore, the sizes of the two *DraI* fragments (2.4 and 1.6 kb, Fig. 3A, lane 1) suggest that each fragment contains a single gene. The *Cab-6B* gene was indeed localized to the 1.6 kb *DraI* fragment by restriction digest and nucleotide sequence determination of the cloned DNA (Figs. 1, 4B). The 2.4 kb *DraI* fragment must therefore contain the *Cab-6A* gene, but since no *DraI* site was found on the cDNA clone of *Cab-6A*, the locations of the two

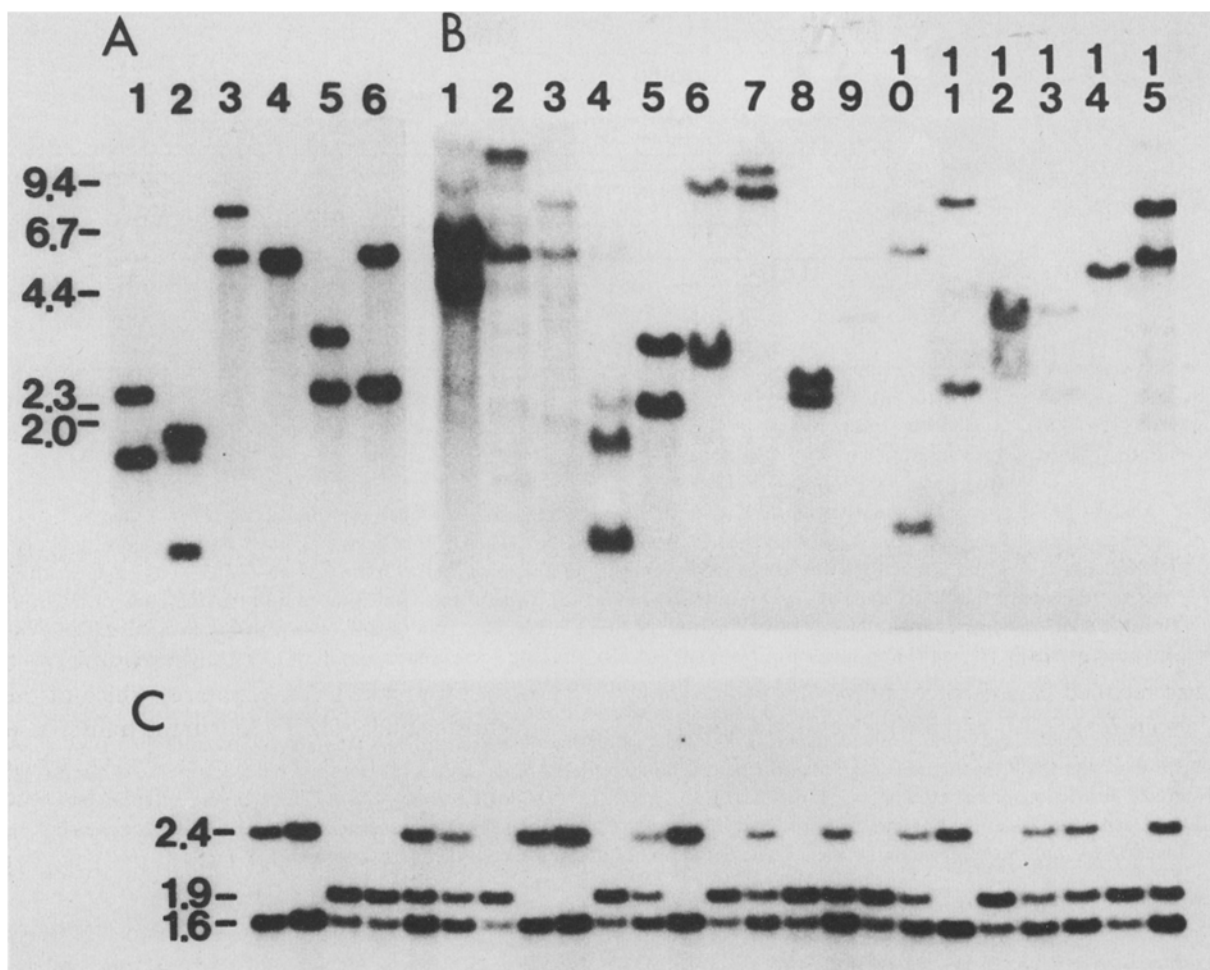


Fig. 3. Southern blot analysis.

A. Survey of restriction fragment length polymorphisms between *L. esculentum* and *L. pennellii*. Odd-numbered lanes are *L. esculentum*, even-numbered lanes are *L. pennellii*. 3 μ g of digested DNA were loaded per lane. Lanes 1 and 2: *Dra*I-digested DNA. Lanes 3 and 4: *Eco*RI digests. Lanes 5 and 6: *Eco*RV digests. A mixture of the *Eco*RI-*Nco*I and *Nco*I-*Nco*I fragments from the coding sequence of the cDNA clone of *Cab-6A* (probes #1 and #2 in Fig. 4b) was used to probe the blot.

B. Survey of 13 species in the family Solanaceae and one species (watermelon) from the Cucurbitae family for the presence of Type I PSI CAB genes. In each case, 3 μ g of a DNA sample from a single individual was digested with *Eco*RI. The probe was the same as in A. Lane 1 - *Lycopersicon hirsutum*, 2 - *L. chmielewskii*, 3 - *L. peruvianum*, 4 - *L. lycopersicoides*, 5 - *Solanum phureja*, 6 - *S. pseudolulo*, 7 - *S. ovalifolium*, 8 - *S. albidum*, 9 - *Capsicum chinense*, 10 - *Datura meteloides*, 11 - *Petunia hybrida*, 12 - *Nicotiana tomentosiformis*, 13 - same species as 12 but an individual from a different accession, 14 - watermelon, 15 - *L. esculentum*.

C. Segregation of F2s of the *L. esculentum* \times *L. pennellii* cross. Each lane was loaded with 3 μ g of *Dra*I-digested DNA from a single plant. The first two samples on the left are *L. esculentum*, the next two are *L. pennellii*, the rest are F2s. The probe was the same as in A. Only three phenotypes are observed: the *L. esculentum* parental phenotype (two dark bands of 2.4 and 1.6 kb), the *L. pennellii* parental phenotype (a dark band of 1.9 kb and a light band of 1.6 kb [The 0.7 kb fragment visible in Fig. 3A, lane 2, ran off the gel in this blot]) and the heterozygous phenotype (three bands of 2.4, 1.9 and 1.6 kb).

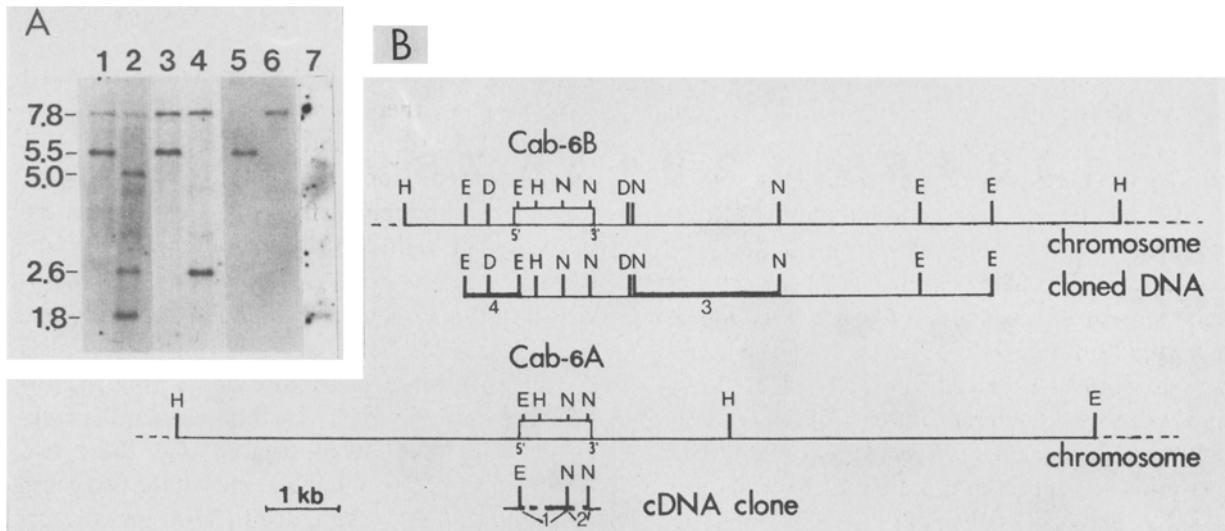


Fig. 4. Analysis of the physical organization of the *Cab-6a* and *Cab-6B* genes in the tomato *Cab-6* locus on chromosome 5.

A. Southern blots. All lanes contain 3 μ g of digested *L. esculentum* DNA. The DNA in lanes 1, 3 and 5 was digested with *EcoRI*, the DNA in lanes 2, 4, 6 and 7 was digested with *HindIII*. Lanes 1 and 2 were probed with the 0.38 kb *EcoRI-NcoI* fragment from the 5' coding sequence of the cDNA clone of *Cab-6A* (probe #1 in Fig. 4B). Lanes 3 and 4 were probed with the 0.28 kb *NcoI* fragment from the 3' coding sequence of the cDNA clone of *Cab-6A* (probe #2 in Fig. 4B). Lanes 5 and 6 were probed with the 2.0 kb *NcoI* fragment downstream (3') from the *Cab-6B* gene (probe #3 in Fig. 4B), and lane 7 was probed with the 0.7 kb *EcoRI* fragment upstream (5') from the *Cab-6B* gene (probe #4 in Fig. 4B). Size of fragments (in kb) is indicated on the left.

B. The arrangement of the *Cab-6A* and *Cab-6B* genes in the *Cab-6* locus deduced from the Southern blot analyses. Locations of restriction sites determined by nucleotide sequencing or by restriction digests of cloned DNA are indicated on both the cloned DNA and chromosomal DNA. Restriction sites deduced from the Southern blots of chromosomal DNA are indicated on the chromosomal DNA only. Restriction sites are defined as: D - *DraI*, E - *EcoRI*, H - *HindIII* and N - *NcoI*. The locations of additional *NcoI* and *DraI* sites more distal to the genes from the sites of each kind indicated have not been determined. The heavy lines marked on the cloned DNA and the numbers under them indicate the probes used in the Southern blots presented in Fig. 4A.

DraI sites in relation to the beginning and end of the *Cab-6A* gene could not be determined. We conclude that there are only two Type I PSI CAB genes in the tomato genome, and thus the minimum number of genes encoding each type of CAB protein identified so far in tomato is two. Unfortunately, we have not yet been able to isolate the *Cab-6A* gene; a clone carrying it is not present in our tomato Charon 4 library, nor in another tomato library in the vector EMBL3 [25] that we have screened.

Multiple restriction fragments were also observed when restricted DNA from other Solanaceae species were probed with the tomato Type I PSI CAB probe (Fig. 3B), suggesting that these species, too, contain multiple genes encoding the Type I PSI CAB protein.

Genetic mapping

Southern blots of *DraI* digests of *L. esculentum* and the congeneric species *L. pennellii* probed with a Type I PSI CAB sequence show different patterns (Fig. 3A, C). The *L. esculentum* phenotype consists of two dark bands of 2.4 and 1.6 kb, while the *L. pennellii* phenotype consists of a dark band of 1.9 kb, a light band of 1.6 kb and a light band of 0.7 kb (the latter band is visible in Fig. 3A, lane 2, but it ran off the gel shown in Fig. 3C). These restriction fragment length polymorphisms (RFLPs) were used to map the *Cab-6A* and *Cab-6B* genes in tomato to each other and to other markers by examining *DraI*-digested DNA samples from 40 F2 individuals of the *L. esculentum* \times *L. pennellii*

cross for their Type I PSI CAB RFLP phenotypes and RFLP phenotypes of other DNA markers.

Only three Type I PSI RFLP phenotypes were observed in the F2s: the *L. esculentum* parental phenotype, the *L. pennellii* parental phenotype, and the double heterozygote (a 2.4 kb band, a 1.9 kb band, and a 1.6 kb band intermediate in intensity between that of *L. esculentum* and that of *L. pennellii*) (Fig. 3C). The lack of recombinant phenotypes thus indicates that *Cab-6A* and *Cab-6B* are tightly linked to each other. Note that because no segregation is observed, it could not be determined whether the *L. esculentum* 2.4 kb *DraI* fragment is allelic to the *L. pennellii* 1.9 kb *DraI* fragment and the *L. esculentum* dark 1.6 kb fragment is allelic to the *L. pennellii* light 1.6 kb fragment, or the other way around. We further note that if, for example, the dark and light 1.6 kb fragments were allelic to each other, certain recombinant classes (those with two 2.4 kb alleles, one dark 1.6 kb allele and one light 1.6 kb allele and those with two dark 1.6 kb alleles, one 1.9 kb allele and one 2.4 kb allele) would be difficult to distinguish from the *L. esculentum* parental phenotype and the double heterozygote, respectively. Similarly, some recombinant classes in the alternative model are also phenotypically not distinguishable from the parental phenotypes or the double heterozygote. However, in both models some of the expected recombinant classes are phenotypically distinguishable from the parental phenotypes and the double heterozygotes, and such phenotypes have not been observed.

Analysis of cosegregation of the Type I PSI *DraI* RFLPs with other markers (data not shown) revealed linkage to two DNA markers, CD78 and CD74A, which have recently been mapped to the end of chromosome 5 (Tanksley, manuscript in preparation). There were 5 recombinants in the F2s of the CD74A × (*Cab-6A*, *Cab-6B*) cross ($\chi^2 = 52.2$, $P < 0.0001$ for the test of independent assortment) and 7 recombinants in the F2s of the CD78 × (*Cab-6A*, *Cab-6B*) cross ($\chi^2 = 34.6$, $P < 0.0001$). The order of the markers was derived as described in [3] and is: CD78 – 6 cM – CD74A – 7 cM – (*Cab-6A*, *Cab-6B*), with the *Cab-6* locus the most distal on the chromosome.

Genomic organization of *Cab-6A* and *Cab-6B*

The cDNA sequence of *Cab-6A* contains no *HindIII* sites; however, there is an *HindIII* site at the beginning of the second intron in *Cab-6B* (Fig. 1). A probe derived from the *EcoRI-NcoI* fragment which spans the region from the first exon to the third exon (Figs. 1, 4B) hybridizes to all the four *HindIII* fragments (of 7.8, 5.0, 2.6 and 1.8 kb) observed in tomato genomic blot (Fig. 4A). A probe consisting of the *NcoI-NcoI* fragment which spans the region from exon III to exon IV (Figs. 1, 4B) hybridizes only to the 7.8 and 2.6 kb *HindIII* fragments (Fig. 4A), demonstrating that these two *HindIII* fragments contain the ends of the two genes and that *Cab-6A*, like *Cab-6B*, also contains an *HindIII* site in an intron. However, because the nucleotide sequence of the coding regions of *Cab-6A* and *Cab-6B* are only 1% divergent from each other (including the 5' and 3' transcribed regions), it is not possible to distinguish with probes from the coding regions of *Cab-6A* or *Cab-6B* which of these *HindIII* fragments carry which gene. A probe consisting of a 2.0 kb *NcoI-NcoI* fragment 3' to *Cab-6B* (Fig. 4B) hybridizes only to the 7.8 kb *HindIII* fragment (Fig. 4A), revealing that this fragment carries the end of the *Cab-6B* gene and also demonstrating that the extensive sequence homology between the two genes extends no further than 400 nucleotides downstream from the stop codon. The 0.7 *EcoRI* fragment containing the promoter region of *Cab-6B* (Fig. 4B) hybridizes only to the 1.8 kb *HindIII* fragment, showing that this *HindIII* fragment carries the beginning of *Cab-6B* and again demonstrating that the extensive sequence homology between the two genes in this region is limited to a short sequence upstream from the first ATG codon. By elimination, the 2.6 kb *HindIII* fragment contains the 3' end of *Cab-6A*, and the 5.0 kb *HindIII* fragment contains the 5' end of *Cab-6A*. The *EcoRI* fragments were similarly assigned to the respective genes.

The map of restriction sites in and around the two genes, deduced from the restriction mapping analysis described in the preceding paragraph and also from restriction site and nucleotide sequence

analyses of cloned DNA, is summarized in Fig. 4B. This extensive restriction mapping analysis further supports the conclusion that the tomato genome contains only two genes, *Cab-6A* and *Cab-6B*, encoding the Type I PSI CAB polypeptide. Whereas the genetic data show that *Cab-6A* and *Cab-6B* are linked, the restriction maps presented in Fig. 4B indicate that they reside at least 7–10 kb apart. This spacing is much greater than that observed in the compound tomato loci *Cab-1* and *Cab-3*, whose genes encode PSII CAB proteins [17].

Expression of the Type I PSI CAB genes

The transcriptional activity of genes encoding the Type I PSII CAB polypeptides has been extensively studied. These genes have been found to be highly

expressed in green leaf tissue and to a lesser extent in other green parts of the plant [5, 14, 21, 26]. In addition, their expression is strongly induced by light [5, 14, 26]. We have employed the Northern blot technique to look at the steady-state level of mRNA transcribed from the *Cab-6A* and *Cab-6B* genes, constituting the entire set of Type I PSI CAB genes of tomato, in different tomato organs (Fig. 5). As a probe we used the 0.38 kb *EcoRI-NcoI* fragment from the 5' coding region of the cDNA clone of *Cab-6A* (Figs. 1, 4). Because *Cab-6A* and *Cab-6B* are almost identical in their coding regions and the proximal 5' and 3' non-coding ends (Fig. 1), we could not differentiate between the mRNAs of the two genes. Our probe is thus "type specific" under the conditions used (hybridization at 65°C, 2 × SSC, 1 × Denhardt's solution, 0.1% SDS; washing sequentially at 65°C

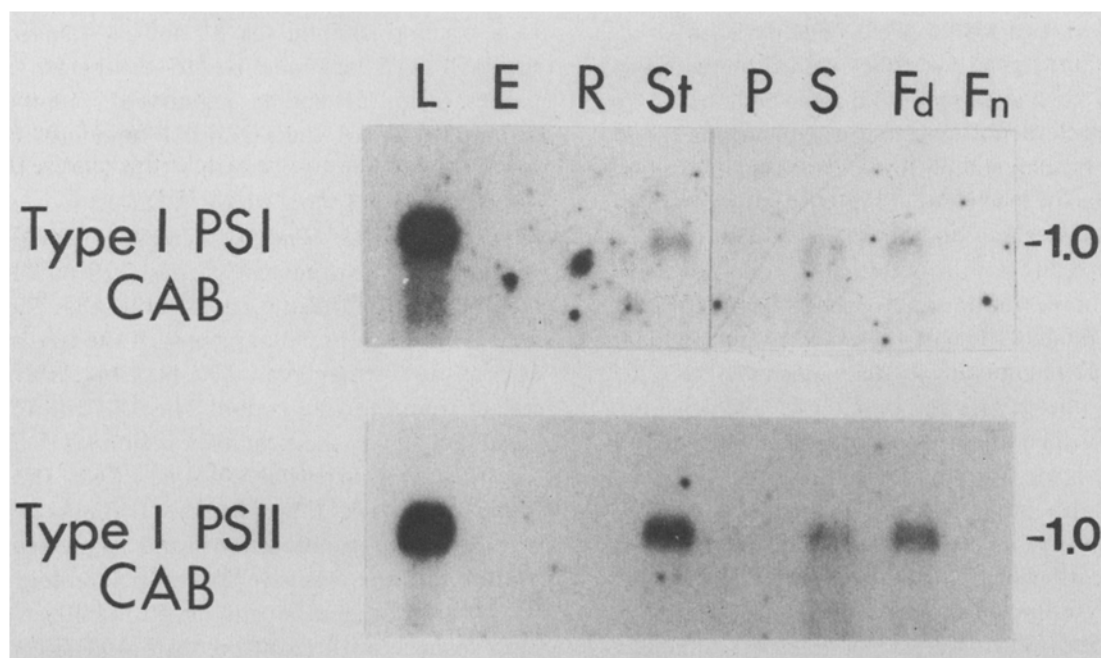


Fig. 5. Northern blots of RNA isolated from different tomato organs. The top panel is an autoradiograph of a blot probed with the 0.38 kb *EcoRI-NcoI* fragment of the cDNA clone of *Cab-6A*. This probe is "type-specific" for the Type I PSI CAB genes (see text). The specific activity of the probe was 4.5×10^6 cpm/ μ g and the X-ray film was exposed for 4 d. The bottom panel is an autoradiograph of a blot probed with a 1.0 kb *HpaI-PstI* fragment from the coding sequence of *Cab-1B*, a Type I PSII CAB gene [17]. This probe is "type-specific" for Type I PSII CAB genes [17, 21]. The specific activity of the probe was 7×10^6 cpm/ μ g and the X-ray film was exposed for 22 h. RNA was isolated from each organ as described [21]. The organs are (from left): L – leaf, E – etiolated seedlings, R – root, St – stem, P – petal, S – sepal, Fd – fruit (15 days after pollination) harvested 7 h after sunrise, Fn – fruit (15 days after pollination) harvested 7 h after sunset. 3 μ g of total RNA was loaded on each well. The size of the Type I PSI CAB mRNA indicated on the right (in kb) is consistent with the size of the full length cDNA clone of *Cab-6A* (897 bases, ref. [11] and Fig. 1) plus the poly A tail. The size of the Type I PSII CAB mRNA is also 1.0 kb [21].

in $2 \times \text{SSC}$ and $1 \times \text{SSC}$) for Type I PSI CAB genes; in DNA-DNA hybridization experiments under less stringent conditions, the same probe hybridized only to the *Cab-6* genes (Fig. 4A). For comparison, we have also probed a replicate Northern blot with a probe which is, under the same conditions, specific for Type I PSII CAB genes (refs. [17, 21]; see legend to Fig. 5).

The Northern blots of Type I PSI CAB mRNA and Type I PSII CAB mRNA from all organs tested are very similar to each other (Fig. 5). Relatively high level of mRNA is observed in green leaf tissue, lower amount is seen in stems and sepals, and no detectable amount is observed in etiolated seedlings, roots or the yellow petals. Furthermore, both types of genes show a diurnal effect – some mRNA observed in the light, no mRNA observed in the dark – both in fruit tissue (Fig. 5) and in leaf tissue (Piechulla, unpublished results). Thus, the well-documented pattern of gene expression of the Type I PSII CAB genes – light-inducible, green tissue-specific – appears to apply to the Type I PSI CAB genes as well.

Despite these similarities, when the specific activities of the probes and the lengths of exposure of the blots shown in Fig. 5 are taken into account, it is clear that the steady-state level of Type I PSI CAB mRNA is much lower than that of Type I PSII CAB mRNA, although precise figures could not be obtained from the present experiments. No doubt the lower level of Type I PSI CAB mRNA is partly due to the fact that tomato has only two such genes, whereas it has 8–10 genes of the Type I PSII CAB. It would be of interest to determine if the transcriptional activity of individual Type I PSI CAB genes is also lower than that of Type I PSII CAB genes, or perhaps other factors in addition to the gene dosage effect account for the lower steady-state level of Type I PSI CAB mRNA. Another interesting observation is that the relative expression of the Type I PSI CAB genes in organs other than leaf is significantly reduced when compared with that of the Type I PSII CAB genes (i.e. the ratio leaf/stem, for example, is much higher for the Type I PSI CAB mRNA).

Evolution of the CAB gene family

Three branches of the CAB gene family have now been identified in tomato, and each branch contains more than one gene. The structural homology of all these genes [11] indicates that they all originated by repeated cycles of duplications followed by structural divergence, although the functional significance of the structural heterogeneity of the CAB polypeptides is not yet understood. Nor is it known at which period in plant evolution each of these branches of the family originated. Furthermore, there are still CAB polypeptides, including the “CP29” protein of PSII [10] and several types of proteins in PSI [6, 9], whose genes have not yet been isolated and characterized.

A summary of comparisons of coding sequence divergences of tomato CAB genes from the three identified branches of the gene family is given in Fig. 6. The tomato Type I PSII CAB genes are 30% divergent from Type II PSII genes, and the Type I PSI CAB genes are 55% divergent from any PSII CAB gene at the equivalent regions of the genes (they also contain two large deletions relative to the PSII CAB genes (Fig. 2; ref. [11])).

Our data concerning the divergence values for CAB genes from the three branches of the CAB gene family examined so far are consistent with the hypothesis that the first division in the CAB gene family occurred by gene divergence for specificity for either PSI or PSII, and that such a division might be as old as the two photosystems themselves. However, sequences of CAB genes from other branches of the family will be required to further test this hypothesis. Furthermore, calculation of the exact time of origin of each branch of the CAB gene family cannot be undertaken until reliable values for rates of nucleotide substitutions are established for plants, and until the degree of divergence among the CAB genes is examined by sophisticated formulae which take into account multiple hits, different substitution rates at different positions in the codons, etc.

Within each branch of the CAB gene family in tomato, further duplications have occurred. However, divergence among tomato genes encoding the same type of CAB polypeptide is limited to the DNA

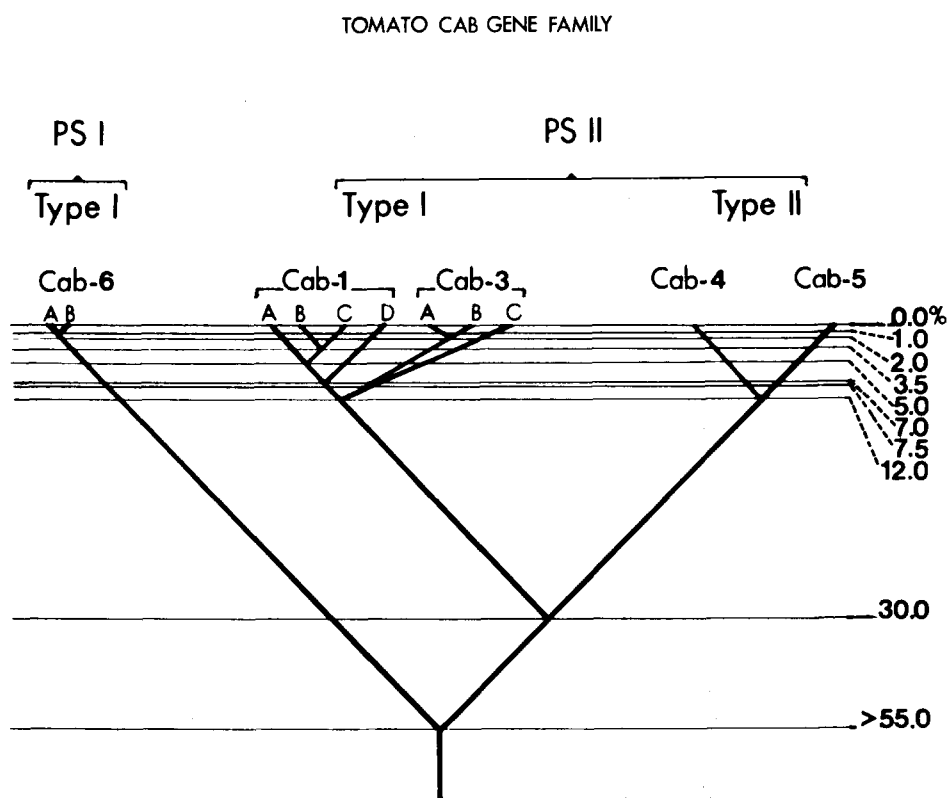


Fig. 6. A phylogenetic tree of genes in the branches of the tomato CAB gene family which have been identified and physically analyzed so far. Numbers on the right indicate percent divergence. Only coding sequences were compared. Deletions/insertions, which are usually very short (2–3 codons) were ignored in these comparisons. However, the Type I PSI genes have two large deletions (10–20 codons each) relative to PSII CAB genes, and the numerical value of their divergence from the PSII genes is therefore indicated as >55%, (55% is the value of divergence of the homologous sequences). The lengths of the branches of these phylogenetic trees are not proportional to the divergence levels and should not be compared with each other. They are only meant to display, schematically, the results of nucleotide sequence comparisons, and as such cannot be directly converted to evolutionary time scales for several reasons. First, the rate of nucleotide substitutions in plant nuclear genes is not yet known. Secondly, the figures for percentage divergence at both extremes of the trees are clearly underestimates: the bottom figures are underestimated because no correction is performed for multiple hits, and the top figures, for within-locus divergence, almost surely reflect the lengths of time since the most recent gene conversion events and not times of origin of the gene duplications within each of the loci.

sequence; the sequences of the proteins encoded are identical or almost identical [17, 19, 20]. When the genes encoding the same CAB type are unlinked, as in the case of *Cab-4* and *Cab-5* and the *Cab-1* genes as compared with the *Cab-3* genes, the nucleotide sequence divergence is always greater than among genes residing in the same locus, as in the case of *Cab-6A* and *Cab-6B*, the *Cab-1* genes, and the *Cab-3* genes (i.e., for genes encoding the same type, intralocus divergence is always smaller than interlocus divergence). A similar situation is observed in

the tomato gene family encoding the small subunit of the enzyme ribulose 1,5-bisphosphate carboxylase/oxygenase (RBCS). Here, too, unlinked genes are more divergent from each other than linked genes, and it has been shown that coding sequence information among genetically unlinked RBCS genes is conserved through selection on the protein level [18] whereas nucleotide sequence divergence among RBCS genes in the same locus is limited by the frequent occurrence of gene conversion [25]. Such intralocus gene conversion events have possi-

bly occurred also in the RBCS and CAB loci in pea [4, 26].

It is not yet known when the gene duplication that gave rise to *Cab-6A* and *Cab-6B* occurred; however, since other Solanaceae species contain similarly duplicated Type I PSI CAB genes, it is likely this gene duplication is not recent and presumably cannot fully explain the high sequence conservation between the two genes. It is also interesting to note that for both RBCS and CAB tomato systems, the highest level of sequence homology between neighboring genes occur in the compound loci *Rbcs-3* and the *Cab-6* whose genes all contain introns, and is lowest in the compound loci *Cab-1* and *Cab-3* where the resident genes do not contain introns.

Acknowledgements

We thank Dr D. Zamir for performing the Southern blot in Fig. 3B. This work was supported by an NIH grant GM38408 to ARC and an NSF grant DCB-8517773 to SDT. EP was supported by an NIH Post-doctoral Training Grant Fellowship and NEH was supported by an NSF Post-doctoral Fellowship.

References

1. Anderson JM: Photoregulation of the composition, function and structure of thylakoid membranes. *Ann Rev Plant Physiol* 37: 93–136 (1986).
2. Bernatzky R, Tanksley SD: Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* 112: 887–898 (1986).
3. Bernatzky R, Tanksley SD: Majority of random cDNA clones correspond to single loci in the tomato genome. *Mol Gen Genet* 203: 8–14 (1986).
4. Cashmore AR: Nuclear genes encoding the small subunit of ribulose biphosphate carboxylase. In: Kosuge T, Meredith C, Hollander A (eds) *Genetic Engineering of Plants. An Agricultural Perspective*. Plenum Press, New York (1983) pp 29–38.
5. Cashmore AR: Structure and expression of a pea nuclear gene encoding a chlorophyll a/b-binding thylakoid polypeptide. *Proc Natl Acad Sci (USA)* 81: 2960–2964 (1984).
6. Darr SC, Sommerville SC, Arntzen CJ: Monoclonal antibodies to the light-harvesting chlorophyll a/b protein complex of photosystem II. *J Cell Biol* 103: 731–740 (1986).
7. Dunsmuir P, Smith SM, Bedbrook J: The major chlorophyll a/b-binding protein of petunia is composed of several polypeptides encoded by a number of distinct nuclear genes. *J Mol Appl Genet* 2: 285–300 (1983).
8. Dunsmuir P: The petunia chlorophyll a/b-binding genes: a comparison of CAB genes of different gene families. *Nucl Acids Res* 13: 2503–2518 (1985).
9. Evans PK, Anderson JM: The chlorophyll a/b-proteins of PSI and PSII are immunologically related. *FEBS Lett* 199: 227–233 (1986).
10. Green BR, Camm EL: The nature of the light-harvesting complex as defined by sodium dodecyl sulfate polyacrylamide gel electrophoresis. *Bioch Biophys Acta* 681: 256–262 (1982).
11. Hoffman NE, Pichersky E, Malik VS, Castresana C, Ko K, Darr SC, Cashmore AR: The nucleotide sequence of a tomato cDNA clone encoding a photosystem I protein with homology to photosystem II chlorophyll a/b-binding polypeptides. *Proc Natl Acad Sci (USA)* (in press).
12. Karlin-Neumann GA, Kohorn BD, Thorner JP, Tobin EM: A chlorophyll a/b protein encoded by a gene containing an intron with characteristics of a transposable element. *J Mol Appl Genet* 3: 45–61 (1985).
13. Kohorn BD, Harel E, Chitnis PR, Thorner JP, Tobin EM: Functional and mutational analysis of the light-harvesting chlorophyll a/b protein of thylakoid membranes. *J Cell Biol* 102: 972–981 (1986).
14. Lamppa GK, Morelli G, Chua N-H: Structure and developmental regulation of a wheat gene encoding the major chlorophyll a/b-binding polypeptide. *Mol Cell Biol* 5: 1370–1378 (1985).
15. Leutwiler LS, Meyerowitz EM, Tobin EM: Structure and expression of three light-harvesting chlorophyll a/b-binding protein genes in *Arabidopsis thaliana*. *Nucleic Acids Res* 14: 4051–4076 (1986).
16. Maxam AM, Gilbert W: Sequencing end-labelled DNA with base-specific chemical cleavage. *Methods Enzymol* 65: 499–560 (1980).
17. Pichersky E, Bernatzky R, Tanksley SD, Breidenbach RB, Kausch AP, Cashmore AR: Molecular characterization and genetic mapping of two clusters of genes encoding chlorophyll a/b-binding proteins in *Lycopersicon esculentum* (tomato). *Gene* 40: 247–258 (1985).
18. Pichersky E, Bernatzky R, Tanksley SD, Cashmore AR: Evidence for selection as a mechanism in the concerted evolution of *Lycopersicon esculentum* (tomato) genes encoding the small subunit of the ribulose 1,5-biphosphate carboxylase/oxygenase. *Proc Natl Acad Sci (USA)* 83: 3880–3884 (1986).
19. Pichersky E, Bernatzky R, Tanksley SD, Malik VS, Cashmore AR: Genomic organization and evolution of the RBCS and CAB gene families in tomato and other higher plants. In: Nevins D (ed.) *Tomato Biotechnology*. Alan Liss, Inc., New York (in press).
20. Pichersky E, Hoffman NE, Malik VS, Bernatzky R, Tanks-

- ley SD, Szabo L, Cashmore AR: The tomato *Cab-4* and *Cab-5* genes encode a second type of CAB polypeptides localized in photosystem II. *Plant Mol Biol* 9: 109–120 (1987).
21. Piechulla B, Pichersky E, Cashmore AR, Gruissem W: Expression of nuclear and plastid genes for photosynthesis-specific proteins during tomato fruit development and ripening. *Plant Mol Biol* 7: 367–376 (1986).
 22. Smeekens S, van Oosten J, de Groot M, Weisbeek P: Silene cDNA clones for a divergent chlorophyll a/b-binding protein and a small subunit of ribulose 1,5-bisphosphate carboxylase. *Plant Mol Biol* 7: 433–440 (1986).
 23. Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98: 503–517 (1975).
 24. Stayton MM, Black M, Bedbrook J, Dunsmuir P: A novel chlorophyll a/b-binding protein gene from petunia which encodes the lower molecular weight *Cab* precursor protein. *Nucl Acids Res* 14: 9781–9796 (1986).
 25. Sugita M, Manzara T, Pichersky E, Cashmore AR, Gruissem W: Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. Submitted.
 26. Timko MP, Kausch AP, Hand JM, Herrera-Estrella L, Van der Broeck G, Van Montagu M, Cashmore AR: The structure and expression of nuclear genes encoding polypeptides of the photosynthetic apparatus. In: Arntzen C, Bogorad L, Bonitz S, Steinbeck K (eds) *Molecular Biology of the Photosynthetic Apparatus*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1985) pp 381–396.
 27. Vallejo's CE, Tanksley SD, Bernatzky R: Localization in the tomato genome of DNA restriction fragments containing sequences homologous to the rRNA (45s), the major chlorophyll a/b-binding polypeptides and the ribulose 1,5-bisphosphate carboxylase genes. *Genetics* 12: 93–105 (1986).