Plant Molecular Biology 12: 257–270, 1989
© 1989 Kluwer Academic Publishers. Printed in Belgium

A new member of the CAB gene family: structure, expression and chromosomal location of *Cab*-8, the tomato gene encoding the Type III chlorophyll a/b-binding polypeptide of photosystem I

Eran Pichersky,[1] Thomas G. Brock,[1] Duc Nguyen,[1] Neil E. Hoffman,[2] Birgit Piechulla,[3] Steven D. Tanksley[4] and Beverley R. Green[5]
[1]*Biology Department, University of Michigan, Ann Arbor, MI 48109 USA;* [2]*Department of Plant Biology, Carnegie Institute of Washington, 290 Panama Street, Stanford, CA 94305, USA;* [3]*Institut für Biochemie der Pflanze der Universität Göttingen, Untere Karspule 2, 34 Göttingen, FRG;* [4]*Department of Plant Breeding and Biometry, Cornell University, Ithaca, NY 14853, USA;* [5]*Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 2Bl, Canada*

Received 28 July 1988; accepted in revised form 31 October 1988

Key words: chloroplast protein import, DNA sequences, *Lycopersicon esculentum*, plant evolution, RFLP markers

Abstract

We have previously reported the isolation and characterization of tomato nuclear genes encoding two types of chlorophyll a/b-binding (CAB) polypeptides localized in photosystem (PS) I and two types of CAB polypeptides localized in PSII. Sequence comparisons shows that all these genes are related to each other and thus belong to a single gene family. Here we report the isolation and characterization of an additional member of the tomato CAB gene family, the single tomato nuclear gene, designated *Cab*-8, which encodes a third type of CAB polypeptide localized in PSI. The protein encoded by *Cab*-8 is 65% and 60% divergent from the PSI Type I and Type II CAB polypeptides, respectively. The latter two are 65% divergent from each other. Only some short regions of the polypeptides are strongly conserved. The *Cab*-8 locus maps to chromosome 10, 9 map units from *Cab*-7, the gene encoding the Type II PSI CAB polypeptide. The *Cab*-8 gene contains two introns; the first intron matches in position the single intron in the Type II PSII CAB genes and the second intron matches in position the second intron in the Type II PSI CAB gene. Like other CAB genes, *Cab*-8 is light-regulated and is highly expressed in the leaf and to a lesser extent in other green organs.

Introduction

The photosynthetic apparatus of plants is made up of two complexes, termed photosystem I (PSI) and photosystem II (PSII), which are located in the thylakoid membranes of chloroplasts (for review see [2, 32]). Some of the pigment-binding proteins in these complexes are chloroplast-encoded [27]. A distinct class of pigment-binding proteins consists of the chlorophyll a/b-binding (CAB) polypeptides which are nuclear-coded and show a range of molecular weights of 22–29 kDa (as determined on SDS-PAGE gels) [9, 11, 12]. These proteins are localized in structures known as the light-harvesting complexes (LHC) I and II, which are associated with PSI and PSII, respectively, but which can be isolated as aggregates distinct from the 'core' PSI and PSII complexes. The CAB protein-bound chlorophyll

molecules in LHCI and LHCII are the primary acceptors of light energy in the form of photons. The captured excitation energy is then transferred to the PSI and PSII reaction centers for conversion into chemical energy. The CAB polypeptides are also believed to play a major role in mediating grana stacking, and the phosphorylation of threonine residues at the N-termini of certain CAB polypeptides has been correlated with this phenomenon [1, 4, 5, 8, 17].

At least two types of CAB proteins have been found in the LHCII complex of higher plants and as many as four different types of CAB proteins are present in LHCI [9, 11, 13]. It is not known at present what distinct functions each type of CAB protein has, what pigments each type binds and the stoichiometry of the pigment-protein interaction. The three-dimensional structure of the CAB polypeptides has not been determined, although models based on theoretical considerations of the primary amino acid sequence and some relevant experimental data have been proposed [3, 13, 14].

We are interested in the structure and function of the CAB polypeptides. In order to obtain the complete primary sequence of the proteins involved, we have been isolating all the CAB genes from the plant species *Lycopersicon esculentum* (tomato) and determining their nucleotide sequences. We have also been investigating CAB gene structure, genomic organization and mode of expression in order to gain insight into the biogenesis of the LHCI and LHCII complexes and the origin of the different types of CAB polypeptides comprising them. We have previously reported the isolation and characterization of eight genes, found in three different genetic loci (designated *Cab*-1, *Cab*-2, and *Cab*-3) encoding the Type I PSII CAB polypeptides [22], two genes (*Cab*-4 and *Cab*-5) encoding the Type II PSII CABs [23], two genes (*Cab*-6A and *Cab*-6B) encoding the Type I PSI CAB protein [13, 24], and one gene (*Cab*-7) encoding the Type II PSI CAB protein [25]. Immunological cross-reactivity experiments carried out in several laboratories had previously indicated that the various CAB polypeptides in both PSI and PSII might be structurally related to each other [9, 11, 33]. However, one study found no cross-reactivity between PSI and PSII CAB polypeptides [34], raising the possibilities that substantial divergence has occurred among these two groups of proteins or that the cross-reactivity observed in the other studies was due to some artifacts. Comparisons of the nucleotide sequences of the tomato PSI and PSII CAB genes characterized so far (and the sequences of the proteins they encode) conclusively demonstrated that the different CAB proteins in both PSI and PSII are all structurally related to each other [13, 22–25]. Thus, the genes encoding many, and perhaps all, CAB polypeptides belong to a gene family and have been created by repeated cycles of gene duplication and divergence. Here we report the isolation of an additional, previously uncharacterized, member of the CAB gene family, the single tomato nuclear gene encoding the Type III PSI CAB polypeptide, and describe its structure, chromosomal location, mode of expression, and its relationship to the other members of this complex gene family.

Materials and methods

Isolation of genomic and cDNA clones and their characterization

All procedures were as described ([22–24] and references therein). Conditions for Southern blots [28] were as described in Pichersky *et al.* [22]. Nucleotide sequences were determined by the chemical method [19]. The genomic clone was isolated from the EMBL3 library described in Sugita *et al.* [30]. cDNA clones were isolated from a tomato cDNA library described in Pichersky *et al.* [23] and Hoffman *et al.* [13]. The cDNA inserts were excised from the original plasmid (pARC7) of the library by cutting at the *Sst* I site of the 5′ polylinker and the *Xba* I site at the 3′ polylinker flanking the insert, and recloned into the plasmid pUC18 for nucleotide sequence determination or into pGEM4 for *in vitro* protein synthesis.

In vitro *transcription, translation, and import into isolated tomato chloroplasts*

These procedures were described by Pickersky *et al.* [23]. In addition to the import of precursor protein

into isolated pea chloroplasts, for which protocols were described in Pichersky *et al.* [23], we also did import experiments with isolated tomato chloroplasts. The tomato chloroplasts were prepared as described for pea chloroplasts, with the following modifications:

- Chloroplast isolation: tomato leaves contained high levels of starch, which destroyed the chloroplasts during the isolation procedure. We therefore used cotyledons from 10–12-day-old seedlings. Polyvinyl pyrrolidone 360 (Sigma) was added to the grinding buffer at a final concentration of 1% (w/v). The density of the Percoll (Pharmacia) gradient was increased from 80% to 85%.
- Import conditions: The total reaction volume was 500 μl, with a total amount of chloroplasts equalling 300 μg chlorophyll. 2000000 cpm of ^{35}S-methionine, ^{35}S-cysteine-labelled preCAB-8 protein was added per reaction. The reaction buffer contained 10 mM ATP, 10 mM $MgCl_2$, 20 mM methionine, 20 mM cysteine, pH 7.5. Chloroplasts were incubated in the dark at 25 °C.
- All other conditions were the same as described for the pea chloroplast protocol [23], including the thermolysin treatment.

Localization of the processed CAB-8 polypeptide within the chloroplast

After import, tomato chloroplasts were reisolated on a Percoll cushion and thylakoid membranes prepared as previously described [23]. The thylakoid membranes were then solubilized for 30 min in the dark using an aqueous solution of surfactant containing 4.5% (w/v) heptyl-β-D-thioglucopyranoside (Calbiochem), 4.5% (w/v) octyl-β-D-thioglucoside (Calbiochem) and 1% (w/v) SDS at a final chlorophyll : surfactant ratio of 1 : 10. The solubilized membrane preparations were next run on the deriphat PAGE gel system of Peter *et al.* [20]. In this gel system, chlorophyll-protein complexes are visible on the gel after the run as green bands; one of these green bands consists mostly of PSI proteins, whereas others consist mostly of PSII proteins. Per lane 25 μg chlorophyll were loaded. The gel was run at 200 V constant voltage for 30 min at 4 °C. Unlike the procedure described in Peter *et al.* [20], no SDS was added to the electrophoresis buffer.

After the solubilized thylakoids were run on the deriphat gel system, each lane was cut and soaked in 1× SDS-PAGE sample buffer [16], then placed in an SDS gel [16] at 90° to the direction of the original run. The SDS gel was run at 5 W constant power. After this second run, gels were stained with Coomassie Blue, photographed, and then fluorographed.

The PSI and the PSII (LHCII) fractions shown in Fig. 3E were obtained by fractionation of thylakoid membranes using the sucrose density gradient procedure of Mullet *et al.* [18] further described in Pichersky *et al.* [23].

Genetic mapping

Linkage analysis of cDNA markers was as previously described [6, 7, 22–24].

Northern blots

mRNA was isolated from the different organs and Northern blots were performed as previously described [26, 24]. The 0.2 kb *Spe* I-*Bgl* II fragment from the 3′ nontranslated region of *Cab*-8 (Fig. 1) was used as a probe. This fragment hybridizes exclusively to *Cab*-8 DNA sequences under the conditions used in the Northern blots (65 °C, 2×SSC, wash at 1×SSC) (data not shown).

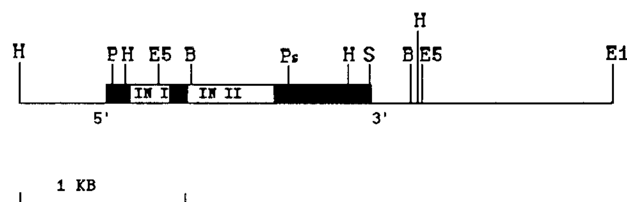

Fig. 1. Restriction map of the cloned *Cab*-8 gene and its surrounding regions. Sites are defined as: B – *Bgl* II; E1 – *Eco* RI; E5 – *Eco* RV; H – *Hind* III; P – *Pvu* II; Ps – *Pst* I; S – *Spe* I.

```
AAGCTTATCGTTAATTCTAGTCATTTTACATTGTTAGGTTCAACATTTATTAATAAATTTCCTAACGCAAATACAGAAA    79
TTTAAATCAAAATTCTTGAATTGAGCTAAACATGTAACATACGTATATCTCCGCCTTGTGTATCGTATTAACTTGTAGT   158
TATCACAAGAAACACAAATTTCAAATTCAGAATCCGTCACGAGAAAAAACTATATACCCTAGTAAATCTATGAGAAGGC   237
AAGGTGGCAATACCAAACAAGAGTATCTAAGATTTTCATTTGTGACTATAGGAATATAATATCTCTTATCTGATTTAAT   316
GAATCCACATGTTCACTTCTCATTTGTCCACAAGATCACAACTTTATCTTCAATATTCACAACTTGTTATATCAACCAC   395
AACAATTTCTATTCTTTTCACTCAGTCCCACAAAATACTTTGTTCCCTTATTTGCCACCTTTTGTATTTAATTTATTCT   474
TTGTGGAGCTAAGTGTTTATATTATTCTTCTTCTCAAAAAAACAAAAACAAACAAAAAAGAGAAGAGAAATT ATG GCA 552
                                                                          M   A
ACT CAA GCA TTG ATC TCT TCA TCA TCA ATT AGC ACT TCA GCT GAA GCT GCA AGA CAA ATT   612
 T   Q   A   L   I   S   S   S   S   I   S   T   S   A   E   A   A   R   Q   I
ATT GGA TCA AGA ATT TCA CAA TCT GTT ACT AGA AAA GCT TCT TTT GTT GTA AGG GCT GCT   672
 I   G   S   R   I   S   Q   S   V   T   R   K   A   S   F   V   V   R   A   A
TCC ACT CCA CCA GTC AAG GTAATTTTCTATACTCGGGGTCGTTTCGTTCAAAACAAGGATATACTGAAATCTA   745
 S   T   P   P   V   K
TGCTACTCGGACGCTTTAAAGATAATGTGTATACTTATTGGAATCCTCCAGAAATACACTATTTTTTGAAGGATTCGAT   824
AAAAGTCATCTACATTCTTTAAGTGTCTGAGCAAGATATCCGTGATCCATAGTTATTTTGAATTTGTGTGCAAGTGTGA   903
GTTGTTGTTACCTTAGTTGCTCGGACTCTTTAAAGAAATTACAAAAATTGTATAGAATAATTGATGAAAACTTTAAAAT   982
AATACACAAGTACTCATACACTTGATTCTATTTTACAG CAA GGA GCA AAT AGG CAA CTC TGG TTT GCA  1050
                                        Q   G   A   N   R   Q   L   W   F   A
TCC AAA CAA AGC CTT TCA TAC TTG GAT GGA AG GTAAGATCTATGAGAATTGGTGTGACTTATAATAGC 1118
 S   K   Q   S   L   S   Y   L   D   G   S
TCTTTCAATAGCCTTGTCGAGTTTATGTTTTGTACACTGACATAAAAGATAAAAATCGACTAGCAACAAATCTAGGTAA 1197
CACTCAATATTAAGCCTGATATAGTAACCTAGAAAATTGCACATATAGTTTAACATGGTGCATATATACAATAATCGAC 1276
AAGAGTTTATGTTTCGTGCATTGACAGTGTAAAGAAGGTTAAGTTGATTTACAACAATTTGTAGGTCAAACCTAATATA 1355
GTATCCTGGAAATGGATTTATATCTTGCTACAATCAATTGAATTGCACACATAGTACATGGTGCACATAAATGGTGTAA 1434
TTGACAAGAATTTATGTTCTATGCACTAACAGTGTAAAGACAGTTAAATTGATTTATAACAAATATAGGTCGCGAATCA 1513
ACCCTAATATAGTATAAAAACTCATGCTGCACATATAAGGTAATCGACAAGAGTCTATTTTATGCACTTACATTGCACA 1592
GATGATATATAACTTGAGTGAACTTACTAAATGATTGTTTGCTGATCCACAG T CTT CCC GGT GAC TTT GGT  1663
                                                        L   P   G   D   F   G
TTT GAT CCC TTG GGA CTC TCG GAC CCT GAA GGC ACA GGA GGT TTC ATC GAG CCG AAA TGG 1723
 F   D   P   L   G   L   S   D   P   E   G   T   G   G   F   I   E   P   K   W
CTA GCA TAT GGC GAG GTC ATC AAT GGT CGG TTC GCC ATG TTA GGA GCT GCA GGG GCA ATA 1783
 L   A   Y   G   E   V   I   N   G   R   F   A   M   L   G   A   A   G   A   I
GCA CCA GAG ATT CTT GGA AAA GCT GGT CTC ATC CCA CAA GAA ACA GCA CTT GCA TGG TTC 1843
 A   P   E   I   L   G   K   A   G   L   I   P   Q   E   T   A   L   A   W   F
CAA ACT GGT GTT ATC CCA CCA GCA GGA ACA TAC AAC TAC TGG GCT GAC AAT TAC ACA TTG 1903
 Q   T   G   V   I   P   P   A   G   T   Y   N   Y   W   A   D   N   Y   T   L
TTT GTT CTT GAA ATG GCA CTC ATG GGC TTT GCT GAG CAC AGG AGA TTC CAA GAT TGG GCC 1963
 F   V   L   E   M   A   L   M   G   F   A   E   H   R   R   F   Q   D   W   A
AAA CCT GGT TCA ATG GGG AAA CAA TAC TTC TTG GGT CTT GAA AAG GGT TTG GGT GGG TCA 2023
 K   P   G   S   M   G   K   Q   Y   F   L   G   L   E   K   G   L   G   G   S
GGT GAT CCA GCA TAC CCA GGT GGC CCA TTG TTT AAC CCA CTT GGA TTT GGA AAA GAT GAA 2083
 G   D   P   A   Y   P   G   G   P   L   F   N   P   L   G   F   G   K   D   E
AAG TCT ATG AAG GAA TTG AAG CTT AAG GAG ATT AAG AAC GGA AGA CTT GCT ATG TTG GCT 2143
 K   S   M   K   E   L   K   L   K   E   I   K   N   G   R   L   A   M   L   A
ATT TTG GGA TAC TTT ATT CAA GCA TTG GTT ACT GGT GTT GGA CCT TAC CAA AAC CTT CTT 2203
 I   L   G   Y   F   I   Q   A   L   V   T   G   V   G   P   Y   Q   N   L   L
GAT CAT TTG GCT GAT CCA GTA AAC AAC AAT GTC TTG ACT AGT CTC AAG TTT CAC TAAACTA 2264
 D   H   L   A   D   P   V   N   N   N   V   L   T   S   L   K   F   H
AGTAGAGCAAAAGCAATCCACTCTTTCTTCTAAATCTCCCCCCCTCCCCTTCTATGTGTAATGTAACATTATGTGCAGA 2343
TAGTATTTTTCATCTCATCTGTAATGCTAAAACAAGGACATGTCATGTTTAAAAGGCCTGATATATATTTTTAGAGGAA 2422
ATCATCTAGTATTCTGATCTTACTGATTCGACTAATTCAGATTCATGTCACCTAAGATCTAATTAAAGAAGGAAAATAC 2501
TTCCTACCAAAGCTTGAACTCGAGATATCTAGTTAAGGGTGGAGGGGGATCATATTCATCCCACCACACCCTTGAGGTA 2580
GTCTAAAATCATATATCAGGTGCATTGATGAAGGTAATTTCAAATGAAAAACACTATTAGACATAAAAACAAAATCCTA 2659
```

Fig. 2. Nucleotide sequence of the *Cab*-8 gene and cDNA clones derived from its mRNA, and the polypeptide they encode. The sequence shown extends from the *Hin*d III site 0.55 kb upstream to the start codon to about 150 nucleotides downstream from the 3′-most *Hin*d III site (Fig. 1). The underlined nucleotide at position 499 is the 5′-most nucleotide of cDNA clone pCAB8-52, and the underlined nucleotide in the first intron (position 1014) is the first nucleotide of cDNA clone pCAB8-14. The underlined nucleotides at the 3′ end of the gene are the last nucleotides in pCAB8-52 (position 2359), pCAB8-14 (2393; also another cDNA clone terminates there) and four other cDNA clones (2371, 2420, 2429 and 2432).

Results

Isolation and characterization of Cab-*8 gene and cDNA clones*

We have recently isolated the tomato gene *Cab*-7 which encodes the Type II PSI CAB polypeptide [25]. In Southern blots of tomato genomic DNA, a probe derived from the *Cab*-7 gene (a full-length clone, pCAB7-1B [25]) hybridized, under low-stringency conditions (6×SSC, 60 °C, wash at 2×SSC), to several fragments in addition to the fragment carrying the *Cab*-7 gene (data not shown). We identified some of these fragments as carrying the *Cab*-1 and *Cab*-3 genes, which we had already cloned [22]. However, some of the hybridizing fragments did not carry previously cloned CAB genes. We therefore used the full-length cDNA clone derived from *Cab*-7 to probe (under the same low-stringency conditions) a tomato genomic library in the phage vector EMBL3 [30]. As expected, some of the positively hybridizing phages contained the *Cab*-1 and *Cab*-3, as well as *Cab*-7, genes. A single positively hybridizing recombinant phage which contained neither of these genes was also isolated. The tomato DNA insert from this phage was subcloned into pUC18 plasmids for further characterization.

A restriction map of the tomato DNA insert in this phage was prepared (Fig. 1). Southern blots of subcloned fragments, using short probes derived from different parts of the *Cab*-7 coding region, identified a region hybridizing to the 3′ coding region of the *Cab*-7 gene. The appropriate segments were then subjected to nucleotide sequence determination. Sequence comparisons with the *Cab*-7 and other CAB genes identified a region which encodes the C-terminus of a protein with strong homology to CAB C-terminal sequences. We then proceeded to determine the nucleotide sequence of the new gene, which we designated *Cab*-8, to more than 2 kb upstream from the stop codon of the coding region (Fig. 2). The potential coding information of this region was compared with previously determined CAB sequences, and some regions of homology were easily discerned. However, because of a lack of discernable homology in most of the middle and the N-terminal coding regions, and the observation that *Cab*-8 must contain introns since there was no single uninterrupted reading frame throughout the sequence, the precise coding region of *Cab*-8 could only be established from a comparison with cDNA clones.

We used the 0.2 kb *Spe* I-*Bgl* II fragment from the 3′ non-coding region of *Cab*-8 (Fig. 1) to isolate cDNA clones derived from the *Cab*-8 gene. We isolated over 200 cDNA clones by screening approximately 10^5 recombinant plasmids from a cDNA library prepared from leaf mRNA [23]. However, the majority of these clones were very short, and only three were longer than 500 bp. The three longest clones and four shorter clones were subjected to nucleotide sequence determination. The longest clone, designated pCAB8-52, contains 972 nucleotides (excluding the poly(A) tail) with an open reading frame (beginning with a methionine codon) encoding a protein of 273 amino acids (Fig. 2). That pCAB8-52 contains the complete coding information for this protein was established by the production of this protein *in vitro* and the subsequent import experiments (see below). Thus, pCAB8-52 also contains a 48 bp 5′ non-translated leader and a 105 bp 3′ non-translated region, followed by a poly(A) tract (Fig. 2). The comparison between the sequences of pCAB8-52 and *Cab*-8 revealed that *Cab*-8 contains two introns of 329 and 562 nucleotides, respectively (Fig. 2).

The other six cDNA clones sequenced are shorter and none contains the entire coding region. Two observations of interest were made:

1. Of the total of seven clones, only two terminate at the same 3′ position. However, since all these clones contain a poly(A) tract of 30–80 bp at their 3′ end, it is not possible to determine exactly where processing of the initial transcript occurred in each case. This is so because the last recognizable nucleotide in the cDNA clones is always followed by an adenine both in the genomic sequence and in the cDNA clone (i.e., the beginning of the poly(A) tract) (Fig. 2). Nonetheless, it is clear that processing occurred at six different positions. Multiple termination/processing sites (two or three) have been previously observed for

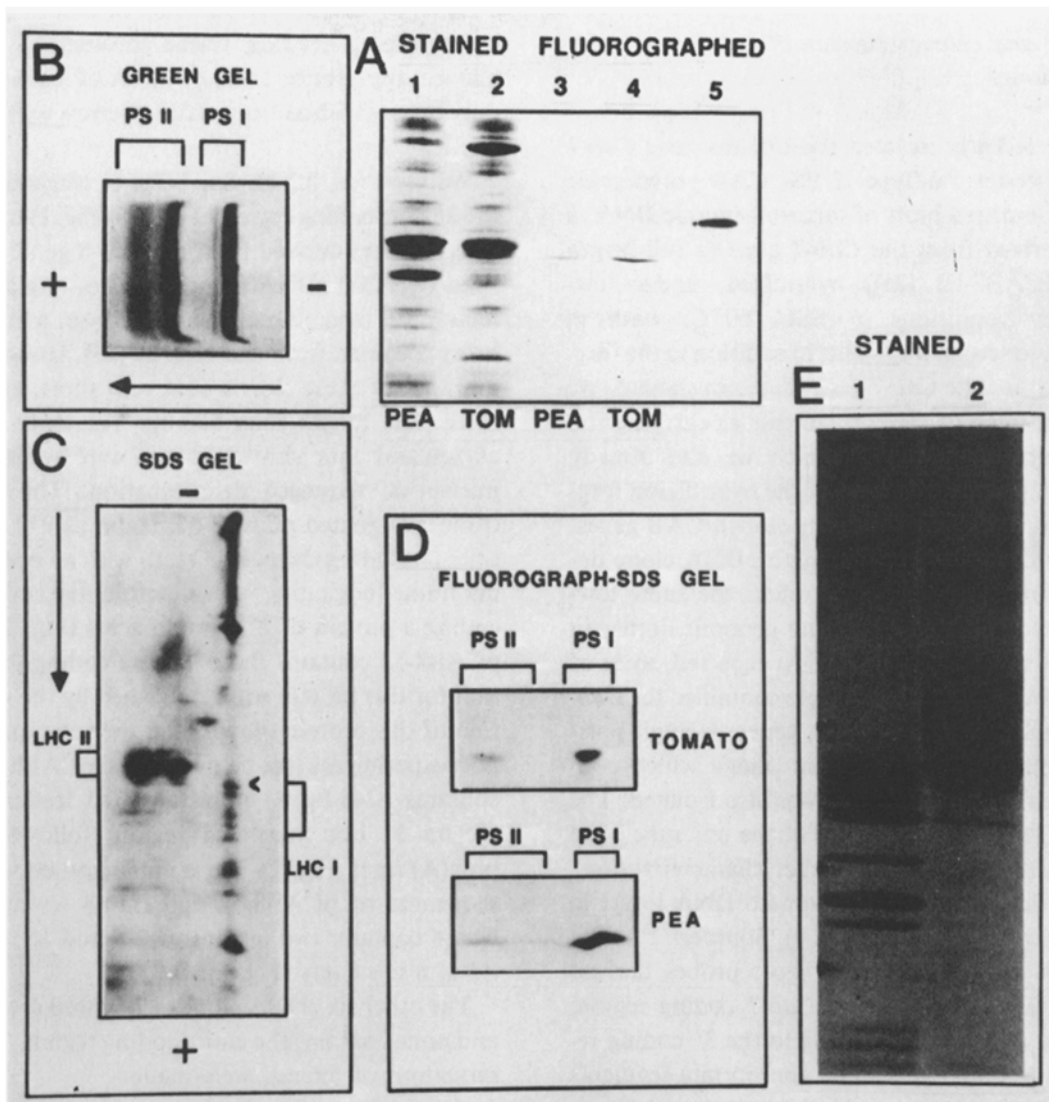

Fig. 3. Import of preCAB-8 and localization of the processed protein in the thylakoid membranes. A. Lanes 1 and 2, Coomassie blue-stained SDS gel of pea and tomato thylakoids. Lanes 3 and 4, fluorograph of lanes 1 and 2, respectively. Lane 5, fluorograph of SDS gel of the CAB-8 precursor. B. Deriphat PAGE separation of chlorophyll complexes from detergent-solubilized tomato thylakoids. This is a picture of the *unstained gel* (the bands observed are green). Similar results were obtained from pea thylakoids (data not shown). C. SDS gel analysis of proteins within the chlorophyll complexes resolved in B. After the deriphat-PAGE, samples were run on SDS gels (without urea) in a second dimension 90° to the direction of the first electrophoresis. A similar result was obtained for the pea thylakoids (data not shown). Bands corresponding to LHCI and LHCII CAB polypeptides [13, 23] are marked in the figure. The arrowhead indicates the migration position of the processed CAB-8 protein (i.e. the radioactive band observed in Fig. 3D; see also Fig. 3A). D. Fluorography of the 2D gel of tomato shown in Fig. 3C and the 2D gel of the pea sample (not shown). The areas corresponding to the green bands of PSI and PSII are indicated. E. SDS gels of tomato PSI and PSII (LHCII) fractions obtained by the sucrose density gradients method of Mullet *et al.* [18] as described in Pichersky *et al.* [23].

transcripts from the same CAB gene [10, 25], but the situation in the *Cab*-8 system appears to be an extreme case.

2. One clone, pCAB8-14, is co-linear with pCAB8-52 up to the first nucleotide of exon 2; upstream from this position, pCAB8-14 contains seven more nucleotides, but these are the last seven nucleotides of the first intron (Fig. 2).

Localization of the processed CAB-8 polypeptide within the chloroplast

The *in vitro* synthesized preCAB-8 polypeptide was imported into isolated tomato and pea chloroplasts and processed to a single, smaller molecular form (Fig. 3A). The import efficiency was at the same level observed for other CAB precursors (data not shown). After import, the thylakoid membranes were isolated, partially solubilized, and run on the deriphat gel system [20]. Two major green bands, and a third, more diffuse green band, were observed at the end of the run (Fig. 3B). The protein contents of these bands was analyzed by running them in a second dimension in the SDS-PAGE gel system of Laemmli [16] and staining with Coomassie Blue (Fig. 3C). The most slowly migrating green band from the first dimension contained mostly PSI proteins (cf. Fig. 3E, lane 1), whereas the other two bands had the prominent LHCII polypeptides (cf. Fig. 3E, lane 2). The LHCII polypeptides were not present in the PSI band. Fluorography of the SDS gel (Fig. 3D) revealed that over 90% of the counts were present in a single polypeptide which comigrated with the PSI green band in the first dimension. This processed CAB-8 protein comi-

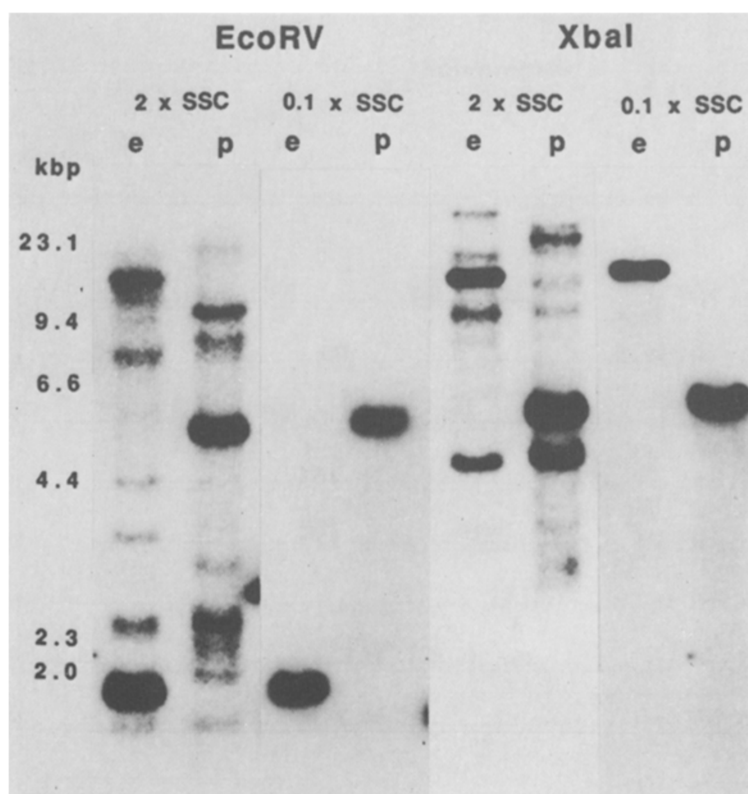

Fig. 4. Southern blot analysis. *L. esculentum* (e) and *L. pennellii* (p) DNA (3 μg per lane) were digested with *Eco* RV and *Xba* I restriction enzymes, run on 1% agarose gel, and blotted onto nylon membranes. The probe was the fragment extending from the *Pst* I site at the beginning of the third exon to the *Spe* I site at the end of the third exon. For each restriction digest, the two left lanes were washed at low stringency (2× SSC) and the two right lanes were washed at high stringency (0.1× SSC). All other conditions were as described in the text.

grates in the Laemmli SDS-PAGE system with a PSI polypeptide (arrow in Fig. 3C; see also Fig. 3A) previously identified as the slowest-mobility LHCI CAB polypeptide [13]. Most of the rest of the radioactivity was found in another band of identical mobility but which was localized in the diffuse green band which contained mostly LHCII proteins. Note that the sharper (more slowly migrating) green PSII band did not contain the processed CAB-8. The diffuse green band may represent a mixture of completely dissociated, individual LHCI and LHCII polypeptides each still bound with pigment molecules. We think it is unlikely that CAB-8 is also specifically associated with LHCII to some extent, in addition to its major association with PSI, since we have observed similar distribution of the other PSI CAB proteins (Type I and Type II) in the deriphat gel system (N.E. Hoffman, in preparation).

Genetic mapping

Restriction fragment length polymorphism (RFLP) mapping was used to locate the chromosomal position of *Cab*-8. The 0.5 kb *Pst* I-*Spe* I fragment from pCAB8-52 was used to probe filters containing *Eco* RV and *Xba* I digested DNA from *L. esculentum* and *L. pennellii* (Fig. 4). Under low stringency conditions (hybridization at 6× SSC, 65 °C, wash at 2× SSC), the probe remained hybridized to several fragments, but under high-stringency (hybridization conditions the same, wash at 0.1× SSC), the probe

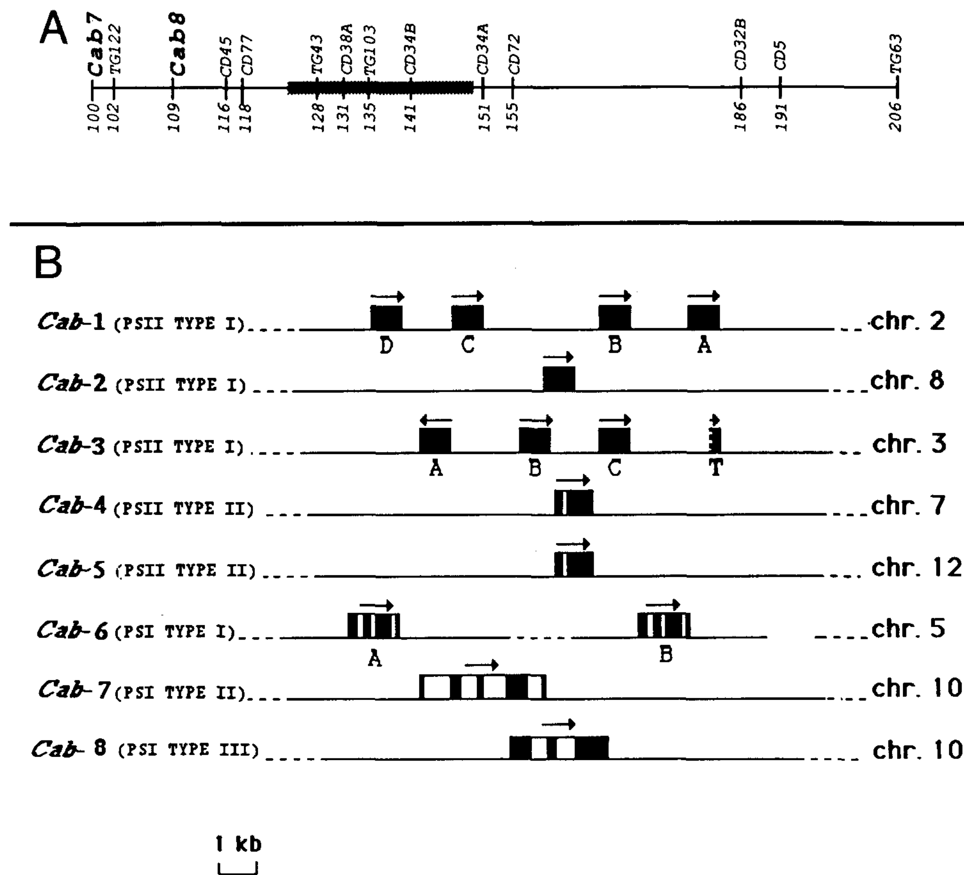

Fig. 5. Linkage map of chromosome 10 showing location of the *Cab*-8 locus relative to the *Cab*-7 locus and other mapped loci. Shaded region indicates approximate position of centromere. B. The distribution of members of the CAB gene family throughout the tomato genome. The *Cab*-6A and *Cab*-6B genes are genetically closely linked to each other but are at least 7 kb apart. Data for the *Cab*-1 through *Cab*-7 loci are from refs. 22, 23, 24, and 25].

remained hybridized to a single fragment in the genome of each species.

Both the *Eco* RV and *Xba* I digests revealed RFLP for the single fragment hybridizing at high stringency (Fig. 4). An F2 population, derived from a cross between these two species, was analyzed for segregation for the *Xba* I and *Eco* RV fragments as well as for a number of other previously mapped RFLP markers [31]. The *Cab*-8 fragments segregated into the expected monogenic classes (two homozygous parental genotypes and heterozygotes; data not shown) and mapped to a point near the end of chromosome 10 (Fig. 5A), approximately 9 map units from the *Cab*-7 locus which is at the terminal end of this linkage group [25]. The segregation analysis indicated that *Cab*-8 maps to the same position as a previously isolated and mapped, but not identified, cDNA clone, designed CD56 [31]. We therefore used the *Cab*-8 probe in a Southern blot of CD56 DNA. The *Cab*-8 probe indeed bound to CD56 DNA under high-stringency conditions (data not shown), indicating that CD56 is a *Cab*-8 cDNA clone and confirming the map assignment of *Cab*-8.

Expression of the Cab-*8 gene*

We have used the Northern blot technique to look at the steady-state level of mRNA encoding the CAB-8 protein in several tomato organs and under different light regimes (Fig. 6). The pattern of *Cab*-8 mRNA levels parallels that of mRNAs encoding the

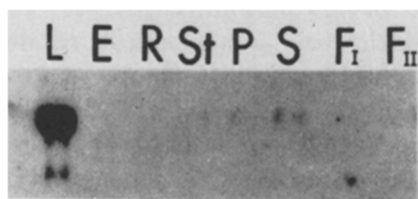

Fig. 6. Northern blots of RNA from different tomato organs. The organs are (from left): L – leaf, E – etiolated seedlings, R – root, St – stem, P – petal, S – sepal, F_I – green fruit (15 days after pollination), F_{II} – mature fruit (45 days after pollination). The probe used was the 0.2 kb *Spe* I-*Bgl* II fragment from the 3′ end of the *Cab*-8 gene (Fig. 1). The specific activity of the probe was 2.8×10^6 cpm/μg DNA and the X-ray film was exposed for 14 days.

Type I and Type II PSI CAB polypeptides and the Type I and Type II PSII CAB polypeptides [24, 25]. From comparisons of the specific activities of the probes and the times of exposure of the blots (see legend to Fig. 6), we calculate that the steady-state level of *Cab*-8 mRNA in the leaf is approximately equal to that of the Type I and Type II PSI CAB mRNAs, which in turn have been shown to be several-fold lower than the steady-state level of PSII CAB mRNAs in all organs examined and under all light regimes [24, 25]. As with the other two types of PSI CAB genes [24, 25], very little *Cab*-8 message is observed in green tissues other than leaf (relative to their abundance in the leaf), while PSII CAB mRNAs are quite abundant in such tissues, especially in the stem [24]. In fact, there is more *Cab*-8 mRNA in sepals than in stems (Fig. 6).

Discussion

Cab-*8 encodes the Type III PSI CAB polypeptide*

The import experiments (Fig. 3) indicate that the protein encoded by the tomato gene *Cab*-8 is localized in PSI. Although some label was associated with the diffuse green band which contained mostly LHCII proteins, over 90% of the counts were found in the single green band whose protein profile shows it is composed of PSI proteins. The fact that the preCAB-8 was imported and properly processed into a single protein associated with PSI indicates that the cDNA clone used to prepare the transcript for *in vitro* protein synthesis contained the entire coding region for the protein. The processed CAB-8 polypeptide comigrates on SDS-PAGE gels with one of several proteins in PSI which cross react with a monoclonal antibody made against a PSII CAB protein, thus identifying them as CAB-related polypeptides [9, 13]. The genes encoding two other of these proteins, designated Type I and Type II PSI CABs, have previously been isolated and characterized [24, 25]. Thus, *Cab*-8 encodes a third PSI CAB polypeptide, which we designate Type III PSI CAB. The Southern blots (Fig. 4) indicate that the tomato genome does not contain additional genes with strong homology to *Cab*-8. Thus, the Type III PSI

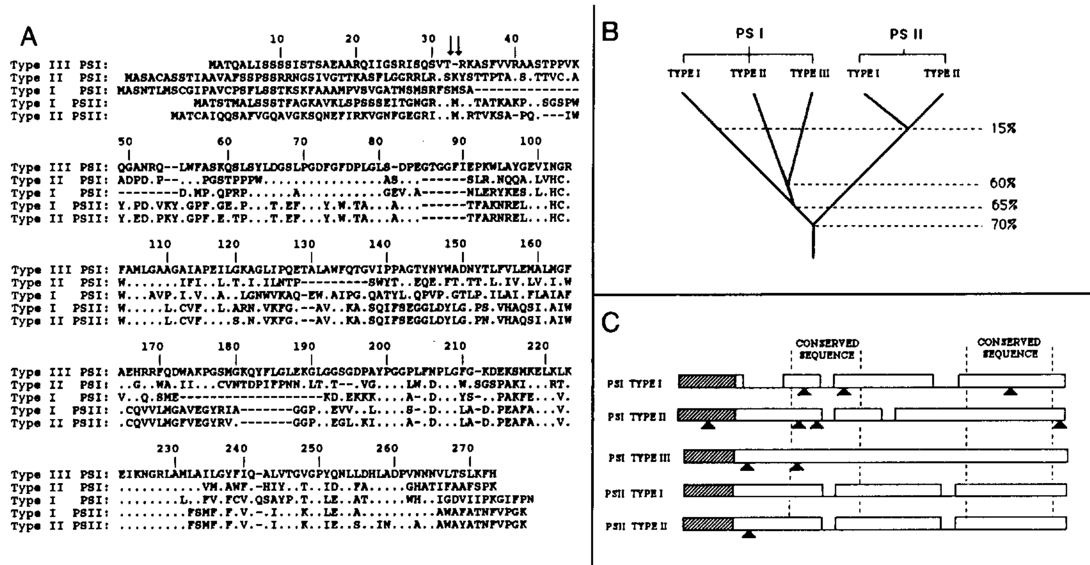

Fig. 7. Protein and gene evolution in the tomato CAB gene family. A. Comparison of protein sequences. The Type III PSI sequence is that of preCAB-8 (this report). Type II PSI – preCAB-7 [25], Type I PSI – preCAB-6A [13, 24], Type I PSII – preCAB-3C [22] and Type II PSII – preCAB-4 [23]. The numbering system refers to the preCAB-8 polypeptide. A dot represents identity with the preCAB-8 sequence (no matching was done in the region of the transit peptide), a dash represents a gap introduced to maximize homology. The arrows indicate putative sites of processing [13, 17, 22, 23]. B. A phylogenetic tree of the tomato CAB proteins. Numbers indicate percent divergence of the amino acid sequences. Only homologous sequences are compared; deletions/insertions are ignored in these comparisons. Lengths of branches are *not* proportional to degree of divergence. C. A schematic diagram of the main structural features of the five 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 273 amino acids. The shaded box is the transit peptide (32–45 amino acids) which is removed after import into the chloroplasts. Open boxes represent mature protein, with interruptions to indicate deletions relative to other CAB sequences (deletions of three amino acids or less are not indicated). Segments which are strongly conserved in all five types are outlined. The locations of introns in the genes with respect to the protein sequences are indicated with triangles.

CAB protein, like the Type II PSI CAB protein, is encoded by a single gene in tomato. All other types of CAB polypeptides identified to date are encoded by multiple genes in tomato (Fig. 5B).

The exact location of the cleavage of the transit peptide has not been determined for preCAB-8 (nor for any other CAB precursors). However, the processing site for the Type I PSII CAB precursor has been tentatively identified [17], and sequence alignment suggests that the transit peptide of preCAB-8 contains 32 or 33 amino acids (Fig. 7A). The mature portion of the CAB-8 protein therefore contains 240–241 amino acids, with a formula-calculated molecular weight of 26.1 kDa. The molecular weight of CAB-8 calculated from its *M* in SDS-PAGE is 27 kDa (see, for example, Fig. 1 in [13]). Other CAB polypeptides also show a discrepency between formula-calculated molecular weight and estimation of molecular weight from SDS-PAGE: the PSI Type I and Type II have formula-calculated molecular weights of 22.0 and 24.9 kDa, respectively, but migrate on SDS-PAGE as 23.7 and 26 kDa proteins, respectively, and PSII CAB proteins have formula-calculated molecular weights of 24.9–25.4 kDa but migrate on SDS-PAGE in the 27–28 kDa range [13, 22–25, 29]. The retardation in migration of all CAB polypeptides in SDS-PAGE is most likely due to their hydrophobic character. The degree of retardation relative to the formula-calculated molecular weight is smaller for the PSI CAB polypeptides, consistent with their being less hydrophobic. The reduced overall hydrophobicity of the PSI CAB polypeptides is due mainly to the existence of a strongly hydrophilic region between their second and third hydrophobic, membrane-spanning regions (see below), whereas

the corresponding segment in the PSII CAB polypeptides is only slightly hydrophilic [13, 29].

The protein encoded by *Cab*-8 is clearly structurally related to other CAB polypeptides. Sequence alignments (Fig. 7A) reveal that the two strongly conserved regions among all the CAB polypeptides analyzed to date [13, 24] are also conserved in the CAB-8 protein. These two regions, amino acids 62–120 and 199–265 (Fig. 7A), are each thought to consist of a membrane-spanning region (MSR) and a hydrophilic sequence N-terminal to the MSR, and it has been hypothesized that the conserved sequences in these MSRs are involved in chlorophyll binding [13, 21]. The CAB-8 protein, like all other tomato CAB sequences, also has a third predicted MSR situated between the two conserved MSRs, but it is not highly conserved in sequence among CAB proteins. However, CAB-8 does share some limited sequence homology in this region with CAB-7, the gene encoding the second most slowly migrating PSI CAB polypeptide [25].

The first conserved MSR in CAB-8 differs from all the other CAB proteins in having Asn and Gly at position 102–103 (Fig. 7A) instead of His and Cys. This variation is all the more interesting because the two conserved MSRs show some sequence similarity to each other [13], a similarity which, however, has not included these two positions in previously obtained CAB sequences. Thus, in the equivalent positions in the second conserved MSR, all the polypeptides have Asn and Gly rather than His and Cys; CAB-8 is so far the only CAB polypeptide which contains Asn and Gly in the equivalent positions in both the first and second conserved MSRs. In view of the established role of His residues in binding chlorophyll in prokaryotes and their postulated role in binding chlorophyll in plants [13], and the lack of His residues in the second conserved MSR in all CAB sequences determined to date, the apparent equivalency of His and Asn is intriguing (see also [21] for further discussion of probable ligands involved in pigment binding).

It was initially thought that CAB proteins are strongly conserved, because greater than 90% sequence identity was observed for CAB proteins from monocots and dicots [14, 15, 22]. It is now clear that such conservation is limited to the same type of CAB polypeptides. Thus, the Type I and Type II PSII CAB polypeptides are indeed over 90% identical in monocots and dicots *within the type*, but are only 85% identical to each other even within the same species [23]. The overall sequence homology of the Type III PSI CAB protein encoded by *Cab*-8 to the Type I and Type II PSI CAB proteins is 35–40%, which is only marginally greater than its homology to PSII CAB sequences (30%) (Fig. 7A, B). Thus, the sequences of the three types of PSI CAB polypeptides determined to date are substantially divergent from each other, as well as from the PSII CAB proteins. The large differences in primary sequence among the CAB polypeptides might reflect their different functions in the photosynthetic apparatus, especially in view of the strong conservation of PSII CAB proteins *within the type* among different plant species (for the PSI CAB sequences, with one exception [29] no sequences are yet available other than those for tomato). Alternatively, these differences might simply reflect the phylogentic relationships among the different members of the CAB gene family.

Expression of Cab-*8*

The pattern of steady-state level of *Cab*-8 mRNA in the different tomato organs and under different light regimes (Fig. 6) was very similar to that of other types of tomato PSI CABs. As estimated from the specific activity of the probe and the time of exposure of the blot (legend to Fig. 6), the steady-state level of *Cab*-8 mRNA is on the same order of magnitude of that of Type I and Type II PSI CAB. As is the case for other CAB mRNAs, the highest level of *Cab*-8 mRNA is seen in illuminated leaf tissue. Much less message is observed in other green organs such as stem and sepal, and no mRNA is observed in etiolated seedlings and roots.

The isolation of a cDNA clone, pCAB8-14, with some of the first intron sequence still intact and no sequence from the first exon present raises the possibility of alternate splicing. Alternatively, it is possible that this cDNA clone was simply made from an incompletely processed mRNA template (which was either prematurely degraded or incompletely copied

into a cDNA). We favor the latter explanation for the following reasons: if pCAB8-14 represents a functional, full-length message, the protein made from this mesage will be very short and also contain no discernable transit peptide; there is no evidence for such a short CAB polypeptide in the plant cell. If pCAB8-14 is an incomplete cDNA clone of a message that was processed differently from the pCAB8-52 message, inspection of the genomic sequence for possible 5′ and 3′ consensus splice reveals that such a message will code for a protein of at least 313 amino acids (including a transit peptide of 32–33 amino acids), and again there is no evidence for the existence of such a CAB polypeptide. Also, we have recently isolated a cDNA clone for a new type of CAB polypeptide in which an entire intron is present (E. Pichersky and B.R. Green, unpublished). Since that intron contains stop codons in-frame, clearly the message from which that cDNA clone was made was not yet completely processed.

Chromosomal location of Cab-8

Cab-8 is located on chromosome 10, approximately 9 map units from *Cab*-7, the gene encoding the Type II PSI CAB polypeptide (Fig. 5A). This is the first example of genes encoding different types of CAB polypeptides found on the same chromosome in tomato. Although the *Cab*-8 gene was isolated using a probe derived from the *Cab*-7 gene, and the CAB-7 protein indeed shows greater degree of homology to CAB-8 protein than any other tomato CAB sequence does (Fig. 7A, B), this degree of homology (40%) is still very low and it is therefore not clear whether their linkage is fortuitous or reflects a more recent evolutionary origin of these loci by gene duplication. The Southern blots indicate that the coding region of *Cab*-8 DNA hybridizes, at low stringency, to several tomato nuclear fragments in addition to itself (Fig. 4). We have identified some of these fragments as carrying the *Cab*-7, *Cab*-1, and *Cab*-3 genes, and are investigating whether any of the others carry previously uncloned CAB genes.

Structure of Cab-*8 and intron evolution in the CAB gene family*

Comparison of the sequence of the cDNA clone pCAB8-52 with the sequence of the *Cab*-8 gene reveals that *Cab*-8 contains two introns (Figs. 1, 2). The first intron occurs at or close to the position of the first intron in the Type II PSII CAB genes (because of the lack of sequence homology in this region, it is not possible to align the sequences there with certainty). The second intron in *Cab*-8 occurs at the exact position of the second intron in *Cab*-7; an intron occurs three codons downstream from this position in the tomato *Cab*-6A and *Cab*-6B genes (encoding the Type I PSI CAB protein). A summary of intron positions in different types of CAB genes is given in Fig. 7C. One observation is that PSII CAB genes, which are more highly expressed, have fewer introns (0 or 1) than PSI CAB genes (2, 3, and 4). The fact that many of the introns found in one type of CAB genes do not correspond in position to introns found in the other CAB types strongly suggests that some of them may have been introduced into the CAB genes in recent evolutionary times. Alternatively, there may have been a massive loss of introns in different lineages of the CAB gene family.

The second intron of *Cab*-8 does seem to be ancestral since it is also found in *Cab*-7. Moreover, in the tomato *Cab*-6 genes an intron occurs three codons downstream from the position of the second intron of *Cab*-8, and in a recently characterized PSII CAB gene (E. Pichersky and B.R. Green, unpublished) and intron occurs four codons downstream from this position. The first intron of *Cab*-8 might also be ancestral if it is indeed the same intron found in the Type II PSII CAB genes, but this identity is not certain. A better understanding of intron evolution in the CAB gene family will come when additional types of genes are examined and when the same types of CAB genes from different plant species are isolated, analyzed, and compared.

Acknowledgements

This work was supported by USDA grant 88-37262-3878 and a University of Michigan's Rackham

Faculty Research Grant to EP, and by an NSF grant DCB-8517773 to SDT.

References

1. Allen JF, Bennett J, Steinbeck KE, Arntzen CJ: Chloroplast protein phosphorylation couples plastoquinone redox state to distribution of excitation energy between photosystems. Nature 291: 25–29 (1981).
2. Anderson JM: Photoregulation of the composition, function and structure of thylakoid membranes. Ann Rev Plant Physiol 37: 93–136 (1986).
3. Anderson JM, Goodchild DJ: Transbilayer organization of the main chlorophyll a/b-protein of photosystem II of thylakoid membranes. FEBS Lett 213: 29–33 (1987).
4. Bennett J: Phosphorylation of chloroplast membrane polypeptides. Nature 291: 344–346 (1977).
5. Bennett J, Steinbeck KE, Arntzen CJ: Chloroplast phosphorylation: regulation of excitation energy transfer by phosphorylation of thylakoid membrane proteins. Proc Natl Acad Sci USA 77: 5253–5257 (1980).
6. Bernatzky R, Tanksley SD: Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112: 887–898 (1986).
7. Bernatzky R, Tanksley SD: Majority of random cDNA clones correspond to single loci in the tomato genome. Mol Gen Genet 203: 8–14 (1986).
8. Canaani O, Barber J, Malkin S: Evidence that phosphorylation and dephosphorylation regulate the distribution of excitation energy between the two photosystems of photosynthesis *in vivo*: photoacoustic and fluorimetric study of an intact leaf. Proc Natl Acad Sci USA 81: 1614–1618 (1984).
9. 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).
10. Dean C, Tamaki S, Dunsmuir P, Favreau M, Katayama C, Dooner H, Bedbrook J: mRNA transcripts of several plant genes are polyadenylated at multiple sites *in vivo*. Nucl Acids Res 14: 2229–2240 (1986).
11. Evåns PK, Anderson JM: The chlorophyll a/b-proteins of PSI and PSII are immunologically related. FEBS Lett 199: 227–233 (1986).
12. Green BR, Camm EL: The nature of the light-harvesting complex as defined by sodium dodecyl sulfate polyacrylamide gel electrophoresis. Biochim Biophys Acta 681: 256–262 (1982).
13. 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 84: 8844–8848 (1987).
14. Karlin-Neumann GA, Kohorn BD, Thornber 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).
15. Kohorn BD, Harel E, Chitnis PR, Thornber 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).
16. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685 (1970).
17. Mullet JE: The amino acid sequence of the polypeptide segment which regulates membrane adhesion (grana stacking) in chloroplasts. J Biol Chem 258: 9941–9948 (1983).
18. Mullet JE, Burke JJ, Arntzen CJ: Chlorophyll proteins of PSI. Plant Physiol 65: 814–822 (1980).
19. Maxam AM, Gilbert W: Sequencing end-labelled DNA with base-specific chemical cleavages. Meth Enzymol 65: 499–560 (1980).
20. Peter GF, Machold Q, Thornber JP: Identification and isolation of Photosystem I and Photosystem II pigment proteins from higher plants. In: Hawood JL, Walton TJ (eds). Plant Membranes, Structure and Function. The Biomedical Society, London (in press).
21. Peter GF, Thornber JP: Antenna components of photosystem II with emphasis on the major pigment-protein, LHC IIb. In: Scheer H, Schneider S (eds). Photosynthetic light-harvesting systems – structure and function, pp. 175–186. De Gruyter, Berlin (1988).
22. 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).
23. Pichersky E, Hoffman NE, Malik VS, Bernatzky R, Tanksley SD, Szabo L, Cashmore AR: The tomato *Cab*-4 and *Cab*-5 genes encode a second type of CAB polypeptide localized in photosystem II. Plant Mol Biol 9: 109–120 (1987).
24. Pichersky E, Hoffman NE, Bernatzky R, Piechulla B, Tanksley SD, Cashmore AR: Molecular characterization and genetic mapping of DNA sequences encoding the Type I chlorophyll a/b-binding polypeptide of photosystem I in *Lycopersicon esculentum* (tomato). Plant Mol Biol 9: 205–216 (1987).
25. Pichersky E, Tanksley SD, Piechulla B, Stayton MM, Dunsmuir P: Nucleotide sequence and chromosomal location of *Cab*-7, the tomato gene encoding the Type II chlorphyll a/b-binding polypeptide of Photosystem I. Plant Mol Biol 11: 69–71 (1988).
26. 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).
27. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Cunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdon N, Shimoda H, Sugiura M: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5: 2043–2049 (1986).

28. Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503–517 (1975).
29. Stayton MM, Brosio P, Dunsmuir P: Characterization of a full-length petunia cDNA clone encoding a polypeptide of the light-harvesting complex associated with photosystem I. Plant Mol Biol 10: 127–137 (1987).
30. 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. Mol Gen Genet 209: 247–256 (1987).
31. Tanksley SD, Miller J, Paterson A, Bernatzky R: Molecular mapping of plant chromosomes. In: Gustafson JP, Appels R (eds). Chromosome Structure and Function, pp. 157–173. Plenum Press, New York (1988).
32. Thornber JP: Chlorophyll-proteins: Light-harvesting and reaction center components of plants. Annu Rev Plant Physiol 26: 127–158 (1975).
33. White MJ, Green BR: Antibodies to the photosystem I chlorophyll a+b antenna crossreact with polypeptides of CP29 and LHCII. Eur J Biochem 163: 545–551 (1987).
34. Williams RS, Ellis RJ: Immunological studies on the light-harvesting polypeptides of PSI and PSII. FEBS Lett 203: 295–300 (1986).