

The tomato *Cab-4* and *Cab-5* genes encode a second type of CAB polypeptides localized in Photosystem II

E. Pichersky,^{1,4*} N. E. Hoffman,² V. S. Malik,¹ R. Bernatzky,³ S. D. Tanksley,³ L. Szabo¹ & A. R. Cashmore²

¹Laboratory of Cell Biology, The Rockefeller University, New York, NY 10021, USA; ²Plant Science Institute, Biology Department, University of Pennsylvania, Philadelphia, PA 19104, USA; ³Department of Plant Breeding and Biometry, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA; ⁴Permanent address: Biology Department, University of Michigan, Ann Arbor, MI 48109, USA (*author for correspondence)

Received 26 February 1987; in revised form 12 April 1987; accepted 13 April 1987

Key words: *Lycopersicon esculentum*, multigene family, cDNA clones, plant evolution

Abstract

The photosynthetic apparatus of plant chloroplasts contains two photosystems, termed Photosystem I (PSI) and Photosystem II (PSII). Both PSI and PSII contain several types of chlorophyll a/b-binding (CAB) polypeptides, at least some of which are structurally related. It has been previously shown that multiple genes encoding one type of PSII CAB polypeptides exist in the genome of many higher plants. In tomato, there are at least eight such genes, distributed in three independent loci. Genes encoding a second type of CAB polypeptides have been isolated from several plant species, but the precise location of the gene products has not been determined. Here we show that tomato has two unlinked genes encoding this second type and that this type of CAB polypeptide is also localized in PSII.

Introduction

Photosystems I and II (PSI and PSII), localized in the thylakoid membranes of plant chloroplasts, each contain several types of chlorophyll a/b-binding (CAB) polypeptides (for review see ref. [2]). The CAB polypeptides of each photosystem, together with the non-covalently bound chlorophyll and other pigment molecules, form aggregates known as Light Harvesting Complex I (LHCI) and LHCII (of PSI and PSII, respectively) which harvest light energy for the process of photosynthesis. In the absence of protein sequence data, the CAB proteins have been shown to be a heterogenous mixture by several techniques [10, 14,

28, 34]. Immunological comparisons have demonstrated that several CAB proteins in both PSI and PSII are structurally related to each other [10, 14] although substantial structural divergence between the PSI CAB proteins, on one hand, and the PSII CAB proteins, on the other hand, has occurred [10, 14, 34].

Recently it has been demonstrated that, in both monocotyledonous and dicotyledonous plants, one type of PSII CAB polypeptides is encoded by a set of 3–20 nuclear genes per haploid genome [5, 6, 9, 11, 12, 18, 20, 21]. The tomato genome contains 8–10 genes encoding this type of PSII CAB polypeptide [26, 27]. Genes encoding a second type of CAB polypeptides have been described in several

species [17, 26, 27, 29, 31], but the precise location of their products in the thylakoid has not been determined. We report here that two genes of this second type exist in the genome of the dicot plant *Lycopersicon esculentum* (tomato). We have used a full-length cDNA clone of one of these tomato genes to synthesize a precursor CAB polypeptide *in vitro*. The precursor was imported into isolated pea chloroplasts, and their thylakoid membranes were then fractionated to separate the two photosystems and determine the suborganellar location of the processed, mature protein. The results demonstrated that this second type of CAB polypeptides is localized in PSII. Cloning and sequencing of CAB genes, coupled with the assignment of the gene products to either PSI or PSII, will provide us with the amino acid sequences of PSI and PSII CAB proteins, and this structural information will hopefully help in elucidating the mechanistic details of the functions of these polypeptides.

Materials and methods

cDNA cloning

The construction of one tomato leaf cDNA library was previously described [4]. A second tomato leaf cDNA library was a gift from Dr Danny Alexander, ARCO Plant Research Institute, Dublin, CA, USA. It was constructed according to the procedure of Alexander *et al.* [1].

Library screening

The tomato cDNA library of Bernatzky and Tanksley [4] was first screened under non-stringent conditions ($6 \times \text{SSC}$, 60°C , washing at $2 \times \text{SSC}$, 60°C) with a mung-bean CAB cDNA probe. The mung-bean cDNA, which encodes a major component of LHCII [32], was used initially because at this stage of our investigation cloned tomato CAB sequences were not yet available. A tomato cDNA clone which hybridized weakly, relative to other tomato CAB cDNA clones, to the mung-bean probe was isolated. This clone, designated p3-48,

was shown by DNA sequencing to be an incomplete cDNA clone, and it was then used, under stringent conditions (hybridization: $4 \times \text{SSC}$, 70°C ; washing: $0.1 \times \text{SSC}$, 70°C), as a probe to isolate full-length cDNA clones from the tomato cDNA library provided to us by Dr D. Alexander. Two clones, designated pCAB4 and pCAB5, isolated from this library are the subject of this report.

DNA isolation, Southern blots, restriction digest and subcloning

The procedures employed are described and referenced in Pichersky *et al.* [26].

Nucleotide sequence determination

The chemical method [22] was employed as described in Pichersky *et al.* [26].

Genetic mapping

Linkage analysis of isozyme and cDNA markers with CAB-related restriction fragments was as previously described [3, 4, 33]. F2 progeny DNA, isolated from a cross between *L. esculentum* (a derivative of cv. VF36) and *L. pennellii*, was digested with *EcoRV* (pCAB4 probe) or *DraI* (pCAB5 probe), Southern blotted [30] and hybridized to nick-translated insert as described in Pichersky *et al.* [26]. After hybridization, the blots were washed under high stringency conditions ($0.1 \times \text{SSC}$, 68°C).

SP6 vectors and in vitro transcription and translation

An *EcoRV* – *EcoRI* fragment containing the *Cab-1B* gene [26] was cloned into SP64 vector [23]. To remove an ATG triplet upstream from the initiating ATG codon [26], 22 base pairs from the *EcoRV* site were eliminated using *Bal-31* exonuclease prior to cloning into the SP64 vector.

The *Cab-3C* gene, contained on a 1.3-kb *Hind*III–*Xba*I fragment (this fragment has an internal *Hind*III site; the 1.3-kb fragment was obtained by partial *Hind*III digest) [26] was also cloned into SP64. *Cab-4* was cloned into SP64 making use of the *Hind*III site on the cDNA (Fig. 2) and the *Xba*I site from the polylinker of the plasmid used in the construction of the cDNA library [1].

In vitro transcription was carried out according to Melton *et al.* [23] except that the rNTP concentration was reduced to 250 μ M and the dinucleotide GpppG was added to a final concentration of 500 μ M. The addition of GpppG resulted in the ‘capping’ of the mRNA produced. The *in vitro* synthesized mRNA was translated in the wheat germ system described by Erickson and Blobel [13] using ³⁵S-methionine to label the proteins. With this wheat germ system, we have found that capping of the mRNA was essential for translation.

Import of CAB precursors into chloroplasts and localization of processed CAB polypeptides within the chloroplasts

Isolation of pea chloroplasts and import of labelled precursors into isolated chloroplasts were carried out according to Schmidt *et al.* [28]. After import, the chloroplast mix (0.5 mg chlorophyll/ml) was incubated with thermolysin (final concentration 100 μ g/ml) at 0°C for 30 min. Isolation of thylakoid membranes and fractionation to PSI and LHCII were according to Mullet *et al.* [25] and Darr *et al.* [10] and references therein. Western blots were performed as described in Darr *et al.* [10] except that after incubation with the polyclonal antibodies, Protein A conjugated to alkaline phosphatase was used instead of the anti-mouse antibodies.

Results and discussion

cDNA isolation and genetic mapping

In previous studies, five genetic loci (designated

Cab-1 through *Cab-5*) containing CAB-related sequences were identified and mapped in the tomato genome [33]. Two of these, *Cab-1* on chromosome 2 and *Cab-3* on chromosome 3, were each shown [26] to contain several genes which encode proteins with over 90% homology to pea PSII CAB polypeptides [6] and also to reported CAB sequences from petunia [12], *Arabidopsis thaliana* [21], wheat [20] and *Lemna gibba* [18] (the latter two species are monocots). The tomato loci *Cab-4* (on chromosome 7) and *Cab-5* (on chromosome 12) were determined to be single-gene loci on the basis of the minimum size of DNA fragments from these loci which hybridized to the CAB probe [33]. The *Cab-4* and *Cab-5* genes also did not hybridize as strongly as *Cab-1*, *Cab-2*, and *Cab-3* genes did to probes derived from either a mung-bean CAB gene or from tomato *Cab-1* and *Cab-3* genes [26, 27, 33].

We have now isolated cDNA clones which were synthesized from mRNA transcribed from the *Cab-4* and *Cab-5* genes. Figure 1A shows the *Eco*RV-restricted DNA of segregating F2 progeny (see Materials and methods, section ‘Genetic mapping’ for description of the cross) observed on a Southern blot when the insert in plasmid pCAB4 (see Material and methods, section ‘Library screening’ for designation of plasmids) was used as a probe. Under the high-stringency hybridization conditions used (see Materials and methods, ‘Genetic mapping’), over 80% of the bound probe was hybridized to a DNA fragment of 3.6 kb (the *L. esculentum* allele), to a DNA fragment of slightly faster mobility (the *L. pennellii* allele) or to both (the heterozygous F2), thus identifying these fragments as carrying the gene from which the insert in pCAB4 was obtained. Previous co-segregation analysis with isozyme and cDNA markers has identified the *L. esculentum* 3.6-kb *Eco*RV fragment as constituting the *Cab-4* locus on chromosome 7 (see ref. [33] for details on the linkage). Fig. 1B shows segregation analysis of F2 progeny DNA digested with *Dra*I, with the insert in pCAB5 as a probe under the same stringent conditions used in Fig. 1A. Under these conditions, the probe bound almost exclusively to a DNA fragment of 2.5 kb (the *L. esculentum* allele), a slightly smaller fragment (the *L. pennellii* allele) or to both (the heterozygous F2).

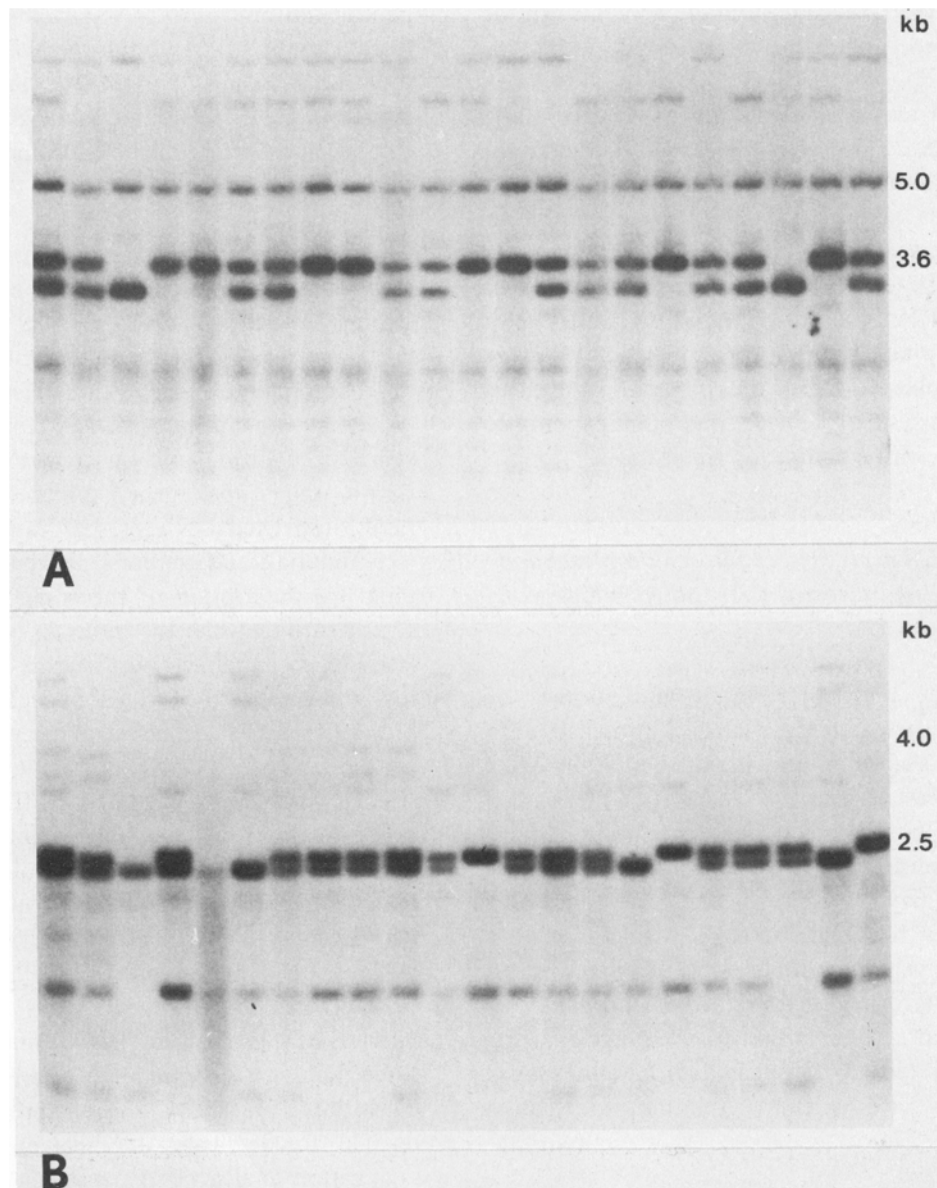


Fig. 1. F2 segregation analyses. A: F2 progeny DNA digested with *EcoRV* and hybridized to the insert of pCAB4. The numbers at right refer to size of restriction fragments in kilobases (kb). B: F2 progeny DNA digested with *DraI* and hybridized to the insert of pCAB5. Hybridization conditions were as described in the 'Genetic mapping' section in Materials and methods. See text for details.

The 2.5-kb *DraI* fragment, which thus carries the gene from which the insert in pCAB5 was made, has been previously mapped to the *Cab-5* locus on chromosome 12 [33].

Sequences of pCAB4, pCAB5 and encoded proteins

The nucleotide sequences, together with the predicted amino acid sequences, of the cDNA clones corresponding to *Cab-4* and *Cab-5* are shown in Fig. 2. pCAB4 encodes a CAB polypeptide of 265 amino acids, and also contains 58 base pairs (bp) upstream from the initiating methionine codon and 39 bp downstream from the termination codon TGA plus a poly-A tail of about 50 bases

(Figs. 2, 3). The coding region in pCAB4 is approximately 70% homologous to tomato CAB genes from the *Cab-1* and *Cab-3* loci. pCAB5 is an incomplete cDNA clone; its sequence specifies the entire mature polypeptide and eight amino acids of the transit peptide (Figs. 2, 3). Within the overlapping CAB-coding sequences of pCAB4 and pCAB5 there is 88% homology at the nucleotide level, and the mature CAB polypeptides (229 amino acids) differ by only four amino acids (1.7% divergence).

CAB protein localization

To determine whether the CAB polypeptides encoded by the divergent lineage of CAB genes are

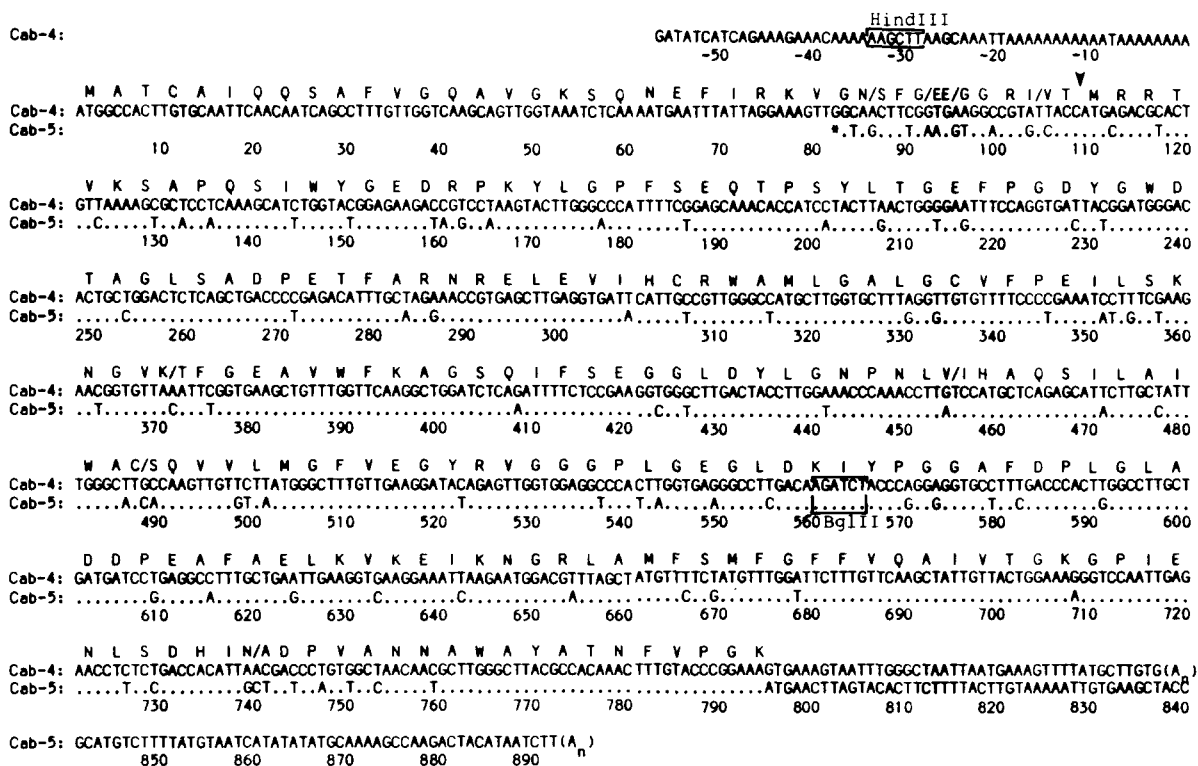


Fig. 2. Nucleotide sequences of pCAB4 and pCAB5 cDNAs and the predicted amino acid sequences of the encoded CAB polypeptides. All non-coding sequences, and the pCAB4 coding sequence, are shown in full. The coding sequence of pCAB5 is shown only when it differs from that of pCAB4. The amino acid sequence (one-letter symbols) is shown above, and, when different, the amino acid specified by pCAB4 is on the left, the amino acid specified by pCAB5 is on the right. The locations of the *HindIII* site in pCAB4 and the *BglIII* sites in pCAB4 and pCAB5 are also indicated. The arrow indicates the putative processing site of the precursor in the chloroplast. The pCAB5 clone is incomplete; an asterisk indicates the extent of the sequence available.

```

      1      10      20      30      40      50
CAB-1B MAAATMALSSPSFAGQAVKLSPSAS----EISGNGRITMRKAY--AKSAPSSSPW
CAB-3C ..TS.....ST...K.....S-----T...V...TATK..P.S.G...
LGAB30 ...-S.....LV.K...A.A.S----VF.E..VS...TAGKP.PVS.G...
CAB-4  ...-C.IQQSA.V...GK.QNEFIRKVGNF.E.....RT.---...-Q.I.
CAB-5  .....*SFEG..V...RT.---...-Q.I.
LGAB19 ...SAI---.SA.....L.-QRDELVRKVGSG---.FS..RT.---AVS-Q.I.

      60      70      80      90      100
CAB-1B YGPDORVKYLGPFSGESPSYLTGEFPGDYGWDTAGLSADPETFAKNRELEVIHCRW
CAB-3C .....
LGAB30 .....A.....A.....
CAB-4  ..E..P.....EQT.....R.....
CAB-5  ..E..P.....EQT.....R.....
LGAB19 ..A..P.F.....EQT.....S...

      110     120     130     140     150     160
CAB-1B AMLGALGCVFPELLARNGVKFGEAVWFKAGSQIFSEGGLDYLGNPVLVHAQSILA
CAB-3C .....
LGAB30 .....
CAB-4  .....I.SK.....N.....
CAB-5  .....I.SK...T.....N.I.....
LGAB19 .....I....SK...Q.....A.....N.....

      170     180     190     200     210
CAB-1B IWACQVVLMGAVEGYR IAGGPLGEVVDPLYPGGSFDPLGLAEDPEAFELKVKEI
CAB-3C .....D.....
LGAB30 ..T.....V.....D.....
CAB-4  .....F...VG...GL.KI...A...D.....
CAB-5  ..S...F...VG...GL.KI...A...D.....
LGAB19 ..T...L...V...G...L...A...D.....

      220     230     240     250     260
CAB-1B KNGRLAMFSMFGFFVQA IYTGKGPLLENLADHLADPVNNNAWAFATNFVPGK
CAB-3C .....
LGAB30 .....
CAB-4  .....S..IN..A...Y.....
CAB-5  .....S..I...A...Y.....
LGAB19 .....I...A.....

```

Fig. 3. CAB protein sequence comparisons. The predicted CAB protein sequences encoded by the tomato *Cab-1B* and *Cab-3C* [26], *Cab-4* and *Cab-5* (this report), and the *Lemna gibba* genes *LGAB30* [18] and *LGAB19* [17] are shown. Only the top sequence is shown in full; other sequences are shown only when they differ from the top sequence. A dash represents a gap introduced to maximize homology. The numbering refers to the longest polypeptide (CAB-3C). The CAB-5 sequence is derived from an incomplete cDNA clone; an asterisk indicates the extent of the sequence available. The arrow indicates the putative processing site of the precursor in the chloroplast.

localized in PSII or PSI, we imported both types of tomato CAB precursors into isolated pea chloroplasts and studied their distribution inside the organelle. The import of protein precursors by isolated chloroplasts is now a routine procedure [8, 15] and it is also well established that chloroplasts of pea and other higher plants will import, process, and assemble with fidelity precursors from diverse higher plant species [7, 24, 28]. The CAB protein precursors were obtained by cloning the full-length

cDNA from pCAB4 and the intronless *Cab-1B* and *Cab-3C* genes into the *in vitro* transcription system SP6 [23] in order to obtain functional RNA, followed by *in vitro* translation (see Material and methods).

Import of *in vitro* synthesized, radioactively labelled tomato CAB-1B, CAB-3C and CAB-4 precursor proteins by isolated pea chloroplasts was followed by fractionation into PSI and LHCII, the light-harvesting complex of PSII (see Materials and

methods). The fractionation method used, linear sucrose gradient, results in a major green band at the top of the gradient and a second green band at the bottom of the gradient; the upper band contains the LHCII complex, the lower band contains the entire PSI fraction [25]. Essentially all TCA-

precipitable radioactivity was located in fractions corresponding to the upper chlorophyll peak (Fig. 4A, B).

To determine the degree of separation of PSI and LHCII, we used the CAB polypeptides of the two photosystems as markers; these proteins were de-

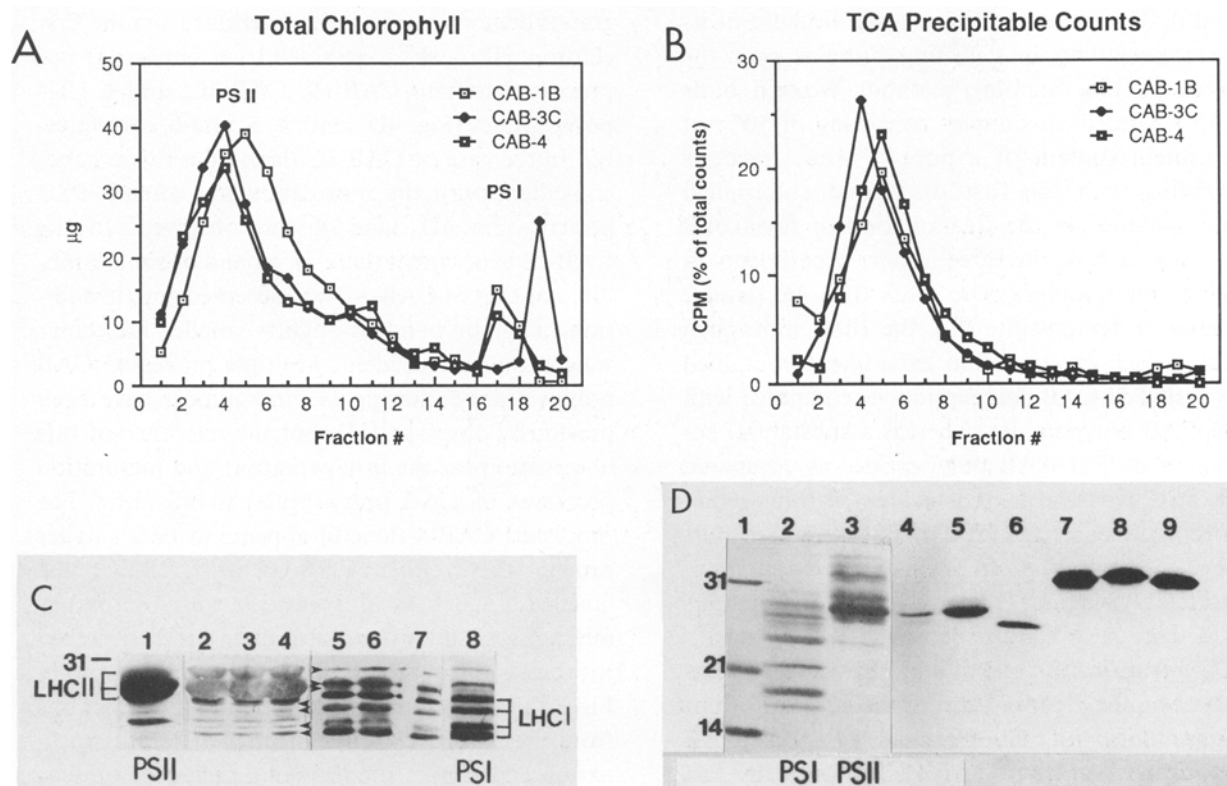


Fig. 4. Localization of CAB-1B, CAB-3C and CAB-4 polypeptides in the chloroplast. Import of *in vitro* synthesized CAB-1B, CAB-3C and CAB-4 precursors into pea chloroplasts was followed by isolation of thylakoid membranes and further fractionation into PSI and PSII complexes (see text for details).

A: chlorophyll content of fractions (0.5 ml) obtained after sucrose gradient centrifugation of solubilized thylakoid membranes following import experiments; fraction 1 is the top of the gradient.

B: TCA-precipitable counts of the 20 fractions from each of the three thylakoid fractionations following import.

C: SDS-PAGE of fractions of pea thylakoid fractionation experiments. Lane 1 – LHCII (the light-harvesting complex of PSII). Lane 8 – PSI. Lanes 1 and 8 are samples *not* from an import experiment and are stained with Coomassie blue. Lanes 2 through 7, Western blots with anti-CAB antibodies of fractionated thylakoids after import experiments (see text for details). The samples are from: Lane 2 – first chlorophyll peak of the CAB-1B experiment (fractions 4–6), lane 3 – 1st chlorophyll peak of the CAB-3C experiment (fraction 3–5), lane 4 – 1st chlorophyll peak of the CAB-4 experiment (fraction 3–5), lane 5 – second chlorophyll peak of the CAB-1B experiment (fraction 16–18), lane 6 – 2nd chlorophyll peak of the CAB-3C experiment (fraction 18–20), lane 7 – 2nd chlorophyll peak of the CAB-4 experiment (fraction 16–18). The arrows point to the PSI CAB proteins in the PSII fractions and the PSII CAB proteins in the PSI fractions.

D: SDS-PAGE. Lane 1 – molecular weight markers and their size, in kDa, lane 2 – tomato PSI fraction and lane 3 – tomato LHCII fraction, all stained with Coomassie blue. Lanes 4, 5 and 6 – autoradiography of samples from the first chlorophyll peak of the CAB-1B, CAB-3C, and CAB-4 import experiments, respectively. Lane 7, 8 and 9 – autoradiography of *in vitro* synthesized CAB-1B, CAB-3C, and pCAB-4 precursors, respectively.

ected by the Western blotting technique using antibodies made against a *Vicia faba* PSI CAB polypeptide (a gift from Dr Kenton Ko). The antibodies in this polyclonal antibody preparation recognize the same proteins in pea PSII and PSI as does an anti-pea CAB monoclonal antibody, as well as one additional protein in PSII (compare our Fig. 4C, lanes 1 and 8, with Darr *et al.* [10], Fig. 4 lanes g, h and i). This technique was chosen because of its greater sensitivity in detecting proteins over the Coomassie blue staining method. Western blots were performed on samples consisting of 50% of the protein content of a pool of three fractions coinciding with the first or second chlorophyll peaks obtained in the fractionation of thylakoids following each of the three import experiments (a total of six samples). The blots (Fig. 4C, lane 2 through 7) demonstrate that the first chlorophyll peak in each fractionation experiment contained very little PSI CAB polypeptides as compared with PSII CAB polypeptides, whereas a substantial enrichment of PSI CAB polypeptides as compared with PSII CAB proteins is evident in the second chlorophyll peaks (the ratio of PSII complex units to PSI complex units in thylakoid membranes is considerably greater than 1; see, for example, Fig. 4, lane A in Darr *et al.* [10]).

Our identification of PSI and PSII CAB proteins is based on the identity between the patterns of our Western blots with that obtained with an anti-CAB monoclonal antibody ('MLH12', see text and Fig. 4, lanes g, h and i in Darr *et al.* [10]). Note, however, that MLH12 and our polyclonal antibodies react to a different extent with some of the CAB proteins in PSI. Because of the different degree of cross-reactivity with different CAB polypeptides exhibited by our polyclonal antibody preparation as well as by the monoclonal, the intensity of staining of the different CAB proteins in Fig. 4C lanes 2–7 is not an indication of their abundance *relative to each other* (in addition, some uneven fading of the stain used in the blotting technique in conjunction with the alkaline phosphatase has occurred, especially in the wide bands). We have observed with duplicate CAB protein samples stained with Coomassie blue and Western blotted (data not shown) that the width of each protein band is a bet-

ter indication of their relative abundance.

Samples from the fractions of the two chlorophyll peaks of each of the three import experiments were also subjected to SDS-PAGE followed by autoradiography. In this experiment, only one tenth of the protein content of the corresponding fraction was used in order to avoid distortion of bands caused by overloading the gel. The autoradiographs demonstrate that the fractions from the first chlorophyll peaks contained the imported and processed tomato CAB-1B, CAB-3C, and CAB-4 polypeptides (Fig. 4D, lanes 4, 5, and 6, respectively). In the case of CAB-3C (lane 5), a single band, coincident with the most abundant tomato PSII protein (Fig. 4D, lane 3), was observed. In the CAB-1B experiment (lane 4), a band of similar mobility as that of CAB-3C was observed and, in addition, a minor band of slightly smaller molecular weight was also evident. Multiple processed CAB polypeptides following *in vitro* import have been previously observed [18], but the relevance of this observation for the *in vivo* import and maturation processes of CAB polypeptides is not clear. The processed CAB-4 (lane 6) appears to be a smaller protein than either CAB-1B or CAB-3C; the labelled CAB-4 band coincides with the faster-migrating of the two minor protein bands observed just below the abundant CAB polypeptide band in Fig. 4D, lane 3. Autoradiography of the fractions from the second chlorophyll peak of each fractionation experiment for the same period of time as for the fractions of the first peaks (10 d) revealed no labelled polypeptides (data not shown), consistent with the profiles of the TCA-precipitable counts.

Although the fractions of the first chlorophyll peaks are slightly contaminated with PSI, there are clearly more PSI CAB proteins in the second peak than in the first peak in every fractionation experiment. Yet, the labelled protein is associated with the first chlorophyll peak, which is highly enriched in LHCII, and not the second one. Thus, our results indicate that the CAB-1B, CAB-3C and CAB-4 proteins are all localized in LHCII, the light-harvesting complex of PSII. Although we have not tested CAB-5 and each protein encoded by the rest of the CAB genes in the *Cab-1* and *Cab-3* loci, be-

cause of their strong structural homologies to the tested CAB proteins it is reasonable to assume that they too are localized in LHCII.

It should be noted that Kohorn and Tobin [19] have shown that the transit peptide of a *Lemna gibba* CAB protein (encoded by the gene *LGABI9*, see next section) similar to the tomato CAB-4 and CAB-5 could direct the import of an *in vitro* altered CAB protein into the chloroplasts and its subsequent targeting to the thylakoid membrane. Furthermore, some of the processed protein was associated with PSII. However, the majority of the processed forms of this aberrant CAB polypeptide was found elsewhere in the thylakoids. The failure to demonstrate exclusive association of this protein with PSII was attributed to the nature of the changes induced in it [19]. In the import experiments reported here, we have used naturally occurring CAB precursors, and in all cases, including the one involving the import of the CAB-4 precursor, most of the imported and processed proteins were associated with the PSII complex. Furthermore, unlike the situation in the study by Kohorn and Tobin, no additional molecular forms of the proteins were found elsewhere in the thylakoids; the autoradiographs of SDS-PAGE of unfractionated thylakoids from each of our import experiments gave identical patterns (data not shown) to that obtained in autoradiographs of SDS-PAGE of the corresponding PSII fractions (Fig. 4D, lanes 4–6).

In pea, the major fraction of the PSII CAB polypeptides is encoded by a set of genes with high homology to the tomato *Cab-1* and *Cab-3* genes [5, 6, 9, 26]. A minor fraction of slightly smaller molecular weight PSII CAB polypeptides is also observed [28]; however, at present it is not known whether this fraction is encoded by gene(s) homologous to the tomato *Cab-4* and *Cab-5* genes.

Protein sequence comparisons

CAB-4 and CAB-5 proteins show 85% homology in the mature part of the protein with the tomato CAB-1 or CAB-3 proteins, and less than 50% homology in the transit peptides (Fig. 3). In comparison, CAB-4 and CAB-5 are 92% homologous in

the mature part with the protein encoded by a CAB gene, designated *LGABI9*, from the monocot species *Lemna gibba* [17]. Another *Lemna gibba* CAB gene, *LGAB30* [18], encodes a protein with over 90% homology to the tomato CAB-1 and CAB-3 proteins but only 83% homology to the CAB-4 and CAB-5 (Fig. 3). These comparisons suggest that the tomato genes in the *Cab-1* and *Cab-3* loci and the *Lemna gibba* gene *LGAB30* belong to one CAB gene lineage, which we designate Type I PSII CAB genes, and the tomato genes *Cab-4* and *Cab-5* and the *Lemna gibba* gene *LGABI9* belong to a second CAB gene lineage, which we designate Type II PSII CAB genes. These results further suggest that the two lineages diverged from each other before the split of the monocots and dicots. A similar conclusion was reached independently by Smeekens *et al.* [29] based on a partial sequence of cDNA clone for Type II PSII CAB gene from the dicot *Silene pratensis*.

Although the function of the CAB polypeptides is understood, the precise roles of the distinct types of CAB polypeptides is not well understood. Experimental evidence and theoretical considerations have been used to predict the conformation of the CAB polypeptides in the thylakoid membranes [17, 18]. Sequence comparisons of CAB variants can identify (1) segments of the CAB proteins which are highly conserved and might possibly carry out an invariable function such as binding chlorophyll a and b, and (2) other regions which are less conserved and might serve other functions. For this analysis to produce optimal results, however, we first need to know the exact function of each CAB polypeptide. In comparisons of the Type I and Type II CAB polypeptides (Fig. 3) segments of relatively high sequence conservation are clearly discernable (e.g., aa 70–120 and 190–240). It is interesting to note that the second half of each of these two regions is predicted to span the thylakoid membrane and the first half of each is on the stromal side of the membrane in the model of Karlin-Neumann *et al.* [17]. Furthermore, these two regions are also conserved (although to a lesser extent) in a PSI CAB polypeptide whose cDNA clone we have recently isolated [16]. Other regions are less conserved, and in particular the N-termini of the

mature CAB polypeptides show considerable divergence (see also ref. 26). This is an interesting observation in view of the fact that the N-termini of pea CAB polypeptides have been assigned both structural and functional roles (see discussion in Pichersky *et al.* [26] and references therein).

Structure of the *Cab-4* gene

The *Lemna gibba* CAB gene *LGAB19*, but not *LGAB30*, contains an intron [17, 18]. The tomato *Cab-1* and *Cab-3* genes are known to lack introns [26]. If the tomato *Cab-4* and *Cab-5* genes belong to the same lineage as *LGAB19*, they might also be expected to contain introns. Since we have not been successful in our attempts to isolate genomic clones containing the *Cab-4* and *Cab-5* genes (our genomic libraries do not contain these genes), we tested this possibility in the following way. pCAB4 was restricted with the enzymes *Hind*III and *Bgl*II and run on agarose gel side by side with tomato genomic DNA restricted with the same enzymes. The gel was then blotted and probed with the 0.6-kb *Hind*III–*Bgl*II fragment from pCAB4 (Fig. 2) followed by washing under stringent conditions ($0.1 \times \text{SSC}$, 68°C) for a total of 1 h. The 0.6-kb *Hind*III–*Bgl*II fragment of pCAB4 encompasses the coding sequence of the N-terminal 75% of the gene. In *LGAB19*, the single 84-bp intron is found about 20% into the coding sequence [17]. Two fragments of 0.7 and 1.5 kb were observed in the genomic blot (Fig. 5 lane B), both of which are larger than the 0.6-kb *Hind*III–*Bgl*II fragment of pCAB4 (Fig. 5 lane A). Prolonged washing under the same stringent conditions resulted in much lower radioactivity in the 1.5-kb band with only minor reduction in the radioactivity bound to the 0.7-kb fragment (data not shown), suggesting that the 1.5-kb fragment corresponds to the *Cab-5* gene and the 0.7-kb fragment corresponds to the *Cab-4* gene.

The observation that the genomic *Hind*III–*Bgl*II fragment derived from *Cab-4* is approximately 100 bp longer than the 0.6-kb *Hind*III–*Bgl*II fragment in the cDNA clone pCAB4 (Fig. 5) demonstrates that *Cab-4* does possess intron(s). This result cannot determine un-

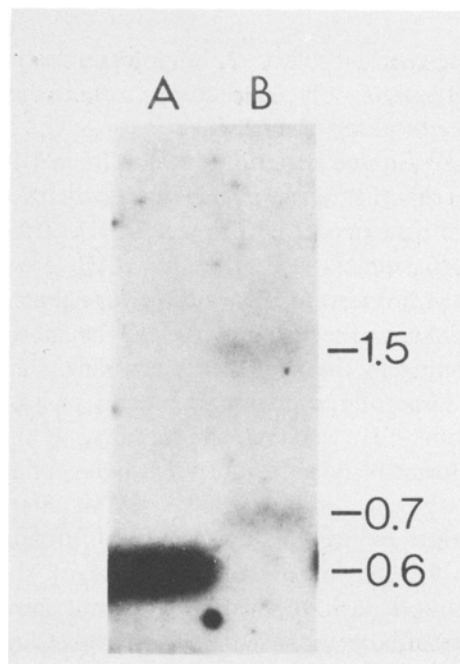


Fig. 5. Southern blot analysis. Samples were run on 1% agarose gel, blotted to nitrocellulose paper, and probed with ^{32}P -labelled 0.6 kb *Hind*III–*Bgl*II fragment from pCAB4 under stringent conditions (see Materials and methods). Lane A, pCAB4 restricted with *Hind*III and *Bgl*II, Lane B, tomato genomic DNA restricted with the same enzymes. The size of the observed fragments, in kilobase, are shown on the right. The 0.7-kb fragment in lane B is from *Cab-4* (see text).

quivocally the number and exact position of introns, but because of the small difference in size between the genomic and cDNA fragments it is likely that only a single intron is present in this part of the *Cab-4* gene. At present we do not know whether *Cab-5* contains introns. A petunia Type II PSII CAB gene whose sequence was recently reported also contains a single intron in the same position as in the *Lemna giba* *LGAB19* gene [31]. This petunia gene encodes a protein whose mature part differs by four residues from that of CAB-5 and seven residues from that of CAB-4 (1.7% and 3.1% divergence, respectively), a level that is consistent with the taxonomic relationship between petunia and tomato (both are in family Solanaceae) and the generally high level of Type I PSII CAB protein sequence conservation among diverse plant species.

The data presented in this paper, together with data from other laboratories, indicate that there is

substantial heterogeneity for PSII CAB protein sequences specified by the genetic system. Whether, and to what extent, additional diversity is generated by post-translational modification of CAB polypeptides *in vivo* is still an open question. Both monocots and dicots contain at least two types of PSII CAB polypeptides (as defined here), Type I and Type II, and these are substantially divergent from each other but strongly conserved, within types, across species. These observations strongly suggest some differences in function for the two types.

Acknowledgements

We wish to thank Dr Danny Alexander, Plant Research Institute, ARCO, Dublin, CA, USA for his gift of the tomato leaf cDNA library, Dr Kenton Ko for his gift of anti-CAB antibodies, and Dr Sylvia C. Darr for communicating results before publication. This work was supported by a DOE (DE-FG02-87ER13680) and a NIH (GM38408) grant to ARC and NSF grant DCB-8517773 to SDT. EP is a recipient of an NIH Postdoctoral Training Grant Fellowship, and NEH is a recipient of an NSF post-doctoral fellowship.

References

- Alexander DC, McKnight TD, Williams BG: A simplified and efficient vector-primer cDNA cloning system. *Gene* 31: 79–89, 1984.
- Anderson JM: Photoregulation of the composition, function and structure of thylakoid membranes. *Ann Rev Plant Physiol* 37: 93–136, 1986.
- Bernatzky R, Tanksley SD: Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* 112: 887–898, 1986.
- Bernatzky R, Tanksley SD: Majority of random cDNA clones correspond to single loci in the tomato genome. *Mol Gen Genet* 203: 8–14, 1986.
- Brogliè R, Bellemare G, Bartlett SG, Chua N-H, Cashmore AR: Cloned DNA sequences complementary to mRNA encoding precursors to the small subunit of ribulose 1,5-bisphosphate carboxylase and a chlorophyll a/b-binding polypeptide. *Proc Natl Acad Sci USA* 78: 7304–7308, 1981.
- 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.
- Chua N-H, Schmidt GW: Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose 1,5-bisphosphate carboxylase. *Proc Natl Acad Sci USA* 75: 6110–6114, 1978.
- Cline K, Werner-Washburne M, Lubben TH, Keegstra K: Precursors of two nuclear-encoded chlorophyll proteins bind to the outer envelope membrane before being imported into chloroplasts. *J Biol Chem* 260: 3691–3696, 1985.
- Corozzi G, Brogliè R, Cashmore AR, Chua N-H: Nucleotide sequences of two pea cDNA clones encoding the small subunit of ribulose 1,5-bisphosphate carboxylase and the major chlorophyll a/b-binding thylakoid polypeptide. *J Biol Chem* 258: 1399–1402, 1983.
- Darr SC, Sommerville SC, Arntzen CJ: Monoclonal antibodies to the light-harvesting chlorophyll a/b protein complex of photosystem II. *J Cell Biol* 103: 733–740, 1986.
- 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.
- Dunsmuir P: The petunia chlorophyll a/b-binding genes: a comparison of CAB genes of different gene families. *Nucleic Acids Res* 13: 2503–2518, 1985.
- Erickson AH, Blobel G: Cell-free translation of mRNA in a wheat germ system. *Methods Enzymol* 96: 38–50, 1983.
- Evans PK, Anderson JM: The chlorophyll a/b-proteins of PSI and PSII are immunologically related. *FEBS Lett* 199: 227–233, 1986.
- Grossman AR, Bartlett SG, Schmidt GW, Mullet JE, Chua N-H: Optimal conditions for post-translational uptake of proteins by isolated chloroplasts. *J Biol Chem* 257: 1558–1563, 1982.
- 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.
- 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.
- 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.
- Kohorn BD, Tobin EM: Chloroplast import of light-harvesting chlorophyll a/b-proteins with different amino termini and transit peptides. *Plant Physiol* 82: 1172–1174, 1986.
- 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.
- 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.

22. Maxam AM, Gilbert W: Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol* 65: 499–560, 1980.
23. Melton DA, Krieg PA, Rebagliati MR, Maniatis T, Zinn K, Green MR: Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res* 12: 7035–7056, 1984.
24. Mishkind ML, Wessler SR, Schmidt GW: Functional determinants in transit sequences: import and partial maturation by vascular plant chloroplasts of the ribulose 1,5-bisphosphate carboxylase of *Chlamydomonas*. *J Cell Biol* 100: 226–234, 1985.
25. Mullet JE, Burke JJ, Arntzen CJ: Chlorophyll proteins of photosystem I. *Plant Physiol* 65: 814–822, 1980.
26. 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.
27. 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.
28. Schmidt GW, Bartlett SG, Grossman AR, Cashmore AR, Chua N-H: Biosynthetic pathways of two polypeptide subunits of the light harvesting chlorophyll a/b protein complex. *J Cell Biol* 91: 468–478, 1981.
29. 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.
30. Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98: 503–517, 1975.
31. 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. *Nucleic Acids Res* 14: 9781–9796, 1986.
32. Thompson WF, Everett M, Polans NO, Jorgensen RA: Phytochrome control of RNA levels in developing pea and mung-bean leaves. *Planta* 158: 487–500, 1983.
33. Vallejos 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* 11: 93–105, 1986.
34. Williams RS, Ellis RJ: Immunological studies on the light-harvesting polypeptides of PSI and PSII. *FEBS Lett* 203: 295–300, 1986.