



HAL
open science

Molecular characterization of a second copy of holocarboxylase synthetase gene (*hcs2*) in *Arabidopsis thaliana*

Laurence Denis, Marie Grossemy, Roland Douce, Claude Alban

► **To cite this version:**

Laurence Denis, Marie Grossemy, Roland Douce, Claude Alban. Molecular characterization of a second copy of holocarboxylase synthetase gene (*hcs2*) in *Arabidopsis thaliana*. *Journal of Biological Chemistry*, 2002, 277 (12), pp.10435 - 10444. 10.1074/jbc.M111110200 . hal-02678731

HAL Id: hal-02678731

<https://hal.inrae.fr/hal-02678731>

Submitted on 31 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Molecular Characterization of a Second Copy of Holocarboxylase Synthetase Gene (*hcs2*) in *Arabidopsis thaliana**

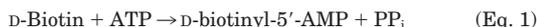
Received for publication, November 20, 2001
Published, JBC Papers in Press, January 9, 2002, DOI 10.1074/jbc.M111110200

Laurence Denis, Marie Grossemy, Roland Douce, and Claude Alban‡

From the Laboratoire mixte CNRS/INRA/Aventis (UMR 1932), Aventis crops science, 14–20 rue Pierre Baizet, 69263 Lyon CEDEX 9, France

Holocarboxylase synthetase (HCS), catalyzing the covalent attachment of biotin, is ubiquitously represented in living organisms. Indeed, the biotinylation is a post-translational modification that allows the transformation of inactive biotin-dependent carboxylases, which are committed in fundamental metabolisms such as fatty acid synthesis, into their active *holo* form. Among other living organisms, plants present a peculiarly complex situation. In pea, HCS activity has been detected in three subcellular compartments and the systematic sequencing of the *Arabidopsis* genome revealed the occurrence of two *hcs* genes (*hcs1* and *hcs2*). *Hcs1* gene product had been previously characterized at molecular and biochemical levels. Here, by PCR amplification, we cloned an *hcs2* cDNA from *Arabidopsis thaliana* (Ws ecotype) mRNA. We observed the occurrence of multiple cDNA forms which resulted from the alternative splicing of *hcs2* mRNA. Furthermore, we evidenced a nucleotide polymorphism at the *hcs2* gene within the Ws ecotype, which affected splicing of *hcs2* mRNA. This contrasted sharply with the situation at *hcs1* locus. However, this polymorphism had no apparent effect on total HCS activity *in planta*. Finally, *hcs2* mRNAs were found 4-fold less abundant than *hcs1* mRNA and the most abundant *hcs2* mRNA spliced variant should code for a truncated protein. We discuss the possible role of such a multiplicity of putative HCS proteins in plants and discuss the involvement of each of *hcs* genes in the correct realization of biotinylation.

Biotin (vitamin H) is a cofactor for some carboxylases and decarboxylases (1). The biotinylation of these enzymes is a post-translational modification, which allows the transformation of inactive apo-proteins into their active *holo* forms. The covalent attachment of biotin is catalyzed by biotin protein ligase also called holocarboxylase synthetase (HCS).¹ D-Biotin is incorporated to a specific Lys residue of newly synthesized apo enzyme, via an amide linkage between the biotin carboxyl group and a unique ϵ -amino group of Lys residue (2). It occurs in two steps, Equations 1 and 2 as follows,



* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF414935-AF414947.

‡ To whom correspondence should be addressed. Tel.: 33-472-85-28-42; Fax: 33-472-85-22-97; E-mail: claud.alban@aventis.com.

¹ The abbreviations used are: HCS, holocarboxylase synthetase; ORF, open reading frame; RACE, rapid amplification of cDNA ends; RT, reverse transcriptase; SNP, single nucleotide polymorphism.

D-Biotinyl-5'-AMP + apocarboxylase

→ holocarboxylase + AMP (Eq. 2)

In *Escherichia coli*, there is only one biotin-dependent carboxylase, the acetyl-CoA carboxylase. It is biotinylated by HCS also called BirA. The *BirA* gene has been cloned and the protein purified and extensively studied (for a review, see Ref. 3). This is a model system for the biochemical study of biotinylation. However, if the ligases from other bacterial sources are very similar to the *E. coli* enzyme, their eukaryotic counterparts appear quite different. On the one hand, there is more than only one biotin-dependent carboxylase in eukaryotic cells, and hence, there are different substrates for the biotinylation reaction. This raises the issue of substrate specificity of this reaction. On the other hand, the cytoplasm of eukaryotic cells is compartmentalized and the occurrence of biotin-dependent carboxylases in different subcellular compartments complicates the understanding of their biotinylation. In mammalian cells, HCS activity is located in both cytosol and mitochondria (4). In human cells, there are four carboxylases requiring biotin as a prosthetic group (acetyl-CoA carboxylase in the cytosol and pyruvate carboxylase, 3-methylcrotonyl-CoA carboxylase, and propionyl-CoA carboxylase in mitochondria). Only one gene coding for HCS had been evidenced (5, 6). As a consequence, in humans, mitochondrial and cytosolic HCS activities should result from alternative splicing of the corresponding mRNA. Furthermore, different HCS polypeptides had been detected in cytosol that could result from the alternative use of three translation initiation codons (7). In yeast, only one *hcs* gene has been isolated and code for the cytosolic activity of HCS, enabling the biotinylation of both carboxylases detected in yeast cytosol (acetyl-CoA carboxylase and pyruvate carboxylase) (8).

In the plant kingdom, HCS activity has been detected in three subcellular compartments (cytosol, mitochondrion, and plastid) from pea leaves (9). This is therefore the most complex situation described. These locations are in agreement with the presence of carboxylases in each of the subcellular compartments (acetyl-CoA carboxylase and geranyl-CoA carboxylase in plastid, acetyl-CoA carboxylase in cytosol, and methylcrotonyl-CoA carboxylase in mitochondria) (for a review, see Ref. 10). A cDNA encoding HCS has been isolated and characterized in *Arabidopsis thaliana* (9). The product of the gene has been biochemically studied and localized in plastids (11). The cytosolic and mitochondrial HCS activities remain unexplained. In the present study, we intended to clone a cDNA resulting from the expression of a new *hcs* gene (*hcs2*) using RT-PCR. We observed an unexpected diversity of messenger variants resulting from the transcription of this gene. They come from *hcs2* mRNA alternative splicing. We investigated the genomic organization of *hcs2*. We evidenced different allelic isoforms of the *hcs2* gene and studied the effects of this polymorphism on *hcs1* and *hcs2* expression and HCS activities.

TABLE I
Synthetic oligonucleotides used

Primer name	Primer sequence	Annealing temperature ^a
<i>hcs1</i> and <i>hcs2</i> cDNA cloning		
<i>hcs2.met</i>	5'-CGTTATTGAACAGTTTCGTGG-3'	
<i>hcs2.stop</i>	5'-GCTAAATAACATCTGCGTTCACC-3'	
<i>hcs1.met</i>	5'-ATGGAAGCAGTTTCGTTCAACAACAACC-3'	
<i>hcs1.stop</i>	5'-CAACGCACAACATCTACTTTACCAGC-3'	
RACE experiment		
<i>hcs2.E3</i>	5'-GAGACACAACTGAACCAACTGG-3'	
<i>hcs1.E4</i>	5'-CCATACATTCTTTGTTCTGC-3'	
<i>hcs2</i> gene characterization		
<i>hcs2.E7</i>	5'-GGAGCTTACTATAGGACATGGC-3'	
Real-time PCR		
SNP.5	5'-GTTGTGACTATCCAGGTAGGATG-3'	55 °C
SNP.ag.3	5'-CCCGAAGAAGTCAAACCCCT-3'	55 °C
SNP.gg.3	5'-CCCGAAGAAGTCAAACCCC-3'	55 °C
Actin.5	5'-GTTTTGCTGGGGATGATGC-3'	55 °C
Actin.3	5'-GGATTGAGCTTCATCGCC-3'	55 °C
<i>hcs2.Q.5</i>	5'-CCAGTTGGTTTCAGTTTGTGTCTCTGATATCCAGT-3'	60 °C
<i>hcs2.Q.3</i>	5'-GGCACTACTCGACCATCTCCATTTCTAAT-3'	60 °C
<i>hcs1.Q.5</i>	5'-CCAGTTGGTTTCAGTTTGTGTCACTGATATCCAGT-3'	64 °C
<i>hcs1.Q.3</i>	5'-GGCAGACTCGACCATCTCCATTTCTAGA	64 °C
E2.AA	5'-CACTCACGATGTTGTTTCTCATAA-3'	50 °C
E2.TT	5'-CACTCACGATGTTGTTTCTCATT-3'	50 °C
E4.al	5'-AAGACAACCTTTTGGAGATTCCCAA-3'	50 °C
E8.GG	5'-AAAATGTTGTGACTATCCAGGG-3'	50 °C
E8.CT	5'-AAAATGTTGTGACTATCCAGCT-3'	50 °C
E10	5'-GTTACCAAAAGATGCTTC-3'	50 °C
<i>cac1</i> cloning and expression		
5'-CAC1	5'-GGTTAGTACTGGATCCTCATCAAATGCTGC-3'	
3'-CAC1	5'-CAGAGAATGAATTTACTCGAGCTTCTTTTCTCC-3'	

^a The annealing temperature used for the real-time PCR reactions is indicated.

EXPERIMENTAL PROCEDURES

Media, Bacterial Strains, and Chemicals—D-[8,9-³H]Biotin (33 Ci/mmol) was purchased from Amersham Bioscience Inc. Isopropyl-1-β-D-galactopyranoside was obtained from Bioprobe Systems. All other biochemicals were obtained from Sigma Chimie SARL.

Temperature-sensitive *E. coli* *birA215* mutant (strain BM4050) was generously provided by Dr. A. M. Campbell. Mutations in the *birA* gene affect the biotin ligase function of the BirA protein resulting in biotin auxotrophy. The mutant had been lysogenized with the helper phage (λDE3) harboring a copy of the T7 RNA polymerase gene, using the λDE3 lysogenization kit from Novagen (11). The resulting *E. coli* *birA215* (DE3) could be used to express target cDNAs cloned in the pET vector under the control of the T7 promoter.

Plant Material—*A. thaliana* (ecotype Wassilewskija) plants were grown in soil under greenhouse conditions (23 °C with a 16-h photoperiod, and a light intensity of 200 μmol of photons m⁻² s⁻¹) until harvested for analysis.

Synthetic Oligonucleotides—The oligonucleotide primers utilized in this study are listed in Table I.

Cloning of *Arabidopsis hcs2* and *hcs1* cDNAs—Poly(A) mRNAs from *Arabidopsis* leaves were prepared using the Straight A's mRNA Isolation System, according to the manufacturer's instructions (Novagen). First strand cDNA was synthesized from 250 ng of mRNA in a final volume of 20 μl using Oligo(dT)₂₀ primers (Thermoscript RT-PCR System, Invitrogen). PCR reaction was then conducted using 1-μl aliquots of the RT reaction and *hcs2*- or *hcs1*-specific primer, in the presence of 1 unit of Pwo polymerase (Roche Molecular Biochemicals). *Hcs2*-specific oligonucleotides (*hcs2.met*, *hcs2.stop*) and *hcs1*-specific oligonucleotides (*hcs1.met*, *hcs1.stop*) were chosen according to *hcs2* gene (accession number AC007505.4, AAF19230) and *hcs1* cDNA (9) sequences. RT-PCR products were cloned into pPCRscript (Stratagene) and sequenced.

Identification of *hcs1* and *hcs2* Transcription Start Site—The transcription start site of *hcs1* and *hcs2* was determined by 5'-RACE using *Arabidopsis* leaves cDNAs (GeneRacer Kit, Invitrogen). Two sequential PCRs were performed with two pairs of nested adaptor primers and gene-specific primers in exon 3 of *hcs2* (*Hcs2.E3*) and gene-specific primer in exon 4 of *hcs1* (*Hcs1.E4*). The resulting PCR products were cloned into pPCRscript and sequenced.

Cloning and Analysis of the *Arabidopsis hcs1* and *hcs2* Genes—Full-length *hcs1* and *hcs2* genes were amplified by PCR from *Arabidopsis* leaves total DNA extracted by the CTAB method (12). *Hcs1* was amplified using *hcs1.met* and *hcs1.stop* primers and *hcs2* using *hcs2.met* and *hcs2.stop* primers. The resulting PCR products were cloned into pCR4Blunt-TOPO vector (Invitrogen) and sequenced. Partial *hcs2* gene fragment was amplified by PCR using *hcs2.exon7* and *hcs2.stop* primers and used for further restriction pattern polymorphism analysis (*Hpa*II and *Bst*NI digestions).

Sequencing and Analysis of Data—DNA sequencing was performed on both strands using the Prism Kit with fluorescent dideoxynucleotides (Applied Biosystems, Genome Express, Meylan, Grenoble, France) and T3 and Reverse universal primers. In addition, specific oligonucleotide primers were used for further sequencing. VectorNTI (InforMax) software package was used for sequence analyses.

Real-time PCR—Relative quantification experiments were done by real-time PCR using Lightcycler System (Roche Biomolecular Biochemicals) and Lightcycler-Faststart DNA Master SYBR Green I kit. For each measurement 1 μl of DNA preparation was used as a template in a standard 10-μl LightCycler PCR with appropriate primers (used at a final concentration of 1 μM) and convenient MgCl₂ concentration (3 mM for amplification of genomic DNA and 5 mM for amplification of cDNA). Amplification and detection were performed using the following profile: 95 °C/8 min followed by 40 cycles of 95 °C/10 s, × °C/5 s and 72 °C/5 s. The annealing temperature (× °C) which is primer dependent is indicated in Table I. The specificity of the reaction was verified by melting curve analysis obtained by increasing temperature from 55 to 95 °C (0.1 °C/s).

***Hcs2* Alleles Quantification**—The relative quantity of *hcs2-ag* and *hcs2-gg* alleles in a DNA solution prepared from a mixture of *Arabidopsis* plants was determined by real-time PCR using the PCR primers SNP.5 and SNP.ag.3 (*hcs2ag* allele specific) and SNP.5 and SNP.gg.3 (*hcs2gg* allele specific) (Table I). DNA preparations from plant 1 (*hcs2gg* allele) and DNA preparations from plant 7 (*hcs2ag* allele) were used as standards for quantification.

***hcs1* and *hcs2* Expression**—Poly(A) mRNAs from different organs (leaves, stems, roots, flowers, siliques, seeds) of *Arabidopsis* were prepared using Straight A's mRNA Isolation System according to the

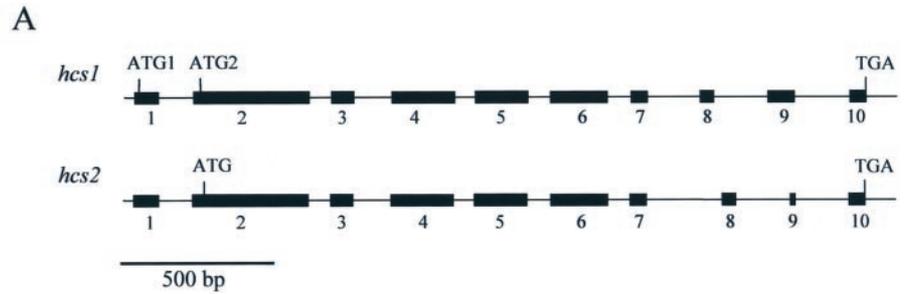
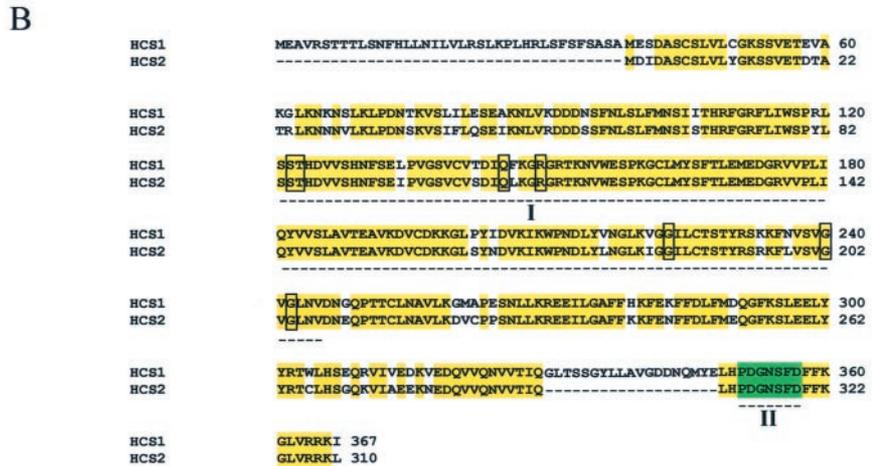


FIG. 1. The genomic structure of the *hcs1* and *hcs2* genes. A, schematic diagram of *hcs1* and *hcs2* loci. Boxes and lines are exons and introns. The translation start site (ATG) and stop codon (TGA) are indicated. B, amino acid sequence comparison between *Arabidopsis* HCS1 (accession number AAC49706) and HCS2 (conceptual translation of the putative mRNA). Identical amino acids are shaded in yellow. Region I corresponds to the catalytical core domain. Boxed amino acids are implied in biotin binding (identified in BirA, *E. coli*) (17). Region II, shaded in green, corresponds to the PDGNSF domain conserved in all eukaryotic HCS.



manufacturer's instructions (Novagen). First strand cDNA was synthesized from 1 µg of mRNA in a final volume of 20 µl using Oligo(dT)₂₀ primers (Thermoscript RT-PCR System, Invitrogen). The relative amount of *hcs1* and *hcs2* mRNAs was determined by real-time PCR using the PCR primers *hcs2*.Q.5, *hcs2*.Q.3, *hcs1*.Q.5, and *hcs1*.Q.3 (Table I). Dilutions of linear plasmid pPCRscript-*hcs1* and pPCRscript-*hcs2* were used as standards for quantification. Amplification of actin cDNA (accession number U39449) with actin.3 and actin.5 primers was used as internal standard of mRNA integrity and cDNA preparation.

Hcs2 Splicing Variants Quantification—Poly(A) mRNAs from leaves of plant line Ws-0-1 and plant line Ws-0-7 of *Arabidopsis* were prepared. First strand cDNA was synthesized and used for real-time PCR analyses as described above. Exon2/exon3 splicing variants were discriminated using E2.AA/E4 and E2.TT/E4 primers. Exon8/exon9 splicing variants were discriminated using E8.GG/E10 and E8.CT/E10 primers (Table I). Dilutions of linear pPCRscript-*hcs2a* and pPCRscript-*hcs2d* were used as standards for quantification.

Activity Measurements—Samples of *Arabidopsis* plants were harvested, ground in liquid nitrogen with a mortar and pestle, and homogenized in 1 volume of chilled buffer A composed of 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM dithiothreitol, 5 mM ε-aminocaproic acid, 1 mM benzamidine-HCl, and 1 mM phenylmethylsulfonyl fluoride. After 15 min centrifugation (18,000 × g), the supernatant was collected and used for further analysis. The protein concentration of the extract was determined using the method of Bradford (13). HCS activity in the protein extracts was measured according to the protocol described by Tissot *et al.* (14) using purified CAC1 as the apocarboxylase substrate.

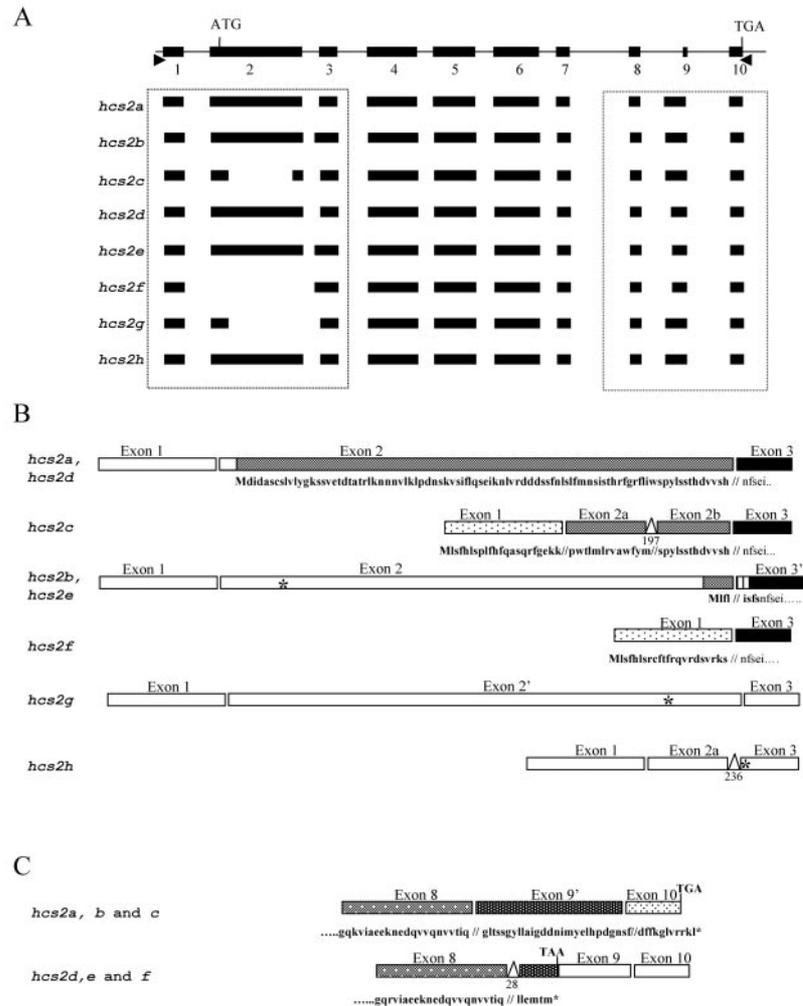
Cloning, Expression, and Purification of Recombinant Apo-CAC1 Substrate—The pET-29a+-CAC1 plasmid encoding the *Arabidopsis* biotin carboxyl carrier protein was constructed by means of PCR using the pYES-R cDNA library (15). Primers used for amplification (5'-CAC1 and 3'-CAC1) were synthesized in accordance with the published cDNA sequence (Ref. 16, accession number U23155). They introduced a *Bam*HI restriction site and a *Xho*I restriction site. After digestion with *Bam*HI and *Xho*I, the PCR fragment was subcloned into pET29a+ vector (Novagen) digested with the same enzymes. This construct encodes a fusion protein containing the CAC1 protein devoid of the chloroplast transit peptide and a 32-residue S-Tag presequence. The *E. coli*

birA215(DE3) strain transformed with pET29-CAC1 was grown and induced at 37 °C in Luria-Bertani medium supplemented with kanamycin (50 µg/ml). When A₆₀₀ reached 0.6, 1 mM isopropyl-1-β-D-galactopyranoside was added and growth was continued at 37 °C for 3–4 h. Cells were collected by centrifugation and resuspended in buffer A. They were then disrupted by sonication for 15 min at 0 °C and lysates were centrifuged at 40,000 × g for 15 min. Supernatant was incubated with S-protein-agarose in batch processing for 1 h at 25 °C with gentle shaking. Purification of the recombinant protein by cleavage of the leader sequence was achieved with a biotinylated thrombin according to the manufacturer's instructions (Novagen).

RESULTS

Isoforms of *hcs2* mRNA—The systematic sequencing of the *Arabidopsis* genome revealed a second *hcs* gene (*hcs2*, accession number AAF19230) which presents strong homologies with the gene encoding the plastidial HCS (*hcs1*, accession number AAD31371), confirming our prediction of the occurrence of distinct HCS isoforms in the plant kingdom (9, 11). The genomic organization of *hcs2* is very close to that of *hcs1* (Fig. 1A). Like *hcs1*, the *hcs2* gene is predicted to be divided in 10 exons. The putative protein resulting from the conceptual translation of the putative *hcs2* mRNA shares strong similarities with HCS1 (displaying 82% amino acid identity) and possesses the conserved domains of biotin ligases (Fig. 1B). Moreover, it does not present any predictable transit peptide. It could then catalyze the cytosolic HCS activity. To obtain the messenger resulting from the transcription of the *hcs2* gene, mRNAs were isolated from *Arabidopsis* leaves and analyzed by RT-PCR. Specific oligonucleotide primers (designed *hcs2*.met and *hcs2*.stop in the 5' and the 3' ends of the *hcs2* gene, and chosen according to ORF prediction and homologies with *hcs1* gene) were used to amplify the product of *hcs2* gene. Surprisingly, we noticed the occurrence of multiple cDNAs for *hcs2*.

FIG. 2. Cloning of cDNAs encoding *Arabidopsis* HCS2 isoforms. *A*, schematic representation of *hcs2* mRNAs and their gene organization. The positions and directions of the primers used for RT-PCR analyses are indicated on the bottom. 5' and 3' variable regions are boxed. *B*, *Hcs2* isoforms variations in the 5' terminus of the cDNA. Exons are represented by boxes. The 5'-untranslated regions are indicated in white. The amino acid corresponding sequence is indicated above. A slash indicates a junction between exons. The triangle symbol marks a deletion and the corresponding size is indicated below. A striped box represents the 10 nucleotides insertion at the beginning of exon 3 in *hcs2b* and *hcs2e* variants. An asterisk represents upstream nonsense codons. Any ORF was detected for clones *hcs2g* and *hcs2h*. *C*, *hcs2* isoforms variations in the 3' terminus of the cDNA. The 3'-untranslated regions are indicated in white. The amino acid corresponding sequence is indicated above. A slash indicates a junction between exons. Locations of stop codons (TGA and TAA) are shown.



PCR-assisted cloning enabled us to isolate eight distinct cDNAs (Fig. 2A). They exhibited sequence deletions or additions compared with the predicted *hcs2* mRNA and were distinct from each other in 5' and 3' ends. The central part of the cDNA sequence was invariant. Six cDNAs were found to encode a family of putative HCS2 proteins designated HCS2a to HCS2f. Two of these cDNAs (*hcs2g* and *hcs2h*) did not present any ORF. Sequence alignment of the putative HCS2 proteins shows that they all share the central sequence (from exon 3 to exon 8) corresponding to the catalytic core, whereas their COOH-terminal and NH₂-terminal are divergent. Fig. 2B shows the complex structure of the NH₂-terminal from the HCS2 protein family. HCS2a and HCS2d protein sequences corresponded to the predicted one. But, HCS2b and HCS2e differed from this sequence. Indeed, in *hcs2b* and *hcs2e* cDNA, we observed a 10-nucleotide insertion at the beginning of exon 3 from the sequence of intron 2 (according to the predicted structure of *hcs2* gene). This created a frameshift and resulted in the production of a shorter ORF, the first initiating Met being localized at the end of exon 2. The putative HCS2f protein presented a large deletion in the NH₂-terminal. Indeed, the entire exon 2 sequence was absent from the *hcs2f* cDNA sequence. The first Met to initiate the *hcs2f* mRNA was localized in exon 1. Finally, HCS2c presented an original sequence in the NH₂-terminal. Indeed, 197 nucleotides were deleted from exon 2 (position 39–235 of the exon 2). The first Met to initiate the *hcs2c* mRNA was at the beginning of exon 1 and the first 35 amino acids (exon 1-exon 2a) were completely different from those of other HCS2 putative protein sequences. None of the observed HCS2

COOH-terminal corresponded to the predicted one. Indeed, the *hcs2* cDNA sequence surrounding exon 9 was different from the expected one. Two HCS2 putative protein families were identified (Fig. 2C). The first one comprised HCS2a, HCS2b, and HCS2c and the second, HCS2d, HCS2e, and HCS2f. Indeed, 28 nucleotides at the beginning of exon 9 of *hcs2a*, *hcs2b*, and *hcs2c* cDNAs were missing in *hcs2d*, *hcs2e*, and *hcs2f* cDNAs. This resulted in a frameshift and created a premature termination signal in exon 9. Thus, the resulting putative proteins were shorter and they did not present the PDGNSF domain conserved among eukaryotic biotin ligases.

Besides these large variations which modify the 3' and the 5' ends of the *hcs2* cDNAs and resulting ORFs, we observed a number of SNP. *Hcs2d* and *hcs2e* sequences were in agreement with *hcs2* gene sequence referenced in the data bank whereas *hcs2a*, *hcs2b*, *hcs2c*, and *hcs2f* presented SNP compared with these sequences. Indeed, within the sequenced cDNA region, 12 polymorphic sites were observed (listed in Table II). Of these, six resulted in silent substitutions, two involved conservative amino acid change, three involved nonconservative amino acid changes, and one resulted in a change from a Trp (TGG, position 235, exon 2) to a termination triplet (TAG). This last modification, observed in *hcs2g* cDNA, precluded the determination of an ORF for this sequence.

The diversity of *hcs2* mRNA splice variants contrasts with the occurrence of only one *hcs1* mRNA isoform previously isolated by functional complementation of *E. coli* mutants (9). To investigate the possible occurrence of other mRNA isoforms of *hcs1*, we intended to clone its corresponding cDNA by means of

TABLE II
Nucleotide polymorphism at *hcs2* cDNA

Nucleotides are numbered beginning at the start of each indicated exon. cDNAs possessing the same nucleotide as putative cDNA sequence are marked by dots at that site. Nucleotide changes are indicated by the appropriate letter and nucleotide deletion by a "d." For partial cDNA sequences, absence of data is indicated by a "n.s." (not sequenced). The putative cDNA sequence used as reference is reported according to the systematic sequencing of *Arabidopsis* genome (18).

cDNA	Exon 1		Exon 2		Exon 4		Exon 5		Exon 6	Exon 7	Exon 8	Exon 9	
Putative	21	39	8	97	235	3	119	14	114	118	39	9	46
<i>hcs2 a</i>	G	G	T	A	G	A	T	A	C	G	G	G	G
<i>hcs2 b</i>	.	A	C	.	A	C	C	T	T	.	C	A	A
<i>hcs2 c</i>	.	.	C	d	d	C	C	T	T	T	.	A	A
<i>hcs2 d</i>
<i>hcs2 e</i>
<i>hcs2 f</i>	d	.	d	d	d
<i>hcs2 g</i>	.	.	C	.	A _s ^a	C	C	n.s	n.s	n.s	n.s	n.s	n.s
<i>hcs2 h</i>	.	.	C	d	d	C	C	T	T	n.s	n.s	n.s	n.s

^a A_s indicates the occurrence of a termination signal.

PCR. mRNAs were isolated from *Arabidopsis* leaves and analyzed by RT-PCR, as described for *hcs2*. Specific oligonucleotide primers (designed *hcs1.met* and *hcs1.stop* in the 5' and the 3' ends of the *hcs1* gene) were used to amplify the product of the expression of *hcs1* gene. Interestingly, under these conditions we did not evidence any other *hcs1* mRNA form than that previously characterized.

Identification of *hcs2* Transcription Start Site—To accurately determine the translation initiation of *hcs2*, we intended to clone the entire 5' sequence of *hcs2* mRNA from the transcription start. By means of a 5'-RACE experiment we cloned *hcs2* 5' end (Fig. 3B). This analysis evidenced two additional exons before the first ATG but did not reveal any other upstream ATG. To complete *hcs1* and *hcs2* mRNA studies, we also determined the *hcs1* transcription start by 5'-RACE (Fig. 3A). Surprisingly, we observed an *hcs1* 5' end different from the one previously characterized (entire *hcs1* cDNA isolated by means of functional complementation (9)). A 101-nucleotide sequence from the 5'-untranslated region was absent in the clone obtained by RACE. The borders of this deleted sequence matched with the GT/AG rules of splicing in the *hcs1* gene. As a consequence, two different 5'-untranslated regions of *hcs1* are generated by alternative splicing of the same *hcs1* pre-messenger RNA.

Are Sequence Variations of *hcs2* mRNA Generated by Alternative Splicing?—We determined the 5' donor and 3' acceptor splice sites of each of the *hcs2* cDNA clones identified (listed in Table III). The sequence of the introns was obtained from the *hcs2* gene sequence (accession number AAF19230). The nucleotide sequence at the 3' acceptor splice sites of intron 2 (position 559) and intron 2' (position 549) demonstrated that both sites are consistent with canonical GT/AG rules. This strongly suggests that *hcs2a* and *hcs2b* mRNAs are generated by alternative splicing events. We could put forward a similar alternative splicing mechanism to explain the occurrence of *hcs2f* mRNA. The splicing of a cryptic intron (intron 2bis) inside exon 2 could explain the deletion of 197 nucleotides of the exon 2. But, the 3' acceptor splice site (position 437) of the intron 2bis do not match the AG/GT rule. Moreover, the occurrence of both 3' ends of *hcs* cDNA clones could be explained by an alternative use of two 3' acceptor splice sites in intron 8 (positions 2055 and 2083 in the *hcs2* gene). However, the first one (tccgg²⁰⁵⁷) does not match the consensus sequence (Table III). The 3' splice site (GT) and 5' splice site dinucleotides (AG) are almost invariable in higher plant introns (19). As a result, we could not explain all the nucleotide variations observed, by alternative mRNA splicing mechanisms.

Identification of Allelic Variants of *hcs2* Gene—To understand the origin of the unusual spliced mRNAs, we decided to

study in more detail the sequence of the *hcs2* gene. Specific oligonucleotide primers (designed *hcs2.E7* and *hcs2.stop*, Table I) were used to amplify from genomic DNA, a 844-nucleotide fragment of the *hcs2* gene, including the nonconsensual intron 8/exon 9 border. After cloning and sequencing, we identified two different sequences, designated *hcs2gg* (type I 844-nucleotide fragment) and *hcs2ag* (type II 844-nucleotide fragment) (Fig. 4A). These two sequences provided us an explanation for the occurrence of two different 3' ends of *hcs2* cDNA. Indeed, a SNP in intron 8 of *hcs2ag* created a new 3' acceptor splice site (TCCGG sequence in *hcs2gg* gene became TCCAG in *hcs2ag* gene) 28 nucleotides upstream from the former. This modification accounts for the unusual splicing event observed in *hcs2a*, *hcs2b*, and *hcs2c* cDNAs. Consequently, *hcs2a*, *hcs2b*, and *hcs2c* mRNAs should result from the transcription of a second *hcs2* gene (*hcs2ag*) presenting some SNP compared with the first form (*hcs2gg*). To identify the relationship between *hcs2ag* and *hcs2gg*, we analyzed the distribution of these two *hcs2* forms in an *Arabidopsis* population. We used the endonuclease restriction profile polymorphism created by the GG/AG SNP (Fig. 4B). We analyzed 11 different *Arabidopsis* plants. Among these 11 plants, four presented the type I digestion pattern and seven presented the type II digestion pattern. None of them presented a mixed digestion pattern. As a consequence, all the analyzed plants presented only one of the two *hcs2* gene isoforms evidenced. Systematic sequencing confirmed that there was only one *hcs2* gene copy in *Arabidopsis* genome. We could then conclude that the two different isoforms of *hcs2* gene are allelic *hcs2* variants. Moreover, 11 inbred lines (Ws-0-1 to Ws-0-11) were generated from the Ws-0 accession for further analysis by planting self-seeds from the original Ws-0. We verified that each of the resulting lines preserved the digestion pattern of the parental plant. Besides, we measured the relative distribution of these alleles by real-time quantitative PCR. In a preparation of genomic DNA from a population of *Arabidopsis* plants, we quantified 45% of *hcs2ag* allele (data not shown). This ratio confirmed the nearly equal distribution of *hcs2ag* and *hcs2gg* in the *Arabidopsis* seed stock analyzed.

To evidence the effect of *hcs2* gene polymorphism on *hcs2* mRNA splicing, we studied the relative distribution of exon 9 splicing variants resulting from the use of two 3' splicing acceptors (splicing type I and II, Fig. 5A). Using the real-time RT-PCR approach, we quantified each of them (Fig. 5B). In a preparation of mRNAs from whole *Arabidopsis* plants ground from our seed stock, we detected both splicing types but type II was slightly in the majority representing ~70% of *hcs2* mRNA. We analyzed the relative amount of each of these transcripts in different organs of *Arabidopsis* plants, with similar results (data not shown). In plant Ws-0-1 (*hcs2gg* allele), type II

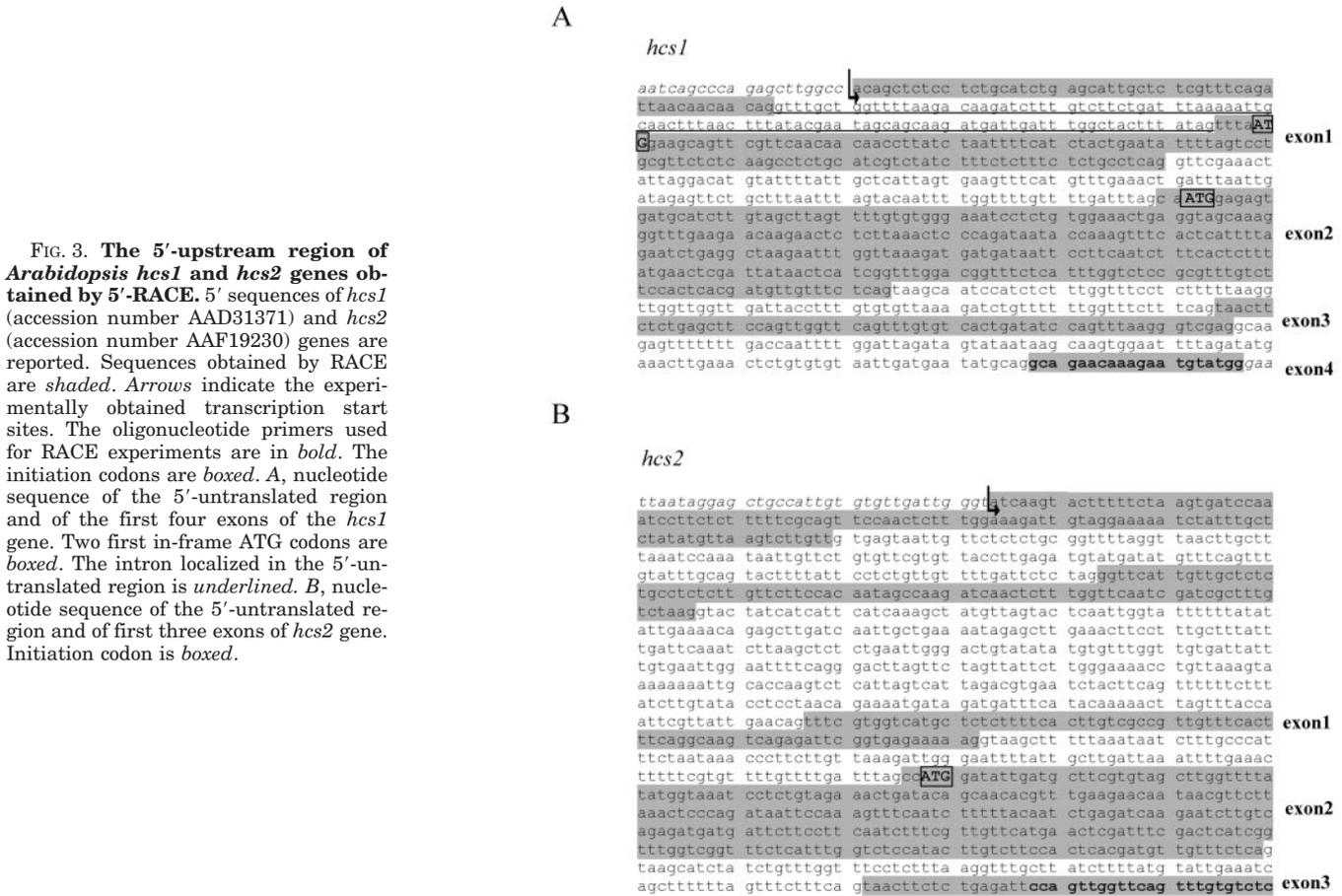


FIG. 3. The 5'-upstream region of *Arabidopsis hcs1* and *hcs2* genes obtained by 5'-RACE. 5' sequences of *hcs1* (accession number AAD31371) and *hcs2* (accession number AAF19230) genes are reported. Sequences obtained by RACE are shaded. Arrows indicate the experimentally obtained transcription start sites. The oligonucleotide primers used for RACE experiments are in bold. The initiation codons are boxed. A, nucleotide sequence of the 5'-untranslated region and of the first four exons of the *hcs1* gene. Two first in-frame ATG codons are boxed. The intron localized in the 5'-untranslated region is underlined. B, nucleotide sequence of the 5'-untranslated region and of first three exons of *hcs2* gene. Initiation codon is boxed.

TABLE III
Exon-intron junction sequence of the Arabidopsis hcs2 gene

The nucleotide sequences flanking the 5' donor and 3' acceptor splice sites of *hcs2* are shown. These arrangements were deduced by comparing the cDNA sequences with the sequence of *hcs2* gene referenced in the data bank. Exon sequences are indicated by capital letters and intron sequences by lower-case letters. Consensus introns bordure sequences are underlined. Sequences marked by an asterisk indicate sites at which consensus is not respected. Nucleotides flanking the introns are numbered.

Exon	5' Donor splice site	Intron	3' Acceptor splice site	Exon	Occurrence in cDNA
1	AG ⁹⁰ /gtaagc	1	tttag ²⁰³ /CC	2	All
2	CA ⁴⁷⁷ /gtaagc	2'	tttag ²⁴⁹ /TT	3'	<i>hcs2b</i> , <i>2e</i>
2	CA ⁴⁷⁷ /gtaagc	2	ttcag ⁵⁵⁹ /TA	3	<i>hcs2a</i> , <i>2c</i> , <i>2d</i> , <i>2f</i> , <i>2g</i> , <i>2h</i>
2a	TG ²⁴⁰ /gtaaat	2bis	*tttgg ⁴³⁷ /TC	2b	<i>hcs2c</i> , <i>2h</i>
3	AG ⁶²¹ /gtaagt	3	tgtag ⁷¹² /GC	4	All
4	AG ⁸⁶¹ /gtagc	4	ttcag ⁹³⁸ /GG	5	All
5	TG ¹⁰⁵⁹ /gtaata	5	tgtag ¹¹⁵² /GT	6	All
6	AG ¹³⁰⁸ /gtaatg	6	ttcag ¹⁴⁶³ /GT	7	All
7	AG ¹⁵⁰⁹ /gtaatg	7	cgcag ¹⁹⁰⁷ /TG	8	All
8	AG ¹⁹⁷⁴ /gtagga	8	gctag ²⁰⁸³ /CT	9	<i>hcs2d</i> , <i>2e</i> , <i>2f</i>
8	AG ¹⁹⁷⁴ /gtagga	8'	*tccgg ²⁰⁵⁵ /GG	9'	<i>hcs2a</i> , <i>2b</i> , <i>2c</i> , <i>2g</i> , <i>2h</i>
9	AG ²¹³⁴ /gtgata	9	tacag ²¹⁹² /TT	10	All

amount reached 90% of total *hcs2* mRNA, whereas splicing type I level did not exceed the background noise of PCR. Thus, in "*hcs2gg*" plants, there was only one detectable splicing event that created a premature termination signal. On the contrary, in Ws-0-6 plants (*hcs2ag* allele), splicing type I was significantly detected. Its relative amount exceeded the level of splicing type II and raised 60% of total *hcs2* mRNA. As a result, in "*hcs2ag*" plants, the intron 8 SNP (nucleotide 2055) created a new splice site which entitled the detection of splicing type I of exon 9 in *hcs2* mRNA. The second splice site (the only one used in *hcs2gg* plants) kept on being used and explained the detection of splicing type II of exon 9 in *hcs2* mRNA. Consequently, in these plants and contrarily to *hcs2gg* plants, alternative splicing of exon 9 created two different mRNA 3' ends and hence, two different putative NH₂-terminal HCS2 proteins. In

summary, a nucleotide polymorphism at the *hcs2* gene created an allele-dependent polymorphism of exon 9. To quantify this polymorphism, we sequenced the entire *hcs2* gene.

Polymorphism at *hcs2* Gene—A total of 2322 nucleotides of the *hcs2* gene were sequenced for five inbred lines of *A. thaliana* Ws ecotype (Ws-0-1, Ws-0-4, Ws-0-5, Ws-0-6, and Ws-0-10). This encompassed the 2251 nucleotides containing the HCS2 protein coding region plus nine introns and a portion of the flanking DNA including 23 nucleotides at the 5' end and 48 nucleotides at the 3' end of the ORF. We confirmed the results obtained by restriction patterns and obtained three different *hcs2* sequences (Ws-0-4/Ws-0-6/Ws-0-10, Ws-0-1, and Ws-0-5). Furthermore, within the entire sequenced region, 21 polymorphic sites were observed. The position of all nucleotide changes found among the Ws-0-4 and Ws-0-5 *hcs2* genes (*hcs2ag* allele)

FIG. 4. Analysis of the 3' terminal of the *hcs2* gene. A, schematic representation of the two 3' sequences of the *hcs2* gene and their corresponding spliced mRNA. A 844-nucleotide PCR fragment (from the beginning of intron 7 to the termination signal) was cloned and sequenced. Two different sequences were evidenced: type I sequence (*hcs2gg* allele) and type II sequence (*hcs2ag* allele). The cDNA clones matching the putative mRNA sequence are indicated at the beginning of the sequence. Lines and boxes in the *hcs2* gene represent exons and introns, respectively. Arrowheads indicate sites of polymorphism. B, schematic representation of type I and type II fragment digestion profiles. Genomic DNA was extracted from 11 individual *Arabidopsis* plants. For each of these plants, a 844-nucleotide PCR fragment was tested for *Bst*NI and *Hpa*II digestion. The size of the theoretical digested fragment is indicated. The digestion patterns corresponding to the Ws-0-3 plant and Ws-0-5 plant are presented. For each of them, other plants presenting the same profile are indicated.

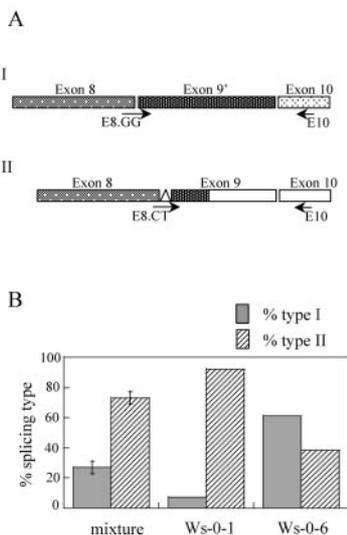
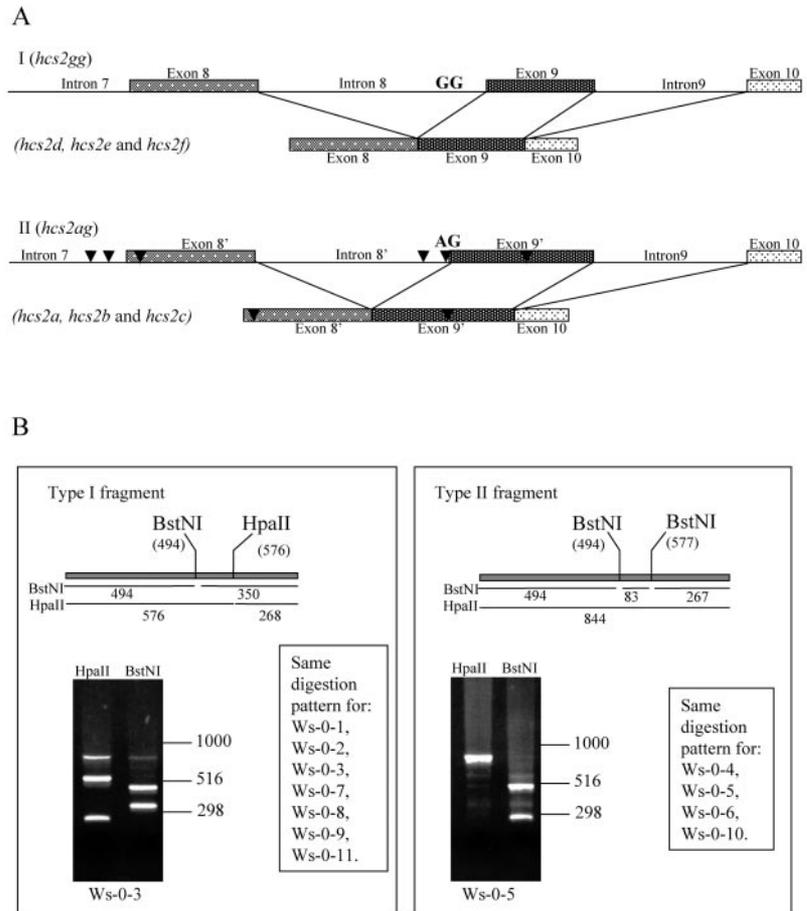


FIG. 5. Real-time RT-PCR quantification of spliced variants of exon 9. A, a schematic representation of E8.GG, E8.CT, and E10 oligonucleotides positions on exon 8/exon 9 junction and on exon 10. E8.GG oligonucleotide enabled the quantification of type I splicing event. E8.CT primer oligonucleotide enabled the quantification of type II splicing event. B, real-time RT-PCR experiment on poly(A)⁺ RNA from *Arabidopsis* aerial parts from a mixture of plants, from Ws-0-1 and Ws-0-6 plants using splicing type-specific primers. Data (percentage of the total amplification) are means of three independent experiments performed with four cDNA dilutions ± S.D. β-Actin has been used as internal standard of RNA integrity and quantity.

and 10 were observed in coding sequence. Of these, five resulted in silent substitutions, three involved conservative amino acids changes, one involved nonconservative amino acid change (substitution of Ile with Asn), and one resulted in a change from a Trp to a stop codon (position 437 (exon 2)). Two SNP resulted in the creation of splice sites (position 437 (exon 2) and 2055 (intron 8)). These observations elucidated the unexplained splicing of exon 2 (in *hcs2c*) and unexplained splicing of exon 9 (*hcs2a*, *hcs2b*, and *hcs2c*). The Ws-0-1 *hcs2* gene did not present any polymorphism compared with the Col-0 *hcs2* gene sequence. The polymorphisms observed in the entire *hcs2* gene sequence are consistent with the previously observed polymorphisms in *hcs2* cDNA sequences. Indeed, the *hcs2* gene from Ws-0-4 type plants presented the same exonic SNP that those previously observed in *hcs2a*, *hcs2b*, and *hcs2c* cDNA clones (see Table II).

To determine whether *hcs2* gene polymorphism only reflected non-selective genomic variations inside the Ws ecotype or if it was *hcs2* gene-specific, we investigated *hcs1* gene polymorphism. A total of 2385 nucleotides of the *hcs1* gene were sequenced in Ws-0-1 and Ws-0-6 plants. Ws-0-1 *hcs1* gene sequence was identical to the Col-0 *hcs1* gene sequence. Within the entire sequenced region, only two polymorphic sites were observed in Ws-0-6 *hcs2* gene. Only one was observed in coding sequence (position 83 in exon 1), and resulted in a conservative amino acid change (substitution of His with Arg). As a result, 0.084% of nucleotides are polymorphic within *hcs1* and 0.04% of the observed polymorphisms result in a change in amino acid composition. By way of comparison, 0.9% of nucleotides are polymorphic within *hcs2* and 0.17% of the observed polymorphisms result in a change in amino acid composition.

We measured HCS activity in Ws-0-1 to Ws-0-11 plant lines. Although *hcs2* gene sequence polymorphism have severe con-

is listed in Table IV. Ws-0-6 and Ws-0-10 *hcs2* genes presented the same 20 polymorphic sites than those reported in Ws-0-4. Of the 21 SNP observed, 11 were present in intron sequence

TABLE IV
Nucleotide polymorphism at *hcs2* gene

Nucleotides are numbered beginning at the start of the ORF. The position of the nucleotides is indicated, I1: intron 1, E2: exon 2. When the site of polymorphism is exonic, its position in the exon is specified above. Nucleotide changes are indicated by the appropriate letter. The Col-0 *hcs2* gene sequence is used as reference and reported according to the systematic *Arabidopsis* genome sequencing (18).

Ecotype	104 I1	158 I1	210 E2 8	437 ^a E2 235	665 I3	714 E4 3	830 E4 119	951 E5 14 ^b	1051 E5 114	1145 I5	1270 E6 118 ^b	1336 I6	1430 I6	1452 I6	1480 E7	1788 I7	1881 I7	1915 E8 9 ^b	2050 I8	2055 ^a I8	2102 E9 46 ^b
Col-0	T	G	T	G		A	T	A	C	C	G	T	C	G	G	A	A	G		G	G
Ws-0-4	A	T	C	A	C	C	C	T	T	A	T	A	T	A	G	G	C	A	G	A	A
Ws-0-5	A	T	C	A	C	C	C	T	T	A	T	A	T	A	A	G	C	A	G	A	A

^a Nucleotides marked indicate sites at which splice site changes occur.

^b Nucleotides marked indicate sites at which amino acid changes occur in the deduced protein sequence. These changes occur at position 14 of exon 5 (substitution of Asn with Ile), position 118 of exon 6 (substitution of Lys with Asn), position 39 of exon 7 (substitution of Cys with Trp), position 9 of exon 8 (substitution of Arg with Lys), and position 46 of exon 9 (substitution of Val with Ile).

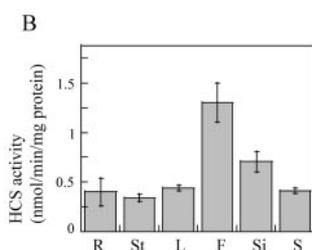
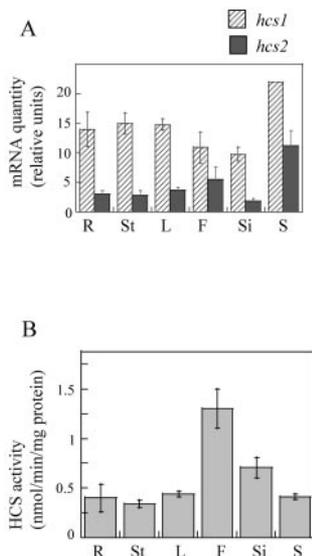


FIG. 6. **Expression of *hcs1* and *hcs2* genes.** Real-time RT-PCR experiment on poly(A)⁺ RNA from various *Arabidopsis* organs (roots (R), stems (St), leaves (L), flowers (F), siliques (Si), and seeds (S)) using *hcs1*- and *hcs2*-specific primers (A). Data are means of three independent experiments performed with four cDNA dilutions \pm S.D. β -Actin has been used as internal standard of RNA integrity and cDNA preparation. HCS activity measurement has been performed on protein extracted from the same organs (B).

sequences on *hcs2* mRNA splicing and COOH-terminal HCS2 protein sequence, the *in vitro* assay performed with CAC1 as acceptor substrate did not evidence any significant variation of total HCS activity correlated with the nature of *hcs2* allele (data not shown). Moreover, plants had no discernible phenotype, whatever *hcs2* allele they bore.

Expression of *hcs* Genes—To have insight into the physiological function of each *hcs* gene, we intended to study their expression patterns in various plant organs. By means of quantitative RT-PCR measurements, we evaluated *hcs1* and *hcs2* transcripts abundance (Fig. 6A). *Hcs1* and *hcs2* mRNAs were detected in all analyzed organs (roots, stems, leaves, siliques, flowers, and seeds). *Hcs1* mRNA amount was relatively constant except in the seeds where it was twice that in other organs. Total *hcs2* mRNA level was also constant in vegetative organs, but was increased in flowers and seeds. The relative amount of *hcs1* mRNA was greater (about 4-fold higher) than that of total *hcs2* mRNA. The *hcs1/hcs2* mRNA ratio was slightly different in flowers and seeds where there was only twice as many *hcs1* mRNA as *hcs2*. Interestingly, the higher level of *hcs2* mRNA in flowers was correlated with an increase of HCS activity in these organs (about 3-fold, Fig. 6B).

To determine more accurately the *hcs2* expression profile, we studied the distribution of the two alternative splice variants of intron 2/exon 3 border (Fig. 7). Two different 3' acceptor splice

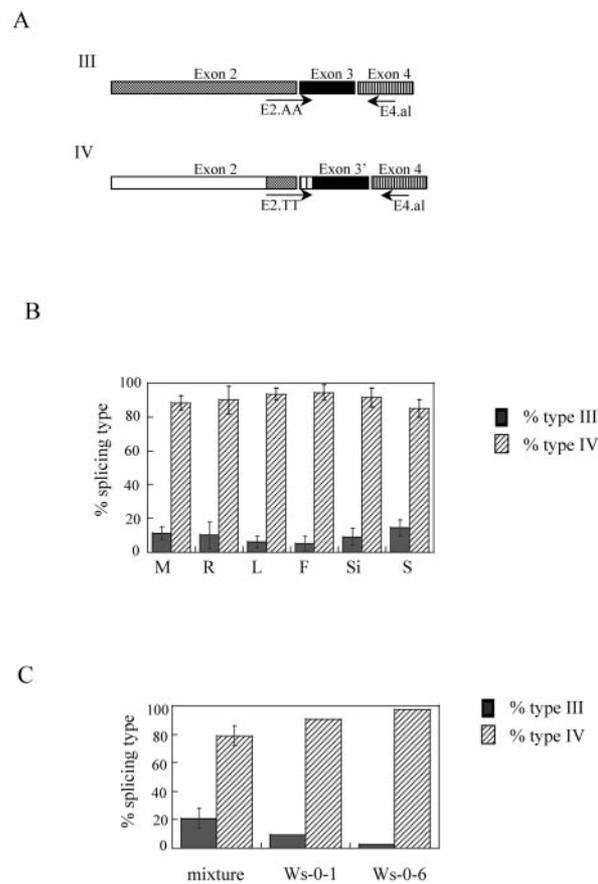


FIG. 7. **Real-time PCR quantification of spliced variants of exon 3.** A, a schematic representation of E2.AA, E2.TT, and E4.al oligonucleotide positions on exon 2/exon 3 junction and on exon 4. E2.AA oligonucleotide enabled the quantification of type III splicing event. E2.TT oligonucleotide enabled the quantification of type IV splicing event. B, real-time RT-PCR experiment on poly(A)⁺ RNA from various *Arabidopsis* organs (aerial parts (M), roots (R), stem (St), leaves (L), flowers (F), siliques (Si), and seeds (S)) using splicing type specific primers. Data (percentage of the total amplification) are means of three independent experiments performed with four cDNA dilutions \pm S.D. β -Actin has been used as an internal standard of RNA integrity. C, real-time RT-PCR experiment on poly(A)⁺ RNA from *Arabidopsis* aerial parts from a mixture of plants, from Ws-0-1 and from Ws-0-6 plants using splicing type-specific primers. Data (percentage of the total amplification) are means of three independent experiments performed with four cDNA dilutions \pm S.D. β -Actin has been used as internal standard of RNA integrity.

sites distant from 10 nucleotides are alternatively used. We quantified each of the resultant messengers by way of real-time PCR. Surprisingly, type III was the most abundant splicing type, resulting in the creation of a 10-nucleotide longer exon 3 (exon 3') and in a frameshift (see Fig. 2). The putative proteins associated with this splicing are NH₂-terminal truncated

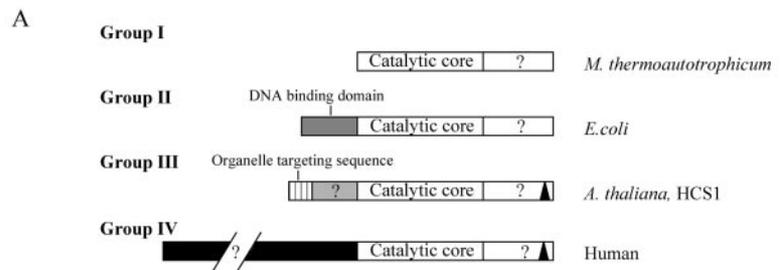


FIG. 8. Comparison of *Arabidopsis* HCS2 putative proteins and HCSs from various organisms. A, classification of HCSs (adapted from Ref. 20). HCSs shown (organism, accession number) are as follows: *Methanobacterium thermoautotrophicum* strain Δ H, AAB86373; *E. coli*, M101123; *A. thaliana*, AAC49706; and *Homo sapiens*, P50747. ?, the function of this region is unknown. The triangle represents the PDGNSF domain. B, relative classification of HCS2 proteins. The associated allele is indicated. The putative subcellular localizations are predictions from chloroP and targetP prediction servers.

B

	Putative HCS2 protein	Related group	Putative localization
<i>Hcs2ag</i> alleles	HCS2b	?	Vesicular pathway
	HCS2c	III	Cytosol
	HCS2a	III	Cytosol
<i>Hcs2gg</i> alleles	HCS2e	I	Vesicular pathway
	HCS2d	?	Cytosol
	HCS2f	?	Mitochondria

(HCS2b and HCS2e). This splicing distribution did not vary with the tissue experienced (Fig. 7B) and was not dependent on the *hcs2* gene allele (Fig. 7C).

DISCUSSION

When making an inventory of the parties involved in biotinylation, we observed a peculiarly complex situation in plants. HCS activity had been detected in three subcellular compartments (9) and the systematic sequencing of *Arabidopsis* genome revealed the occurrence of two *hcs* genes (*hcs1* and *hcs2*) (18). *Hcs1* was previously characterized and found to encode plastidial isoform (9, 11). Here, we focused our attention on the *hcs2* gene and its products. The salient facts of our investigation are (i) *hcs2* gene is expressed and subjected to alternative splicing; (ii) *hcs2* gene is characterized by high level of nucleotide polymorphism inside Ws ecotype; (iii) the level of polymorphism at *hcs2* locus contrasts sharply with that observed at *hcs1* locus; (iv) the observed polymorphism affects splicing of *hcs2* mRNA but does not have any effect on HCS activity level in whole plant extracts nor on plant phenotype; (v) total *hcs2* mRNA level is 4-fold lesser than *hcs1* mRNA; and (vi) the most abundant *hcs2* mRNA spliced variant should encode a truncated protein.

Are HCS2 Putative Proteins New Forms of Eukaryotic HCS?—We evidenced different isoforms of *hcs2* mRNA which could encode putative proteins variable in amino- and carboxyl-terminal. We were not able to test the occurrence of these proteins in a cellular extract of *Arabidopsis* although we had at our disposal the HCS1 antibody. This antibody cross-reacted with HCS2 recombinant proteins (data not shown) but the HCS proteins concentration in plant extracts was too low to be accurately detected. These putative proteins can be classified according to their domains organization (20) (Fig. 8A). What-

ever *hcs2* allele, the most represented putative HCS2 protein (according to its corresponding mRNA relative abundance (*hcs2b* for *hcs2ag* and *hcs2e* for *hcs2gg*)) should be NH₂-terminal truncated. It is an original eukaryotic HCS organization and this NH₂-terminal structure is related to some bacterial ligases which do not present the helix-turn-helix domain of BirA (involved in DNA fixation and biotin operon regulation) and present no more than the catalytic core domain in NH₂-terminal. We evidenced two different putative COOH-terminal HCS2 protein sequences. One presents the PDGNSF eukaryotic specific domain (observed in HCS2a, HCS2b and HCS2c) while the other is devoid of this conserved sequence. This partition is allele-dependent, as far as *hcs2gg* allele plants could not produce the COOH-terminal complete HCS2 protein. The function of this domain is not determined nor its exact involvement in HCS activity, but human HCS deleted of its last 31 amino acids (comprising the PDGNSF sequence) lost all activity (21). However, for the *Arabidopsis* enzyme, which is closer to its bacterial counterpart, the HCS2 proteins depleted from this sequence (HCS2d, -e, and -f) could remain active.

The significance of several splice variants of *hcs2* is unknown. As this alternative splicing affects the 5' of *hcs2* mRNA and should modify the NH₂ terminus of HCS2 protein, it could produce proteins with different subcellular distributions (Fig. 8). Indeed, prediction programs such as Psort, chloroP, and mitoP give putative HCS2a, HCS2c, and HCS2d a cytosolic localization and HCS2f a mitochondrial localization. These supposed localizations are in good agreement with HCS activities found in plant cell cytosol and mitochondria (9) and matches well with sites of accumulation of some biotin-containing proteins in plant cells (10). According to the prediction programs, HCS2b and HCS2e should present a vesicular path-

way localization (Fig. 8). It is interesting to notice that in *Arabidopsis*, two "cytosolic" ACCase isoforms have been described (22). The corresponding gene sequences are very close, but one of them (accession number AAG40564) should present a NH₂-terminal extension which is predicted to be membranous. Moreover, prediction programs (targetP, chloroP) give this protein a reticular localization. Thus, the HCS2 NH₂-terminal truncated isoform could be its specific ligase.

Alternatively, the active domain of the *hcs2* gene could come down to the sequence between exon 3 and exon 9. So, exon 2 and exon 10, subjected to deleterious alternative splicing could be useless in HCS2 protein. This would give the HCS2 protein peculiar properties (substrate specificity and kinetic properties) different from HCS1. It could influence substrate specificity of each of them, in particular with respect to biotinylation sites different from the consensus one. Indeed, a cytosolic seed-specific biotinylated protein (SBP65) with no carboxylase activity had been evidenced in pea (23). This protein presents an atypical biotinylation site. A SBP65 homologue has been found in *Arabidopsis* seeds (24). It is also possible that other original biotinylated proteins exist in *Arabidopsis* that could explain the diversity of HCS isoforms in plant cells.

Hcs2 Gene Presents a Nucleotide Polymorphism—When comparing Col-0 and Ws-0-4 *hcs2* sequence, 0.9% of nucleotides are different and 0.17% of the observed polymorphisms result in a change in amino acid composition. It is a significant level of intraspecific sequence polymorphism. By way of comparison, such level of polymorphism was found within the *RPS2* gene (1.26%) (25) or at the RPP5 disease resistance locus (26). These are examples of sequence polymorphism affecting plant Resistance (R) genes which are currently subjected to such genetic variation (27). Bergelson *et al.* (28) examined levels of polymorphism for three nuclear genes within and among the population of *Arabidopsis*. For these nuclear genes, the average nucleotide diversity was 0.04% within the *Arabidopsis* population and 0.14% between populations. The polymorphism observed in the *hcs2* gene could have deleterious effects on HCS protein integrity because of its effect on *hcs2* mRNA splicing. Despite these severe consequences on *hcs2* mRNA and HCS2 protein sequences, the ratio of each *hcs2* alleles was found to be around 50% in the Ws population we had at our disposal, and all the plants analyzed were homozygous for this trait. Moreover, there were no apparent consequences of this allelic diversity on HCS activity and plant phenotypes, which probably indicate relatively weak functional constraint on the *hcs2* gene. A possibility is that protein HCS2 function is redundant with that of HCS1. These two genes have very close sequence homology. Thus, one cannot exclude that *hcs2* could be an inactive pseudogene in *Arabidopsis*. It would be the relic of a former duplication event of the *hcs* gene consistent with the general duplication of *Arabidopsis* genome (18) leaving, after evolution, only one resulting copy (*hcs1*) active. This possibility would be consistent with the observation that some of the spliced HCS2 isoforms are allele-dependent (HCS2a, -c, -d, -f; Fig. 8B). We have isolated by RACE experiment, an *hcs1* mRNA 5' end different from the formerly studied. The observed deletion could influence the ATG choice. Indeed, two in-frame ATG can be used as initiation codons (9) in the *hcs1* mRNA (see also Fig. 2A). In the present case, the observed deletion covered a zone of repeated sequence that can create some secondary structure. The structure of the 5'-untranslated region can influence the initiation of translation (29). The first ATG gives HCS1 a plastidial localization (11), the second could give HCS1 a cytosolic localization. As a result, *hcs1* gene could be sufficient to explain cytosolic and plastidial HCS activities. This situation had been observed in human, where only one *hcs* gene and two

subcellular localizations of HCS activity (mitochondrial and cytosolic) were evidenced (6).

Another possibility is that HCS2 is useful on occasion, but such occasions may not occur during the lifetime of the average plant. For example, *Arabidopsis adh* is a nonessential gene that is thought to be subjected to balancing selection for two haplotypes (30). Likewise, Henikoff and Comai (31) observed an allelic diversity of *CMT1*, some of the alleles presenting deleterious mutations. They put forward such an hypothesis to explain the maintenance of both alleles. In our case, *hcs2* could encode an isoform implied, for example, in specific activation of cytosolic ACCase, committed in flavonoids or long chain fatty acid synthesis, restricted in the epidermal cells. As a result, HCS2 deficiency could have any effect apart from the occurrence of an environmental stress such as UV exposition or pathogen attack.

Studies of the biochemical and kinetic properties of HCS2 recombinant proteins and of their subcellular localization in the plant cell, which are currently in progress in our laboratory, will entitle us to have insight into their physiological function. Moreover, reverse genetic studies and particularly characterization of knockout *hcs1* and *hcs2* mutants, which have been initiated, will give us some elements to unravel the role of each of *Arabidopsis hcs* gene in the realization of protein biotinylation.

Acknowledgments—We thank Dr. Renaud Dumas, Dr. Michel Matringe, and Dr. Dominique Job for useful discussions and suggestions.

REFERENCES

- Knowles, J. R. (1989) *Annu. Rev. Biochem.* **58**, 195–221
- Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C., and Wood, H. G. (1988) *J. Biol. Chem.* **263**, 6461–6464
- Perham, R. N. (2000) *Annu. Rev. Biochem.* **69**, 961–1004
- Chang, H. L., and Cohen, N. D. (1983) *Arch. Biochem. Biophys.* **225**, 237–247
- Suzuki, Y., Aoki, Y., Ishida, Y., Chiba, Y., Iwamatsu, A., Kishino, Y., Niikawa, N., Matsubara, Y., and Narisawa, K. (1994) *Nat. Genet.* **8**, 122–128
- Leon-Del-Rio, A., Leclerc, D., Akerman, B., Wakamatsu, N., and Gravel, R. A. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 4626–4630
- Hiratsuka, M., Sakamoto, O., Li, X., Suzuki, Y., Aoki, Y., and Narisawa, K. (1998) *Biochim. Biophys. Acta* **1385**, 165–171
- Cronan, J. E., Jr., and Wallace, J. C. (1995) *FEMS Microbiol. Lett.* **130**, 221–230
- Tissot, G., Douce, R., and Alban, C. (1997) *Biochem. J.* **323**, 179–188
- Alban, C., Job, D., and Douce, R. (2000) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 17–47
- Tissot, G., Pépin, R., Job, D., Douce, R., and Alban, C. (1998) *Eur. J. Biochem.* **258**, 586–596
- Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983) *Plant Mol. Biol. Reporter* **1**, 19–21
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Tissot, G., Job, D., Douce, R., and Alban, C. (1996) *Biochem. J.* **314**, 391–395
- Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M., and Davis, R. W. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1731–1735
- Choi, J. K., Yu, F., Wurtele, E. S., and Nikolau, B. J. (1995) *Plant Physiol.* **109**, 619–625
- Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J., and Matthews, B. W. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 9257–9261
- The *Arabidopsis* Genome Initiative (2000) *Nature* **408**, 796–815
- Simpson, G. G., and Filipowicz, W. (1996) *Plant Mol. Biol.* **32**, 1–41
- Mukhopadhyay, B., Purwantini, E., Kreder, C. L., and Wolfe, R. S. (2001) *J. Bacteriol.* **183**, 3804–3810
- Campeau, E., and Gravel, R. A. (2001) *J. Biol. Chem.* **276**, 12310–12316
- Yanai, Y., Kawasaki, T., Shimada, H., Wurtele, E. S., Nikolau, B. J., and Ichikawa, N. (1995) *Plant Cell Physiol.* **36**, 779–787
- Duval, M., DeRose, R. T., Job, C., Faucher, D., Douce, R., and Job, D. (1994) *Plant Mol. Biol.* **26**, 265–273
- Job, C., Laugel, S., Duval, M., Gallardo, K., and Job, D. (2001) *Seed Sci. Res.* **11**, 149–161
- Caicedo, A. L., Schaal, B. A., and Kunkel, B. N. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 302–306
- Noel, L., Moores, T. L., van der Biezen, E. A., Parniske, M., Daniels, M. J., Parker, J. E., and Jones, J. D. G. (1999) *Plant Cell* **11**, 2099–2111
- Bergelson, J., Kreiman, M., Stahl, E., and Tian, D. (2001) *Science* **292**, 2281–2285
- Bergelson, J., Stahl, E., Dudek, S., and Kreiman, M. (1998) *Genetics* **148**, 1311–1323
- Botto, J. M., Vincent, J. P., and Mazella, J. (1997) *Biochem. J.* **324**, 389–393
- Hanfstingl, U., Berry, A., Kellogg, E. A., Costa, J. T., 3rd, Rudiger, W., and Ausubel, F. M. (1994) *Genetics* **138**, 811–823
- Henikoff, S., and Comai, L. A. (1998) *Genetics* **149**, 307–318